



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ**

**ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΥΓΕΙΑΣ**

**ΤΜΗΜΑ ΒΙΟΧΗΜΕΙΑΣ ΚΑΙ ΒΙΟΤΕΧΝΟΛΟΓΙΑΣ**

**ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ**

**ΕΛΕΝΗ Α. ΓΑΛΛΙΟΠΟΥΛΟΥ**

**Μοριακή Βιολόγος - Γενετίστρια**

**Λειτουργική ανάλυση του γονιδίου *FRA10A1* του ανθρώπου στο  
πλαίσιο του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου  
συναρμογής**

Λάρισα, 2018

**Λειτουργική ανάλυση του γονιδίου *FRA10A1* του ανθρώπου στο πλαίσιο του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής**

### **Τριμελής Συμβουλευτική Επιτροπή**

#### **Σαραφίδου Θεολογία (Επιβλέπουσα Καθηγήτρια)**

Επίκουρος Καθηγήτρια Μοριακής Γενετικής Ζωικών Οργανισμών, Τμήμα Βιοχημείας και Βιοτεχνολογίας, Πανεπιστήμιο Θεσσαλίας

#### **Μαμούρης Ζήσης**

Καθηγητής Γενετικής Ζωικών Πληθυσμών, Τμήμα Βιοχημείας και Βιοτεχνολογίας, Πανεπιστήμιο Θεσσαλίας

#### **Μοσχονάς Νικόλαος**

Καθηγητής Γενικής Βιολογίας-Ιατρικής Μοριακής Γενετικής, Ιατρική Σχολή, Πανεπιστήμιο Πατρών

### **Υπόλοιπα μέλη Επταμελούς Εξεταστικής Επιτροπής**

#### **Κλάπα Μαρία**

Ερευνήτρια Β', Ινστιτούτο Επιστημών Χημικής Μηχανικής, Ίδρυμα Τεχνολογίας και Έρευνας (ΙΕΧΜΗ/ΙΤΕ)

#### **Λιαδάκη Καλλιόπη**

Επίκουρος Καθηγήτρια Βιοχημικής Φαρμακολογίας, Τμήμα Βιοχημείας και Βιοτεχνολογίας, Πανεπιστήμιο Θεσσαλίας

#### **Μπαλατσός Νικόλαος**

Επίκουρος Καθηγητής Βιοχημείας, Τμήμα Βιοχημείας και Βιοτεχνολογίας, Πανεπιστήμιο Θεσσαλίας

#### **Γιακουντής Αντώνιος**

Επίκουρος Καθηγητής Μοριακής Βιολογίας-Γονιδιοματικής, Τμήμα Βιοχημείας και Βιοτεχνολογίας, Πανεπιστήμιο Θεσσαλίας

**ΕΛΕΝΗ Α. ΓΑΛΛΙΟΠΟΥΛΟΥ**

**Λειτουργική ανάλυση του γονιδίου *FRA10AC1* του ανθρώπου στο  
πλαίσιο του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου  
συναρμογής**

ΕΛΕΝΗ Α. ΓΑΛΛΙΟΠΟΥΛΟΥ

2018

Λειτουργική ανάλυση του γονιδίου *FRA10AC1* του ανθρώπου στο πλαίσιο του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής

«Διδακτορική Διατριβή, Πανεπιστήμιο Θεσσαλίας»

Αριθμός προκαταρκτικών σελίδων: 19  
Συνολικός αριθμός σελίδων: 329  
Αριθμός πινάκων: 24  
Αριθμός εικόνων: 85  
Αριθμός παραρτημάτων: 2  
Αριθμός βιβλιογραφικών παραπομπών: 838

## ΠΕΡΙΛΗΨΗ

Το γονίδιο *FRA10AC1* εντοπίζεται στην σπάνια εύθραυστη χρωμοσωματική θέση *FRA10A*, η οποία επάγεται απουσία φυλλικού οξέος στο θρεπτικό μέσο καλλιέργειας των κυττάρων, στην χρωμοσωματική περιοχή 10q23.3 του ανθρώπου. Η *FRA10A* αποτελεί την πιο συχνά εμφανιζόμενη σπάνια εύθραυστη θέση στο γονιδίωμα του ανθρώπου με συχνότητα που εκτιμάται στους 1/500. Οι φορείς της *FRA10A* εμφανίζουν νοητική και αναπτυξιακή υστέρηση, όπως δυσμορφικά χαρακτηριστικά, μικρό ανάστημα, υποσπαδία και δυσχέρεια λόγου. Η μοριακή βάση της κυτταρογενετικής εμφάνισής της είναι η επέκταση, κατά τουλάχιστον ~200 φορές, τρινουκλεοτιδικών επαναλήψεων του τύπου (CGG)*n* που εντοπίζονται στην 5' μη μεταφραζόμενη περιοχή του γονιδίου *FRA10AC1*. Αυτή η επέκταση οδηγεί στην υπερμεθυλίωση της περιοχής αυτής και στη μεταγραφική καταστολή του αντίστοιχου αλληλομόρφου.

Το γονίδιο *FRA10AC1* είναι μεταγραφικά ενεργό σε όλους τους ιστούς ενήλικων ατόμων παρουσιάζοντας υψηλότερα επίπεδα έκφρασης στον εγκέφαλο, την καρδιά, τους σκελετικούς μύες και το ήπαρ. Το κύριο μετάγραφο του γονιδίου *FRA10AC1* κωδικοποιεί μια πρωτεΐνη 315 αμινοξέων που εμφανίζει αποκλειστικά πυρηνική τοπολογία. Η *FRA10AC1* είναι μια συντηρημένη πρωτεΐνη καθώς εμφανίζει ορθόλογα μόρια σε πλήθος ευκαρυωτικών, πολυκύτταρων ή μονοκύτταρων οργανισμών, αλλά όχι σε προκαρυωτικούς οργανισμούς. Έχει ταυτοποιηθεί επανειλημμένα ως συστατικό του μείζονος σωματίου συναρμογής και των επιμέρους υποσυμπλόκων του, *B act* (activated), *C* και *P*. Επίσης, έχει δειχθεί, με γενετικά πειράματα στο φύκος *Chlamydomonas reinhardtii*, ότι συνεισφέρει στην αναγνώριση των θέσεων συναρμογής, και με διαφορετικές πειραματικές μεθόδους, ότι αλληλεπιδρά με συστατικά του σωματίου συναρμογής (*DGCR14*, *SF3B2*). Βάσει των παραπάνω υποδεικνύεται η συμμετοχή της στις διαδικασίες επεξεργασίας του mRNA.

Με στόχο την κατανόηση του λειτουργικού ρόλου της *FRA10AC1* έγινε, αρχικά, η ανασύσταση του δικτύου πρωτεϊνικών αλληλεπιδράσεων της και, ακολούθως, ολόκληρου του σωματίου συναρμογής. Για την ανακατασκευή των δικτύων χρησιμοποιήθηκε η μεταβάση *PICKLE* η οποία περιλαμβάνει καταχωρήσεις από πέντε βάσεις δεδομένων πρωτεϊνικών αλληλεπιδράσεων του ανθρώπου, καθώς και οι βάσεις δεδομένων *DroID* και *Worm Interactome Database* που περιλαμβάνουν πρωτεϊνικές αλληλεπιδράσεις από τους οργανισμούς-μοντέλα, *D. melanogaster* και *C. elegans*. Πραγματοποιήθηκε η ανάκτηση των δεδομένων που αφορούσαν μόνο άμεσες πρωτεϊνικές αλληλεπιδράσεις με βασικό κριτήριο, στην περίπτωση του δικτύου του σωματίου συναρμογής, τη βιοχημική ταυτοποίηση των πρωτεϊνών ως συστατικά του ριβονουκλεοπρωτεϊνικού αυτού συμπλόκου. Η ανάλυση των τοπολογικών παραμέτρων του δικτύου του σωματίου συναρμογής και της γονιδιακής οντολογίας της ευρύτερης «γειτονιάς» της *FRA10AC1* συνεισέφερε στην εξαγωγή συμπερασμάτων σχετικά με την πιθανή βιολογική λειτουργία της *FRA10AC1*. Παράλληλα, δημιουργήθηκε ένα κυτταρικό μοντέλο *HeLa* με μόνιμα κατεσταλμένη την έκφραση του γονιδίου *FRA10AC1*, μέσω της χρήσης κλώνων *shRNA* που στοχεύουν στη κωδική περιοχή του ενδογενούς γονιδίου. Στο κυτταρικό μοντέλο εφαρμόστηκαν ομικές

προσεγγίσεις στα τρία βασικά επίπεδα κυτταρικής λειτουργίας (μεταγραφικό, πρωτεϊνικό, μεταβολικό) ώστε να ελεγχθούν οι επιδράσεις που προκαλούσε η καταστολή της έκφρασης του γονιδίου στην κυτταρική φυσιολογία. Η μελέτη των αλλαγών αυτών στο πλαίσιο του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής αλλά και εκτός αυτού συνεισέφερε στην ανάπτυξη προτεινόμενων μοντέλων λειτουργίας της FRA10AC1 εντός του σωματίου συναρμογής.

## Ευχαριστίες

Γράφοντας αυτές τις λέξεις αναρωτιέμαι ποια λέξη θα μπορούσε να χαρακτηρίσει την πορεία ενός διδάκτορα;...ένα μεγάλο ΣΧΟΛΕΙΟ γνώσεων, χαρακτήρων, συμπεριφορών! Στο τέλος αυτής της πορείας δε μπορώ να μην εκφράσω την ευγνωμοσύνη μου, σε ανθρώπους που ήταν εκεί, και μου προσέφεραν ο καθένας με τον δικό του τρόπο, ...πολλά!

Ο πρώτος άνθρωπος που θα ήθελα να ευχαριστήσω ειλικρινά είναι η επιβλέπουσα καθηγήτρια της διατριβής, Επίκουρος καθηγήτρια Θεολογία Σαραφίδου. Είναι ο άνθρωπος που, από τα χρόνια του μεταπτυχιακού μου, με ενέπνευσε να εισέλθω στον επιστημονικό κόσμο και μου έδωσε την ευκαιρία να ξεκινήσω την πορεία μου. Με δίδαξε, με συμβούλεψε, με εμπιστεύτηκε, με βοήθησε στις δύσκολες κατάστασεις που συναντήσαμε μπροστά μας όλα αυτά τα χρόνια.

Ειλικρινείς ευχαριστίες θα ήθελα να εκφράσω στον Καθηγητή Νικόλαο Μοσχονά, που αν και η έδρα του είναι στο εργαστήριο Γενικής Βιολογίας του τμήματος Ιατρικής του Πανεπιστημίου Πατρών παρακολούθησε όλη την πορεία μου και συνέβαλε στην εξέλιξη της διατριβής μου με ουσιαστικές παρατηρήσεις και σχόλια, και με φιλοξένησε στο εργαστήριο του για την παρακολούθηση των πειραμάτων που πραγματοποιήθηκαν εκεί.

Ένα μεγάλο ευχαριστώ θέλω να εκφράσω στον Καθηγητή Ζήση Μαμούρη (και πλέον Πρύτανη του Πανεπιστημίου Θεσσαλίας) που στάθηκε άγρυπνος συμπαραστάτης στην πορεία μου στο εργαστήριό του και μου έδωσε την πολύτιμη βοήθειά του σε κάποιες κρίσιμες στιγμές που νοιώθεις ότι έπεσες και δεν μπορείς να σηκωθείς.

Επιπλέον, ευχαριστώ θερμά το μέλος της επταμελούς εξεταστικής επιτροπής, την Ερευνήτρια Β' και Επικεφαλής του εργαστηρίου Μεταβολικής Μηχανικής και Συστημικής Βιολογίας του ΙΤΕ/ΙΕΧΜΗ, Μαρία Κλάπα, για τις πολύτιμες συμβουλές της και την καθοδήγησή της στον επιστημονικό κόσμο των δικτύων και των μεταβολικών μονοπατιών. Καθώς επίσης και τα άλλα μέλη της επταμελούς εξεταστικής επιτροπής, την Επίκουρη καθηγήτρια Καλλιόπη Λιαδάκη και τον Επίκουρο καθηγητή Νικόλαο Μπαλατσό για τις συμβουλές τους σε θέματα κυτταροκαλλιιεργειών και για τη συμμετοχή τους στην εξεταστική επιτροπή και τον Επίκουρο καθηγητή Αντώνιο Γιακουντή για την προθυμία του να συμμετάσχει στην εξεταστική επιτροπή της διατριβής.

Πολλές και θερμές ευχαριστίες θα ήθελα να εκφράσω για τα μέλη του εργαστηρίου Γενετικής, Συγκριτικής και Εξελικτικής Βιολογίας με πρώτο από όλους την «κολώνα»-lab manager αυτού, τον Δρ. Κώστα Σταμάτη που ήταν δίπλα μου πάντα όταν τον χρειαζόμουν και σε πρακτικά και ανθρώπινα ζητήματα που προέκυπταν (...και δεν ήταν λίγα!!). Τι να πω για τον «senior» υποψήφιο διδάκτορα του εργαστηρίου ή αλλιώς Θέμη επικίνδυνο(!) Γιαννούλη, για τις πολλές στιγμές γέλιου (στην καζούρα πρώτος) που ζήσαμε στην Πλούτωνος (#ποιο\_νέο κτίριο) αλλά και για την συνεισφορά του γενικά σε επιστημονικά θέματα, αλλά, κυρίως, πληροφορικής (άσχετη γαρ!!). Το δίδυμο του εργαστηρίου έρχεται να συμπληρώσει η υποψήφια διδάκτορας, Κωνσταντίνα Σαρρή που συνεισέφερε (με το δικό της τρόπο πάντα!) σε όλες τις διασκεδαστικές στιγμές αλλά και που ήταν δίπλα μου στο μεγαλύτερο κομμάτι της πορείας μου. Και φυσικά, το νέο σχετικά αίμα του εργαστηρίου (από τους υπολοίπους), τους υποψήφιους διδάκτορες Ανδρέα Τσιπουρλιάνο και Μαρία Μαρκαντώνη, οι οποίοι, παρόλο που το «bonding» μπορεί να πει κανείς ότι άργησε να γίνει, μου έδωσαν λύσεις και ώθηση να συνεχίσω όταν ήμουν χαμένη (ευχαριστώ Μαρία) και μου θύμισαν πόσο ωραίο είναι να υπάρχει επικοινωνία με τους συνεργάτες σου (και τους δυο) όταν κιόλας ο άλλος αγχώνεται περισσότερο από



σένα (δεν το περίμενα ότι υπάρχει, Ανδρέα). Επίσης, σε πολλούς προπτυχιακούς (αγαπημένο μου Πανδωράκι) και μεταπτυχιακούς φοιτητές που πέρασαν ή συνεχίζουν να είναι στο εργαστήριο οι οποίοι προσέφεραν την θετική τους νότα στην πορεία μου όλα αυτά τα χρόνια. Επιπλέον, θέλω να ευχαριστήσω τα μέλη του εργαστηρίου Δομικής και Λειτουργικής Βιολογίας για τη συνεργασία μας και κυρίως τον υποψήφιο διδάκτορα και φίλο Γιώργο Στραβοδήμο, που ήταν εκεί όταν τον χρειαζόμουν δίχως δεύτερη σκέψη.

Μιλώντας για τις φίλιες μου με το εργαστήριο Βιοχημείας, το πιο σημαντικό μέλος του για μένα, είναι η πολυαγαπημένη μου φίλη Δρ. Δήμητρα Χατζηλεοντιάδου, που όχι απλά με στήριξε, ήταν δίπλα μου, αλλά και ζήσαμε πολλές ωραίες και αξέχαστες στιγμές (ξέρεις εσύ!). Θα σ'αγαπώ σε όποια μεριά του πλανήτη κι αν πας!

Ένα άλλο αναπόσπαστο κομμάτι στην πορεία μου ήταν η συμμετοχή μου στη δευτεροβάθμια εκπαίδευση. Ένα μεγάλο ευχαριστώ σε όλα τα παιδιά που με έκαναν «κυρία» από νωρίς και ήταν το αγχολυτικό μου για έξι συναπτά έτη. Το σημαντικότερο που κέρδισα από αυτήν την εμπειρία ήταν τα κορίτσια μου, οι φίλες μου, Κάκια Ψύρρα και Γιούλη Πετρούλη τις οποίες ευχαριστώ από καρδιάς για τη στήριξη, τις αμέτρητες συζητήσεις και συμβουλές τους. Οι φίλοι λένε πρέπει να βρίσκονται και να μιλάνε συχνά, αλλά οι πραγματικοί φίλοι με μία κουβέντα τους, μία αγκαλιά την κατάλληλη στιγμή κάνουν τη σχέση ουσιαστική και νοιώθεις σαν μην πέρασε μια μέρα, όσο καιρό και αν έχεις να τους δεις. Τέτοιες φίλιες είχα την τύχη να κάνω, την πρώτη από το δημοτικό, και για όσα χρόνια περάσουν, το Μητσάκι μου, που όλο όσα μου έχει δώσει, όχι μόνο αυτά τα χρόνια αλλά και γενικότερα στη ζωή μου, είναι αμέτρητα. Την Βασούλα μου, σαν λυκειόπαιδο πια, με την οποία ζήσαμε παράλληλα την πορεία μας ως υποψήφιοι διδάκτορες και, μιλώντας την ίδια γλώσσα, οι συμβουλές της και οι συζητήσεις μας ήταν βάλαμο για μένα. Την Βάσια και την Ελένη, από τα φοιτητικά μου χρόνια, το τρελοκόριτσο και η ισορροπημένη, εγώ η φωνή της λογικής! Τι να πω γι' αυτά τα κορίτσια η μία σε κάθε ευκαιρία μου θύμιζε ότι πρέπει να ζω την κάθε στιγμή, χωρίς να χάνω λεπτό, και η άλλη, πέρα από τη στήριξη της, μου έκανε, το καλύτερο δώρο, να βάλω λάδι στη μικρή μας Λυδία, που στο ακουσματά της έπαιρνα δύναμη για να ολοκληρωθεί η συγγραφή αυτής της διατριβής.

Δε μπορώ να εκφράσω με λόγια τα ευχαριστώ και την αγάπη μου στους γονείς μου, Αντωνία και Αλέξανδρο, που ήταν στο πλάι μου και με παρότρυναν σε κάθε μου βήμα, και ανέχτηκαν τα νεύρα και τις ανεξήγητες αντιδράσεις μου, απλά.... υπομένοντας και πιστεύοντας σε μένα.

Όλοι αυτοί με βοήθησαν να καταλάβω τη δύναμη που κρύβω μέσα μου και ότι μπορώ να καταφέρω πολλά αρκεί να πιστέψω στον εαυτό μου!!...και τους ΕΥΧΑΡΙΣΤΩ!

*Στη δύναμη που κρύβουμε μέσα μας.....*

## ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

1. ΕΙΣΑΓΩΓΗ.....	1
1.1. Το γονίδιο <i>FRA10AC1</i> του ανθρώπου.....	1
1.1.1. Κλωνοποίηση του γονιδίου <i>FRA10AC1</i> .....	1
1.1.2. Γονιδιωματική οργάνωση του γονιδίου <i>FRA10AC1</i> .....	1
1.1.3. Πρότυπο έκφρασης και εναλλακτικά μετάγραφα του γονιδίου <i>FRA10AC1</i> .....	3
1.2. Εύθραυστες χρωμοσωματικές θέσεις.....	4
1.2.1. Συνήθεις εύθραυστες θέσεις.....	6
1.2.2. Σπάνιες εύθραυστες θέσεις.....	7
1.2.3. Κλινική σημασία των εύθραυστων θέσεων.....	8
1.2.4. Συσχετισμός εύθραυστης θέσης <i>FRA10A</i> και γονιδίου <i>FRA10AC1</i> με νοσήματα... 9	
1.3. Χαρακτηριστικά της πρωτεΐνης <i>FRA10AC1</i> .....	11
1.3.1. Η πρωτεΐνη <i>FRA10AC1</i> σχετίζεται με την επεξεργασία του πρόδρομου mRNA... 12	
1.4. Μηχανισμός συναρμογής.....	17
1.4.1. Πειραματικές μελέτες απομόνωσης και ταυτοποίησης των συστατικών του σωματίου συναρμογής.....	22
1.4.2. Λειτουργική σχέση της συναρμογής με άλλες συναφείς βιολογικές διαδικασίες.....	23
1.5. Δικτυακή Βιολογία υπό το πρίσμα της Συστημικής Βιολογίας.....	25
1.5.1. Δίκτυα πρωτεϊνικών αλληλεπιδράσεων.....	28
1.5.1.1. Ταυτοποίηση πρωτεϊνικών αλληλεπιδράσεων.....	28
1.5.1.1.1. Περιορισμοί των μεθόδων ταυτοποίησης πρωτεϊνικών αλληλεπιδράσεων.....	31
1.5.1.2. <i>In silico</i> προσέγγιση κατασκευής πρωτεϊνικών δικτύων.....	32
1.5.1.3. Κατασκευή πρωτεϊνικών δικτύων μέσω της ανάκτησης δεδομένων από βάσεις πρωτεϊνικών αλληλεπιδράσεων.....	33
1.5.1.4. Ανάλυση πρωτεϊνικών δικτύων.....	34
1.5.1.5. Σύνολο γενετικών νοσημάτων στην προοπτική των δικτύων (human diseasome).....	34
1.5.2. Συναρμογή και ασθένειες.....	37
1.5.3. Δίκτυα πρωτεϊνικών και ρυθμιστικών αλληλεπιδράσεων των συστατικών του σωματίου συναρμογής.....	38
1.6. Αξιοποίηση των ομικών προσεγγίσεων στη Συστημική Βιολογία.....	52
1.6.1. Μεταγραφωμική (Transcriptomics).....	60
1.6.2. Πρωτεϊνωμική (Proteomics).....	61
1.6.3. Μεταβολομική (Metabolomics).....	63
ΑΝΤΙΚΕΙΜΕΝΟ ΜΕΛΕΤΗΣ.....	65
2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ.....	66
2.1. Βάσεις Δεδομένων.....	66
2.1.1. Βάσεις Δεδομένων για χαρακτηριστικά των γονιδίων.....	66
2.1.1.1. Πηγή πληροφοριών σε επίπεδο γονιδίου (Gene) ( <a href="https://www.ncbi.nlm.nih.gov/gene">https://www.ncbi.nlm.nih.gov/gene</a> ).....	66
2.1.1.2. Διερεύνηση ορθόλογων μορίων (HomoloGene) ( <a href="https://www.ncbi.nlm.nih.gov/Homologene/">https://www.ncbi.nlm.nih.gov/Homologene/</a> ).....	66
2.1.2. Βάσεις δεδομένων για χαρακτηριστικά, λειτουργία και οντολογία των πρωτεϊνών.....	67
2.1.2.1. Βάση Δεδομένων Γονιδιακής Οντολογίας (Gene Ontology Consortium)	

( <a href="http://www.geneontology.org/">http://www.geneontology.org/</a> ).....	67
2.1.2.2. Βάση Δεδομένων για την περιγραφή των πρωτεϊνών (UniProt, Universal Protein Resource) ( <a href="http://www.uniprot.org/">http://www.uniprot.org/</a> ).....	68
2.1.2.3. Βάση Δεδομένων DAVID (Database for Annotation, Visualization and Integrated Discovery) ( <a href="http://david.abcc.ncifcrf.gov/">http://david.abcc.ncifcrf.gov/</a> ).....	69
2.1.2.4. Βάση Δεδομένων για την υποκυτταρική τοπολογία πρωτεϊνών και τα επίπεδα έκφρασης γονιδίων (Human Protein Atlas) ( <a href="http://www.proteinatlas.org/">http://www.proteinatlas.org/</a> ).....	70
2.1.2.5. Βάση Δεδομένων για τα συστατικά/πρωτεΐνες του σωματίου συναρμογής (Spliceosome Database) ( <a href="http://spliceosomedb.ucsc.edu/">http://spliceosomedb.ucsc.edu/</a> ).....	71
2.1.3. Βάσεις Δεδομένων για εύρεση βιβλιογραφικών αναφορών και για γενετικά νοσήματα.....	72
2.1.3.1 Βάση Δεδομένων βιβλιογραφικών αναφορών (PubMed) ( <a href="https://www.ncbi.nlm.nih.gov/pubmed/">https://www.ncbi.nlm.nih.gov/pubmed/</a> ).....	72
2.1.3.2 Βάση Δεδομένων για γενετικές διαταραχές του ανθρώπου (OMIM) ( <a href="http://www.ncbi.nlm.nih.gov/omim/">http://www.ncbi.nlm.nih.gov/omim/</a> ).....	72
2.1.4 Βάσεις δεδομένων πρωτεϊνικών αλληλεπιδράσεων του ανθρώπου.....	72
2.1.4.1. BioGRID (Biological General Repository for Interaction Datasets) ( <a href="https://thebiogrid.org/">https://thebiogrid.org/</a> ).....	73
2.1.4.2. HPRD (Human Protein Reference Database) ( <a href="http://www.hprd.org/">http://www.hprd.org/</a> ).....	73
2.1.4.3. IntAct (Database of protein interaction data) ( <a href="http://www.ebi.ac.uk/intact/">http://www.ebi.ac.uk/intact/</a> ).....	74
2.1.4.4. MINT (Molecular INTeraction database) ( <a href="http://mint.bio.uniroma2.it/mint/">http://mint.bio.uniroma2.it/mint/</a> )..	75
2.1.4.5. DIP (Database of Interacting Proteins) ( <a href="http://dip.doe-mbi.ucla.edu/">http://dip.doe-mbi.ucla.edu/</a> ).....	75
2.1.4.6. Μετα-βάση PICKLE ( <a href="http://www.pickle.gr/">http://www.pickle.gr/</a> ).....	76
2.1.5. Βάσεις δεδομένων πρωτεϊνικών αλληλεπιδράσεων σε οργανισμούς-μοντέλα.....	77
2.1.5.1. DroID (Drosophila Interactions Database) ( <a href="http://droidb.org/">http://droidb.org/</a> ).....	77
2.1.5.2. Worm Interactome Database ( <a href="http://interactome.dfc.harvard.edu/">http://interactome.dfc.harvard.edu/</a> ).....	77
2.2. Εργαλεία οπτικοποίησης και ανάλυσης δικτύων.....	78
2.2.1. Πρόγραμμα οπτικοποίησης δικτύων, Cytoscape ( <a href="http://www.cytoscape.org">www.cytoscape.org</a> , version 3.2.1).....	78
2.3. Καλλιέργεια και μετασχηματισμός βακτηρίων.....	79
2.3.1. Βακτηριακά Στελέχη.....	79
2.3.2. Καλλιέργεια βακτηρίων.....	79
2.3.3. Δεκτικά βακτηριακά κύτταρα <i>DH5a</i> .....	80
2.3.3.1. Διαδικασία παρασκευής δεκτικών βακτηριακών κυττάρων <i>DH5a</i> .....	80
2.3.4. Μετασχηματισμός βακτηρίων με θερμικό σοκ.....	81
2.4. Χειρισμός και τροποποίηση κυττάρων θηλαστικών <i>HeLa</i> .....	82
2.4.1. Κυτταρικές σειρές θηλαστικών.....	82
2.4.2. Καλλιέργεια κυττάρων θηλαστικών.....	82
2.4.2.1. Συνθήκες και βασικές αρχές κυτταρικής καλλιέργειας.....	82
2.4.2.2. Διατήρηση κυτταροκαλλιιεργειών.....	83
2.4.2.3. Κρυοδιατήρηση κυττάρων.....	84
2.4.2.4. Επανακαλλιέργεια κυττάρων.....	85
2.4.2.5. Καταμέτρηση κυττάρων για επίστρωση τρυβλίων καλλιέργειας.....	85
2.4.3. Δημιουργία μόνιμα διαμολυσμένης κυτταρικής σειράς.....	87
2.4.4. Πρωτόκολλο επεξεργασίας και συλλογής κυττάρων για ομικές αναλύσεις.....	90
2.5. Απομόνωση νουκλεϊκών οξέων & τεχνικές ανασυνδυασμένου DNA.....	92
2.5.1. Απομόνωση γονιδιωματοικού DNA.....	92
2.5.2. Απομόνωση πλασμιδιακού DNA.....	93

2.5.2.1. Πλασμιδιακοί φορείς.....	95
2.5.2.2. Πλασμιδιακές κατασκευές.....	96
2.5.3. Απομόνωση RNA.....	101
2.5.3.1. Επεξεργασία του RNA με DNAση.....	102
2.5.3.2. Αντίστροφη μεταγραφή.....	102
2.5.4. Τεμαχισμός με περιοριστικές ενδονουκλεάσες.....	103
2.5.5. Αντίδραση σύνδεσης.....	103
2.5.6. Ηλεκτροφόρηση σε πηκτή αγαρόζης.....	104
2.5.7. Απομόνωση τεμαχίων DNA και καθαρισμός προϊόντων PCR και πέψης.....	105
2.5.8. Αλυσιδωτή αντίδραση πολυμεράσης (PCR).....	105
2.6. Απομόνωση πρωτεϊνών και ανάλυση western.....	107
2.6.1. Απομόνωση και ποσοτικοποίηση πρωτεϊνών.....	108
2.6.2. Ηλεκτροφόρηση πρωτεϊνών σε αποδιατακτική πηκτή πολυακρυλαμίδης.....	109
2.6.3. Ηλεκτρομεταφορά πρωτεϊνών σε μεμβράνη.....	110
2.6.4. Ανοσοαποτύπωση.....	111
2.6.5. Ανοσοανίχνευση.....	112
3. ΑΠΟΤΕΛΕΣΜΑΤΑ.....	113
3.1. <i>In silico</i> ανακατασκευή δικτύου άμεσων πρωτεϊνικών αλληλεπιδράσεων της πρωτεΐνης FRA10AC1 του ανθρώπου.....	113
3.1.1. Συλλογή πρωτεϊνικών αλληλεπιδράσεων από βάσεις δεδομένων.....	113
3.1.2. Κριτήρια αξιολόγησης των δεδομένων για τις άμεσες πρωτεϊνικές αλληλεπιδράσεις της FRA10AC1.....	117
3.1.3. Κριτήρια αξιολόγησης των δεδομένων για τις έμμεσες πρωτεϊνικές αλληλεπιδράσεις της FRA10AC1.....	124
3.1.4. <i>In silico</i> ανακατασκευή του δικτύου πρωτεϊνικών αλληλεπιδράσεων της FRA10AC1.....	126
3.2. <i>In silico</i> ανακατασκευή του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής (spliceosome) και των υποσυμπλόκων του.....	128
3.2.1. <i>In silico</i> ανακατασκευή του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής.....	128
3.2.1.1. Δημιουργία του συνόλου δεδομένων των συστατικών του σωματίου συναρμογής.....	128
3.2.1.1.1. Λειτουργική κατηγοριοποίηση των συστατικών του σωματίου συναρμογής.....	132
3.2.1.1.2. Η πρωτεϊνική σύσταση των υποσυμπλόκων του σωματίου συναρμογής.....	137
3.2.1.1.3. Λειτουργική κατηγοριοποίηση των συστατικών των υποσυμπλόκων του σωματίου συναρμογής.....	137
3.2.1.1.4. Προσδιορισμός και αφθονία των πρωτεϊνών με καθολική παρουσία κατά τη συναρμογή.....	143
3.2.1.1.5. Προσδιορισμός των πρωτεϊνών που εμφανίζονται μόνο σε συγκεκριμένα υποσύμπλοκα.....	146
3.2.1.2. Δημιουργία του συνόλου δεδομένων των πρωτεϊνικών αλληλεπιδράσεων των συστατικών του σωματίου συναρμογής.....	146
3.2.1.2.1. Αλληλοεπικάλυψη βάσεων δεδομένων σε πρωτεΐνες, αλληλεπιδράσεις και ερευνητικές εργασίες.....	150
3.2.1.3. Οπτικοποίηση του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής.....	154
3.2.2. <i>In silico</i> ανακατασκευή των δικτύων πρωτεϊνικών αλληλεπιδράσεων των υποσυ-	

μπλόκων του σωματίου συναρμογής.....	164
3.2.3. Διερεύνηση της τοπολογίας της πρωτεΐνης FRA10AC1 στο συνολικό δίκτυο του σωματίου συναρμογής και στα δίκτυα των υποσυμπλόκων.....	170
3.2.4. Ανάλυση τοπολογικών παραμέτρων του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής.....	173
3.2.4.1. Σύνθετες (complex) παράμετροι.....	176
3.2.4.2. Οι επιδράσεις της απομάκρυνσης των πρωτεϊνών-«hubs» και της FRA10AC1 στις τοπολογικές παραμέτρους το δικτύου του σωματίου συναρμογής και των υποσυμπλόκων του.....	178
3.2.4.3. Ανάλυση ικανότητας σχηματισμού συμπλεγμάτων πρωτεϊνών εντός του δικτύου του σωματίου συναρμογής.....	181
3.2.5. Συσχετισμός των συστατικών του σωματίου συναρμογής με γενετικά νοσήματα.....	184
3.2.6. Πρωτεΐνες στην «γειτονιά» της πρωτεΐνης FRA10AC1 που σχετίζονται με γενετικά νοσήματα.....	187
3.3. Έλεγχος επιλεγμένων πρωτεϊνικών αλληλεπιδράσεων του ανακατασκευασμένου δικτύου της πρωτεΐνης FRA10AC1.....	189
3.3.1. Υποκλωνοποίηση των cDNAs των γονιδίων <i>SAP30BP</i> , <i>THOC1</i> και <i>MFAP1</i> του ανθρώπου σε πλασμιδιακό φορέα έκφρασης σε κύτταρα θηλαστικών pmCherry-C1.....	190
3.3.2. Έλεγχος πρωτεϊνικών αλληλεπιδράσεων με τη μέθοδο Μεταφοράς Ενέργειας Φθορισμού μέσω Συντονισμού (FRET).....	192
3.4. Συνδυαστικές ομικές αναλύσεις της τροποποιημένης έκφρασης του <i>FRA10AC1</i> σε κυτταρικό μοντέλο.....	195
3.4.1. Επεξεργασία και συλλογή κυττάρων <i>HeLa</i> για ομικές αναλύσεις.....	196
3.4.1.1. Μεταβολομική ανάλυση κυττάρων <i>HeLa</i> που συλλέχθηκαν με διαφορετικά πρωτόκολλα.....	197
3.4.1.2. Το πρωτόκολλο 3B, το PBS ως ρυθμιστικό διάλυμα έκπλυσης και τα κύτταρα της 2 <sup>ns</sup> ανακαλλιέργειας αποτελούν το βέλτιστο τρόπο επεξεργασίας και συλλογής κυττάρων για ομικές αναλύσεις.....	204
3.4.2. Καταστολή της έκφρασης του γονιδίου <i>FRA10AC1</i> στα κύτταρα <i>HeLa</i> .....	205
3.4.3. Συλλογή δειγμάτων για ομικές αναλύσεις.....	211
3.4.4. Εφαρμογή ομικών συνδυαστικών αναλύσεων του κυτταρικού μοντέλου με τροποποιημένη την έκφραση του <i>FRA10AC1</i> .....	213
3.4.4.1. Ανάλυση μεταγραφικού προτύπου.....	213
3.4.4.2. Ανάλυση πρωτεϊνικού προτύπου.....	214
3.4.4.3. Ανάλυση μεταβολικού προτύπου.....	216
3.5. Πρωτεΐνες του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής με τροποποιημένη έκφραση σε πρωτεϊνικό και μεταγραφικό επίπεδο, μετά την καταστολή της έκφρασης του γονιδίου <i>FRA10AC1</i> .....	217
3.6. Δημιουργία κυτταρικής σειράς με επαγωγίμη έκφραση της πρωτεΐνης DGCR14....	222
3.6.1. Δημιουργία σταθερά διαμολυσμένης κυτταρικής σειράς <i>HeLa</i> που εκφράζει τον καταστολέα της τετρακυκλίνης.....	223
3.6.2. Δημιουργία κυτταρικής σειράς με επαγωγίμη έκφραση του γονιδίου <i>DGCR14</i> ..	225
4. ΣΥΖΗΤΗΣΗ.....	228
4.1. Το δίκτυο της FRA10AC1 περιέχει άμεσες και αξιόπιστες πρωτεϊνικές αλληλεπιδράσεις που υποδεικνύουν την συμμετοχή της στις διαδικασίες επεξεργασίας του mRNA.....	228

4.1.1. Η χρήση της μεθόδου FRET για τον έλεγχο των αλληλεπιδράσεων της FRA10AC1 με πρωτεΐνες που συμμετέχουν στις διαδικασίες επεξεργασίας του mRNA.....	230
4.2. <i>In silico</i> ανακατασκευή του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής.....	231
4.2.1. Οι 630 πρωτεΐνες του σωματίου συναρμογής αποτελούν ένα πλήρες, αξιόπιστο και μη επικαλυπτόμενο σύνολο δεδομένων.....	232
4.2.2. Ογδόντα πρωτεΐνες είναι παρούσες καθ' όλη τη διάρκεια συναρμολόγησης του σωματίου συναρμογής.....	235
4.2.3. Το δίκτυο πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής και των υποσυμπλόκων του περιέχει αποκλειστικά άμεσες αλληλεπιδράσεις και είναι ιδιαίτερα αξιόπιστο.....	237
4.2.4. Ο προσδιορισμός των μοναδικών πρωτεϊνικών αλληλεπιδράσεων στα διαδοχικά στάδια συναρμολόγησης του σωματίου συναρμογής δίνει μια προκαταρκτική εικόνα της δυναμικής του.....	241
4.2.5. Ο αυτοματοποιημένος προσδιορισμός των οντολογικών κατηγοριών (GO terms) των πρωτεϊνών απαιτεί επιπρόσθετη αξιολόγηση.....	243
4.2.6. Ο προσδιορισμός των GO terms των πρωτεϊνών του σωματίου συναρμογής εμπλουτίζει το σύνολο δεδομένων με πρωτεΐνες που συμμετέχουν και σε άλλες διαδικασίες της επεξεργασίας και μεταφοράς του mRNA, πλην της συναρμογής.....	245
4.2.7. Τα ποσοστά ενσωμάτωσης των πρωτεϊνών του δικτύου στα GO terms συνάδουν με τα αντίστοιχα του συνόλου δεδομένων του σωματίου συναρμογής, ενώ μειώνεται το ποσοστό των πρωτεϊνών που δεν έχουν χαρακτηρισμένο λειτουργικό ρόλο...	249
4.2.8. Η λειτουργική κατηγοριοποίηση των πρωτεϊνών των υποσυμπλόκων του σωματίου συναρμογής και των αντίστοιχων δικτύων τους συνάδει με την πορεία συναρμολόγησής τους στο πρόδρομο μετάγραφο.....	250
4.2.9. Η ανάλυση των τοπολογικών παραμέτρων του δικτύου συνεισφέρει στον προσδιορισμό των ιδιοτήτων και της αρχιτεκτονικής του.....	251
4.2.10. Η εύρεση συμπλεγμάτων πρωτεϊνών (clusters) συνεισφέρει στον προσδιορισμό της λειτουργίας πρωτεϊνών με μη χαρακτηρισμένο βιολογικό ρόλο.....	257
4.2.11. Μεγάλος αριθμός πρωτεϊνών του δικτύου του σωματίου συναρμογής και της «γειτονιάς» της FRA10AC1 συσχετίζονται με νοσήματα.....	259
4.3. Περιορισμοί και προοπτικές του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής.....	261
4.3.1. Η εισαγωγή των αλληλεπιδράσεων των πρωτεϊνικών ισομορφών του δικτύου θα συνεισέφερε στην τροποποίηση του δικτύου.....	262
4.3.2. Τα ιστο-ειδικά δίκτυα πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής θα προσομοιάζαν περισσότερο την <i>in vivo</i> κατάσταση.....	264
4.3.3. Η διερεύνηση της αλλαγής του «προτύπου» των αλληλεπιδράσεων κατά τη διάρκεια συναρμολόγησης του σωματίου συναρμογής θα συνεισέφερε στην ανάκτηση πληροφοριών για τη δυναμική του.....	264
4.3.4. Η εισαγωγή των αλληλεπιδράσεων RNA-RNA και RNA-πρωτεϊνών στο δίκτυο του σωματίου συναρμογής θα συνέβαλλε στη δημιουργία ενός πιο ολοκληρωμένου δικτύου.....	266
4.4. Η ανάλυση του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής σε συνδυασμό με τις ομικές αναλύσεις του κυτταρικού μοντέλου με τροποποιημένη έκφραση του γονιδίου <i>FRA10AC1</i> παρέχουν σημαντικές ενδείξεις για τη βιολογική λειτουργία της <i>FRA10AC1</i> .....	267
4.4.1. Η εφαρμογή ομικών αναλύσεων συνεισφέρει στην εύρεση των αλλαγών που προκα-	

λούνται μετά την καταστολή της <i>FRA10AC1</i> στην κυτταρική φυσιολογία.....	268
4.4.2. Πιθανοί μοριακοί μηχανισμοί μέσω των οποίων η καταστολή της δράσης της <i>FRA10AC1</i> τροποποιεί την έκφραση των άλλων μορίων.....	289
4.4.3. Πιθανά μοντέλα λειτουργίας της <i>FRA10AC1</i> εντός του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής.....	295
Summary.....	303
Βιβλιογραφία.....	304

## Κατάλογος Εικόνων

Εικ. 1: Χαρτογράφηση του γονιδίου <i>FRA10AC1</i> στην χρωμοσωματική περιοχή 10q23.3 του ανθρώπου.....	2
Εικ. 2: Γονιδιωματική οργάνωση του γονιδίου <i>FRA10AC1</i> του ανθρώπου.....	2
Εικ. 3: Πρότυπο έκφρασης του γονιδίου <i>Fra10ac1</i> του ποντικού.....	3
Εικ. 4: Εναλλακτικά μετάγραφα του γονιδίου <i>FRA10AC1</i> και αντίστοιχες πρωτεϊνικές ισομορφές.....	4
Εικ. 5: Απεικόνιση των συνήθων εύθραυστων χρωμοσωματικών θέσεων <i>FRA3B</i> , <i>FRA16D</i> , <i>FRA7H</i> .....	5
Εικ. 6: Κυτταρογενετική έκφραση της σπάνιας εύθραυστης χρωμοσωματικής θέσης <i>FRA10A</i> .....	9
Εικ. 7: Πρωτεϊνική ακολουθία <i>FRA10AC1</i> (NP_660289.2).....	11
Εικ. 8: Υποκυτταρικός εντοπισμός της ανασυνδυασμένης πρωτεΐνης <i>FRA10AC1</i> σε νεφρικά κύτταρα πιθήκου ( <i>COS-7</i> ).....	12
Εικ. 9: Σχηματική απεικόνιση και διαφορές των ιντρονίων τύπου U2 και τύπου U12.....	18
Εικ. 10: Σχηματική αναπαράσταση των δυο αντιδράσεων μετεστεροποίησης.....	18
Εικ. 11: Συναρμολόγηση του μείζονος σωματίου συναρμογής.....	21
Εικ. 12: Οργάνωση των συστατικών του κυττάρου ως ένα δίκτυο αλληλεπιδράσεων.....	26
Εικ. 13: Ενδεικτικό δίκτυο πρωτεϊνικών αλληλεπιδράσεων.....	27
Εικ. 14: Ενδεικτική απεικόνιση της ανοσοκατακρήμνισης.....	29
Εικ. 15: Ενδεικτική απεικόνιση του συστήματος δυο υβριδίων στο σακχαρομύκητα.....	30
Εικ. 16: Απεικόνιση του Human Diseasesome.....	36
Εικ. 17: Ενδεικτική απεικόνιση δικτύων πρωτεϊνικών αλληλεπιδράσεων πρωτεϊνών core (A) και non-core (B).....	41
Εικ. 18: Κατηγοριοποίηση πρωτεϊνών του σωματίου συναρμογής σε modules.....	44
Εικ. 19: Δίκτυο λειτουργικών συνδέσεων του σωματίου συναρμογής.....	47
Εικ. 20: Δίκτυα πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής στον οργανισμό <i>S. cerevisiae</i> .....	49
Εικ. 21: Αναπαράσταση μιας διαδικασίας στο δίκτυο Petri.....	51
Εικ. 22: Ροή πληροφορίας από το γονιδίωμα μέχρι τον τελικό κυτταρικό φαινότυπο μέσω των δεδομένων των ομικών αναλύσεων.....	54
Εικ. 23: Διαδικασία παρεμβολής μέσω RNA (RNAi).....	57
Εικ. 24: Α: Πλάκα Neubauer Β: Τετραγωνισμένο πλέγμα στο αιμοκυτταρόμετρο Neubauer για την μέτρηση των Κυττάρων.....	86
Εικ. 25: Αδρός φυσικός χάρτης του φορέα pm-Cherry.....	95
Εικ. 26: Χάρτης πλασμιδιακού φορέα pcDNA <sup>TM</sup> 6/TR.....	96
Εικ. 27: Αδρός φυσικός χάρτης του πλασμιδιακού φορέα pLKO.1-puro.....	97
Εικ. 28: Στόχοι κλώνων shRNA στο cDNA του γονιδίου <i>FRA10AC1</i> .....	98
Εικ. 29: Χάρτης πλασμιδιακού φορέα pcDNA 3.1 His.....	100
Εικ. 30: Χάρτης πλασμιδιακού φορέα pcDNA 4/TO.....	101
Εικ. 31: Διαδικασία ηλεκτρομεταφοράς πρωτεϊνών σε μεμβράνη.....	111
Εικ. 32: Δίκτυο άμεσων πρωτεϊνικών αλληλεπιδράσεων της πρωτεΐνης <i>FRA10AC1</i> .....	127
Εικ. 33: Δημιουργία συνόλου δεδομένων των συστατικών του σωματίου συναρμογής.....	131
Εικ. 34: Λειτουργική ομαδοποίηση των 630 πρωτεϊνών-Ubiquitin IDs του σωματίου συναρμογής.....	135
Εικ. 35: Ομαδοποίηση των κατηγοριών γονιδιακής οντολογίας του συνόλου δεδομένων του σωματίου συναρμογής.....	136
Εικ. 36: Ομαδοποίηση των κατηγοριών γονιδιακής οντολογίας των επιμέρους υποσυμπλόκων του σωματίου συναρμογής.....	139
Εικ. 37: Απόλυτες τιμές πρωτεϊνών κάθε οντολογικής κατηγορίας στα επιμέρους υποσύμπλοκα του σωματίου Συναρμογής.....	141
Εικ. 38: Διαδικασία ανάκτησης και αξιολόγησης των δεδομένων των PPIs που προέκυψαν από τις βάσεις δεδομένων του ανθρώπου και των οργανισμών μοντέλων.....	149
Εικ. 39: Αλληλοεπικάλυψη πρωτεϊνών-αλληλεπιδράσεων-ερευνητικών εργασιών των 5 βάσεων δεδομένων PPIs	



μέσω διαγράμματος Venn.....	152
Εικ. 40: Δίκτυο πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής.....	155
Εικ. 41: Λειτουργική ομαδοποίηση των 444 πρωτεϊνών-κόμβων του δικτύου του σωματίου συναρμογής.....	160
Εικ. 42: Ομαδοποίηση των πρωτεϊνών σε κατηγορίες GO στο δίκτυο πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής.....	161
Εικ. 43: Δίκτυο πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής και επιμέρους δίκτυα των κύριων οντολογικών κατηγοριών.....	163
Εικ. 44: Δίκτυα PPI των υποσυμπλόκων του σωματίου συναρμογής ανεξάρτητα από το συνολικό δίκτυο PPI....	165
Εικ. 45: Δίκτυα PPI των υποσυμπλόκων του σωματίου συναρμογής διατηρώντας την τοπολογία των πρωτεϊνών κόμβων του συνολικού δικτύου PPI.....	166
Εικ. 46: Ομαδοποίηση των κατηγοριών GO των δικτύων PPI των επιμέρους συμπλόκων του σωματίου συναρμογής.....	168
Εικ. 47: Ομαδοποίηση των πρωτεϊνών των δικτύων PPI των επιμέρους συμπλόκων του σωματίου συναρμογής σε επτά οντολογικές κατηγορίες.....	169
Εικ. 48: Δίκτυα PPI της «γειτονιάς» της FRA10AC1 στο συνολικό δίκτυο PPI του σωματίου συναρμογής και στα δίκτυα των υποσυμπλόκων B act, C και P.....	172
Εικ. 49: Κατανομή του αριθμού των αλληλεπιδράσεων των πρωτεϊνών-κόμβων (node degree distribution) του δικτύου του σωματίου συναρμογής.....	175
Εικ. 50: Διαγράμματα σύνθετων παραμέτρων του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής.....	178
Εικ. 51: Συμπλέγματα πρωτεϊνών του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής.....	183
Εικ. 52: Κατηγορίες γενετικών νοσημάτων που σχετίζονται με μεταλλάξεις συστατικών του σωματίου συναρμογής.....	185
Εικ. 53: Συσχετισμός των πρωτεϊνών-κόμβων με γενετικά νοσήματα στο πλαίσιο του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής.....	185
Εικ. 54: Συσχετισμός των πρωτεϊνών-κόμβων των δύο συμπλεγμάτων του δικτύου συναρμογής με γενετικά Νοσήματα.....	186
Εικ. 55: Συσχετισμός πρώτων και δευτέρων γειτόνων της πρωτεΐνης FRA10AC1 με γενετικά νοσήματα εντός του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής.....	188
Εικ. 56: Προϊόντα PCR για τα cDNAs των γονιδίων <i>MFAP</i> , <i>SAP30BP</i> και <i>THOC1</i> .....	191
Εικ. 57: Έλεγχος ανασυνδυασμένων πλασμιδίων με τεμαχισμό με ένζυμο περιορισμού και PCR.....	192
Εικ. 58: Μέτρηση αποτελεσματικότητας της ανάλυσης FRET σε συνεστιακό μικροσκόπιο σάρωσης με ακτίνες Laser με τη μέθοδο της φωτοεύκανσης.....	194
Εικ. 59: Γραμμοποίηση των τεσσάρων διαγονιδίων και του διαγονιδίου που χρησιμοποιήθηκε ως αρνητικός έλεγχος (MSH) με το ένζυμο περιορισμού KpnI (nt 1189).....	206
Εικ. 60: Έλεγχος μυκοπλάσματος με PCR.....	207
Εικ. 61: Καθαρισμός αντισώματος α-FRA10AC1.....	208
Εικ. 62: Ανάλυση western των επιπέδων έκφρασης της πρωτεΐνης <i>FRA10AC1</i> σε κλώνους κυττάρων <i>HeLa</i> , μόνιμα διαμολυσμένους με shRNA που στοχεύουν στο γονίδιο <i>FRA10AC1</i> .....	209
Εικ. 63: Ανάλυση western των επιπέδων έκφρασης της πρωτεΐνης <i>FRA10AC1</i> των κλώνων shRNA(3)_5 και shRNA(4)_4.....	209
Εικ. 64: Ημιποσοτικός προσδιορισμός των επιπέδων mRNA του <i>FRA10AC1</i> στους κλώνους υποέκφρασης.....	210
Εικ. 65: Ηλεκτροφόρηση ολικού RNA των δυο τεχνικών επαναλήψεων (A) και (B) του κλώνου υποέκφρασης (4_4), και του αρνητικού ελέγχου του, C (MSH 2).....	212
Εικ. 66: Πρωτεΐνες του δικτύου αλληλεπιδράσεων του σωματίου συναρμογής με τροποποιημένη έκφραση μετά την καταστολή της έκφρασης του <i>FRA10AC1</i> .....	218
Εικ. 67: Πρωτεΐνες της «γειτονιάς» της <i>FRA10AC1</i> με τροποποιημένη έκφραση μετά την καταστολή της έκφρασής της.....	219
Εικ. 68: Πρωτεΐνες με τροποποιημένη έκφραση μετά την καταστολή της έκφρασης του γονιδίου <i>FRA10AC1</i> στην ευρύτερη «γειτονιά» της <i>FRA10AC1</i> εντός του δικτύου του σωματίου συναρμογής.....	220
Εικ. 69: Δίκτυο της ευρύτερης «γειτονιάς» της <i>FRA10AC1</i> εντός του σωματίου συναρμογής με όλες τις 39 πρωτεΐνες που παρουσίασαν τροποποιημένη έκφραση μετά την καταστολή της έκφρασης του γονιδίου <i>FRA10AC1</i> .....	221
Εικ. 70: Επαγώγιμο σύστημα έκφρασης του γονιδίου <i>DGCR14</i> .....	223
Εικ. 71: Ηλεκτροφόρηση ολικού RNA από τους κλώνους κυττάρων <i>HeLa</i> #3, 4 και 5 (A) και προϊόντα PCR του cDNA του γονιδίου TetR για τους κλώνους αυτούς (B).....	225
Εικ. 72: Ανάλυση western των επιπέδων υπερέκφρασης της ανασυνδυασμένης πρωτεΐνης <i>DGCR14</i> έπειτα από επαγωγή της έκφρασης με δοξκυκλίνη για 24h.....	226

Εικ. 73: Ανάλυση western των επιπέδων υπερέκφρασης της ανασυνδυασμένης πρωτεΐνης DGCR14 έπειτα από επαγωγή της έκφρασης με δοξυκυκλίνη.....	227
Εικ. 74: Μεταβολικές προσαρμογές των καρκινικών κυττάρων.....	270
Εικ. 75: Η διαδικασία της γλουταμινόλυσης και οι μεταβολικές αντιδράσεις που έπονται αυτής.....	271
Εικ. 76: Μεταβολικές αλλαγές που παρατηρούνται στα πρωτόκολλα 1 και 2.....	273
Εικ. 77: Πορεία αποκατάστασης της γλυκόζης.....	275
Εικ. 78: Μεταβολικές αλλαγές του πρωτοκόλλου 3A.....	275
Εικ. 79: Διαδικασία σύνθεσης και διάσπασης του γλυκογόνου.....	277
Εικ. 80: Μεταβολίτες με σημαντικά αυξημένη ή μειωμένη συγκέντρωση στους κλώνους υποέκφρασης του γονιδίου FRA10AC1 και τα μεταβολικά μονοπάτια στα οποία ενσωματώθηκαν.....	280
Εικ. 81: Μεταβολίτες που αυξήθηκαν ή μειώθηκαν κατά την καταστολή της έκφρασης του γονιδίου FRA10AC1 και το μονοπάτι των πολυολών στο οποίο ενσωματώθηκαν.....	281
Εικ. 82: Μεταβολίτες που αυξήθηκαν στους κλώνους υποέκφρασης του γονιδίου FRA10AC1 και το μονοπάτι παραγωγής της N-ακετυλογλυκοζαμίνης που είναι υπεύθυνο για τη γλυκοζυλίωση των πρωτεϊνών.....	281
Εικ. 83: Πιθανός βιολογικός ρόλος της FRA10AC1 ως πρωτεΐνη-προσαρμογέας μεταξύ των αρχικών (αναγνώριση θέσεων συναρμογής) και καταλυτικών σταδίων (υποσύμπλοκα B και B act) της συναρμογής βάσει των «γειτόνων» της στο δίκτυο αλληλεπιδράσεων του σωματίου συναρμογής.....	297
Εικ. 84: Πιθανός βιολογικός ρόλος της FRA10AC1 ως πρωτεΐνη-προσαρμογέας μεταξύ των αρχικών (αναγνώριση θέσεων συναρμογής) και καταλυτικών σταδίων (υποσύμπλοκα C και P) της συναρμογής βάσει των «γειτόνων» της στο δίκτυο αλληλεπιδράσεων του σωματίου συναρμογής.....	298
Εικ. 85: Πιθανός βιολογικός ρόλος της FRA10AC1 ως πρωτεΐνη-προσαρμογέας μεταξύ της συναρμογής και της εξόδου του mRNA από τον πυρήνα ή της αποικοδόμησης μη κατάλληλων mRNAs.....	300

## Κατάλογος Πινάκων

Πιν. 1: Στατιστικά στοιχεία της βάσης δεδομένων BioGRID.....	73
Πιν. 2: Εκκινητές για την ενίσχυση του cDNA του γονιδίου FRA10AC1 και DGCR14.....	100
Πιν. 3: Αλληλουχίες εκκινητών.....	107
Πιν. 4: Οι πρώτοι «γειτόνες» της πρωτεΐνης FRA10AC1 και οι αλληλεπιδράσεις μεταξύ αυτών.....	115
Πιν. 5: Έμμεσες αλληλεπιδράσεις της πρωτεΐνης FRA10AC1.....	116
Πιν. 6: Δεκαπέντε από τα αλληλεπιδρώντα μόρια της πρωτεΐνης FRA10AC1 αποτελούν συστατικά του σωματίου συναρμογής.....	121
Πιν. 7: Θερμικός πίνακας (heat-map) της αφθονίας της FRA10AC1 και των πρώτων «γειτόνων» της στα υποσύμπλοκα A, B, B act και C.....	122
Πιν. 8: Ερευνητικές εργασίες απομόνωσης και ταυτοποίησης των συστατικών είτε ολόκληρου του σωματίου συναρμογής ή των επιμέρους υποσυμπλόκων του.....	129
Πιν. 9: Θερμικός πίνακας (heat-map) των πρωτεϊνών με καθολική παρουσία στα υποσύμπλοκα A, B, B act και C του σωματίου συναρμογής.....	145
Πιν. 10: Συγκριτική συνεισφορά των βάσεων δεδομένων PPIs στον άνθρωπο.....	153
Πιν. 11: Ομαδοποίηση των πρωτεϊνών-κόμβων του δικτύου σε οντολογικές κατηγορίες σε σύγκριση με το σύνολο δεδομένων του σωματίου συναρμογής.....	157
Πιν. 12: Κατηγορίες οντολογίας των πολυλειτουργικών πρωτεϊνών-κόμβων σε σύγκριση με αυτές του συνόλου δεδομένων του σωματίου συναρμογής.....	158
Πιν. 13: Ομαδοποίηση των πρωτεϊνών σε GO του δικτύου PPI σε σύγκριση με το σύνολο δεδομένων του σωματίου συναρμογής.....	162
Πιν. 14: Τιμές απλών τοπολογικών παραμέτρων του δικτύου PPI του σωματίου συναρμογής.....	173
Πιν. 15: Οι δέκα πρωτεΐνες του δικτύου του σωματίου συναρμογής με τον υψηλότερο αριθμό συνδέσεων.....	176
Πιν. 16: Τιμές απλών τοπολογικών παραμέτρων του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής έπειτα από την απομάκρυνση των πρωτεϊνών-«hubs» CSNK2A1 και HNRNPK με τις περισσότερες συνδέσεις και της πρωτεΐνης FRA10AC1.....	179
Πιν. 17: Τιμές απλών τοπολογικών παραμέτρων των δικτύων πρωτεϊνικών αλληλεπιδράσεων των υποσυμπλόκων B act και C πριν και μετά την απομάκρυνση της FRA10AC1.....	180
Πιν. 18: Αποτελεσματικότητα της ανάλυσης FRET.....	195
Πιν. 19: Ομοιότητες και διαφορές των τεσσάρων πρωτοκόλλων συλλογής κυττάρων.....	198
Πιν. 20: Πειραματικές συνθήκες των πρωτοκόλλων 1 και 2.....	199
Πιν. 21: Πειραματικές συνθήκες των πρωτοκόλλων 3A και 3B.....	199
Πιν. 22: Ποιοτικός και ποσοτικός έλεγχος του RNA των δειγμάτων που προορίζονταν για μεταγραφωμική ανάλυση.....	212

Πιν. 23: Μετρήσεις πρωτεϊνικού περιεχομένου των 2 δειγμάτων των κλώνων shRNA(4)_4, shRNA(3)_5, shRNA(3)_5 και MSH 2.....	213
Πιν. 24: Ποσοτικοποίηση του RNA των κλώνων #3, #4 και #5, μετά από την επεξεργασία με DNAση.....	224

# **1. ΕΙΣΑΓΩΓΗ**

## **1.1. Το γονίδιο *FRA10AC1* του ανθρώπου**

### **1.1.1. Κλωνοποίηση του γονιδίου *FRA10AC1***

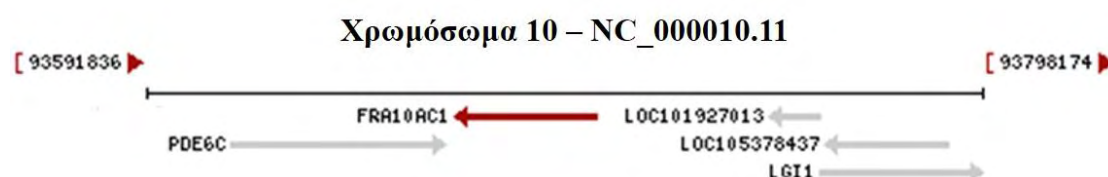
Το γονίδιο *FRA10AC1* εντοπίζεται στην σπάνια εύθραυστη χρωμοσωματική θέση *FRA10A* η οποία επάγεται απουσία φυλλικού οξέος στο θρεπτικό μέσο καλλιέργειας των κυττάρων και χαρτογραφείται στην περιοχή 10q23.3 του ανθρώπου (Sarafidou et al. 2004). Οι φορείς της *FRA10A* εμφανίζουν νοητική και αναπτυξιακή υστέρηση, όπως δυσμορφικά χαρακτηριστικά, μικρό ανάστημα, υποσπαδία και δυσχέρεια λόγου (Manrou et al. 1991, Petit et al. 1986, Sarafidou et al. 2004). Οι εύθραυστες χρωμοσωματικές θέσεις είναι χρωμοσωματικές περιοχές οι οποίες εμφανίζουν «χάσματα» ή θραύσεις κατά τη μετάφαση της μίτωσης, όταν τα κύτταρα καλλιεργούνται υπό συγκεκριμένες συνθήκες και διακρίνονται σε σπάνιες και συνήθεις, όπως θα αναλυθεί περαιτέρω στην Ενότητα 1.2. (Lukusa & Fryns 2008).

Η ταυτοποίηση του γονιδίου *FRA10AC1* έγινε βάσει της χρωμοσωματικής του θέσης λαμβάνοντας υπόψη ότι η μοριακή βάση της κυτταρογενετικής εμφάνισης των σπάνιων εύθραυστων θέσεων που είναι ευαίσθητες στο φυλλικό οξύ είναι η επέκταση τρινουκλεοτιδικών επαναλήψεων του τύπου (CGG)<sub>n</sub>. Ειδικότερα, με *in silico* διερεύνηση της χρωμοσωματικής περιοχής 10q23.3 εντοπίστηκαν τέτοιου τύπου επαναλήψεις στην 5' μη μεταφραζόμενη περιοχή του γονιδίου (Sarafidou et al. 2004). Ο πολυμορφισμός του αριθμού των επαναλήψεων δείχθηκε με γονοτύπηση 81 ατόμων του γενικού πληθυσμού. Έτσι, βρέθηκαν τέσσερα διαφορετικά αλληλόμορφα στα οποία ο αριθμός των επαναλήψεων κυμαίνεται από οκτώ έως δεκατέσσερις. Το πιο κοινό αλληλόμορφο είναι αυτό που φέρει εννιά επαναλήψεις CGG (Sarafidou et al. 2004). Η επιβεβαίωση της συσχέτισης της δραματικής επέκτασης των συγκεκριμένων επαναλήψεων με την κυτταρογενετική έκφραση της εύθραυστης θέσης *FRA10A* πραγματοποιήθηκε με αντίστοιχη ανάλυση στους φορείς της *FRA10A*. Έτσι βρέθηκε ότι στους φορείς οι επαναλήψεις επεκτείνονται κατά τουλάχιστον ~200 φορές. Επιπλέον, με ανάλυση μεθυλίωσης μέσω της χρήσης περιοριστικών ενδονουκλεασών, ευαίσθητων ή μη στη μεθυλίωση, και ακόλουθης υβριδοποίησης με ειδικούς ανιχνευτές (Sarafidou et al. 2004) ταυτοποιήθηκε ότι η αντίστοιχη περιοχή είναι υπερμεθυλιωμένη. Επίσης, η συνέπεια αυτής της υπερμεθυλίωσης είναι η μεταγραφική καταστολή του αντίστοιχου αλληλομόρφου. Η καταστολή αποδείχθηκε μέσω αντίστροφης μεταγραφής συζευγμένης με PCR (Reverse Transcription-Polymerase Chain Reaction, RT-PCR) που στόχευε στα μετάγραφα του *FRA10AC1* σε άτομα ετερόζυγα για τρεις πολυμορφισμούς σε συνδυασμό με αλληλούχηση (Sarafidou et al. 2004).

### **1.1.2. Γονιδιωματική οργάνωση του γονιδίου *FRA10AC1***

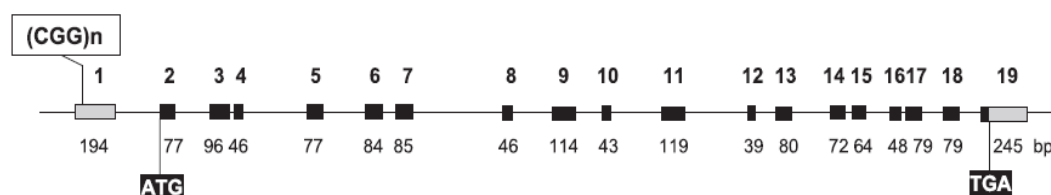
Το γονίδιο *FRA10AC1* εκτείνεται σε έκταση περίπου 35kb και εντοπίζεται ανάμεσα στο γονίδιο της φωσφοδιεστεράσης 6C (*PDE6C*) και στο γονίδιο *LGII* (Leucine Rich Glioma Inactivated 1) με κατεύθυνση μεταγραφής από το τελομερές προς το κεντρομερές (Sarafidou et al. 2004) (Εικ. 1). Το γονίδιο *PDE6C* κωδικοποιεί μία

υδρολάση κυκλικών νουκλεοτιδίων (cAMP, cGMP) που συμμετέχει στα σηματοδοτικά μονοπάτια ανίχνευσης του φωτός από τους φωτοϋποδοχείς στον αμφιβληστροειδή του ανθρώπου (Arshavsky 2002, Conti & Beavo 2007, Majumder et al. 2015). Μεταλλάξεις στο γονίδιο αυτό συνδέονται με διάφορες αυτοσωματικές επικρατείς και υπολειπόμενες ασθένειες (συγγενείς αχρωματοψίες, δυστροφία κωνίων και ατροφία ωχράς κηλίδας) (OMIM #600827, Katagiri et al. 2015). Το γονίδιο *LGII* συμμετέχει στην ωρίμανση των διεγερτικών συνάψεων (Lovero et al. 2015) και μεταλλάξεις του προκαλούν μία αυτοσωματική επικρατή μορφή επιληψίας του κροταφικού λοβού (Morante-Redolat et al. 2002).



**Εικ. 1:** Χαρτογράφηση του γονιδίου *FRA10A1* στην χρωμοσωματική περιοχή 10q23.3 του ανθρώπου. Με κόκκινο βέλος απεικονίζεται το γονίδιο *FRA10A1* και, εκατέρωθεν, με γκρι βέλη τα γονίδια *LGII* και *PDE6C*. Επίσης, φαίνεται ο κωδικός και η αρίθμηση της αντίστοιχης νουκλεοτιδικής αλληλουχίας. Ο προσανατολισμός της μεταγραφής συμβολίζεται με τη φορά του βέλους. LOC: μη χαρακτηρισμένοι γενετικοί τόποι (<https://www.ncbi.nlm.nih.gov/gene/118924>).

Το γονίδιο *FRA10A1* περιλαμβάνει δεκαεννέα εξόνια (Εικ. 2), πέντε από τα οποία (13, 14, 15, 16 και 17) υπόκεινται σε εναλλακτική συναρμογή (splicing). Όλα τα όρια των ιντρονίων περιέχουν τις συντηρημένες αλληλουχίες AG/GT στις δυο θέσεις συναρμογής (5' και 3' του ιντρονίου). Οι τρινουκλεοτιδικές επαναλήψεις (CGG)<sub>n</sub> εντοπίζονται στην 5' μη μεταφραζόμενη περιοχή του στο εξόνιο 1 (Εικ. 2) και αποτελούν μέρος των νησίδων CpG του γονιδίου (CpG island). Το cDNA πλήρους μήκους του γονιδίου *FRA10A1* έχει μήκος 1344bp. Το κωδικόνιο έναρξης της μετάφρασης βρίσκεται στην αρχή του εξονίου 2 και το κωδικόνιο λήξης στο τελευταίο εξόνιο (νουκλεοτίδιο 195 και 1140 του cDNA NM\_145246.4, αντίστοιχα) (Εικ. 2) (Sarafidou et al. 2004).

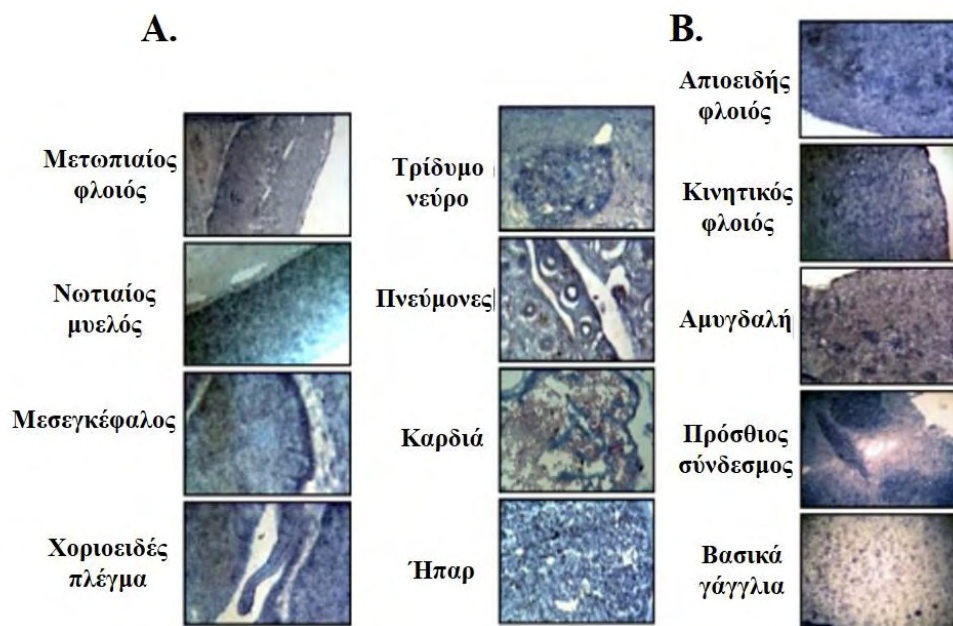


**Εικ. 2:** Γονιδιωματική οργάνωση του γονιδίου *FRA10A1* του ανθρώπου. Φαίνονται η αρίθμηση και το μέγεθος των εξονίων (μαύρα πλαίσια) καθώς και τα ιντρόνια (λεπτές μαύρες γραμμές). Με γκρι πλαίσιο απεικονίζονται οι 5' και 3' μη μεταφραζόμενες περιοχές. Επισημαίνονται οι τρινουκλεοτιδικές επαναλήψεις (CGG)<sub>n</sub>, το κωδικόνιο έναρξης (ATG) και λήξης (TGA) της μετάφρασης (Sarafidou et al. 2004).

### 1.1.3. Πρότυπο έκφρασης και εναλλακτικά μετάγραφα του γονιδίου *FRA10AC1*

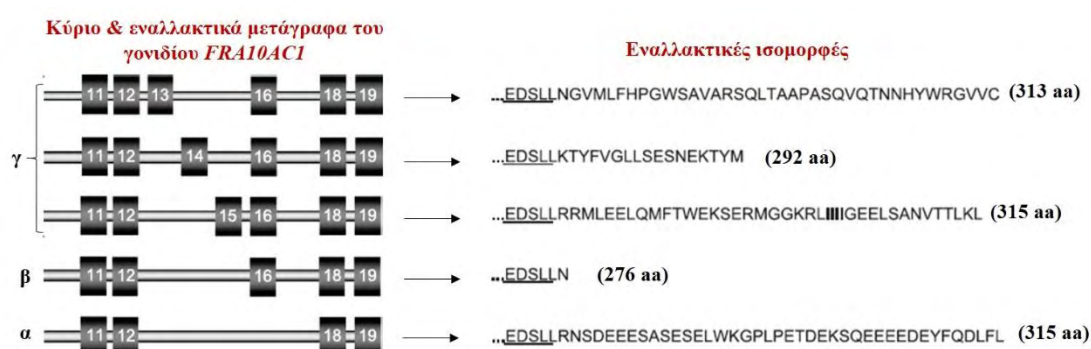
Η διερεύνηση του προτύπου έκφρασης του γονιδίου *FRA10AC1* έχει πραγματοποιηθεί με εναλλακτικές πειραματικές διαδικασίες: α) με ανάλυση Northern, κατά την οποία χρησιμοποιήθηκαν δυο διαφορετικοί ανιχνευτές που αναγνωρίζουν 616 νουκλεοτίδια της 5' κωδικής περιοχής του γονιδίου και 129 νουκλεοτίδια του εξονίου 19 και της 3' αμετάφραστης περιοχής β) με RT-PCR χρησιμοποιώντας εκκινητές που υβριδοποιούνται στα εξόνια 11 και 19 και γ) με RNA *in situ* υβριδοποίηση σε τομές εμβρύων ποντικού και τομές εγκεφάλου ενηλίκων ποντικών με τη χρήση σημασμένου ανιχνευτή RNA ο οποίος αντιστοιχεί στα νουκλεοτίδια 250-743 του cDNA του γονιδίου *Fra10ac1* του ποντικού (BN000292) (Sarafidou et al. υπό συγγραφή, Sarafidou et al. 2004).

Έτσι, έχει δειχθεί ότι το γονίδιο *FRA10AC1* είναι μεταγραφικά ενεργό σε όλους τους ιστούς ενήλικων ατόμων παρουσιάζοντας υψηλότερα επίπεδα έκφρασης σε όργανα με υψηλή μεταγραφική δραστηριότητα όπως είναι ο εγκέφαλος, η καρδιά, οι σκελετικοί μύες και το ήπαρ. Αντίθετα, χαμηλότερα επίπεδα έκφρασης εμφανίζει στο θύμο αδένα, στο σπλήνα, στο λεπτό και στο παχύ έντερο, στον πλακούντα, στον πνεύμονα, στις γονάδες και πολύ χαμηλά επίπεδα έκφρασης στον προστάτη και στα λευκοκύτταρα (Sarafidou et al. 2004). Με βάση την ανάλυση σε ποντικό φαίνεται ότι το γονίδιο *Fra10ac1* εκφράζεται σε συγκεκριμένες περιοχές του αναπτυσσόμενου νευρικού συστήματος (π.χ. μετωπιαίος φλοιός, νωτιαίος μυελός), στους πνεύμονες, στην καρδιά και στο ήπαρ (Εικ. 3). Συγκεκριμένα, στα ενήλικα ποντίκια εμφανίζει υψηλά επίπεδα έκφρασης σε συγκεκριμένα μέρη του νευρικού συστήματος (απιοειδής φλοιός, κινητικός φλοιός, αμυγδαλή, πρόσθιος σύνδεσμος δηλαδή νευρικά ινίδια που ενώνουν τους δυο κροταφικούς λοβούς των ημισφαιρίων του εγκεφάλου και σε βασικά γάγγλια) (Εικ. 3) (Sarafidou et al. υπό συγγραφή).



**Εικ. 3: Πρότυπο έκφρασης του γονιδίου *Fra10ac1* του ποντικού.** (A) τομές εμβρύων ποντικού και (B) τομές εγκεφάλου ενήλικων ποντικών. Το σήμα φαίνεται με σκούρο μπλε χρώμα (Sarafidou et al. υπό συγγραφή).

Μέσω της RT-PCR ταυτοποιήθηκαν, εκτός από το κύριο μετάγραφο, και τέσσερα εναλλακτικά μετάγραφα (Sarafidou et al. 2004). Με βάση την αλληλούχησή τους προέκυψε ότι το κύριο μετάγραφο περιλαμβάνει τα εξόνια 1-12 και 18-19, ενώ τα εναλλακτικά μετάγραφα περιλαμβάνουν συνδυασμούς των εξονίων 13-16 (Εικ. 4). Ειδικότερα, αποκλειστικά στις γονάδες, ταυτοποιήθηκαν τρία εναλλακτικά μετάγραφα τα οποία περιλαμβάνουν το εξόνιο 16 σε συνδυασμό με τα εξόνια 13, 14 και 15 αντίστοιχα (Εικ. 4) (Sarafidou et al. 2004). Η σύγκριση των διαφορετικών αμινοξικών ακολουθιών που προκύπτουν από τα πέντε παραπάνω μετάγραφα υποδεικνύει ότι οι πιθανές διαφορετικές ισομορφές εμφανίζουν ετερογένεια στο καρβοξυ-τελικό άκρο (Εικ. 4) (Sarafidou et al. 2004). Το εξόνιο 17, που αναφέρθηκε προηγουμένως, δεν ταυτοποιήθηκε πειραματικά αλλά η ύπαρξή του προβλέπεται από το EST GKBABE07 το οποίο περιλαμβάνει το εξόνιο 16, 17, 18 και 189 νουκλεοτίδια από το εξόνιο 19.



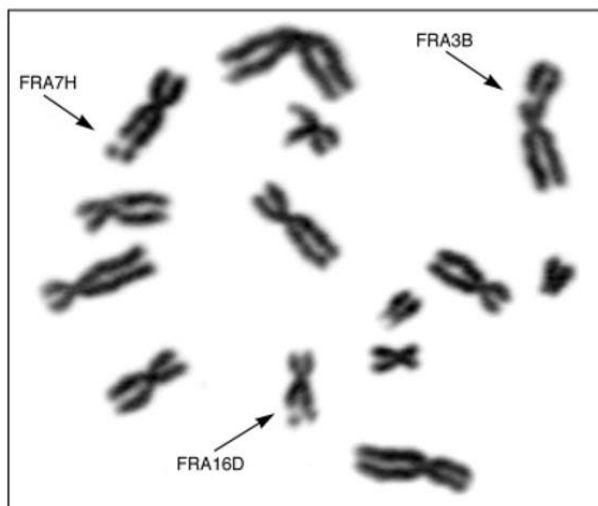
**Εικ. 4: Εναλλακτικά μετάγραφα του γονιδίου *FRA10AC1* και αντίστοιχες πρωτεϊνικές ισομορφές.** Φαίνονται τα εναλλακτικά εξόνια στο κύριο μετάγραφο του γονιδίου *FRA10AC1* (α), στο εναλλακτικό μετάγραφο (β) και στα τρία εναλλακτικά μετάγραφα που ταυτοποιήθηκαν αποκλειστικά στις γονάδες (γ). Δίπλα από κάθε μετάγραφο φαίνεται η αντίστοιχη πρωτεϊνική ακολουθία και τα συνολικά αμινοξέα της προβλεπόμενης πρωτεΐνης, υπογραμμίζοντας τα αμινοξέα που κωδικοποιούνται από το εξόνιο 12 και παραμένουν ίδια σε όλες τις ισομορφές (Sarafidou et al. 2004).

Όπως ήδη αναφέρθηκε, η επέκταση των πολυμορφικών επαναλήψεων στην 5' μη μεταφραζόμενη περιοχή του γονιδίου *FRA10AC1*, είναι η αιτία της κυτταρογενετικής έκφρασης της σπάνιας χρωμοσωματικής εύθραυστης θέσης *FRA10A*, τα χαρακτηριστικά της οποίας αναλύονται στις επόμενες ενότητες σε συνδυασμό με τα γενικά χαρακτηριστικά και των δυο κατηγοριών, συνήθων και σπάνιων εύθραυστων χρωμοσωματικών θέσεων.

## 1.2. Εύθραυστες χρωμοσωματικές θέσεις

Οι εύθραυστες χρωμοσωματικές θέσεις είναι κληρονομήσιμοι γενετικοί τόποι οι οποίοι εμφανίζουν χάσματα ή θραύσεις κατά τη μετάφαση της μίτωσης όταν τα κύτταρα καλλιεργούνται υπό ορισμένες συνθήκες, δηλαδή εκτίθενται σε συγκεκριμένους χημικούς παράγοντες που αναστέλλουν την αντιγραφή του DNA (Durkin & Glover 2007, Sutherland & Richards 1995) (Εικ. 5). Σε καλλιέργειες λεμφοκυττάρων που αναπτύσσονται στις παραπάνω συνθήκες έχει δείχθει ότι σε ποσοστό μεγαλύτερο του 80% των κυττάρων εμφανίζονται θραύσεις σε αυτές τις

θέσεις (Durkin & Glover 2007, Glover et al. 1984). Οι εύθραυστες θέσεις είναι εμπλουτισμένες σε γονίδια μικρο-RNA και χαρτογραφούνται σε υψηλά συντηρημένες θέσεις κατά την εξέλιξη των χρωμοσωμάτων των θηλαστικών. Εγγενές χαρακτηριστικό τους αποτελεί η γονιδιωματική αστάθεια (Ragland et al. 2008, Reshmi et al. 2011, Rossi et al. 2008, Ruiz-Herrera & Robinson 2007).



**Εικ. 5:** Απεικόνιση των συνήθων εύθραυστων χρωμοσωματικών θέσεων *FRA3B*, *FRA16D*, *FRA7H*. Φαίνονται οι εύθραυστες θέσεις στα χρωμοσώματα 3, 7 και 14 στη μετάφαση της μίτωσης (Sutherland et al. 1998).

Με βάση τη συχνότητά τους στον πληθυσμό διακρίνονται σε δυο κατηγορίες, στις σπάνιες και στις συνήθειες (Durkin & Glover 2007). Με βάση τον τρόπο επαγωγής της εμφάνισής τους, που γενικά στοχεύει στην μερική αναστολή της αντιγραφής του DNA, διακρίνονται: α) στις ευαίσθητες στο φυλλικό οξύ, που αποτελούν την κύρια κατηγορία των σπάνιων εύθραυστων θέσεων και β) στις επαγόμενες από δισταμυκίνη (distamycin A) ή βρωμο-δεοξυ-ουριδίνη (BrdU) και αφιδικολίνη ή 5-αζακυτιδίνη. Η δεύτερη κατηγορία συναντάται κυρίως στις συνήθειες εύθραυστες θέσεις (Sutherland & Richards 1995). Όλες αυτές οι χημικές ουσίες αναστέλλουν μερικώς τη δράση της DNA πολυμεράσης ή μειώνουν τα επίπεδα των νουκλεοτιδίων θυμίνης και κυτοσίνης (dTTP, dCTP), εμποδίζοντας την επιμήκυνση της αντιγραφής στις περιοχές των εύθραυστων θέσεων χωρίς όμως να αναστέλλουν τον κυτταρικό κύκλο (Lukusa & Fryns 2008). Η καθυστέρηση στη διαδικασία της αντιγραφής σε αυτές τις θέσεις είναι δυνατόν να προκαλείται από συγκεκριμένα δομικά χαρακτηριστικά (π.χ. ασυνήθειες δευτεροταγείς δομές) με αποτέλεσμα να μην είναι δυνατή η συμπύκνωση της χρωματίνης στα συγκεκριμένα σημεία (Zlotorynski et al. 2003, Pichiorri et al. 2008). Έτσι, η χρωματίνη παραμένει σε αποσυσπειρωμένη μορφή, γεγονός το οποίο απεικονίζεται κυτταρογενετικά ως ασυνέχεια-εύθραυστη θέση σε συγκεκριμένα σημεία των χρωμοσωμάτων (Jiang et al. 2009).



### 1.2.1. Συνήθεις εύθραυστες θέσεις

Οι συνήθεις εύθραυστες θέσεις αποτελούν εγγενές χαρακτηριστικό των χρωμοσωμάτων όλων των ανθρώπων (Lukusa & Fryns 2008). Έως τώρα, έχουν ταυτοποιηθεί κυτταρογενετικά ~200 συνήθεις εύθραυστες θέσεις στο γονιδίωμα του ανθρώπου σε καλλιέργειες λεμφοκυττάρων (Debatisse et al. 2012). Οι είκοσι δυο από αυτές (που έχουν χαρακτηριστεί σε μοριακό επίπεδο) καταλαμβάνουν μεγάλες χρωμοσωματικές περιοχές (της κλίμακας των Mb), περιλαμβάνουν πολύ μεγάλου μήκους γονίδια (Smith et al. 2006, Mrasek et al. 2010), είναι πλούσιες σε αδενίνη και θυμίνη αλλά, σε αντίθεση με τις σπάνιες εύθραυστες θέσεις, δεν εμφανίζουν δορυφορικές επαναλήψεις.

Όπως προαναφέρθηκε, η έκφρασή τους επάγεται υπό την επίδραση συγκεκριμένων χημικών παραγόντων που εμποδίζουν την επιμήκυνση της αντιγραφής χωρίς να διακόπτουν τον κυτταρικό κύκλο. Το γεγονός αυτό υποδεικνύει ότι αυτές οι θέσεις είναι επιρρεπείς στη μη ομαλή εξέλιξη της αντιγραφής διότι μπορεί να περιλαμβάνουν περιοχές που αποτελούν: i) ζώνες αργής εξέλιξης της αντιγραφής, στις οποίες ο αργός ρυθμός της αντιγραφής έχει δείχθει, στο σακχαρομύκητα, ότι προκαλεί θραύσεις στα χρωμοσώματα (Cha & Kleckner 2002). Οι θραύσεις αυτές έχουν προσομοιαστεί με τις συνήθεις εύθραυστες θέσεις των κυττάρων των θηλαστικών στις οποίες έχει παρατηρηθεί αργός ρυθμός της αντιγραφής ή καθυστέρηση της αντιγραφής (Pelliccia et al. 2008, Laird et al. 1993) ii) περιοχές στις οποίες οι μονάδες της μεταγραφής (RNA πολυμεράσες) συναντούν τις μονάδες της αντιγραφής, όπως στην περίπτωση μεγάλων γονιδίων τα οποία μεταγράφονται με αργό ρυθμό, με αποτέλεσμα η μεταγραφή να γίνεται ταυτόχρονα με την αντιγραφή και εμποδίζει την εξέλιξη της τελευταίας (Helmrich et al. 2011) iii) θέσεις με συγκεκριμένα δομικά χαρακτηριστικά (περιοχές πλούσιες σε A-T) τα οποία σχηματίζουν ασυνήθεις δευτεροταγείς δομές (Franchitto 2013, Branzei & Foiani 2010, Cha & Kleckner 2002). Οι δευτεροταγείς αυτές δομές συχνά αποτελούν σημεία χρωμοσωματικών μετατοπίσεων και ελλειμμάτων καθώς και ανταλλαγών μεταξύ των αδελφών χρωματιδίων, χωρίς να υπάρχουν μέχρι τώρα κάποιες ενδείξεις για τη συσχέτιση της δομής αυτών των περιοχών με την ευθραυστότητα τους (Durkin & Glover 2007).

Διάφορες μελέτες έχουν αναφέρει συσχέτιση των συνήθων εύθραυστων θέσεων με τα αρχικά στάδια της καρκινογένεσης και με τη γήρανση. Ειδικότερα, σε μία μελέτη, παρατηρήθηκε, σε ιστούς ποντικών που εμφανίζουν υπερπλασία (αρχικά στάδια της καρκινογένεσης), ότι η έλλειψη ετεροζυγωτίας για ένα σύνολο πολυμορφισμών εμφανίζεται σε μεγαλύτερο ποσοστό στις συνήθεις εύθραυστες θέσεις σε σχέση με τις υπόλοιπες (Tsantoulis et al. 2008). Επιπλέον, στην ίδια εργασία, η καρκινική κυτταρική σειρά *U2OS* (επιθηλιακά κύτταρα οστεοσαρκώματος) που υπερεκφράζει ένα γονίδιο το οποίο συμμετέχει στην επαγωγή των επιδιορθωτικών μηχανισμών των κυττάρων όταν εντοπίζονται δίκλωνες θραύσεις στο DNA, μία διαδικασία γνωστή ως «στρες αντιγραφής-επαγόμενο από ογκογονικό ερέθισμα» (oncogene-induced replication stress) εξετάστηκε, με χρήση μικροσυστοιχίας για ανάλυση πολυμορφισμών, ως προς τις αλλαγές που φέρει στη δομή των χρωμοσωμάτων (π.χ. έλλειμμα ενός τμήματος του χρωμοσώματος). Τα αποτελέσματα της ανάλυσης αυτής έδειξαν ότι το 40% των αλλαγών στη δομή των χρωμοσωμάτων αντιστοιχούν σε συνήθεις εύθραυστες θέσεις

(Tsantoulis et al. 2008). Μια δεύτερη μελέτη ανάλυσης πολυμορφισμών σε τρεις εύθραυστες θέσεις (*FRA3B*, *FRA16D*, *FRA6E*) έγινε σε φυσιολογικά κύτταρα θηλαστικών και σε φυσιολογικά κύτταρα που επάγεται η γήρανσή τους έπειτα από την έκφραση ενός ογκογονιδίου (oncogene-induced senescence), ως συνέπεια της ενεργοποίησης του μηχανισμού της απόκρισης στις βλάβες του DNA (DNA damage checkpoint response). Η ανάλυση αυτή έδειξε ότι η έλλειψη ετεροζυγωτίας είναι στατιστικά σημαντικά αυξημένη στις δυο από τις τρεις εύθραυστες θέσεις (*FRA16D*, *FRA6E*) στα κύτταρα που επάγεται η γήρανσή τους (Di Micco et al. 2006).

### 1.2.2. Σπάνιες εύθραυστες θέσεις

Σε αντίθεση με τις συνήθεις, οι σπάνιες εύθραυστες θέσεις εμφανίζονται σε ένα μικρό μόνο ποσοστό του πληθυσμού (<5%). Έως τώρα έχουν περιγραφεί κυτταρογενετικά 31 σπάνιες εύθραυστες θέσεις, η πλειοψηφία των οποίων (24 από τις 31) εκδηλώνονται όταν τα κύτταρα καλλιεργηθούν απουσία φυλλικού οξέος (Lukusa & Fryns 2008). Έως τώρα έχουν χαρακτηριστεί σε μοριακό επίπεδο έντεκα σπάνιες εύθραυστες θέσεις, εννιά από τις οποίες είναι ευαίσθητες στο φυλλικό οξύ, οι: *FRA1A* (Kremer et al. 1991, Verkerk et al. 1991), *FRA1B* (Knight et al. 1993, Géczy et al. 1997), *FRA1C* (Parrish et al. 1994, Shaw et al. 2002), *FRA16A* (Nancarrow et al., 1994), *FRA10A* (Sarafidou et al., 2004), *FRA11B* (Jones et al. 1995, Jones et al. 1994), *FRA11A* (Debacker et al. 2007), *FRA2A* (Metsu et al. 2014) και *FRA12A* (Winnepenninckx et al. 2007) και δύο που εκδηλώνονται παρουσία δισταμυκίνης ή/και βρωμο-δεοξυ-ουριδίνης [*FRA10B* (Hewett et al. 1998) και *FRA16B* (Yu et al. 1997)].

Η μοριακή βάση όλων των θέσεων που είναι ευαίσθητες στο φυλλικό οξύ είναι η επέκταση πολυμορφικών τρινουκλεοτιδικών επαναλήψεων του τύπου (CGG)<sub>n</sub> σε περιοχές νησίδων CpG. Από τις εννιά θέσεις αυτής της κατηγορίας που έχουν χαρακτηριστεί, οι οκτώ (*FRA1A*, *FRA1B*, *FRA1C*, *FRA10A*, *FRA2A*, *FRA11A*, *FRA11B* και *FRA12A*) εντοπίζονται εντός της 5' μη μεταφραζόμενης περιοχής των γονιδίων *FMR1*, *FMR2*, *FAM11A*, *FRA10AC1*, *AFF3*, *C11orf80*, *CBL2* και *DIP2B*, αντίστοιχα. Ο αριθμός των επαναλήψεων στα φυσιολογικά αλληλόμορφα διαφέρει μεταξύ των εύθραυστων θέσεων, για παράδειγμα υπάρχουν 4-55 αντίγραφα στη θέση *FRA1A*, 6-25 αντίγραφα στη θέση *FRA1B*, 3-20 αντίγραφα στη θέση *FRA2A* και 8-13 αντίγραφα στη θέση *FRA10A* (Lukusa & Fryns 2008, Metsu et al. 2014).

Οι επαναλαμβανόμενες αυτές αλληλουχίες του τύπου (CGG)<sub>n</sub> μπορεί να σχηματίσουν σταθερές δευτεροταγείς δομές, όπως ενδοκλωνικές δομές φουρκέτας (intrastrand hairpins) ή τετραπλή έλικα (quadruplex DNA structures). Εξαιτίας αυτών μπορεί να προκληθεί αποδιάταξη των αλυσίδων DNA στο σημείο που βρίσκονται και να σχηματιστούν δεσμοί υδρογόνου μεταξύ των επαναλαμβανόμενων αλληλουχιών είτε στην ίδια αλυσίδα DNA (slipped strand structure) ή μεταξύ μη ομόλογων επαναλήψεων των δυο αλυσίδων DNA (folded slipped structure) (Sinden 1999, Lukusa & Fryns 2008). Οι δομές αυτές έχει δείχθει ότι μπορεί να διαταράξουν το στάδιο της επιμήκυνσης της αντιγραφής του DNA (Lukusa & Fryns 2008). Η ταυτοποίηση των παραπάνω δομών έγινε *in vitro* με τη χρήση πλασμιδιακών κατασκευών που περιείχαν διαφορετικό αριθμό τρινουκλεοτιδικών επαναλήψεων. Ειδικότερα πραγματοποιήθηκε η γραμμοποίηση των κατασκευών αυτών με κατάλληλα ένζυμα περιορισμού, η

ραδιοσήμανση των άκρων τους, η αποδιάταξη και η επανένωση των δυο αλυσίδων DNA, έτσι ώστε να ελεγχθεί ο σχηματισμός δευτεροταγών δομών από τις τρινουκλεοτιδικές επαναλήψεις μέσω της εξέτασης της κινητικότητας των διαφορετικών μορίων DNA σε πηκτική ακρυλαμίδης (Pearson & Sinden 1996).

Επίσης, οι περιοχές με αυτές τις τρινουκλεοτιδικές επαναλήψεις είναι πιο υπερελικωμένες από την κανονική μορφή του DNA (B-DNA). Έτσι, μπορεί να διαταράσσουν τη διαδικασία της αντιγραφής καθώς αποτελούν σημεία συσσώρευσης υπερελικωμένου DNA που εμποδίζουν την πρόοδο της διχάλας της αντιγραφής. Αν και παρατηρείται παύση της διαδικασίας της αντιγραφής σε αυτό το σημείο, αυτή είναι παροδική και, κατά την επανεκκίνηση της διαδικασίας, μπορεί να λάβει χώρα αύξηση του αριθμού των αντιγράφων των επαναλήψεων λόγω ολίσθησης της πολυμεράσης. Επιπλέον, έχει δειχθεί ότι η επιμήκυνση των επαναλαμβανόμενων αυτών περιοχών διαταράσσει τη συναρμολόγηση των νουκλεοσωμάτων *in vitro*, γεγονός που ενισχύεται από τη μεθυλίωση αυτών των επαναλήψεων (Lukusa & Fryns 2008, Wang & Griffith 1996). Η αποτυχία συμπύκνωσης της χρωματίνης σε αυτά τα σημεία πιθανώς είναι υπεύθυνη για την εμφάνιση των εύθραυστων θέσεων ως «χάσματα» στα χρωμοσώματα. Τέλος, στις επτά από τις οκτώ εύθραυστες θέσεις (εκτός από την *FRA11B*) που εντοπίζονται εντός γονιδίου, η επέκταση των επαναλήψεων τύπου (CGG)<sub>n</sub> έχει ως αποτέλεσμα την υπερμεθυλίωση των γειτονικών νησίδων CpG και κατ' επέκταση τη μεταγραφική καταστολή του αντίστοιχου αλληλομόρφου (Sutherland & Richards 1995, Gécz et al. 1997, Shaw et al. 2002, Sarafidou et al. 2004, Debacker et al. 2007, Metsu et al. 2014, Winnepenninckx et al. 2007).

Η μοριακή ανάλυση των σπάνιων εύθραυστων θέσεων *FRA10B* και *FRA16B* που δεν είναι ευαίσθητες στο φυλλικό οξύ έδειξε ότι σχετίζονται με την επέκταση, έως και αρκετών κιλοβάσεων, μινι-δορυφορικών επαναλήψεων πλούσιων σε AT οι οποίες πιθανώς σχηματίζουν δευτεροταγείς δομές που επίσης αναστέλλουν την αντιγραφή του DNA (Lukusa & Fryns 2008).

### 1.2.3. Κλινική σημασία των εύθραυστων θέσεων

Οι συνήθεις εύθραυστες θέσεις χαρακτηρίζονται από γονιδιωματική αστάθεια, η οποία αυξάνεται σε αυτές τις περιοχές στα καρκινικά κύτταρα (Tsantoulis et al. 2008, Bartkova et al. 2006). Λαμβάνοντας υπόψη το γεγονός ότι εμφανίζονται σε όλους τους ανθρώπους και την πιθανότητα ότι μπορεί να υπάρχουν ομοζυγώτες για αυτές τις θέσεις, υποδεικνύεται ότι δεν έχουν άμεση κλινική σημασία.

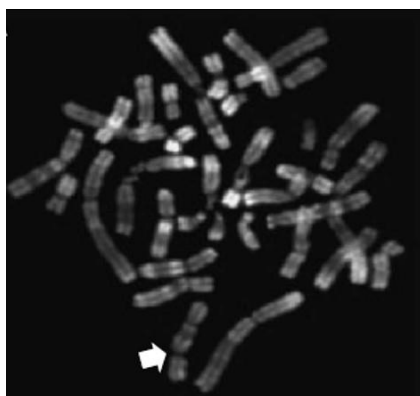
Όσον αφορά τις σπάνιες εύθραυστες θέσεις, μέχρι τώρα, δυο ευαίσθητες στο φυλλικό οξύ έχουν συσχετιστεί με παθολογικό φαινότυπο. Η *FRAXA* και η *FRAXE*, που χαρτογραφούνται στο χρωμόσωμα X, έχουν συσχετιστεί κατ' αντιστοιχία με το σύνδρομο του ευθραύστου X (OMIM # 309550), την πιο συχνή αιτία νοητικής υστέρησης μετά το σύνδρομο Down και με μια πιο σπάνια μορφή νοητικής υστέρησης (Bagni et al. 2012, Kremer et al. 1991, Verkerk et al. 1991, Pieretti et al. 1991, Sofola et al. 2007). Μία τρίτη εύθραυστη θέση, η *FRA11B*, στην οποία χαρτογραφείται το πρωτο-ογκογονίδιο *CBL2*, βρίσκεται στο μεγάλο βραχίονα του χρωμοσώματος 11 στη ζώνη 11q23. Σε ορισμένες περιπτώσεις ασθενών με σύνδρομο Jacobsen (OMIM #147791), που χαρακτηρίζεται από αναπτυξιακή υστέρηση, έχουν βρεθεί ελλείμματα

της χρωμοσωματικής περιοχής κοντά στην εύθραυστη θέση *FRA11B* και εντός αυτής (Penny et al. 1995, Michaelis et al. 1998). Ειδικότερα, σε μία μελέτη, σε έναν ασθενή με το σύνδρομο αυτό, βρέθηκε ότι το αντίστοιχο έλλειμμα εκτείνεται σε μια περιοχή 100kb, που περιλαμβάνει την 5' μη μεταφραζόμενη περιοχή του γονιδίου *CBL2* στην οποία εντοπίζονται επαναλήψεις τύπου (CCG)<sub>n</sub> (Jones et al. 1994). Ωστόσο, δεν έχει δειχθεί σαφής συσχέτιση του ετερόζυγου ελλείμματος του πρωτο-ογκογονιδίου *CBL2* με το σύνδρομο αυτό (Jones et al. 2000). Για παράδειγμα, σε μια άλλη μελέτη, έχει δειχθεί ότι 2 από τους 5 ασθενείς που εξετάστηκαν φέρουν επαναλήψεις τύπου (CCG)<sub>n</sub>, ωστόσο, το έλλειμμα δεν εντοπίζεται σε αυτές τις επαναλήψεις αλλά εκτείνεται σε μια περιοχή 20Mb μακριά από την θέση *FRA11B* (Jones et al. 1995).

Για τις υπόλοιπες σπάνιες εύθραυστες θέσεις, υπάρχουν διάφορες μελέτες που αναφέρουν αυξημένη συχνότητα εμφάνισής τους σε άτομα με διάφορες διαταραχές, κυρίως νοητική υστέρηση, χωρίς όμως να υπάρχει έως τώρα σαφής συσχέτιση γονοτύπου-φαινοτύπου για κάποια από αυτές (Debacker & Kooy 2007). Μάλιστα, τρεις σπάνιες αυτοσωματικές εύθραυστες θέσεις, οι *FRA10B*, *FRA16B*, *FRA17A*, έχουν βρεθεί σε ομοζυγωτία σε άτομα χωρίς παθολογικό φαινότυπο, γεγονός που υποδεικνύει ότι οι συγκεκριμένες εύθραυστες θέσεις δεν επηρεάζουν τη λειτουργία απαραίτητων γονιδίων (Sutherland & Baker 1998). Παρ' όλα αυτά, ορισμένες σπάνιες, ευαίσθητες στο φυλλικό οξύ, αυτοσωματικές εύθραυστες θέσεις δεν έχουν εντοπιστεί σε ομοζυγωτία, αν και, με βάση τη συχνότητα των ετεροζυγωτών, τέτοια άτομα θα ήταν αναμενόμενο να ταυτοποιηθούν. Το γεγονός αυτό πιθανώς υποδηλώνει ότι μία τέτοια κατάσταση δεν είναι συμβατή με τη ζωή (Sutherland & Baker 1998).

#### 1.2.4. Συσχετισμός εύθραυστης θέσης *FRA10A* και γονιδίου *FRA10AC1* με νοσήματα

Η *FRA10A* αποτελεί την πιο συχνά εμφανιζόμενη σπάνια εύθραυστη θέση στο γονιδίωμα του ανθρώπου με συχνότητα που εκτιμάται στο 1/500, σε σύνολο 30.000 περίπου ατόμων που προέρχονται από διαφορετικές μελέτες και πληθυσμιακές ομάδες (Kähkönen et al. 1989) (Εικ. 6). Ωστόσο, η παρατηρούμενη συχνότητα ποικίλλει σημαντικά μεταξύ των διαφορετικών μελετών από 0% (Takahashi et al. 1988) έως 2% (Petit et al. 1986), γεγονός που μπορεί να οφείλεται στην πληθυσμιακή προέλευση των ατόμων. Στις περιπτώσεις που εξετάστηκε η γονεϊκή προέλευση της *FRA10A*, διαπιστώθηκε ότι η κληρονόμησή της ήταν μητρικής προέλευσης (Sutherland 1982, Sutherland 1985).



**Εικ. 6:** Κυτταρογενετική έκφραση της σπάνιας εύθραυστης χρωμοσωματικής θέσης *FRA10A*. Με βέλος απεικονίζεται η εύθραυστη θέση *FRA10A* (Sarafidou et al. 2004).

Έως τώρα, ποικίλες μελέτες δείχνουν ότι η έκφραση της *FRA10A*, και κατ' επέκταση η απλοανεπάρκεια του γονιδίου *FRA10AC1*, μπορεί να σχετίζεται με νοητική υστέρηση, νευρολογικές διαταραχές και αναπτυξιακές ανωμαλίες, όπως δυσμορφικά χαρακτηριστικά, μικρό ανάστημα, υποσπαδία και δυσχέρεια λόγου (Mavrou et al. 1991, Petit et al. 1986, Kähkönen et al. 1989, Sarafidou et al. 2004). Αυτός ο συσχετισμός προέκυψε κυρίως από μελέτες που ελέγξαν την παρουσία εύθραυστων θέσεων στον καρυότυπο ατόμων με νοητική υστέρηση σε σχέση με μη προσβεβλημένα άτομα. Σε ορισμένες από αυτές, η *FRA10A* ανιχνεύθηκε αποκλειστικά σε προσβεβλημένα άτομα (Sutherland 1985, Petit et al. 1986, Mavrou et al. 1991) ενώ σε άλλες εντοπίστηκε και σε ορισμένα υγιή άτομα-φορείς (Sutherland 1982, Kähkönen et al. 1989, Sarafidou et al. 2004). Αυτό μπορεί να υποδηλώνει ότι δεν υπάρχει συσχέτιση της ετερόζυγης κατάστασης της *FRA10A* με κάποιο παθολογικό φαινότυπο αλλά αυτή έχει αναφερθεί λόγω μεροληψίας κατά τη δειγματοληψία διότι στις περισσότερες μελέτες τα άτομα που εξετάστηκαν είχαν προέλθει κατά μεγάλο ποσοστό από ιδρύματα περίθαλψης ατόμων με νοητική υστέρηση, σχολεία ειδικής αγωγής και όχι από τυχαίο δείγμα του γενικού πληθυσμού. Εναλλακτικά, αυτό το γεγονός μπορεί να οφείλεται σε μειωμένη διεισδυτικότητα του παθολογικού φαινοτύπου έτσι ώστε τα κλινικά συμπτώματα να μην παρατηρούνται σε ένα ποσοστό ατόμων-φορέων. Αυτό το φαινόμενο της μειωμένης διεισδυτικότητας έχει δειχθεί στην περίπτωση του συνδρόμου του εύθραυστου X που σχετίζεται με την εμφάνιση της εύθραυστης χρωμοσωματικής θέσης *FRAXA* και την εκδήλωση νοητικής υστέρησης. Ειδικότερα, η διεισδυτικότητα του φαινοτύπου εκτιμήθηκε στο 80% στους άνδρες (Jin & Warren 2000). Επιπλέον, έχει βρεθεί ότι συγκεκριμένου στελέχους ποντίκια knock-out (*C57BL/6*) στα οποία έχει απενεργοποιηθεί το γονίδιο *FMRI*, δεν εμφάνιζαν παθολογικό φαινότυπο, ενώ ένα διαφορετικό στέλεχος knock-out ποντικών (*FVB-129*) εμφάνιζε τα χαρακτηριστικά συμπτώματα του συνδρόμου, γεγονός το οποίο πιθανώς εξηγείται λόγω διαφορετικού γενετικού υπόβαθρου των δυο στελεχών (Dobkin et al. 2000, Kazdoba et al. 2014).

Επίσης, πρόσφατα, στο πλαίσιο μιας μετα-ανάλυσης 1000 ατόμων από δύο μελέτες συσχέτισης πολυμορφισμών σε επίπεδο γονιδιοματικής κλίμακας και επιπέδων βιοδεικτών του εγκεφαλονωτιαίου υγρού σε ασθενείς με Alzheimer (Genome Wide Association Study, GWAS), ταυτοποιήθηκαν δυο πολυμορφισμοί του γονιδίου *FRA10AC1* οι οποίοι εμφάνισαν την υψηλότερη συσχέτιση με τα επίπεδα του βιοδείκτη  $A\beta_{1-42}$ . Ο βιοδείκτης αυτός αφορά τα επίπεδα του β-πεπτιδίου (Amyloid beta-peptides) το οποίο αποτελεί συστατικό των αναπτυσσόμενων αμυλοειδών πλακών που δημιουργούν μεγάλα νευροϊνδιακά συμπλέγματα με αποτέλεσμα την απώλεια των συνδέσεων των νευρικών κυττάρων στον εγκέφαλο των ασθενών με Alzheimer (Van Cauwenbergh et al. 2016). Οι πολυμορφισμοί του γονιδίου *FRA10AC1* εντοπίζονται στο ιντρόνιο 12 και στην περιοχή του υποκινητή του γονιδίου με αντίστοιχους κωδικούς rs10509663 και rs116953792, με τον πρώτο να εμφανίζει την σημαντικότερη συσχέτιση με τα επίπεδα του βιοδείκτη (Li et al. 2015). Οι δυο αυτοί πολυμορφισμοί

του γονιδίου *FRA10AC1* και ένας επιπλέον (rs1503351) ήταν οι μοναδικοί που πέρασαν το κατώφλι στατιστικής σημαντικότητας.

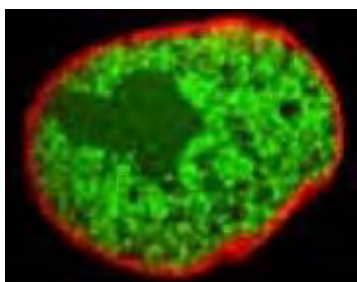
Σε μια άλλη πρόσφατη μελέτη GWAS, που περιλαμβάνει περίπου 10.000 άτομα, στην οποία εξετάστηκαν οι γενετικοί παράγοντες που συσχετίζονται με τον δείκτη αξιολόγησης απόφραξης των αεραγωγών των πνευμόνων FEV1/FVC, ταυτοποιήθηκε στατιστικά σημαντική συσχέτιση με τον πολυμορφισμό rs79105493 που εντοπίζεται ανοδικά του γονιδίου *FRA10AC1* (Lutz et al. 2015).

### 1.3. Χαρακτηριστικά της πρωτεΐνης FRA10AC1

Το κύριο και πιο άφθονο μετάγραφο του γονιδίου *FRA10AC1* κωδικοποιεί μια πρωτεΐνη 315 αμινοξέων, μοριακού βάρους 37,548 kD και ισοηλεκτρικού σημείου 8,25. Το εναλλακτικό μετάγραφο που βρέθηκε στην πλειοψηφία των ιστών παράγει ένα πρωτεϊνικό προϊόν 276 αμινοξέων και τα υπόλοιπα τρία εναλλακτικά μετάγραφα κωδικοποιούν πρωτεΐνες των 315, 292 και 313 αμινοξέων, αντίστοιχα (Sarafidou et al. 2004) (Εικ. 4). Η πιο βασική περιοχή της πρωτεΐνης εντοπίζεται στα αμινοξέα 200-275 (ισοηλεκτρικό σημείο ~10) ενώ η πιο όξινη περιοχή στο καρβοξυτελικό της άκρο (280-315 αμινοξέα) (ισοηλεκτρικό σημείο ~3,70). Σε δυο μελέτες ανάλυσης του φωσφοπρωτεϊνώματος με μεθόδους χρωματογραφίας ιοντοανταλλαγής και φασματομετρία μάζας υγρής φάσης, βρέθηκε ότι η πρωτεΐνη FRA10AC1 φωσφορυλιώνεται σε επτά κατάλοιπα σερίνης στο καρβοξυτελικό της άκρο στις θέσεις 248, 251, 252, 273, 278, 283, 285 (Park et al. 2015, Sharma et al. 2014). Με μεθόδους *in silico* έχει βρεθεί ότι διαθέτει ένα διμερές σήμα πυρηνικού εντοπισμού που περιλαμβάνει τα αμινοξέα 223–238 και 245–262 (μια περιοχή πλούσια σε λυσίνη), το οποίο και επιβεβαιώθηκε πειραματικά (Sarafidou et al. 2004) (Εικ. 7 & 8). Έτσι, τόσο η ανασυνδυασμένη όσο και η ενδογενής πρωτεΐνη, έχει δειχθεί ότι πράγματι εντοπίζονται στο πυρηνόπλασμα είτε σε στικτώδεις δομές (Εικ. 8), είτε σε διάχυτη μορφή (Sarafidou et al. 2004).



**Εικ. 7: Πρωτεϊνική ακολουθία FRA10AC1 (NP\_660289.2).** Με έντονη γραφή φαίνονται οι περιοχές που αποτελούν σήματα πυρηνικού εντοπισμού και η πρωτεϊνική ακολουθία τους. Με κόκκινο χρώμα απεικονίζονται οι θέσεις φωσφορυλίωσης στα επτά κατάλοιπα σερίνης του καρβοξυτελικού άκρου της πρωτεΐνης.



**Εικ. 8:** Υποκυτταρικός εντοπισμός της ανασυνδυασμένης πρωτεΐνης *FRA10AC1* σε νεφρικά κύτταρα πιθήκου (COS-7). Η πρωτεΐνη εντοπίζεται αποκλειστικά στο πυρηνόπλασμα σε στικτώδεις δομές (με κόκκινο φαίνεται η πρωτεΐνη του πυρηνικού φακέλου λαμίνης A/C) (Sarafidou et al. 2004).

Η *FRA10AC1* είναι μια εξελικτικά συντηρημένη πρωτεΐνη καθώς εμφανίζει ορθόλογα μόρια σε πλήθος ευκαρυωτικών, ζωικών και φυτικών οργανισμών, αλλά όχι σε προκαρυωτικούς οργανισμούς. Όπως και στον άνθρωπο, η ορθόλογη πρωτεΐνη στον ποντικό χαρτογραφείται ανάμεσα στα ορθόλογα γονίδια *Pdebc* και *Lgi1* σε συνταινική χρωμοσωματική θέση με αυτή της χρωμοσωματικής περιοχής 10q23–q24 του ανθρώπου και αποτελείται από τον ίδιο αριθμό αμινοξέων. Οι δυο πρωτεϊνικές ακολουθίες εμφανίζουν ταυτότητα της τάξεως του 86% και ομολογία κατά 92%. Επιπλέον, έχουν ταυτοποιηθεί ορθόλογα μόρια στα φυτά με ποσοστό ταυτότητας από 42% έως 55%, στα έντομα με ποσοστό ταυτότητας από 40% έως 57%, και στους οργανισμούς-μοντέλα *Caenorhabditis elegans* και *Drosophila melanogaster* σε ποσοστό 36% και 43%, αντίστοιχα. Υψηλότερο ποσοστό ομοιότητας (45-67%) παρατηρείται στην κεντρική περιοχή των ορθόλογων μορίων (αμινοξέα 105-240 στην πρωτεΐνη του ανθρώπου) γεγονός που πιθανώς υποδηλώνει μια περιοχή σημαντικής λειτουργίας (Sarafidou et al. 2004). Σε μία πρόσφατη ερευνητική μελέτη (Dutcher et al. 2018) ταυτοποιήθηκε λειτουργικό γονίδιο σε έναν απλούστερο και παλιότερα εξελικτικά οργανισμό, στο φύκος *Chlamydomonas reinhardtii*, γεγονός που υποδεικνύει ότι το *FRA10AC1* φαίνεται να έχει συντηρημένη λειτουργία.

### 1.3.1. Η πρωτεΐνη *FRA10AC1* σχετίζεται με την επεξεργασία του πρόδρομου mRNA

Δεδομένου του γεγονότος ότι η πρωτεΐνη *FRA10AC1* έχει απομονωθεί επανειλημμένα ως συστατικό του μείζονος σωματίου συναρμογής σε βιοχημικές μελέτες απομόνωσης της ριβονουκλεοπρωτεϊνικής αυτής μηχανής μπορεί να διατυπωθεί η υπόθεση ότι η πρωτεΐνη αυτή σχετίζεται με τον μηχανισμό επεξεργασίας και ωρίμανσης του πρόδρομου mRNA. Ειδικότερα έχει ταυτοποιηθεί στα επιμέρους υποσύμπλοκα του μείζονος σωματίου συναρμογής B *act* (activated) (Bessonov et al. 2010), C (Bessonov et al. 2008, 2010, Pagan et al. 2013, Schmidt et al. 2014) και P (Pagan et al. 2013), κατά τη συναρμολόγηση των οποίων λαμβάνουν χώρα τα δυο καταλυτικά βήματα της συναρμογής (σύμπλοκο B *act* και C) και η ολοκλήρωση της διαδικασίας (απομάκρυνση ιντρονίου, σύμπλοκο P). Επιπλέον, στο πλαίσιο μίας εργασίας της κύριας ερευνητικής ομάδας που μελετά τη σύσταση του σωματίου συναρμογής, η *FRA10AC1* χαρακτηρίστηκε ως πρωτεΐνη χαμηλής αφθονίας (non-core) λαμβάνοντας υπόψιν την αφθονία της σε σχέση με τα υπόλοιπα συστατικά (Hegele et al. 2012).

Η συσχέτιση της FRA10AC1 με την επεξεργασία του πρόδρομου mRNA και ειδικότερα με τη συναρμογή (splicing) προέρχεται, επίσης, από πειράματα ταυτοποίησης πρωτεϊνικών αλληλεπιδράσεων της (Sarafidou et al. υπό συγγραφή). Αναλυτικότερα, με το σύστημα δυο υβριδίων του σακχαρομύκητα προσδιορίστηκε η αλληλεπίδραση της πρωτεΐνης FRA10AC1 με τις πρωτεΐνες DGCR14 και SF3B2 οι οποίες έχουν χαρακτηριστεί ως συστατικά του σωματίου συναρμογής και έχουν επιβεβαιωμένο ρόλο στην διαδικασία της συναρμογής (βλ. παρακάτω). Οι αλληλεπιδράσεις αυτές επιβεβαιώθηκαν με δοκιμή συν-ανοσοκατακρήμνισης κατά την οποία η πρωτεΐνη FRA10AC1 είχε παραχθεί σε σύντηξη με την τρανσφεράση (S) της γλουταθειόνης (GST) σε βακτήρια ενώ η άλλη πρωτεΐνη του ζεύγους (DGCR14 ή SF3B2) είχε παραχθεί ραδιενεργά σημασμένη (μεθειονίνη S<sup>35</sup>) με *in vitro* μεταγραφή-μετάφραση. Επιπλέον, η αλληλεπίδραση μεταξύ της πρωτεΐνης FRA10AC1 και DGCR14 επιβεβαιώθηκε και με δοκιμή συν-ανοσοκατακρήμνισης των ενδογενών μορίων τους. Ειδικότερα η κατακρήμνιση της πρωτεΐνης FRA10AC1 πραγματοποιήθηκε από πρωτεϊνικό εκχύλισμα κυττάρων θηλαστικών *HeLa* με το αντίσωμα α-FRA10AC1, και ακολούθως ανιχνεύθηκε στο κατακρήμνισμα η πρωτεΐνη DGCR14 με ανάλυση Western. Η ίδια διαδικασία ακολουθήθηκε και αντίστροφα για τις δυο πρωτεΐνες γεγονός που έδειξε ότι η FRA10AC1 και η DGCR14 εντοπίζονται *in vivo* στο ίδιο σύμπλοκο (Sarafidou et al. υπό συγγραφή).

Η αξιοπιστία της αλληλεπίδρασης των πρωτεϊνών FRA10AC1 και DGCR14 ενισχύεται από την ταυτοποίησή της μεταξύ των ορθόλογων μορίων στο είδος *Drosophila melanogaster* στο πλαίσιο μίας μελέτης μεγάλης κλίμακας ταυτοποίησης άμεσων πρωτεϊνικών αλληλεπιδράσεων με το σύστημα δυο υβριδίων στο σακχαρομύκητα (Giot et al. 2003).

Επιπλέον, με συγκριτικά πειράματα RNA *in situ* υβριδοποίησης σε τομές εμβρύων ποντικού και τομές εγκεφάλου ενηλίκων ποντικών μέσω της χρήσης σημασμένων ανιχνευτών RNA για τα τρία γονίδια, παρατηρήθηκε η συν-έκφρασή τους σε όλους τους ιστούς που εξετάστηκαν (π.χ. συγκεκριμένες περιοχές του νευρικού συστήματος, πνεύμονας, καρδιά, ήπαρ) (Sarafidou et al. υπό συγγραφή). Υψηλότερα επίπεδα έκφρασης και των τριών γονιδίων βρέθηκαν σε συγκεκριμένες περιοχές του αναπτυσσόμενου νευρικού συστήματος (π.χ. μετωπιαίος φλοιός, νωτιαίος μυελός), στους πνεύμονες, στην καρδιά και στο ήπαρ εμβρύων ποντικού και σε συγκεκριμένες δομές του νευρικού συστήματος (απιοειδής φλοιός, κινητικός φλοιός, αμυγδαλή, πρόσθιος σύνδεσμος και σε βασικά γάγγλια) σε ενήλικα ποντίκια) (Sarafidou et al. υπό συγγραφή).

Ακόμη, με πειράματα που πραγματοποιήθηκαν σε κύτταρα θηλαστικών *HeLa*, έχει δειχθεί ο υποκυτταρικός συν-εντοπισμός των τριών πρωτεϊνών στο πυρηνόπλασμα των κυττάρων (Sarafidou et al. υπό συγγραφή).

Οι αλληλεπιδράσεις της FRA10AC1 με τις δυο αυτές πρωτεΐνες επιβεβαιώνονται και σε μία άλλη μελέτη μεγάλης κλίμακας ταυτοποίησης άμεσων πρωτεϊνικών αλληλεπιδράσεων μεταξύ των συστατικών του σωματίου συναρμογής. Ειδικότερα, οι συγκεκριμένες αλληλεπιδράσεις εντοπίστηκαν με το σύστημα δυο υβριδίων στο σακχαρομύκητα και με συν-ανοσοκατακρήμνιση των ανασυνδυασμένων μορίων (Hegele et al. 2012).



Η ταυτοποίηση των παραπάνω αλληλεπιδράσεων συνεισφέρει στην εξαγωγή συμπερασμάτων σχετικά με το λειτουργικό ρόλο της FRA10AC1 και την εντάσσει με μεγαλύτερη αξιοπιστία στον μηχανισμό της επεξεργασίας του πρόδρομου mRNA. Η συμμετοχή της FRA10AC1 στη συναρμογή φαίνεται να είναι συντηρημένη καθώς, σε μια πρόσφατη ερευνητική μελέτη στο μονοκύτταρο οργανισμό *Chlamydomonas reinhardtii* (Dutcher et al. 2018), δείχθηκε ότι μια μετάλλαξη απώλειας λειτουργίας του γονιδίου *FRA10AC1* αποκαθιστά τη σωστή εξέλιξη της συναρμογής του mRNA ενός συγκεκριμένου γονιδίου που έχει μεταλλαγμένη 3' θέση συναρμογής, γεγονός που υποδεικνύει τη συνεισφορά της πρωτεΐνης στην αναγνώριση των θέσεων συναρμογής.

Το γονίδιο *DGCR14* είχε μελετηθεί αρχικά μέσω κλωνοποίησης θέσης ως πιθανό υποψήφιο για το σύνδρομο DiGeorge (OMIM #188400) το οποίο χαρακτηρίζεται από αναπτυξιακή υστέρηση και συγγενείς παθήσεις της καρδιάς (Kobrynski & Sullivan 2007, Lindsay et al. 2001). Η κύρια μοριακή αιτία του συνδρόμου είναι ένα έλλειμμα μεγέθους περίπου 3 Mb στη χρωμοσωματική περιοχή 22q11.2. Ωστόσο, σε μεταγενέστερη μελέτη αποδείχθηκε ότι το γονίδιο *DGCR14* δεν σχετίζεται με τα κύρια χαρακτηριστικά του συνδρόμου αλλά ότι το κύριο υπεύθυνο γονίδιο είναι το *TBX-1*, το οποίο κωδικοποιεί έναν μεταγραφικό παράγοντα που συμμετέχει στην ανάπτυξη της καρδιάς, των παραθυροειδών αδένων και του θύμου αδένα (Huh & Ornitz 2010). Η πρωτεΐνη DGCR14 αποτελείται από 476 αμινοξέα. Το καρβοξυτελικό της άκρο φέρει μια δομή σπειροειδούς έλικας (coiled-coil domain), η οποία γενικά συμμετέχει στις αλληλεπιδράσεις με άλλες πρωτεΐνες (Gong et al. 1997). Όπως αναφέρθηκε και προηγουμένως, η DGCR14 εντοπίζεται στο πυρηνόπλασμα (Sarafidou et al. υπό συγγραφή, <http://www.proteinatlas.org/>) και ορθόλογα μόριά της έχουν ταυτοποιηθεί αποκλειστικά σε ευκαρυωτικούς. Πρόσφατες μελέτες υποδεικνύουν τη συμμετοχή της DGCR14 στη ρύθμιση της έκφρασης της ιντερλευκίνης 17α στα προ-φλεγμονώδη T-βοηθητικά λεμφοκύτταρα (T<sub>H</sub>17-λεμφοκύτταρα) (Takada 2015) και στη διαδικασία της συναρμογής (Noma et al. 2014, Dutcher et al. 2018).

Ειδικότερα στην ερευνητική εργασία του Takada (2015) έχει δειχθεί ότι η πρωτεΐνη DGCR14 λειτουργεί ως ρυθμιστής ενός μέλους της οικογένειας των πυρηνικών υποδοχέων (υποδοχέας ROR $\gamma$ ) οι οποίοι δρουν ως μεταγραφικοί παράγοντες. Ο υποδοχέας αυτός ελέγχει την μεταγραφή γονιδίων που ρυθμίζουν τον κερκαδικό ρυθμό, την ομοιοστάση των λιπιδίων και της γλυκόζης (Jetten et al. 2007) και έχει προταθεί ότι είναι βασικός ρυθμιστής της διαφοροποίησης των T<sub>H</sub>17-λεμφοκυττάρων (Ivanov et al. 2006). Σε αυτήν την εργασία, έχει δειχθεί ότι η υπερέκφραση του γονιδίου *DGCR14* ενισχύει την έκφραση της πρωτεΐνης ROR $\gamma$  και την μεταγραφική ενεργότητα γονιδίων που εξαρτώνται από αυτόν τον υποδοχέα. Επιπλέον, παρατήρηθηκαν αυξημένα επίπεδα του mRNA της ιντερλευκίνης 17α (IL17α) και επαγωγή της κυτταρικής διαφοροποίησης των λεμφοκυττάρων T<sub>H</sub>17. Αντίθετα, η καταστολή του γονιδίου *DGCR14* έδειξε ότι μειώνει τα επίπεδα έκφρασης της ιντερλευκίνης 17α (IL17α).

Επιπλέον, στον οργανισμό *C. elegans* έχει δειχθεί η συσχέτιση της πρωτεΐνης DGCR14 με τη διασφάλιση της σωστής συναρμογής του mRNA, σε περιπτώσεις που υπάρχουν μη φυσιολογικές αλληλουχίες στην 3' θέση συναρμογής. Συγκεκριμένα δημιουργήθηκαν διαγονιδιακά *C. elegans* που έφεραν μεταλλάξεις στις 3' θέσεις συναρμογής ορισμένων γονιδίων παρουσία ή απουσία μιας ομόζυγης μετάλλαξης

απώλειας λειτουργίας του ορθόλογου γονιδίου *DGCR14*. Όταν το ορθόλογο μόριο της *DGCR14* είναι μεταλλαγμένο δείχθηκε ότι μειώνονται τα επίπεδα mRNA δυο συγκεκριμένων γονιδίων που έχουν μεταλλάξεις στις 3' θέσεις συναρμογής, γεγονός που υποδεικνύει ότι η πρωτεΐνη *DGCR14* ρυθμίζει την παραγωγή του σωστού μεταγράφου όταν η θέση συναρμογής δεν περιέχει τη συντηρημένη αλληλουχία συναρμογής (Noma et al. 2014). Ακόμη, σε μία άλλη μελέτη, στο μονοκύτταρο οργανισμό *Chlamydomonas reinhardtii*, δείχθηκε ότι μία μετάλλαξη κέρδους λειτουργίας του γονιδίου *DGCR14* αποκαθιστά τη σωστή εξέλιξη της συναρμογής των mRNA ενός συγκεκριμένου γονιδίου που είναι μεταλλαγμένο στην 5' θέση συναρμογής, υποδεικνύοντας τη συντηρημένη λειτουργία της πρωτεΐνης στην αναγνώριση των θέσεων συναρμογής (Dutcher et al. 2018).

Όπως και η πρωτεΐνη *FRA10AC1* έτσι και η *DGCR14* έχει ταυτοποιηθεί επανειλημμένα από βιοχημικές μελέτες σε κύτταρα *HeLa* ως συστατικό του μείζονος σωματίου συναρμογής (Rappsilber et al. 2002, Zhou et al. 2002) και στα ίδια υποσύμπλοκα, το σύμπλοκο B *act* (Bessonov et al., 2010), το σύμπλοκο C (Pagan et al. 2009, 2013, Bessonov et al. 2008, 2010, Schmidt et al. 2014), και το σύμπλοκο P (Pagan et al. 2013). Οι δυο πρωτεΐνες εμφανίζουν και παρόμοια αφθονία αφού έχουν κατηγοριοποιηθεί στις πρωτεΐνες χαμηλής αφθονίας (non-core) σύμφωνα με μία εργασία της κύριας ερευνητικής ομάδας που μελετά τη σύσταση του σωματίου συναρμογής (Hegele et al. 2012).

Έτσι, τα παραπάνω στοιχεία υποδεικνύουν αφενός τον ρόλο της *DGCR14* στη διαδικασία της συναρμογής και αφετέρου την πιθανή συμμετοχή της *FRA10AC1* στο μηχανισμό αυτό με βάση την αξιοπιστία της αλληλεπίδρασής τους και τον συνεντοπισμό τους στα τελικά στάδια του μηχανισμού συναρμογής. Η πιθανή συνεισφορά ή «συνεργασία» των δυο αυτών πρωτεϊνών στη διαδικασία της συναρμογής υποδεικνύεται και από την πρόσφατη μελέτη των Dutcher et al. (2018) στο μονοκύτταρο οργανισμό *Chlamydomonas reinhardtii*, μέσω της οποίας υποδεικνύεται η συντηρημένη βιολογική λειτουργία των δυο πρωτεϊνών στην αναγνώριση των θέσεων συναρμογής ακόμα και σε απλούστερους οργανισμούς, όπως το συγκεκριμένο φύκος.

Πληροφορίες για το πιθανό βιολογικό ρόλο της *DGCR14* μπορούν να ανακτηθούν από μελέτες συσχέτισης και ποντικών knock-out. Ειδικότερα, σε μια μελέτη συσχέτισης σε δείγμα 235 ατόμων, βρέθηκαν δυο πολυμορφισμοί στον υποκινητή του γονιδίου *DGCR14* οι οποίοι έδειξαν στατιστικά σημαντική συσχέτιση με σχιζοφρένεια (Wang et al. 2006). Επιπλέον, ένα μοντέλο ποντικού το οποίο φέρει ένα ετερόζυγο μικροέλλειμμα μεγέθους περίπου 1,5Mb στην χρωμοσωματική περιοχή 16q που περιλαμβάνει 27 γονίδια και μεταξύ αυτών και το *Dgcr14*, έδειξε μειωμένη ικανότητα συγχρονισμού των νευρικών κυττάρων μεταξύ του ιπποκάμπου και του προμετωπιαίου φλοιού κατά τη διάρκεια εκτέλεσης ενός τεστ συμπεριφοράς που σχετίζεται με την χωρική μνήμη. Δεδομένου ότι η συγκεκριμένη λειτουργία διαταράσσεται στη σχιζοφρένεια, μπορεί να υποθεθεί ότι το γονίδιο *DGCR14* συνεισφέρει ως γενετικός παράγοντας κινδύνου στη σχιζοφρένεια (Sigurdsson et al. 2010). Τέλος, σε μία άλλη μελέτη, η δημιουργία ομόζυγων ποντικών knock out για τον μεταγραφικό παράγοντα Goosecoid-like (*Gsc1*), ο οποίος εκφράζεται αποκλειστικά στον μεσοσκελιαίο πυρήνα

στην κοιλιακή περιοχή του ορίου μεσεγκεφάλου-ρομβεγκεφάλου, είχε ως αποτέλεσμα την πλήρη καταστολή της έκφρασης και του γονιδίου *Dgcr14* σε αυτήν την περιοχή, το οποίο ελέγχθηκε λόγω της χρωμοσωματικής γειτνίασης των δυο γονιδίων. Ο συνδυασμός της καταστολής της έκφρασης των δυο γονιδίων οδηγούσε σε προβλήματα στη ρύθμιση του χρόνου παραμονής στα στάδια του ύπνου REM (rapid eye movement) και non-REM (non-rapid eye movement) (Funato et al. 2010).

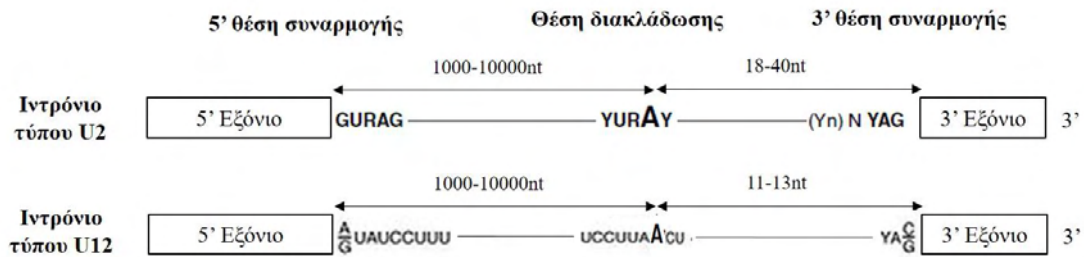
Η πρωτεΐνη SF3B2 αποτελεί την υπομονάδα 2 του πρωτεϊνικού συμπλόκου SF3b του μείζονος σωματίου συναρμογής. Αποτελείται από 872 αμινοξέα και είναι υδρόφιλη. Παρόλο που η πρωτεΐνη αυτή δεν διαθέτει γνωστά μοτίβα πρόσδεσης του RNA έχειδειχθεί ότι συνδέεται με μόρια mRNA (Castello et al. 2012). Επιπλέον, έχειδειχθεί ότι αλληλεπιδρά άμεσα με την υπομονάδα 4 του πρωτεϊνικού συμπλόκου SF3b, SF3B4 (Champion-Arnaud & Reed 1994). Το πρωτεϊνικό σύμπλοκο SF3b μαζί με το πρωτεϊνικό σύμπλοκο SF3a και το U2snRNA σχηματίζουν το μικρό πυρηνικό ριβονουκλεοπρωτεϊνικό σύμπλοκο U2 snRNP. Το σύμπλοκο συναρμογής SF3b/3a προσδένεται στο πρόδρομο mRNA σε μια αλληλουχία ανοδικά της θέσης διακλάδωσης του ιντρονίου και βοηθάει το U2 snRNP να «αγκιστρωθεί» στο πρόδρομο mRNA, έτσι ώστε να σχηματιστεί το υποσύμπλοκο A το οποίο αποτελεί ένα από τα αρχικά στάδια σχηματισμού του σωματίου συναρμογής (prespliceosome) (Golas et al. 2003, Spadaccini 2006). Πέρα από τον καθιερωμένο ρόλο της πρωτεΐνης SF3B2 στη διαδικασία της συναρμογής, έχειδειχθεί ότι συμμετέχει και στη ρύθμιση της εξέλιξης του κυτταρικού κύκλου και ειδικότερα στη μετάβαση από τη φάση G2 στη φάση της μίτωσης (Terada & Yasuda 2006). Πρόσφατες μελέτες έχουν δείξει ότι η μεθυλίωση ενός καταλοίπου αργινίνης στη θέση 508 της πρωτεΐνης SF3B2 είναι απαραίτητη για την αλληλεπίδρασή της με την πρωτεΐνη SMN. Η αλληλεπίδραση αυτή συνεισφέρει στην ωρίμανση και στη συναρμολόγηση των snRNPs καθώς και στη ρύθμιση της εναλλακτικής συναρμογής (Yang et al. 2015). Με πειράματα υποκυτταρικού εντοπισμού έχειδειχθεί ότι η SF3B2 εντοπίζεται στον πυρήνα, σε ειδικές δομές (nuclear speckles) που αποτελούν θέσεις αποθήκευσης των παραγόντων της συναρμογής και ρυθμιστικές θέσεις για τη διαδικασία της μεταγραφής (Pontén et al. 2011). Επιπλέον είναι εξελικτικά συντηρημένη σε πολυκύτταρους ή μονοκύτταρους ευκαρυωτικούς οργανισμούς (Jiang et al. 2009). Όπως είναι αναμενόμενο, έχει χαρακτηριστεί ως συστατικό του μείζονος σωματίου συναρμογής σε όλες τις μελέτες βιοχημικής απομόνωσης των συστατικών ολόκληρου του σωματίου συναρμογής (Rappsilber et al. 2002, Zhou et al. 2002, Makarov et al. 2002, Chen et al. 2007) αλλά και των επιμέρους υποσυμπλόκων του E, A, B, B act, C και P (Makarov et al. 2012, Hartmuth et al. 2002, Behzadnia et al. 2007, Agafonov et al. 2011, Makarova et al. 2004, Herold et al. 2009, Bessonov et al. 2008, 2010; Deckert et al. 2006, Schmidt, et al. 2014; Jurica et al. 2002, Pagan et al. 2009, 2013). Επιπλέον, έχει κατηγοριοποιηθεί στις πρωτεΐνες του U2 snRNP που βρίσκονται σε υψηλή αφθονία (πρωτεΐνες core) (Hegele et al. 2012).

Οι αλληλεπιδράσεις της FRA10AC1 με πρωτεΐνες που συμμετέχουν στη διαδικασία της συναρμογής αποτελώντας συστατικά του σωματίου συναρμογής συνάδουν με την ταυτοποίησή της ως συστατικό των επιμέρους υποσυμπλόκων του σωματίου συναρμογής (B act, C και P) και παρέχουν σημαντικές ενδείξεις για τον πιθανό

βιολογικό της ρόλο σε αυτή τη διαδικασία ή, σε ευρύτερο πλαίσιο, στη διαδικασία της επεξεργασίας του πρόδρομου mRNA. Οι ενδείξεις συμμετοχής της FRA10AC1 στη συναρμογή τείνουν να επιβεβαιώνονται από μία μελέτη στον μονοκύτταρο οργανισμό *Chlamydomonas reinhardtii*, καθώς φαίνεται ότι το *FRA10AC1* είναι ένα αρχαίο εξελικτικά γονίδιο που παρουσιάζει συντηρημένη λειτουργία στη συναρμογή και, πιθανώς, σε κάποια από τις άλλες διαδικασίες της επεξεργασίας του mRNA.

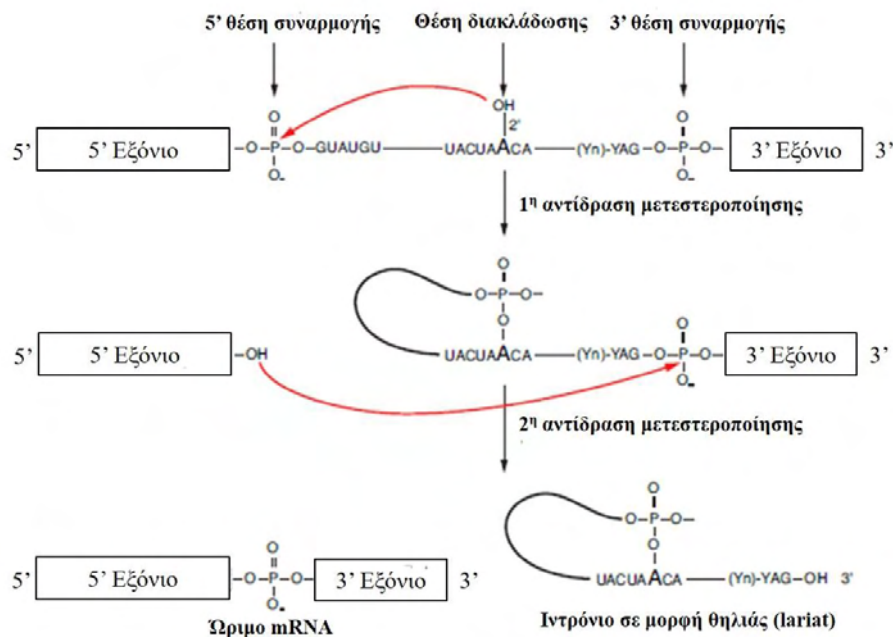
#### 1.4. Μηχανισμός συναρμογής

Τα πρόδρομα μόρια mRNA που προκύπτουν από τα ευκαρυωτικά γονίδια υφίστανται συναρμογή, δηλαδή, απομάκρυνση των ιντρονίων και ένωση των εξόνιων, ώστε να γίνει η ωρίμανση τους για την τελική παραγωγή του πρωτεϊνικού προϊόντος. Τα ιντρόνια απομακρύνονται έπειτα από την αναγνώριση και την κοπή των 5' και 3' θέσεων συναρμογής που βρίσκονται στις 5' και 3' θέσεις των ιντρονίων με τη μεσολάβηση δυο σημαντικών αλληλουχιών, της θέσης διακλάδωσης (branch point sequence) και μιας περιοχής πλούσιας σε πυριμιδίνες (polypyrimidine tract), οι οποίες βρίσκονται ανοδικά της 3' θέσης συναρμογής. Υπάρχουν δυο τύποι ιντρονίων, τα ιντρόνια τύπου U2 που αποτελούν τον κύριο τύπο ιντρονίων και τα ιντρόνια τύπου U12 τα οποία εμφανίζονται σε συχνότητα 0,15-0,34% στον άνθρωπο και συναντώνται κυρίως σε σπονδυλωτά, φυτά, έντομα και ορισμένους μύκητες ενώ απουσιάζουν σε οργανισμούς μοντέλα όπως ο *C. elegans* και ο *Saccharomyces cerevisiae* (Εικ. 9) (Turunen et al. 2013, Sharp & Burge 1997, Burge et al. 1998). Οι διαφορές τους έγκεινται: α) στη συντήρηση των αλληλουχιών των θέσεων συναρμογής, β) στην απόσταση της θέσης διακλάδωσης από την 3' θέση συναρμογής η οποία είναι μικρότερη στα ιντρόνια τύπου U12, και γ) στην περιοχή που είναι πλούσια σε πυριμιδίνες η οποία απουσιάζει εντελώς στα ιντρόνια τύπου U12 (Εικ. 9) (Turunen et al. 2013). Τα γονίδια που περιέχουν ιντρόνια τύπου U12 συναντώνται σε συγκεκριμένες λειτουργικές κατηγορίες γονιδίων και βιολογικά μονοπάτια. Δηλαδή, είναι κατά κύριο λόγο παρόντα σε γονίδια που σχετίζονται με την αντιγραφή και την επιδιόρθωση του DNA, τη μεταγραφή, την επεξεργασία του RNA, τη μετάφραση, την οργάνωση του κυτταροσκελετού, τη μεταφορά μέσω κυστιδίων και τη δράση των τασοελεγχόμενων διαύλων ιόντων. Αντιθέτως, εμφανίζονται σε πολύ μικρό ποσοστό σε γονίδια που σχετίζονται με το μεταβολισμό της ενέργειας και τα βιοσυνθετικά μονοπάτια (Turunen et al. 2013, Alioto 2007, Wu & Krainer 1999). Επίσης, τα γονίδια που διαθέτουν ένα μόνο ιντρόνιο τύπου U12 ανάμεσα σε πολλά ιντρόνια τύπου U2 συμμετέχουν σε διάφορες διαδικασίες που είναι σημαντικές για την ανάπτυξη του κυττάρου και του οργανισμού (He et al. 2011, Younis et al. 2013).



**Εικ. 9: Σχηματική απεικόνιση και διαφορές των ιντρονίων τύπου U2 και τύπου U12.** Φαίνονται οι αλληλουχίες στις 5' και 3' θέσεις συναρμογής των ιντρονίων, στη θέση διακλάδωσης και οι αντίστοιχες διαφορές στους δυο τύπους ιντρονίων. Με έντονη γραφή φαίνεται η αζωτούχος βάση αδενίνη (A) στη θέση διακλάδωσης. N: οποιοδήποτε αζωτούχος βάση, R: πουρίνη, Y: πυριμιδίνη και Yn= περιοχή πλούσια σε πυριμιδίνες (προσαρμοσμένη από Fabrizio & Lührmann 2012).

Η διαδικασία απομάκρυνσης των ιντρονίων λαμβάνει χώρα σε δυο διαδοχικά καταλυτικά βήματα μέσω δυο αντιδράσεων μετεστεροποίησης, δηλαδή ανταλλαγής της οργανικής ομάδας της αλκοόλης ενός νουκλεοτιδίου με την οργανική ομάδα του φωσφορικού εστέρα ενός άλλου νουκλεοτιδίου. Οι αντιδράσεις αυτές περιλαμβάνουν δυο πυρηνόφιλες επιθέσεις-δηλαδή τη μεταφορά ηλεκτρονίων σε θετικά φορτισμένα άτομα-σε δυο βήματα της διαδικασίας. Στην περίπτωση της συναρμογής, αρχικά, η 3' υδροξυλική ομάδα της αδενίνης της θέσης διακλάδωσης πραγματοποιεί πυρηνόφιλη επίθεση στον φωσφορικό εστέρα της 5' θέσης συναρμογής (1<sup>η</sup> αντίδραση μετεστεροποίησης, Εικ. 10), και στη συνέχεια η υδροξυλική ομάδα του 5' άκρου του εξονίου εκτελεί πυρηνόφιλη επίθεση στην 3' θέση συναρμογής (2<sup>η</sup> αντίδραση μετεστεροποίησης, Εικ. 10). Και στα δύο αυτά βήματα διασπώνται οι φωσφοδιεστερικοί δεσμοί μεταξύ των νουκλεοτιδίων που βρίσκονται στα συγκεκριμένα σημεία ώστε να απομακρυνθεί το ιντρόνιο (σε μορφή θηλιάς) και να ενωθούν τα εξόνια (ώριμο mRNA) (Εικ. 10).

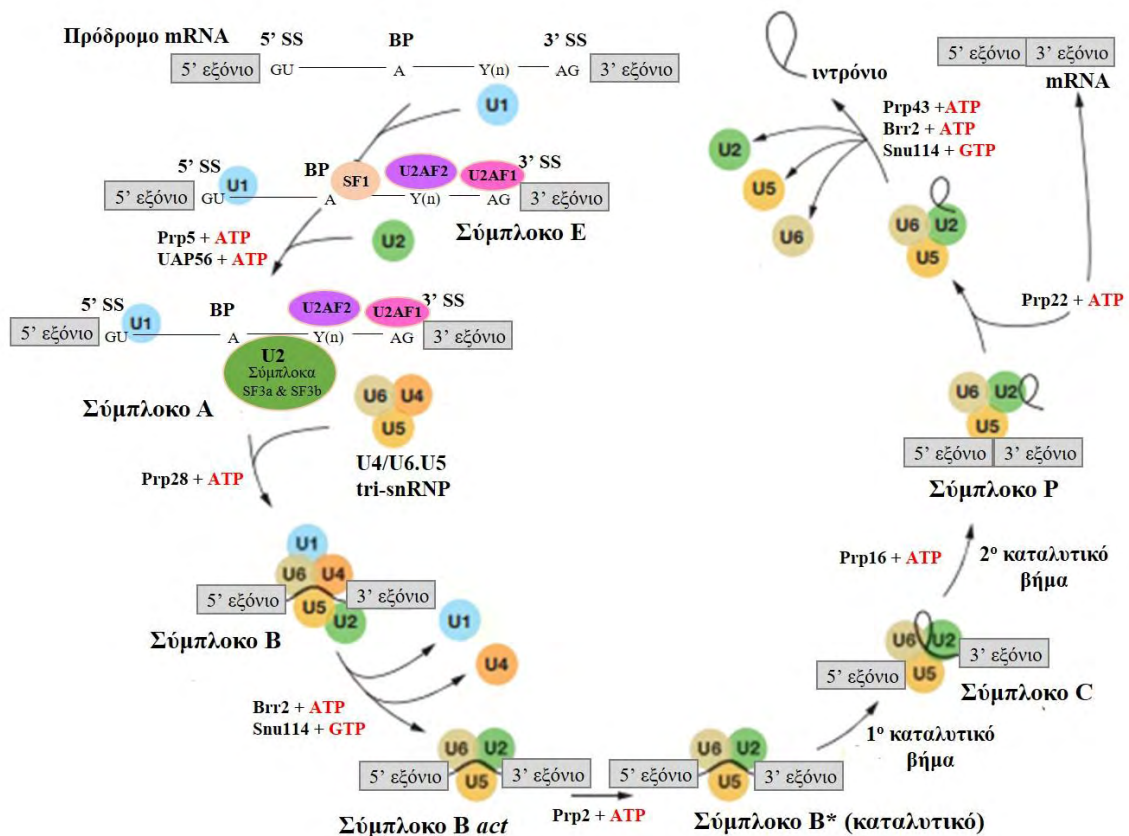


**Εικ. 10: Σχηματική αναπαράσταση των δυο αντιδράσεων μετεστεροποίησης.** Φαίνονται με κόκκινα βέλη οι πυρηνόφιλες επιθέσεις στις 5' και 3' θέσεις συναρμογής, η ακόλουθη διάσπαση των φωσφοδιεστερικών δεσμών, η απομάκρυνση του ιντρονίου και η ένωση των αντίστοιχων εξονίων. Με έντονη γραφή φαίνεται η αζωτούχος βάση αδενίνη (A) στη θέση διακλάδωσης που πραγματοποιεί την πρώτη πυρηνόφιλη επίθεση στην 5' θέση συναρμογής για να σχηματιστεί ο φωσφοδιεστερικός δεσμός σε αυτή τη θέση. R: πουρίνη, Y: πυριμιδίνη και Yn= περιοχή πλούσια σε πυριμιδίνες (προσαρμοσμένη από Fabrizio & Lührmann 2012).

Η διαδικασία της συναρμογής καταλύεται από μία πολύπλοκη και δυναμική ριβονουκλεοπρωτεϊνική μηχανή, το σωματίο συναρμογής, το οποίο είναι το μεγαλύτερο σε μέγεθος ριβονουκλεοπρωτεϊνικό σύμπλοκο (4–5 MDa) ακόμα και σε σύγκριση με το ριβόσωμα (3,3 MDa). Τα ιντρόνια τύπου U2 απομακρύνονται από το μείζον σωματίο συναρμογής (major spliceosome) που χρησιμοποιείται κατά κύριο λόγο από όλους τους ευκαρυώτες (Wahl et al. 2009). Τα ιντρόνια τύπου U12 τα οποία, όπως αναφέρθηκε παραπάνω, εντοπίζονται σε πολύ μικρό ποσοστό, απομακρύνονται από το ελάσσον σωματίο συναρμογής (minor spliceosome) (Schneider et al. 2002). Το κύριο έργο του σωματίου συναρμογής είναι η αναγνώριση και η σύζευξη των σωστών θέσεων συναρμογής μέσα σε ένα πλήθος όμοιων αλληλουχιών. Αυτό επιτυγχάνεται με εξαιρετική ακρίβεια από τον μεγάλο αριθμό των υπομονάδων του που περιλαμβάνουν ποικίλες πρωτεΐνες και μικρά ριβονουκλεοπρωτεϊνικά RNA (snRNA) τα οποία συνεργάζονται και συναρμολογούνται μαζί στο πρόδρομο mRNA (Wahl et al. 2009). Για την επίτευξη αυτής της συναρμολόγησης, λαμβάνουν χώρα αλλαγές στη διαμόρφωση πρωτεϊνών καθώς και πολλές και σημαντικές αλληλεπιδράσεις μεταξύ RNA-RNA, πρωτεΐνης-πρωτεΐνης και RNA-πρωτεΐνης. Αυτές οι αλληλεπιδράσεις είναι συνήθως ασθενείς καθώς η συναρμολόγηση του σωματίου συναρμογής αποτελεί μια δυναμική διαδικασία, αλλά, ταυτόχρονα, ο συνδυασμός πολλών ασθενών αλληλεπιδράσεων συνεισφέρει στη συνολική σταθερότητα του συγκεκριμένου συμπλόκου. Συγκεκριμένα, οι περισσότερες λειτουργικές αλληλεπιδράσεις μεταξύ των μορίων RNA είναι ασθενείς ώστε να απαιτείται η βοήθεια των πρωτεϊνών για την ενίσχυση της σταθερότητάς τους (Wahl et al. 2009). Οι πρωτεΐνες με τη σειρά τους φέρουν και καταλυτικές λειτουργίες υποδεικνύοντας ότι το σωματίο συναρμογής είναι μια ριβονουκλεοπρωτεϊνική μηχανή με ιδιότητες ενζύμου (Abelson 2008, Pyle 2008). Σε αντίθεση με τα ένζυμα, το σωματίο συναρμογής δεν διαθέτει ενεργό κέντρο αλλά, αντιθέτως, το καταλυτικό του κέντρο πρέπει να συναρμολογείται εκ νέου σε κάθε πρόδρομο μετάγραφο μέσω της σταδιακής αλληλεπίδρασης των μικρών ριβονουκλεοπρωτεϊνικών σωματιών (UsnRNPs) τα οποία αποτελούν τους κύριους δομικούς λίθους του συμπλόκου και της πληθώρας των μη ριβονουκλεοπρωτεϊνικών παραγόντων συναρμογής (non-snRNP splicing factors) (Will & Lührmann 2011).

Το μείζον σωματίο συναρμογής αποτελείται από πέντε ριβονουκλεοπρωτεϊνικά σύμπλοκα, U1, U2, U4/U6, και U5 τα οποία προκύπτουν από τη συναρμολόγηση των μικρών ριβονουκλεοπρωτεϊνικών RNAs (snRNAs) με δυο κατηγορίες πρωτεϊνών, τις πρωτεΐνες Sm (B/B', D3, D2, D1, E, F, και G) και, μόνο στην περίπτωση του U6 snRNA, τις Lsm (Sm-like), καθώς και με άλλες πρωτεΐνες που συμμετέχουν στη συναρμολόγηση. Η αλληλουχία των snRNAs-κυρίως οι περιοχές που σχηματίζουν δευτεροταγείς δομές-εμφανίζονται υψηλά συντηρημένες μεταξύ των ειδών (Guthrie &

Patterson 1988). Κατά τη δημιουργία των αρχικών υποσυμπλόκων του σωματίου συναρμογής (prespliceosome), όπως φαίνεται αναλυτικά στην Εικ. 14, το U1 snRNP αλληλεπιδρά με την 5' θέση συναρμογής (5' SS). Η σύνδεσή του ακολουθείται από την ταυτόχρονη σύνδεση πρωτεϊνών πλούσιων σε κατάλοιπα σερίνης και θρεονίνης (SR proteins) (π.χ. U2AF1, U2AF2) και κύριων παραγόντων συναρμογής (π.χ. SF1) στη θέση διακλάδωσης (BP), στην περιοχή που είναι πλούσια σε πολυπυριμιδίνες ( $Y_{(n)}$ ) και στην 3' θέση συναρμογής (3' SS) (Εικ. 11). Έτσι σχηματίζεται το υποσύμπλοκο E (Εικ. 11). Στη συνέχεια, το U2 snRNA ενώνεται με τη θέση διακλάδωσης και η αλληλεπίδραση αυτή σταθεροποιείται μέσω αλληλεπιδράσεων: α) δυο πρωτεϊνικών συμπλόκων του U2 snRNP (SF3a και SF3b) με τη θέση διακλάδωσης, και β) των πρωτεϊνών που προαναφέρθηκαν στο υποσύμπλοκο E τόσο με τα σύμπλοκα SF3a και SF3b όσο και με τη θέση διακλάδωσης, με αποτέλεσμα το σχηματισμό του υποσυμπλόκου A (Εικ. 11). Ακολουθεί η στρατολόγηση του τριμερούς συμπλόκου U4/U6.U5 (U4/U6.U5 tri-snRNP) στο U2 snRNP και ο σχηματισμός του υποσυμπλόκου B κατά τη διάρκεια του οποίου τα U4 και U6 snRNA ενώνονται μεταξύ τους και το U5 snRNA συνδέεται με την 5' θέση συναρμογής (Εικ. 11). Έπειτα από αναδιατάξεις στη διαμόρφωση και στη σύσταση τόσο των πρωτεϊνών όσο και των RNA, συμπεριλαμβανομένης της απελευθέρωσης των U1 και U4 snRNP, δημιουργείται το ενεργοποιημένο υποσύμπλοκο B (*B act*) (Εικ. 11). Αυτό, μέσω συγκεκριμένων αναδιατάξεων (αλληλεπίδραση του U6 snRNA με το U2 snRNA και με την 5' θέση συναρμογής και την επακόλουθη δημιουργία ενδομοριακής φουρκέτας στο U6 snRNA) μετασχηματίζεται σε καταλυτικό υποσύμπλοκο B ( $B^*$ ) (Εικ. 11). Σε αυτό το στάδιο πραγματοποιείται η πρώτη αντίδραση μετεστεροποίησης και σχηματίζεται το υποσύμπλοκο C (Εικ. 11). Στη συνέχεια, έπειτα από επιπρόσθετες αναδιατάξεις, συμπεριλαμβανομένων της αλλαγής της διαμόρφωσης του U2 snRNA και της μετατόπισης του U5 snRNA στην 3' θέση συναρμογής, λαμβάνει χώρα η δεύτερη αντίδραση μετεστεροποίησης με αποτέλεσμα την απελευθέρωση του ιντρονίου σε μορφή θηλιάς και την ένωση των εξονίων. Τα εξόνια παραμένουν συνδεδεμένα με τα U2, U5 και U6snRNP αποτελώντας το υποσύμπλοκο P του σωματίου συναρμογής (post-catalytic complex) (Εικ. 11) (Fabrizio & Lührmann 2012).



**Εικ. 11: Συναρμολόγηση του μείζονος σωματίου συναρμογής.** Φαίνονται τα στάδια συναρμολόγησης του μείζονος σωματίου συναρμογής για την απομάκρυνση των ιντρονίων τύπου U2. Οι πρωτεΐνες SF1, U2AF1 και U2AF2 αντιπροσωπεύουν σημαντικές πρωτεΐνες κατά το σχηματισμό των υποσυμπλόκων E και A. Απεικονίζονται οι ελικάσες τύπου DExD/H (Prp5, UAP56, Prp28, Brr2, Snu114, Prp2, Prp16, Prp22 και Prp43) και το στάδιο συναρμολόγησης και αποσυναρμολόγησης του σωματίου στο οποίο συμμετέχουν. 5' SS: 5' θέση συναρμογής, 3' SS: 3' θέση συναρμογής, BP: θέση διακλάδωσης, Y(n)= περιοχή πλούσια σε πυριμιδίνες, ATP: τριφωσφορική αδενοσίνη, GTP: τριφωσφορική γουανοσίνη (προσαρμοσμένη από Fabrizio & Lührmann 2012).

Το ελάχιστον σωματίο συναρμογής περιέχει τέσσερα ριβονουκλεοπρωτεϊνικά σύμπλοκα U11, U12, U4atac, and U6atac που παρουσιάζουν παρόμοια σύσταση σε πρωτεΐνες με αυτά του μείζονος σωματίου και περιέχουν snRNAs τα οποία επιτελούν ισοδύναμες λειτουργίες με αυτά του μείζονος χωρίς όμως να είναι ταυτόσημα σε επίπεδο αλληλουχίας. Τα ριβονουκλεοπρωτεϊνικά σύμπλοκα U4atac, U6atac και U5 ενώνονται και αυτά σε ένα τριμερές σύμπλοκο (Schneider et al. 2002). Σε αντίθεση με το μείζον σωματίο συναρμογής, τα ομόλογα των ριβονουκλεοπρωτεϊνικών σωματιδίων U1 και U2 snRNPs εμφανίζονται με τη μορφή διμερούς συμπλόκου U11/U12 (U11/U12 di-snRNP) (Wassarman & Steitz 1992) και έχουν διαφορετική πρωτεϊνική σύσταση, περιλαμβάνοντας και πρωτεΐνες που εμφανίζονται μόνο στο ελάχιστον σωματίο συναρμογής (Turunen et al. 2013). Η αναγνώριση των 5' και 3' θέσεων συναρμογής γίνεται από το διμερές σύμπλοκο U11/U12 di-snRNP με τη βοήθεια και άλλων πρωτεϊνών ώστε να σχηματιστούν τα αρχικά στάδια του σωματίου. Ο σχηματισμός του καταλυτικά ενεργού σωματίου συναρμογής ακολουθεί το ίδιο



μονοπάτι με αυτό του μείζονος συμπλόκου (Frilander & Steitz 2001), με τη διαφορά ότι η απομάκρυνση των ιντρονίων τύπου U12 πραγματοποιείται με αργότερο ρυθμό (Singh & Padgett 2009).

Σε κάθε στάδιο συναρμολόγησης και των δυο σωματίων συναρμογής απαιτείται η χρήση μορίων τριφωσφορικής αδενοσίνης (ATP) για κατανάλωση ενέργειας (Εικ. 11).

Αξίζει να σημειωθεί ότι και για τα δυο είδη σωματίων συναρμογής, η δυναμική του δικτύου αλληλεπιδράσεων RNA–RNA και RNA–πρωτεϊνών παίζει κεντρικό ρόλο στην «αποκάλυψη» των περιοχών του πρόδρομου μεταγράφου που είναι απαραίτητες σε κάθε στάδιο συναρμολόγησης, αποσυναρμολόγησης και κατάλυσης του σωματίου. Κύριο ρόλο για την δυναμική αναδιαμόρφωση αυτών των αλληλεπιδράσεων παίζουν οι ελικάσες τύπου DExD/H, οι οποίες συνεισφέρουν στη μετάβαση στο επόμενο βήμα της συναρμολόγησης των σωματίων όταν το προηγούμενο έχει ολοκληρωθεί με επιτυχία (Εικ. 11), απομακρύνοντας τα υποστρώματα RNA (snRNA, mRNA) από τα ίδιου ή διαφορετικού είδους μόρια RNA ή από τις πρωτεΐνες του εκάστοτε βήματος (Liu & Cheng 2015, Cordin & Beggs 2013).

#### 1.4.1. Πειραματικές μελέτες απομόνωσης και ταυτοποίησης των συστατικών του σωματίου συναρμογής

Το σωματίο συναρμογής αποτελεί μία από τις πιο πολύπλοκες «μηχανές» των ευκαρυωτικών κυττάρων (Papasaïkas & Valcárcel 2015) καθώς περιλαμβάνει εκατοντάδες πρωτεΐνες και εννέα μοναδικά μικρά πυρηνικά RNA (snRNAs) που πρέπει να δράσουν συντονισμένα μεταξύ τους και με εξαιρετική ακρίβεια ώστε να γίνει η αποτελεσματική αποκοπή των ιντρονίων (Chen & Moore 2014, Will & Lührmann 2011). Ένας επιπρόσθετος, εξίσου σημαντικός λόγος, που εξηγεί την πληθώρα των πρωτεϊνών του σωματίου συναρμογής, είναι η χρήση των εναλλακτικών θέσεων συναρμογής. Πλέον, είναι γνωστό ότι το ~95% των γονιδίων του ανθρώπου που κωδικοποιούν πρωτεΐνες υφίσταται εναλλακτική συναρμογή (Gerstein et al. 2014, Pan et al. 2008). Έτσι, η ρύθμιση της εναλλακτικής συναρμογής απαιτεί μεγάλη ακρίβεια, γεγονός που αυξάνει την πολυπλοκότητα του σωματίου. Μέχρι τώρα, έχουν δημοσιευτεί αρκετές ερευνητικές εργασίες που περιλαμβάνουν πρωτομικές μελέτες ταυτοποίησης των συστατικών τόσο ολόκληρου του σωματίου συναρμογής (π.χ. Rappsilber et al. 2002, Chen et al. 2007) όσο και των επιμέρους υποσυμπλόκων του (π.χ. Agafonov et al. 2011, Bessonov et al. 2008, 2010, Schmidt et al. 2014). Όλες αυτές οι μελέτες έχουν γίνει αποκλειστικά σε κύτταρα *HeLa*.

Στις μελέτες αυτές χρησιμοποιούνται πρόδρομα μετάγραφα (mRNA), προερχόμενα από αδενοϊό ή από τον οργανισμό μοντέλο *D. melanogaster*, τα οποία περιλαμβάνουν δυο εξόνια στα 5' και 3' άκρα και ένα ενδιάμεσο ιντρόνιο. Ακολουθεί η ανάμειξή τους με κυτταρικό εκχύλισμα *HeLa* ώστε να πραγματοποιηθεί *in vitro* συναρμογή των παραπάνω πρόδρομων μεταγράφων. Στη συνέχεια, το μίγμα τοποθετείται σε στήλη μοριακής διήθησης ή φυγοκεντρείται σε κλίση πυκνότητας γλυκερόλης και στα κλάσματα που προκύπτουν ακολουθεί χρωματογραφία συγγένειας με αντισώματα που αναγνωρίζουν συγκεκριμένες πρωτεΐνες του συμπλόκου (Chen et al. 2007, Zhou et al. 2002). Εναλλακτικά, πριν την ανάμειξή τους με το κυτταρικό εκχύλισμα, τα πρόδρομα μετάγραφα σημαίνονται με την υβριδική πρωτεΐνη-«προσαρμογέα» MS2-MBP

(πρωτεΐνη του καψιδίου του βακτηριοφάγου σε σύντηξη με την πρωτεΐνη που συνδέεται στη μαλτόζη, Maltose Binding Protein) ή με συμπληρωματικά ολιγονουκλεοτίδια που είναι συνδεδεμένα με πρωτεΐνες χαμηλού μοριακού βάρους, όπως η βιοτίνη (Rappsilber et al. 2002, Makarov et al. 2002). Σε αυτές τις περιπτώσεις η χρωματογραφία συγγένειας γίνεται με μαλτόζη ή στρεπταβιδίνη που αναγνωρίζουν αντίστοιχα την MBP ή την βιοτίνη. Σε κάθε περίπτωση ακολουθεί υγρή χρωματογραφία ή ηλεκτροφόρηση δύο διαστάσεων σε συνδυασμό με φασματομετρία μάζας.

Για την απομόνωση των συστατικών των επιμέρους συμπλόκων του σωματίου συναρμογής, αφενός η *in vitro* αντίδραση συναρμογής διακόπτεται σε συγκεκριμένους χρόνους και αφετέρου το πρόδρομο μετάγραφο κατασκευάζεται με τέτοιο τρόπο ώστε να μην περιλαμβάνει απαραίτητες θέσεις για τη διεξαγωγή συγκεκριμένων σταδίων της συναρμογής (π.χ. 3' θέση συναρμογής) ή να διαθέτει μεταλλάξεις σε αυτές (π.χ. εκτεταμένη πολυπυριμιδική περιοχή) (Agafonov et al. 2011, Bessonov et al. 2008, 2010, Schmidt et al. 2014). Και σε αυτή την περίπτωση, ακολουθούν οι βιοχημικές διαδικασίες απομόνωσης και ταυτοποίησης που αναφέρθηκαν παραπάνω.

Στις περισσότερες μελέτες, πριν τη διαδικασία της βιοχημικής απομόνωσης, δεν πραγματοποιείται επεξεργασία με ριβονουκλεάση αλλά τα διάφορα μόρια snRNA απομονώνονται και αναλύονται σε αποδιατακτικό πήκτωμα ακρυλαμίδης.

Από τις μελέτες αυτές προκύπτει ότι ο αριθμός των πρωτεϊνών του σωματίου συναρμογής υπερβαίνει τις 200 είτε πρόκειται για παρασκευάσματα ολόκληρου του σωματίου συναρμογής (π.χ. Rappsilber et al. 2002, Makarov et al. 2002) είτε των επιμέρους υποσυμπλόκων του. Στην περίπτωση των υποσυμπλόκων, περισσότερες πρωτεΐνες έχουν ταυτοποιηθεί στις μελέτες που αφορούν τα υποσύμπλοκα E (267 πρωτεΐνες, Makarov et al. 2012), B (π.χ. 196 πρωτεΐνες, Bessonov et al. 2008), B *act* (π.χ. 211 πρωτεΐνες, Bessonov et al. 2010) και C (π.χ. 231 πρωτεΐνες, Schmidt et al. 2014). Αξιοσημείωτο είναι ότι οι πρωτεΐνες αυτές (τόσο στις μελέτες ολόκληρου του σωματίου όσο και σε αυτές των υποσυμπλόκων) δεν αφορούν αποκλειστικά παράγοντες συναρμογής. Ένα μεγάλο ποσοστό αυτών σχετίζονται με τη μεταγραφή (π.χ. μεταγραφικοί παράγοντες), με την αναδιαμόρφωση της χρωματίνης, την επεξεργασία του 5' και 3' άκρου, την αποικοδόμηση του mRNA όταν αυτό είναι ακατάλληλο για να μεταφραστεί και με την έξοδο του mRNA από τον πυρήνα. Το γεγονός αυτό μπορεί να θεωρηθεί αναμενόμενο διότι η συναρμογή συνδέεται λειτουργικά με τις παραπάνω διαδικασίες, όπως αναλύεται και στην επόμενη ενότητα.

#### 1.4.2. Λειτουργική σχέση της συναρμογής με άλλες συναφείς βιολογικές διαδικασίες

Η ρύθμιση της διεργασίας της συναρμογής αποτελεί ένα απαραίτητο κομμάτι της έκφρασης των ευκαρυωτικών γονιδίων. Εδώ και αρκετά χρόνια είναι πλέον γνωστό ότι η συναρμογή συμβαίνει ταυτόχρονα με τη μεταγραφή (Saldi et al. 2016, Maniatis & Reed 2002). Συγκεκριμένα έχει δειχθεί ότι οι παράγοντες συναρμογής SR συνεισφέρουν στη στρατολόγηση σημαντικών παραγόντων επιμήκυνσης της μεταγραφής στα σημεία παύσης της RNA πολυμεράσης II με αποτέλεσμα τη διευκόλυνση της διαδικασίας και την ταυτόχρονη έναρξη της συναρμογής (Saldi et al. 2016, Lin et al. 2008). Επιπλέον έχει παρατηρηθεί ότι το καρβοξυτελικό άκρο της

μεγαλύτερης υπομονάδας της RNA πολυμεράσης II λειτουργεί ως σημείο «συγκέντρωσης» για ποικίλους παράγοντες επεξεργασίας του RNA συμπεριλαμβανομένων αυτών που συμμετέχουν στην προσθήκη καλύπτρας στο 5' άκρο του μεταγράφου, στη διαδικασία της συναρμογής, στην επεξεργασία του 3' άκρου του μεταγράφου και στην έξοδο του mRNA από τον πυρήνα (Bentley 2005, Fong et al. 2003, Reed 2003). Τα παραπάνω αποτελούν και σημαντικές ενδείξεις για τη σύζευξη των διαδικασιών της μεταγραφής και της συναρμογής με τις λοιπές διεργασίες επεξεργασίας του mRNA (Proudfoot et al. 2002, Maniatis & Reed 2002).

Εκτός από την λειτουργική σύνδεση της συναρμογής με τη μεταγραφή, έχει δειχθεί ότι υπάρχει συσχέτιση της συναρμογής και με την αναδιαμόρφωση της χρωματίνης (Lucio et al. 2010, Gelfman et al. 2013). Οι παράγοντες αναδιαμόρφωσης της χρωματίνης έχει δειχθεί ότι αλληλεπιδρούν με παράγοντες των snRNPs (π.χ. του U2snRNP) οι οποίοι συμμετέχουν κυρίως στα αρχικά στάδια συναρμολόγησης του σωματίου συναρμογής (Allemand et al. 2016, Kfir et al. 2015, Haque & Oberdoerffer 2014). Η πιθανή λειτουργική σχέση της συναρμογής με την αναδιαμόρφωση της χρωματίνης υφίσταται έμμεσα μέσω της διαδικασίας της μεταγραφής (Merkhofer et al. 2014, Moehle et al. 2014, Brzyzek & Świeżewski 2015). Η λειτουργική σχέση της μεταγραφής με τους παράγοντες αναδιαμόρφωσης της χρωματίνης είναι εδραιωμένη. Η αποσυναρμολόγηση των νουκλεοσωμάτων και οι τροποποιήσεις των ιστονών (π.χ. μεθυλίωση, ακετυλίωση) διευκολύνουν την πρόσβαση της RNA πολυμεράσης στις θέσεις έναρξης της μεταγραφής ώστε να ξεκινήσει η διαδικασία αλλά και να στρατολογηθούν οι απαραίτητοι παράγοντες για το στάδιο της επιμήκυνσης της μεταγραφής (Venkatesh & Workman 2015, Lee et al. 2016, Teng et al. 2005, Cho et al. 2013). Ειδικότερα έχει δειχθεί ότι οι επιγενετικές αλλαγές της χρωματίνης (τροποποιήσεις ιστονών) σε συγκεκριμένα ιντρόνια μειώνουν τοπικά τον ρυθμό επιμήκυνσης της μεταγραφής με αποτέλεσμα να στρατολογούνται παράγοντες για την εναλλακτική συναρμογή (de Almeida & Carmo-Fonseca 2012).

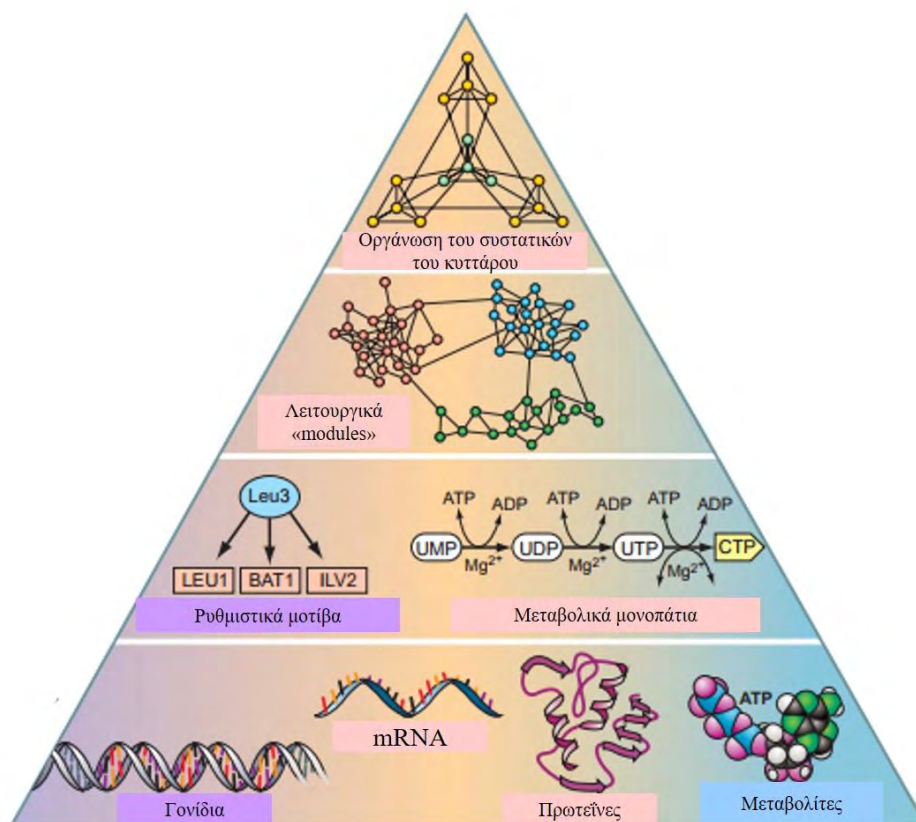
Επιπλέον, η συναρμογή σχετίζεται λειτουργικά με ποικίλες διεργασίες που αφορούν την επεξεργασία του mRNA όπως: α) την προσθήκη καλύπτρας στο 5' άκρο, ειδικότερα, το σύμπλοκο πρωτεϊνών, το οποίο ενώνεται στην 5' καλύπτρα, συνεισφέρει στην στρατολόγηση των U1 snRNP και του U4/U6·U5 snRNP στο πρόδρομο μετάγραφο μέσω αλληλεπιδράσεων με τα συστατικά τους (Pabis et al. 2013, Lewis et al. 1996), β) την επεξεργασία του 3' άκρου, καθώς παράγοντες συναρμογής (π.χ. συστατικά των U1 και U2 snRNP) στρατολογούν στο mRNA πρωτεΐνες οι οποίες είναι απαραίτητες για την αποκοπή και πολυαδενυλίωση του 3' άκρου του mRNA (Davidson & West 2013, Kaida et al. 2010, Kyburz et al. 2006, Niwa et al. 1990), γ) την έξοδο του ώριμου mRNA από τον πυρήνα, ειδικότερα, ένα σύμπλοκο πρωτεϊνών το οποίο στρατολογείται στα αρχικά στάδια συναρμολόγησης του σωματίου συναρμογής αλληλεπιδρά με παράγοντες που είναι υπεύθυνοι για την έξοδο αποκλειστικά μορίων mRNA από τον πυρήνα (Elbarbary & Maquat 2016, Katahira 2015) και δ) την αποικοδόμηση του ώριμου mRNA εάν αυτό δεν είναι κατάλληλο για να μεταφραστεί, καθώς οι πρωτεΐνες του συμπλόκου εξονίου-εξονίου (Exon Junction Complex, EJC) το οποίο στρατολογείται στο mRNA, στον πυρήνα, 24 νουκλεοτίδια ανοδικά των ορίων εξονίου-εξονίου, λειτουργούν ως μόρια «αναγνώρισης» στο κυτταρόπλασμα για

πρωτεΐνες που αλληλεπιδρούν με το ριβόσωμα προκαλώντας την παύση της μετάφρασης στο σημείο που υπάρχει λάθος (Hir et al. 2016, Chuang et al. 2015). Ακόμη, έχει προταθεί πιθανή συσχέτιση με τη ρύθμιση της μετάφρασης στο στάδιο έναρξης της διότι έχει δειχθεί ότι ορισμένοι παράγοντες συναρμογής λειτουργούν ως πρωτεΐνες-προσαρμογείς (adaptors) για τη στρατολόγηση μορίων τα οποία είναι υπεύθυνα για τη μετάφραση συγκεκριμένων μορίων mRNA (Michlewski et al. 2008). Για παράδειγμα, οι πρωτεΐνες του συμπλόκου EJC παραμένουν προσδεσμένες στο ώριμο mRNA και απομακρύνονται μετά τη συναρμολόγηση των ριβοσωμάτων στο κυτταρόπλασμα (Hir et al. 2016). Επίσης υπάρχει πιθανή σύνδεση με την μίτωση καθώς έχει αναφερθεί ότι η απουσία ορισμένων παραγόντων της συναρμογής διαταράσσει τη σύναψη των αδελφών χρωματίδων (sister chromatid cohesion) προκαλώντας προβλήματα στη στοίχιση των χρωμοσωμάτων (chromosome alignment) και στον επακόλουθο σωστό διαχωρισμό των χρωμοσωμάτων, καθυστερώντας την διαδικασία της μίτωσης (van der Lelij et al. 2014, Sundaramoorthy et al. 2014, Watrin et al. 2014, Neumann et al. 2010).

Από τα παραπάνω έχει γίνει κατανοητό αφενός ότι η συναρμογή είναι μια πολύπλοκη διαδικασία η οποία προϋποθέτει τη συνεργασία πολλών μορίων ώστε να πραγματοποιηθεί με ακρίβεια η αφαίρεση των ιντρονίων και αφετέρου ότι συνδέεται λειτουργικά με συναφείς βιολογικές διαδικασίες. Τα τελευταία χρόνια έχουν πραγματοποιηθεί διάφορες ολιστικές προσεγγίσεις για το σωματίο συναρμογής οι οποίες είχαν στόχο την αποσαφήνιση του πολύπλοκου δικτύου των πρωτεϊνικών αλληλεπιδράσεων που λαμβάνουν χώρα κατά τη συναρμολόγηση του σωματίου συναρμογής στο πρόδρομο μετάγραφο, τον έλεγχο της δυναμικής του σωματίου καθώς και τον προσδιορισμό του δικτύου των πρωτεϊνών που ρυθμίζουν την εναλλακτική συναρμογή. Πριν γίνει η περιγραφή των μελετών αυτών, θα προηγηθεί μια εισαγωγή στη Δικτυακή Βιολογία, δηλαδή, στον τομέα της επιστήμης που πραγματεύεται την ανακατασκευή δικτύων αλληλεπιδράσεων, ώστε να γίνει κατανοητή η ανάλυση των δικτύων που δημιουργήθηκαν στις μελέτες αυτές.

### **1.5. Εισαγωγή στη Δικτυακή Βιολογία (Network Biology) υπό το πρίσμα της Συστημικής Βιολογίας**

Για παραπάνω από έναν αιώνα, η βιολογική έρευνα στηριζόταν στην πρακτική της ανάλυσης επιμέρους απλούστερων διαδικασιών που συνδυαστικά συμμετείχαν σε μια πιο πολύπλοκη διαδικασία, με στόχο να ανακτήσει πληροφορίες σχετικά με μεμονωμένα κυτταρικά συστατικά και τις λειτουργίες αυτών ώστε να εξηγήσει αυτές τις διεργασίες που παρουσίαζαν μεγαλύτερη πολυπλοκότητα. Παρ' όλο που ο όγκος και το είδος των πληροφοριών στις επιμέρους αναλύσεις ήταν μεγάλος και αξιόπιστος, γίνεται όλο και περισσότερο σαφές ότι μια διακριτή λειτουργία σπάνια μπορεί να αποδοθεί σε ένα μεμονωμένο μόριο. Αντιθέτως, οι περισσότερες βιολογικές διεργασίες προκύπτουν από σύνθετες αλληλεπιδράσεις μεταξύ των κυτταρικών συστατικών, όπως πρωτεϊνών, νουκλεϊκών οξέων και μικρών χημικών μορίων (Εικ. 12) (Oltvai & Barabási 2002).

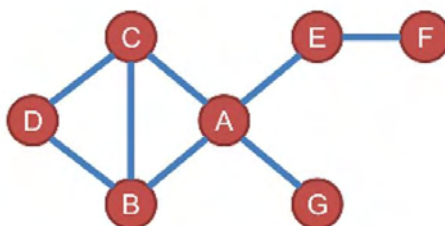


**Εικ. 12: Οργάνωση των συστατικών του κυττάρου ως ένα δίκτυο αλληλεπιδράσεων.** Η βάση της πυραμίδας αντιπροσωπεύει την κυτταρική οργάνωση στα τέσσερα λειτουργικά επίπεδα, σε γονιδιακό, μεταγραφικό, πρωτεϊνικό και μεταβολικό. Η αναπαράσταση της κυτταρικής οργάνωσης σε μορφή δικτύου (4<sup>ο</sup> επίπεδο της πυραμίδας) μπορεί να γίνει μέσω των λειτουργικών συνδέσεων των συστατικών του κυττάρου, σε ρυθμιστικό ή μεταβολικό επίπεδο (2<sup>ο</sup> επίπεδο της πυραμίδας), τα οποία σχηματίζουν λειτουργικά συμπλέγματα μορίων (modules) (3<sup>ο</sup> επίπεδο της πυραμίδας) που ακολούθως ενώνονται μεταξύ τους με στόχο τη διεκπεραίωση των κυτταρικών λειτουργιών (Oltvai & Barabási 2002).

Έτσι, ένας κύριος στόχος της μετα-γονιδιωμιακής βιολογικής έρευνας είναι η κατανόηση της δομής και της δυναμικής του πολύπλοκου δικτύου αλληλεπιδράσεων των κυτταρικών συστατικών το οποίο συνεισφέρει στη δομή και στη λειτουργία των ζωντανών κυττάρων. Η ανάπτυξη τεχνικών μεγάλης ομικής κλίμακας σε όλα τα κυτταρικά επίπεδα, σε συνδυασμό με τις τεχνικές ταυτοποίησης αλληλεπιδράσεων, συνεισφέρουν στον προσδιορισμό του τρόπου και του χρόνου που λαμβάνουν χώρα οι αλληλεπιδράσεις μεταξύ των διαφόρων μορίων. Ποικίλοι τύποι δικτύων αλληλεπιδράσεων μπορούν να προκύψουν από το σύνολο των διαφόρων ειδών συνδέσεων μεταξύ των κυτταρικών συστατικών. Χαρακτηριστικά παραδείγματα αποτελούν τα ρυθμιστικά, τα πρωτεϊνικά, τα μεταβολικά δίκτυα και τα δίκτυα σηματοδότησης. Ο συνδυασμός όλων αυτών των δικτύων προσπαθεί να αναπαραστήσει την κυτταρική συμπεριφορά. Έτσι, ο καθορισμός και η ανάλυση των αρχών οργάνωσης των κυτταρικών δικτύων συνεισφέρει στην κατανόηση του κυττάρου ως ένα σύστημα. Ο τομέας της συστημικής βιολογίας στοχεύει στην ενσωμάτωση του τεράστιου όγκου δεδομένων που παράγονται για τα διαφορετικά επίπεδα κυτταρικής λειτουργίας σε μαθηματικές θεωρίες και μοντέλα συνεισφέροντας στην εξαγωγή συμπερασμάτων για τις λειτουργικές συνδέσεις μεταξύ των κυτταρικών συστατικών, υπό το πρίσμα της θεωρίας της δομής και λειτουργίας του συστήματος

(Wuchty 2014, Barabási & Oltvai 2004, Dotan-Cohen et al. 2009). Η κατανόηση των αρχών οργάνωσης των κυτταρικών δικτύων αποτελεί ένα σημαντικό εργαλείο για τους πειραματικούς βιολόγους, καθώς η συνδεσιμότητα των υπό εξέταση μορίων στα διάφορα είδη δικτύων μπορεί να κατευθύνει τους ερευνητές ως προς τα πειράματα που θα ακολουθήσουν για τη διερεύνηση του βιολογικού ρόλου των εξεταζόμενων μορίων (Wuchty 2014, Barabási & Oltvai 2004, Hakes et al. 2008). Ακόμη, πληροφορίες για ένα μόριο με μη χαρακτηρισμένη λειτουργία μπορούν να εξαχθούν και από τις λειτουργίες των μορίων με τα οποία αλληλεπιδρά στο πλαίσιο ενός δικτύου. Επιπλέον, η ύπαρξη των αλληλεπιδράσεων μεταξύ των μορίων σε οποιοδήποτε κυτταρικό επίπεδο (π.χ. γονίδιο ή πρωτεΐνη ή μεταβολίτης) υποδεικνύει ότι οι επιπτώσεις μιας βιολογικής αλλαγής δεν περιορίζονται μόνο στο μόριο που τη φέρει, αλλά μπορεί να εξαπλωθούν στα αλληλεπιδρώντα μόρια μέσα στο δίκτυο επηρεάζοντας τη λειτουργικότητα των συστατικών του δικτύου.

Η δικτυακή βιολογία, αφορά τη μαθηματική μοντελοποίηση και απεικόνιση των αλληλεπιδράσεων των κυτταρικών συστατικών υπό τη μορφή δικτύων ή, με μαθηματικούς όρους, υπό τη μορφή γράφων. Στα δίκτυα-γράφοις, η πολυπλοκότητα των βιολογικών συστημάτων ανάγεται σε μια απλούστερη μορφή. Έτσι, τα μόρια που εξετάζονται κάθε φορά (π.χ. πρωτεΐνες, γονίδια, RNA, μεταβολίτες) απεικονίζονται με τη μορφή κόμβων (nodes) και οι αλληλεπιδράσεις με τη μορφή ακμών (edges) (Εικ. 13).



**Εικ. 13:** Ενδεικτικό δίκτυο πρωτεϊνικών αλληλεπιδράσεων. Με κόκκινους κύκλους φαίνονται οι κόμβοι (nodes) του δικτύου που αντιστοιχούν στις πρωτεΐνες A, B, C, D, E, F και G και με μπλε γραμμές οι ακμές (edges) που αντιπροσωπεύουν τις πρωτεϊνικές αλληλεπιδράσεις (Magger et al. 2012).

Κάθε ακμή αντιπροσωπεύει μια φυσική, βιοχημική ή λειτουργική αλληλεπίδραση μεταξύ δυο κόμβων (Barabási & Oltvai 2004, Fadhal et al. 2014). Με βάση τη φύση των αλληλεπιδράσεων, τα δίκτυα μπορεί να είναι κατευθυνόμενα ή μη κατευθυνόμενα. Στα κατευθυνόμενα δίκτυα, η αλληλεπίδραση μεταξύ δυο κόμβων έχει ένα συγκεκριμένο προσανατολισμό. Για παράδειγμα, μπορεί να αντιπροσωπεύει την κατεύθυνση της ροής της πληροφορίας από ένα μεταγραφικό παράγοντα στο γονίδιο που ρυθμίζει ή σε μια μεταβολική αντίδραση, την κατεύθυνση της ροής από το υπόστρωμα στο προϊόν. Στα μη κατευθυνόμενα δίκτυα, οι συνδέσεις δεν έχουν έναν καθορισμένο προσανατολισμό (π.χ. δίκτυα πρωτεϊνικών αλληλεπιδράσεων), δηλαδή, μια ακμή αντιπροσωπεύει την αμοιβαία αλληλεπίδραση μεταξύ δυο μορίων (π.χ. πρωτεϊνών) (Fadhal et al. 2014, Barabási & Oltvai 2004).

### 1.5.1. Δίκτυα πρωτεϊνικών αλληλεπιδράσεων

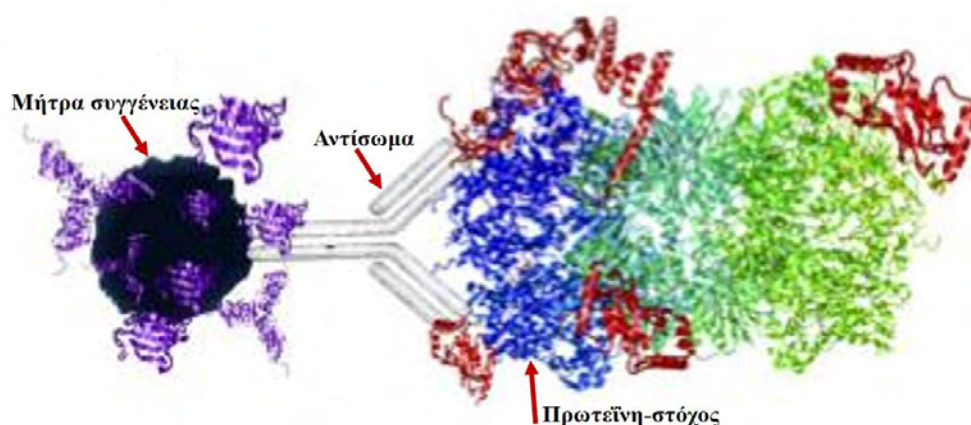
Αν και τα μεταβολικά δίκτυα είναι τα πιο ολοκληρωμένα από άποψη πληροφορίας σε σχέση με τα άλλα βιολογικά δίκτυα καθώς υπάρχει αρκετά μεγάλος αριθμός δεδομένων που αφορούν συγκεκριμένες βιοχημικές αντιδράσεις (οι κόμβοι μπορεί να απεικονίζουν μεταβολίτες ή ένζυμα) (Lee et al. 2008), η ανάπτυξη των τεχνολογιών που αφορούν την ταυτοποίηση των αλληλεπιδράσεων έχει αποδώσει ένα μεγάλο όγκο δεδομένων πρωτεϊνικών αλληλεπιδράσεων τόσο σε οργανισμούς μοντέλα (Simonis et al. 2009, Guruharsha et al. 2011, Schwikowski et al. 2000, Gavin et al. 2006, Vidal 2000, Stuart et al. 2007, Rain et al. 2001) όσο και στον άνθρωπο (Rolland et al. 2014, Stelzl et al. 2005, Rual et al. 2005, Barrios-Rodiles et al. 2005). Τα δεδομένα των αλληλεπιδράσεων σε συνδυασμό με την τοπολογία των πρωτεϊνών στα πρωτεϊνικά δίκτυα καθίστανται απαραίτητα εργαλεία για την κατανόηση των κυτταρικών λειτουργιών και πιθανώς την εύρεση κοινών ομάδων πρωτεϊνών που ευθύνονται για μια ασθένεια (Cannistraci et al. 2013, Barabási et al. 2011). Μέχρι τώρα, εκτιμάται ότι έχει ταυτοποιηθεί ένα μεγάλο ποσοστό (~50%) των πρωτεϊνικών αλληλεπιδράσεων στο σακχαρομύκητα ενώ στον άνθρωπο μόλις το 10-20% του συνόλου των αλληλεπιδράσεων (Rolland et al. 2014, Hein et al. 2015). Το είδος της πληροφορίας που χρησιμοποιείται για την κατασκευή των δικτύων μπορεί να ποικίλλει. Ειδικότερα, οι πληροφορίες μπορεί να προέρχονται είτε από πειραματικές μεθόδους που περιλαμβάνουν πειράματα μικρής ή μεγάλης κλίμακας τα οποία εξετάζουν άμεσες (Hao et al. 2016, Parrish et al. 2006, Ito et al. 2002) και έμμεσες (Hao et al. 2016, Köcher & Superti-Furga 2007) αλληλεπιδράσεις ή από μεθόδους πρόβλεψης (βλ. παρακάτω).

#### 1.5.1.1. Προσδιορισμός πρωτεϊνικών αλληλεπιδράσεων

Σε κυτταρικό επίπεδο είναι γνωστό ότι οι πρωτεΐνες σπάνια λειτουργούν μόνες τους αλλά αλληλεπιδρούν με άλλες πρωτεΐνες για να εκτελέσουν συγκεκριμένες κυτταρικές λειτουργίες. Κάθε λειτουργία μέσα στο κύτταρο λαμβάνει χώρα μέσω της συντονισμένης δράσης πολλών πρωτεϊνικών μορίων. Αυτές οι πρωτεϊνικές μηχανές περιέχουν ιδιαίτερα καλά συντονισμένα κινούμενα μέρη τα οποία λειτουργούν με τέτοιο τρόπο ώστε να εκτελούν τις βιολογικές τους λειτουργίες σε συγκεκριμένο χώρο και χρόνο (Rao et al. 2014, Koh et al. 2012, Alberts 1998, Gavin et al. 2002). Για την ταυτοποίηση των πρωτεϊνών ενός πρωτεϊνικού συμπλόκου καθώς και για τον προσδιορισμό των άμεσων αλληλεπιδράσεων έχουν αναπτυχθεί αρκετές μέθοδοι:

- Μέθοδοι προσδιορισμού πρωτεϊνικών αλληλεπιδράσεων εντός πρωτεϊνικού συμπλόκου

Μια από τις πλέον διαδεδομένες τεχνικές προσδιορισμού πρωτεϊνικών αλληλεπιδράσεων εντός ενός πρωτεϊνικού συμπλόκου είναι η χρωματογραφία συγγένειας κατά την οποία ένα μόριο-συνδέτης ακινητοποιείται σε μια στερεή επιφάνεια (μήτρα) η οποία συνήθως αποτελείται από πολυμερή σφαιρίδια που είναι τοποθετημένα σε μια στήλη. Ακολουθεί η προσθήκη του πρωτεϊνικού εκχυλίσματος στη στήλη κατά την οποία οι πρωτεΐνες που αλληλεπιδρούν με το μόριο-συνδέτη προσκολλώνται σε αυτό, και ακολούθως εκλύονται από τη στήλη. Μπορεί να ακολουθήσει ταυτοποίηση του συνόλου των πρωτεϊνών με φασματομετρία μάζας ή ορισμένων από αυτών με ανάλυση western (Nesvizhskii 2012, Urh et al. 2009). Ένα παράδειγμα χρωματογραφίας συγγένειας, είναι η χρήση κατάλληλου υποστρώματος (π.χ. γλουταθειόνη) που λειτουργεί ως μόριο-συνδέτης. Αυτό το υπόστρωμα συνδέεται με μία ανασυνδυασμένη πρωτεΐνη μέσω ενός επιτόπου που αυτή περιέχει (π.χ. S-τρανσφεράση της γλουταθειόνης). Στη συνέχεια προστίθεται πρωτεϊνικό εκχύλισμα και ταυτοποιούνται οι πρωτεΐνες που αλληλεπιδρούν με την ανασυνδυασμένη πρωτεΐνη. Όταν χρησιμοποιούνται δυο καθαρισμένες και ανασυνδυασμένες πρωτεΐνες στην παραπάνω μέθοδο, τότε πρόκειται για συν-κατακρήμιση η οποία ταυτοποιεί άμεσες πρωτεϊνικές αλληλεπιδράσεις (Einarson 2004). Ένα εξειδικευμένο είδος χρωματογραφίας συγγένειας είναι η ανοσοκατακρήμιση κατά την οποία το μόριο-συνδέτης είναι ένα ειδικό αντίσωμα για την πρωτεΐνη-στόχο, ακολουθεί επώαση με πρωτεϊνικό εκχύλισμα και ταυτοποιούνται οι πρωτεΐνες που αλληλεπιδρούν με την πρωτεΐνη-στόχο (Εικ. 14). Εναλλακτικά, για την ταυτοποίηση μίας πρωτεΐνης αντί ολόκληρου συμπλόκου, χρησιμοποιείται η συν-ανοσοκατακρήμιση κατά την οποία ακολουθείται η ίδια διαδικασία με την ανοσοκατακρήμιση, με τη διαφορά ότι ακολουθεί η ταυτοποίηση μιας συγκεκριμένης πρωτεΐνης που αλληλεπιδρά με την πρωτεΐνη-στόχο μέσω ειδικού αντισώματος (Antrobus & Borner 2011).



**Εικ. 14:** Ενδεικτική απεικόνιση της ανοσοκατακρήμισης. Με κόκκινα βέλη φαίνονται η μήτρα συγγένειας (σφαιρίδια), το ειδικό αντίσωμα και η πρωτεΐνη-στόχος. Οι πρωτεΐνες του συμπλόκου που απομονώνεται φαίνονται σε μορφή σπειροειδούς έλικας με πράσινο ανοιχτό & σκούρο, κόκκινο και μωβ (μη ειδικές αλληλεπιδράσεις) χρώμα (Ten Have et al. 2011).

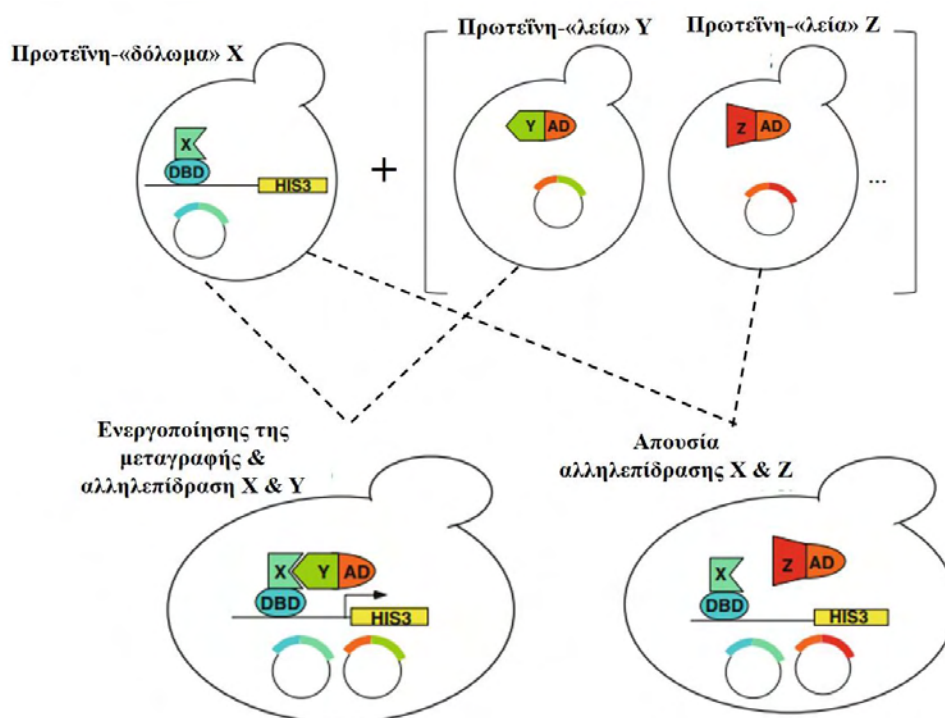
Μια άλλη μέθοδος που χρησιμοποιείται είναι οι κλασματώσεις (co-fractionation), δηλαδή, ο διαχωρισμός των κυτταρικών συστατικών βάσει μεγέθους και πυκνότητας



μέσω υπερφυγοκέντρωσης (Borner et al. 2014, Alberts et al. 2002). Σε αυτή τη μέθοδο ακολουθεί φασματομετρία μάζας για την ταυτοποίηση των πρωτεϊνών.

- Μέθοδοι προσδιορισμού άμεσων πρωτεϊνικών αλληλεπιδράσεων

Μια από τις πιο διαδεδομένες τεχνικές που χρησιμοποιείται για τον έλεγχο των άμεσων πρωτεϊνικών αλληλεπιδράσεων είναι το σύστημα δυο υβριδίων του σακχαρομύκητα τόσο σε επίπεδο μικρής κλίμακας όσο και σε μελέτες μεγάλης κλίμακας. Με αυτή την τεχνική εξετάζονται οι αλληλεπιδράσεις ζευγών πρωτεϊνών που καλούνται «δολώμα» (bait) και «λεία» (prey). Αυτές συντήκονται κατ' αντιστοιχία με μια περιοχή σύνδεσης στο DNA και μια περιοχή ενεργοποίησης της μεταγραφής ενός μεταγραφικού παράγοντα (GAL4) στο σακχαρομύκητα. Όταν λαμβάνει χώρα αλληλεπίδραση μεταξύ του «δολώματος» και της «λείας», αυτές οι περιοχές έρχονται κοντά και ενεργοποιούν τη μεταγραφή ενός ή περισσότερων γονιδίων αναφοράς, για παράδειγμα, τη β-γαλακτοζιδάση η οποία οδηγεί στη σήμανση του κυττάρου του σακχαρομύκητα όταν χρησιμοποιείται χρωματομετρικό υπόστρωμα, και το γονίδιο *HIS3* το οποίο είναι υπεύθυνο για τη βιοσύνθεση της ιστιδίνης και οι μεταλλάξεις αυτού συνεισφέρουν στην επιλογή αποικιών κυττάρων του σακχαρομύκητα απουσία ιστιδίνης (Εικ. 15) (Roberts et al. 2012, Fields & Song 1989, Brückner et al. 2009).



**Εικ. 15: Ενδεικτική απεικόνιση του συστήματος δυο υβριδίων στο σακχαρομύκητα.** Φαίνονται η σύνδεση του «δολώματος» X και της «λείας» Y και η ενεργοποίηση της μεταγραφής του γονιδίου αναφοράς (*HIS3*), καθώς και η απουσία σύνδεσης μεταξύ του ίδιου δολώματος με μία άλλη πρωτεΐνη-«λεία» Z. DBD: περιοχή σύνδεσης στο DNA, AD: περιοχή ενεργοποίησης της μεταγραφής ενός μεταγραφικού παράγοντα (GAL4) στο σακχαρομύκητα (Roberts et al. 2012).

Τέτοια πειράματα έχουν πραγματοποιηθεί σε διάφορους οργανισμούς όπως στους οργανισμούς-μοντέλα *S. cerevisiae* (Uetz et al. 2000), *D. melanogaster* (Formstecher

et al. 2005) και *C. elegans* (Obrdlík et al. 2004) και στον άνθρωπο (Rolland et al. 2014, Rual et al. 2005, Stelzl et al. 2005).

Δυο άλλα είδη βιοφυσικών τεχνολογιών για την μελέτη των άμεσων πρωτεϊνικών αλληλεπιδράσεων είναι: α) η μέθοδος Μεταφοράς Ενέργειας Φθορισμού μέσω Συντονισμού (Fluorescence Resonance Energy Transfer - FRET) η οποία στηρίζεται στη μεταφορά ενέργειας μεταξύ δυο φθορίζοντων μορίων (μόριο-δότης και μόριο δέκτης), λόγω της οποίας εκπέμπεται φθορισμός όταν τα δυο αλληλεπιδρώντα μόρια βρίσκονται σε πολύ κοντινή απόσταση μεταξύ τους (της τάξεως του 1nm) (Clegg 2004). Η μέθοδος αυτή αποτέλεσε μια τεχνική για τον έλεγχο άμεσων πρωτεϊνικών αλληλεπιδράσεων της πρωτεΐνης FRA10AC1, όπως περιγράφεται στα αποτελέσματα, και β) η μέθοδος συντονισμού πλασμονίων (ταλάντωση των ηλεκτρονίων) μέσω επιφάνειας (surface plasmon resonance-SPR) στην οποία χρησιμοποιείται μια μικροσυστοιχία (τσιπ αισθητήρα) που είναι καλυμμένη με μια ακινητοποιημένη πρωτεΐνη-στόχο και ακολουθεί επώαση με ένα πρωτεϊνικό εκχύλισμα. Έπειτα από τη σύνδεση των αλληλεπιδρώντων μορίων στο ακινητοποιημένο υπόστρωμα πραγματοποιείται η ποσοτική μέτρηση της σύνδεσης των εκάστοτε μορίων, δηλαδή η διάρκεια της κάθε αλληλεπίδρασης σε πραγματικό χρόνο. Η ταυτοποίηση των αλληλεπιδρώντων μορίων μπορεί να γίνει με φασματομετρία μάζας (Drescher et al. 2009).

Άλλες μέθοδοι είναι οι μελέτες συν-κρυστάλλωσης, η ανάλυση far western και οι *in vitro* δοκιμές (π.χ. δοκιμές φωσφορυλίωσης ή αποφωσφορυλίωσης).

#### 1.5.1.1.1. Περιορισμοί των μεθόδων προσδιορισμού πρωτεϊνικών αλληλεπιδράσεων

Οι πιο διαδεδομένες τεχνικές από αυτές που αναφέρθηκαν παραπάνω για την ταυτοποίηση πρωτεϊνικών αλληλεπιδράσεων είναι το σύστημα δύο υβριδίων και η ανοσοκατακρήμνιση σε συνδυασμό με φασματομετρία μάζας (Rao et al. 2014, Figeys 2008). Παρ' όλα αυτά, καμία από αυτές τις τεχνικές ή άλλες εναλλακτικές προσεγγίσεις δεν μπορούν από μόνες τους να συμβάλλουν στον πλήρη προσδιορισμό όλου του φάσματος των αλληλεπιδράσεων, καθώς κάθε μία από αυτές καλύπτει ένα μικρό ποσοστό τους (Ewing et al. 2007, Futschik et al. 2007, Regulý et al. 2006, von Mering et al. 2002). Η συμπληρωματική φύση των τεχνικών αυτών μπορεί να αποδοθεί στα εγγενή χαρακτηριστικά τους, συμπεριλαμβανομένου του είδους των αλληλεπιδράσεων που ταυτοποιούνται σε αυτές, σε συνδυασμό με τα ψευδώς θετικά αποτελέσματα που προκύπτουν (Hayes et al. 2016, Figeys 2008).

Όσον αφορά το είδος των αλληλεπιδράσεων, στην περίπτωση των πρωτεϊνικών συμπλόκων μπορούν να ανιχνευθούν αλληλεπιδράσεις, οι οποίες λαμβάνουν χώρα ή σταθεροποιούνται με τη συμμετοχή άλλων μορίων. Επιπλέον, η μέθοδος χρωματογραφίας συγγένειας, σε συνδυασμό με τη φασματομετρία μάζας, εμφανίζει μεροληψία ως προς τις πρωτεΐνες που βρίσκονται σε αφθονία, λόγω ευαισθησίας της μεθόδου. Αντιθέτως με το σύστημα δύο υβριδίων στο σακχαρομύκητα εξετάζεται κάθε φορά ένα ζεύγος πρωτεϊνών (von Mering et al. 2002) και η τεχνική αυτή εμφανίζει μικρότερο ποσοστό μεροληψίας, καθώς οι πρωτεΐνες εκφράζονται σε συγκρίσιμα επίπεδα από έναν ετερόλογο υποκινητή (Dunham et al. 2012, von Mering et al. 2002).

Στις μεθόδους απομόνωσης πρωτεϊνικών συμπλόκων, τα ψευδώς θετικά αποτελέσματα αποδίδονται στη διαταραχή του βιολογικού συστήματος η οποία μπορεί να προκληθεί σε διάφορα στάδια των τεχνικών αυτών. Για παράδειγμα η διαταραχή μπορεί να προκληθεί κατά την προσθήκη ετικετών στην πρωτεΐνη-στόχο (συνέχεια του μορίου συνδέτη) ή στο βήμα της κυτταρική λύσης, το οποίο διαταράσσει τις ασθενείς και παροδικές αλληλεπιδράσεις ή κατά την εφαρμογή πολλαπλών πλύσεων στα βήματα καθαρισμού (Sahni et al. 2015, Taipale et al. 2014, Berger et al. 2013, Xu et al. 2010, Li et al. 2010).

Στη μέθοδο των δυο υβριδίων εμφανίζονται δύο είδη ψευδώς θετικών αποτελεσμάτων, τεχνικής και βιολογικής φύσεως. Η πρώτη κατηγορία μπορεί να είναι απόρροια διαφόρων παραγόντων: i) μεταλλάξεων στα γονίδια αναφοράς, ii) λανθασμένης αναδίπλωσης πρωτεϊνών στο μη φυσικό τους περιβάλλον, iii) έλλειψης απαραίτητων μετα-μεταφραστικών τροποποιήσεων και τέλος iv) ενεργοποίησης του γονιδίου αναφοράς χωρίς να έχει προηγηθεί αλληλεπίδραση του «δολώματος» (bait) με τη «λεία» (prey) («self-activating baits») (Rao et al. 2014, Snider et al. 2015, Scott & Barton 2007). Αντιθέτως η δεύτερη κατηγορία αντιπροσωπεύει αλληλεπιδράσεις, οι οποίες δεν λαμβάνουν χώρα *in vivo* στους ίδιους κυτταρικούς τύπους ή στα ίδια οργανίδια αλλά πραγματοποιούνται όταν οι δυο πρωτεΐνες εκφραστούν σε ετερόλογο σύστημα (Yu & Finley 2009, Ge et al. 2003, Ito et al. 2001).

Οι παραπάνω περιορισμοί αυξάνονται σε πειράματα μεγάλης κλίμακας (high-throughput) καθώς ο αριθμός των εξεταζόμενων πρωτεϊνών είναι μεγάλος με αποτέλεσμα να μην είναι συνήθως εφικτή η επιβεβαίωση των πρωτεϊνικών αλληλεπιδράσεων με εναλλακτικές πειραματικές προσεγγίσεις και κατ' επέκταση να αυξάνεται ο αριθμός των ψευδώς θετικών αποτελεσμάτων (Ma et al. 2017, Snider et al. 2015, Yu & Finley 2009, Scott & Barton 2007). Αντιθέτως, στα πειράματα μικρής κλίμακας πραγματοποιείται στοχευμένη εξέταση συγκεκριμένων πρωτεϊνικών αλληλεπιδράσεων με περισσότερες από μία πειραματικές μεθόδους οι οποίες καθιστούν εφικτό τον έλεγχο της εγκυρότητας των αποτελεσμάτων περιορίζοντας σε μεγάλο βαθμό τα ψευδώς θετικά αποτελέσματα.

Έτσι, παρ' όλο που οι μελέτες μεγάλης κλίμακας αποτελούν μία πλούσια πηγή πληροφοριών για τις πρωτεϊνικές αλληλεπιδράσεις, καθίσταται επιτακτική η ανάγκη πιστοποίησής τους με βιοχημικές στοχευμένες μελέτες μικρής κλίμακας (Hayes et al. 2016) ώστε να αυξηθεί η αξιοπιστία τους. Επιπλέον, στην αύξηση της αξιοπιστίας των αλληλεπιδράσεων συνεισφέρει η εύρεση βιβλιογραφικών πηγών σχετικών τόσο με τα δεδομένα έκφρασης και τοπολογίας των πρωτεϊνών στο κύτταρο, όσο και με την κυτταρική τους λειτουργία (Figeys 2008).

#### 1.5.1.2. In silico προσέγγιση κατασκευής πρωτεϊνικών δικτύων

Αν και οι πειραματικές μέθοδοι που αναφέρθηκαν παραπάνω αποτελούσαν μέχρι πρόσφατα την κύρια επιλογή για την εξέταση των πρωτεϊνικών αλληλεπιδράσεων, τα τελευταία χρόνια έχει παρατηρηθεί και μια μετάβαση από τις πειραματικές (proteomics-based) προς τις υπολογιστικές προσεγγίσεις (*in silico*) (Ngounou et al. 2014). Οι *in silico* προσεγγίσεις προσφέρουν τη δυνατότητα πρόβλεψης όλων των πιθανών πρωτεϊνικών αλληλεπιδράσεων και στηρίζονται στη χρήση ειδικών εργαλείων

βιοπληροφορικής και στην εφαρμογή αλγορίθμων. Ειδικότερα η πρόβλεψη των αλληλεπιδράσεων μπορεί να γίνει με βάση (Zahiri et al. 2013): α) την ομοιότητα της δομής των πρωτεϊνών που εξετάζονται με πρωτεΐνες που είναι γνωστό ότι αλληλεπιδρούν, ώστε να υπάρχει πιθανότητα τα μέλη του πρώτου ζεύγους πρωτεϊνών να αλληλεπιδρούν μεταξύ τους (Smith & Sternberg 2002), β) την εύρεση κοινών χαρακτηριστικών περιοχών (domains) στις εξεταζόμενες πρωτεΐνες με ζεύγη πρωτεϊνών που είναι γνωστό ότι αλληλεπιδρούν (Sprinzak & Margalit 2001), γ) την ύπαρξη αλληλεπιδράσεων μεταξύ των ορθόλογων μορίων των πρωτεϊνών που εξετάζονται (Matthews et al. 2001), και δ) το κοινό πρότυπο έκφρασης των γονιδίων που κωδικοποιούν τις πρωτεΐνες που εξετάζονται (Greenbaum et al. 2002). Με βάση τα παραπάνω κριτήρια συχνά ορίζεται μια τιμή αξιοπιστίας των πιθανών αλληλεπιδράσεων ώστε να επιλεγθούν αυτές με τον μεγαλύτερο βαθμό αξιοπιστίας για να πραγματοποιηθούν στοχευμένα πειράματα.

#### 1.5.1.3. Κατασκευή πρωτεϊνικών δικτύων μέσω της ανάκτησης δεδομένων από βάσεις πρωτεϊνικών αλληλεπιδράσεων

Τα τελευταία χρόνια τα πειραματικά δεδομένα για τις πρωτεϊνικές αλληλεπιδράσεις ενσωματώνονται συνεχώς σε δημόσιες βάσεις δεδομένων, οι οποίες είτε στηρίζονται στη βιβλιογραφία είτε περιέχουν προβλεπόμενες αλληλεπιδράσεις (Cusick et al. 2009). Οι πρώτες μπορεί να περιέχουν δεδομένα τόσο για πειράματα μεγάλης κλίμακας όσο και μικρής κλίμακας, είτε για άμεσες ή για έμμεσες αλληλεπιδράσεις. Στις δεύτερες, τα δεδομένα προκύπτουν από μεθόδους πρόβλεψης (prediction methods) μέσω της χρήσης ειδικών εργαλείων βιοπληροφορικής. Η συλλογή των δεδομένων που προκύπτουν και από τα δυο είδη βάσεων μπορεί να οδηγήσει στην κατασκευή δικτύων τα οποία μπορούν να αξιολογηθούν περαιτέρω ανάλογα με το ερώτημα που θέλει ο μελετητής να απαντήσει κάθε φορά (de Las Rivas & Fontanillo 2010).

Παρόλο που ο όγκος της πληροφορίας που είναι καταχωρημένη στις βάσεις πρωτεϊνικών αλληλεπιδράσεων είναι μεγάλος, παρατηρείται ποικιλομορφία στα διαθέσιμα σύνολα δεδομένων των βάσεων, η οποία αποδίδεται κυρίως στην ενσωμάτωση πληροφοριών από διαφορετικές βιβλιογραφικές αναφορές (Turinsky et al. 2010, Orchard et al. 2012). Συγκεκριμένα το περιορισμένο ποσοστό αλληλοεπικάλυψης τους έχει αποδοθεί: α) στα διαφορετικά κριτήρια ανάκτησης των δεδομένων (literature mining criteria), ακόμα και από τις ίδιες βιβλιογραφικές αναφορές, β) στις διαφορετικές μεθόδους επιλογής, αξιολόγησης και ενημέρωσης (curation methods) των πρωτεϊνικών αλληλεπιδράσεων, και γ) στις διαφορές των ποσοστών ενσωμάτωσης δεδομένων από πειραματικές μελέτες μικρής κλίμακας (Klapa et al. 2013, Turinsky et al. 2010, Cusick et al. 2009, Klingström & Plewczynski 2011, Mathivanan et al. 2006). Επιπλέον, αρκετά συχνά, στις πληροφορίες που ανακτώνται από τις βάσεις δεδομένων, παρουσιάζονται αντιφάσεις και ποσοστά λάθους (Cusick et al. 2009). Το μεγαλύτερο ποσοστό λάθους αποδίδεται κυρίως στην αυτοματοποιημένη ανάκτηση των δεδομένων είτε από τις περιλήψεις είτε από ολόκληρα τα κείμενα των ερευνητικών εργασιών, μέσω της χρήσης λέξεων κλειδιών (text-mining) από τους επιμελητές της κάθε βάσης. Αυτή η προσέγγιση αποτελεί κυρίαρχο τρόπο άντλησης των πληροφοριών (Krallinger et al. 2008). Για να

αποφευχθεί αυτό, και με δεδομένο το μεγάλο ποσοστό ψευδώς θετικών αποτελεσμάτων των μελετών μεγάλης κλίμακας, απαιτείται συνδυαστικά και η λεπτομερής διερεύνηση της βιβλιογραφίας που είναι σχετική με το αντικείμενο μελέτης του εκάστοτε ερευνητή (Turinsky et al. 2010).

Μια εναλλακτική επιλογή ανάκτησης δεδομένων πρωτεϊνικών αλληλεπιδράσεων, δεδομένου του περιορισμένου ποσοστού αλληλοεπικάλυψης των καταχωρήσεων των βάσεων δεδομένων που προαναφέρθηκε, είναι η χρήση μετα-βάσεων οι οποίες ενσωματώνουν όλες τις διαθέσιμες πληροφορίες που είναι καταχωρημένες στις πρωτογενείς βάσεις δεδομένων. Ένα παράδειγμα μίας μετα-βάσης είναι αυτό της PICKLE που περιλαμβάνει αποκλειστικά δεδομένα πρωτεϊνικών αλληλεπιδράσεων (Klapa et al. 2013, Gioutlakis et al. 2017) (βλ. Ενότητα 2.1.4.6).

#### 1.5.1.4. Ανάλυση πρωτεϊνικών δικτύων

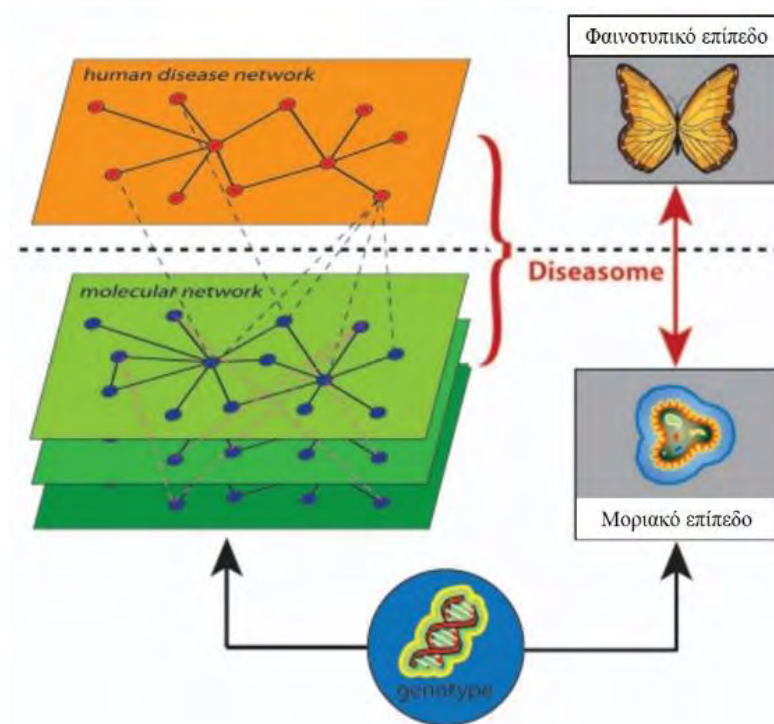
Με δεδομένο ότι η «αρχιτεκτονική» των δικτύων μπορεί να αποκαλύψει σημαντικές αρχές της κυτταρικής οργάνωσης και λειτουργίας, η ανάλυση των πρωτεϊνικών δικτύων γίνεται με υπολογιστικές μεθόδους της θεωρίας των γράφων και στοχεύει: α) στην κατανόηση της τοπολογικής τους οργάνωσης (Pržulj et al. 2004, Valente & Cusick 2006) β) στην εύρεση στενά συνδεδεμένων συμπλεγμάτων πρωτεϊνών (modules), τα οποία μπορεί να έχουν κοινές λειτουργίες (Spirin & Mirny 2003) και γ) στην πρόβλεψη της λειτουργίας μη χαρακτηρισμένων πρωτεϊνών βάσει της τοπολογίας τους στο δίκτυο και των μορίων που αλληλεπιδρούν εφαρμόζοντας ή όχι κατάλληλους αλγόριθμους (Letovsky & Kasif 2003, Hu et al. 2005).

#### 1.5.1.5. Σύνολο γενετικών νοσημάτων στην προοπτική των δικτύων (human diseasome)

Μια εναλλακτική εφαρμογή των δικτύων αφορά τον τομέα της Ιατρικής, και ειδικότερα τη διερεύνηση της μοριακής βάσης ασθενειών του ανθρώπου. Δεδομένης της ύπαρξης κοινών γενετικών αιτιών για την εκδήλωση ασθενειών, τα τελευταία χρόνια επικρατεί η άποψη ότι οι γενετικές μελέτες που αφορούν ασθένειες του ανθρώπου πρέπει να εξεταστούν στην προοπτική των δικτύων (Sahni et al. 2015, Barabási et al. 2011, Goh et al. 2007). Ένα χρήσιμο εργαλείο αυτής της προσέγγισης είναι η αναδυόμενη έννοια του «human diseasome», ενός δικτύου που αποτελείται από δυο επιμέρους δίκτυα τα οποία είναι αλληλοεξαρτώμενα μεταξύ τους, καθώς το πρώτο απεικονίζει τις συνδέσεις μεταξύ γονιδίων (human disease genome) που προκαλούν ή σχετίζονται με συγκεκριμένα νοσήματα και το δεύτερο συνδέει τα νοσήματα αυτά μεταξύ τους (human disease phenome) εφόσον εμφανίζουν κοινή γενετική αιτία, μια πληροφορία που απεικονίζεται στο πρώτο επιμέρους δίκτυο (Goh et al. 2007, Goh et al. 2012). Αυτά τα δυο δίκτυα είναι εκτεταμένα και έχουν πολλαπλές συνδέσεις λόγω του φαινομένου της ετερογένειας γενετικού τόπου που χαρακτηρίζει τα μενδελικά νοσήματα και λόγω της πληθώρας των γονιδίων που συσχετίζονται με ένα συγκεκριμένο πολυπαραγοντικό νόσημα ή τύπο καρκίνου. Η πρώτη προσέγγιση κατασκευής του «human diseasome» έγινε από τους Goh et al. (2007) οι οποίοι στηρίχτηκαν στις πληροφορίες που προσφέρει η βάση δεδομένων OMIM (<https://www.ncbi.nlm.nih.gov/omim/>) για όλες τις γενετικές ασθένειες του ανθρώπου που έχουν γνωστή μοριακή βάση. Στην εργασία αυτή πραγματοποιήθηκε, στο πρώτο

δίκτυο, η ομαδοποίηση των γονιδίων που προκαλούν ή σχετίζονται με την ίδια κατηγορία ασθενειών, και στο δεύτερο δίκτυο η ομαδοποίηση των ασθενειών, που συνδέονται μεταξύ τους, με βάση την παθοφυσιολογία τους. Η ανάλυση του πρώτου δικτύου η οποία περιλάμβανε τη λειτουργική κατηγοριοποίηση (GO analysis), τον έλεγχο του προτύπου έκφρασης στους ιστούς και τις αλληλεπιδράσεις των προϊόντων των γονιδίων έδειξε ότι η παραπάνω ομαδοποίηση συνάδει με κοινούς βιολογικούς ρόλους, κοινά πρότυπα έκφρασης και ότι τα προϊόντα αυτών των γονιδίων τείνουν να αλληλεπιδρούν μεταξύ τους. Οι παρατηρήσεις αυτές επιβεβαιώνονται και από άλλες μελέτες οι οποίες δείχνουν ότι γονίδια που προκαλούν ή συσχετίζονται με την ίδια ασθένεια ή με ασθένειες που εμφανίζουν παρόμοιους φαινότυπους τείνουν να έχουν παρόμοιους βιολογικούς ρόλους, να συνεκφράζονται και να κωδικοποιούν πρωτεΐνες που αλληλεπιδρούν μεταξύ τους (Lage et al. 2007, Oti et al. 2006, Di Pietro & Dell Angelica 2005, Macé et al. 2005, Brunner & Van Driel 2004, Rodgers et al. 2008). Η μελέτη των Goh et al. (2007) αποτέλεσε την απαρχή σε μια σειρά εργασιών οι οποίες χρησιμοποιούν τη βασική προσέγγιση του «human diseasesome» αξιοποιώντας επιπλέον και δίκτυα πρωτεϊνικών αλληλεπιδράσεων (Εικ. 16). Σκοπός αυτών των προσεγγίσεων, στις οποίες λαμβάνεται υπόψη και ο λειτουργικός ρόλος των πρωτεϊνών, είναι η πρόβλεψη νέων υποψήφιων γονιδίων για τις διαφορετικές ασθένειες, βάσει της υπόθεσης ότι ο «γείτονας» μιας πρωτεΐνης σε ένα δίκτυο είναι πιθανό να προκαλεί ή να συσχετίζεται με το ίδιο νόσημα ή νόσημα με παρόμοιο φαινότυπο (Wu et al. 2009, Suthram et al. 2010, X. Wang et al. 2011, Vanunu et al. 2010, Zhang et al. 2011, Bauer-Mehren et al. 2011, Yang et al. 2011, Chan & Loscalzo 2012, Emmert-Streib et al. 2013, Gustafsson et al. 2014). Με παρόμοιες προσεγγίσεις έχουν προταθεί υποψήφια γονίδια για τη νόσο Huntington (Goehler et al. 2004), για αυτοάνοσα και καρδιαγγειακά νοσήματα (Baranzini 2009, Chan & Loscalzo 2012), για νευροεκφυλιστικές ασθένειες (Ahmed et al. 2011), για τον καρκίνο του μαστού (Pujana et al. 2007) κ.α.. Επιπλέον, από μια πρόσφατη μελέτη της ερευνητικής ομάδας που δημιούργησε τη μεταβάση πρωτεϊνικών αλληλεπιδράσεων PICKLE (Klapa et al. 2013, Gioutlakis et al. 2017) δείχθηκε ότι ποσοστό ~98% των 3021 μενδελικών, κατά κανόνα, νοσημάτων που είναι καταχωρημένα στη βάση δεδομένων OMIM οφείλονται σε 2257 γονίδια που κωδικοποιούν πρωτεΐνες του πρωτεϊνικού δικτύου της PICKLE (το 15%) οι οποίες απέχουν το πολύ έναν κόμβο στο δίκτυο αλληλεπιδράσεων. Η μελέτη περιλάμβανε την ανάλυση πρωτεϊνικών δικτύων τα οποία απεικόνιζαν τις πρωτεΐνες που κωδικοποιούνται από γονίδια που είναι υπεύθυνα για μενδελικές διαταραχές και οι οποίες αλληλεπιδρούν άμεσα μεταξύ τους ή είναι τουλάχιστον «δεύτεροι» γείτονες, σε συνδυασμό με δίκτυα νοσημάτων που συνδέονταν μεταξύ τους όταν τα γονίδια που ήταν υπεύθυνα για αυτά τα νοσήματα κωδικοποιούσαν πρωτεΐνες που αλληλεπιδρούσαν είτε άμεσα μεταξύ τους είτε μέσω της μεσολάβησης ενός πρώτου «γείτονα».

Εναλλακτικά το «human diseasesome» έχει συνδυαστεί και με μεταβολικά δίκτυα (Εικ. 16). Ειδικότερα δυο νοσήματα συνδέονται όταν τα γονίδια που είναι υπεύθυνα ή συσχετίζονται με αυτά κωδικοποιούν ένζυμα που καταλύουν διαδοχικές μεταβολικές αντιδράσεις, δηλαδή, μία μπορεί να έπεται ή να προηγείται της άλλης (Lee et al. 2008, Emmert-Streib et al. 2013).



**Εικ. 16: Απεικόνιση του Human Diseasome.** Φαίνονται οι συνδέσεις (κάθετες διακεκομμένες γραμμές) μεταξύ των μοριακών δικτύων (πράσινο), που αφορούν διαφορετικά μοριακά επίπεδα (π.χ. γονίδια, πρωτεΐνες, μεταβολίτες), και του δικτύου που περιλαμβάνει τους φαινοτύπους (human disease network, πορτοκαλί), (προσαρμοσμένη από Emmert-Streib et al. 2013).

Μια διαφορετική προσέγγιση για την εύρεση υποψήφιων γονιδίων για διάφορες γενετικές ασθένειες πραγματοποιήθηκε από τους Liu & Cheng (2015) και περιλάμβανε αρχικά την κατασκευή ενός δικτύου πρωτεϊνικών αλληλεπιδράσεων για 1830 πρωτεΐνες του ανθρώπου. Σε αυτό το δίκτυο δημιουργήθηκαν συμπλέγματα πρωτεϊνών βάσει της τοπολογίας τους. Σε αυτά πραγματοποιήθηκε λειτουργική ανάλυση (GO analysis) των πρωτεϊνών τους και, εφόσον περισσότερες από δυο πρωτεΐνες εμφάνιζαν κοινή λειτουργία, τότε αυτές αποτελούσαν ένα ή περισσότερα λειτουργικά συμπλέγματα εντός του αρχικού τοπολογικού συμπλέγματος. Επιπλέον, οι πρωτεΐνες των τοπολογικών συμπλεγμάτων εξετάστηκαν ως προς τις ασθένειες που προκαλούν ή συσχετίζονται, και αν περισσότερες από δυο συσχετίζονταν με τον ίδιο τύπο νοσήματος, τότε δημιουργούσαν ένα ή περισσότερα συμπλέγματα εντός του τοπολογικού συμπλέγματος. Ακολούθως, εφόσον τα λειτουργικά συμπλέγματα εμφάνιζαν σημαντική αλληλοεπικάλυψη με τα συμπλέγματα των ασθενειών, οι υπόλοιπες πρωτεΐνες των αντίστοιχων τοπολογικών συμπλεγμάτων προτάθηκαν ως πιθανά υποψήφια για τον τύπο νοσήματος που εμφανίζονταν εντός του συμπλέγματος.

Μέχρι τώρα δεν έχει αναφερθεί καμία μελέτη που αφορά το «diseasome» του σώματιου συναρμογής, αν και υπάρχουν αρκετές ασθένειες που έχουν συσχετιστεί με διαταραχές της διαδικασίας της συναρμογής (βλ. επόμενη ενότητα).

### 1.5.2. Συναρμογή και ασθένειες

Το μεταγράφομα του ανθρώπου αποτελείται από ένα τεράστιο αριθμό μορίων RNA, τα οποία υφίστανται περαιτέρω τροποποιήσεις κατά τη διαδικασία της συναρμογής. Η ανίχνευση, σε μια πληθώρα αλληλουχιών, μόνο κάποιων συγκεκριμένων, μικρών σε μέγεθος νουκλεοτιδικών αλληλουχιών, σε συνδυασμό με το μεγάλο ποσοστό της εναλλακτικής συναρμογής των γονιδίων, ιδιαίτερα στον άνθρωπο (Pan et al. 2008, Wang et al. 2008, Gerstein et al. 2014) υποδεικνύουν την πολυπλοκότητα τόσο των ρυθμιστικών στοιχείων των γονιδίων που υφίσταται συναρμογή ή εναλλακτική συναρμογή (*cis*-elements), όσο και των παραγόντων οι οποίοι δρουν *in trans* (*trans-acting*) και ρυθμίζουν το μηχανισμό της συναρμογής. Οι μεταλλάξεις τόσο σε στοιχεία *cis* όσο και σε παράγοντες «*trans-acting*» μπορεί να αποτελούν άμεση αιτία μιας ασθένειας, να συσχετίζονται με ένα πολυπαραγοντικό νόσημα ή να συμβάλλουν (σωματικές μεταλλάξεις) σε κάποιο είδος καρκίνου (Scotti & Swanson 2016, Singh & Cooper 2012). Η διαφορά μεταξύ των επιδράσεων των μεταλλάξεων σε στοιχεία *cis* και σε παράγοντες «*trans-acting*» είναι ότι οι επιπτώσεις των πρώτων έχουν άμεση επίδραση στην έκφραση ενός μόνο γονιδίου, ενώ οι επιδράσεις των δεύτερων αφορούν ένα μεγάλο φάσμα γονιδίων διότι αυτοί ρυθμίζουν τη συναρμογή πολλών γονιδίων. Οι λίγες περιπτώσεις μεταλλάξεων απώλειας λειτουργίας σε παράγοντες «*trans-acting*» μπορεί να υποδηλώνει ότι οι συνέπειες των μεταλλάξεων οδηγούν σε θνησιμότητα κατά την εμβρυική ανάπτυξη (Scotti & Swanson 2016, Ward & Cooper 2010, Cooper et al. 2009).

Όσo όσο υπάρχουν παραδείγματα μενδελικών ασθενειών που οφείλονται σε μεταλλάξεις γονιδίων, τα οποία κωδικοποιούν παράγοντες «*trans-acting*» (συστατικά του σωματίου συναρμογής). Δύο από τα πιο καλά χαρακτηρισμένα είναι: α) η περίπτωση της μελαγχρωστικής αμφιβληστροειδοπάθειας, μια κληρονομήσιμη μορφή τύφλωσης (εκφυλισμός ωχράς κηλίδας, του κεντρικού τμήματος του αμφιβληστροειδή) η οποία προκαλείται από μεταλλάξεις στα συστατικά του τριμερούς συμπλόκου U4/U6.U5 (PRPF3, PRPF4, PRPF6, PRPF8, PRPF31 και SNRNP200) (Tanackovic et al. 2011, Liu & Zack 2013) και β) ένα καλά χαρακτηρισμένο νευρομυϊκό νόσημα (Spinal Muscular Atrophy), που προκαλείται από μεταλλάξεις στην πρωτεΐνη SMN1, η οποία απαιτείται για τη συναρμολόγηση των πρωτεϊνών Sm και Lsm των snRNPs στο κυτταρόπλασμα πριν από την τελική ωρίμανσή τους και την είσοδό τους στον πυρήνα (Winkler et al. 2005, Wang & Cooper 2007). Οι δυο παραπάνω ασθένειες επηρεάζουν διαφορετικά είδη νευρώνων. Για την πρώτη ασθένεια η ευαισθησία των νευρώνων-φωτοϋποδοχέων μπορεί να οφείλεται στην ανεπαρκή παραγωγή των φυσιολογικά υψηλά εκφραζόμενων μεταγράφων (π.χ. ροδοψίνη) λόγω λανθασμένης συναρμογής. Επίσης ο αμφιβληστροειδής του ανθρώπου εκφράζει τα υψηλότερα επίπεδα συγκεκριμένων ιδιοσυστατικά εκφραζόμενων γονιδίων (γονίδια κυτταρικής επιμέλειας) και snRNAs και εμφανίζει ενισχυμένα επίπεδα συναρμογής και εναλλακτικής συναρμογής (Tanackovic et al. 2011). Μια άλλη πιθανή αιτία εμφάνισης των συμπτωμάτων της ασθένειας, ειδικά στους φωτοϋποδοχείς, είναι η δημιουργία συσσωματωμάτων από μεταλλαγμένες πρωτεΐνες (κοινό χαρακτηριστικό των νευροεκφυλιστικών ασθενειών) τα οποία πιθανόν είναι τοξικά ειδικά για τους φωτοϋποδοχείς (Mordes et al. 2007). Όσον αφορά τη δεύτερη ασθένεια, η έλλειψη της



πρωτεΐνης SMN1 οδηγεί στην αποτυχία συναρμολόγησης των snRNPs, υποδεικνύοντας ότι η πρωτεΐνη αυτή διαδραματίζει ρόλο-κλειδί στη διαδικασία της συναρμογής (Cooper et al. 2009). Ωστόσο δεν έχουν βρεθεί ακόμη οι λόγοι για τους οποίους τα φαινοτυπικά χαρακτηριστικά της ασθένειας εκδηλώνονται αποκλειστικά στους κινητικούς νευρώνες. Εικάζεται ότι σε κάθε ιστό υπάρχουν ιστοειδικοί παράγοντες οι οποίοι λειτουργούν σε συνεργασία με την SMN1 και καθορίζουν τη σύσταση των snRNPs. Έτσι, η έλλειψη της SMN1 οδηγεί σε αλλαγή της σύστασής τους με ιστοειδικό τρόπο με αποτέλεσμα να αλλάζουν και οι διαταραχές στη διαδικασία της συναρμογής (Jangi et al. 2017, Cooper et al. 2009).

Επιπλέον, πειραματικές μελέτες αλληλούχισης ολόκληρου του γονιδιώματος ή του εξονιώματος, στο πλαίσιο μελέτης της γενετικής βάσης διαφόρων τύπων καρκίνου όπως οι μυελοειδείς νεοπλασίες και η χρόνια μυελοκυτταρική λευχαιμία έχουν ταυτοποιήσει συχνά σωματικές μεταλλάξεις σε πρωτεΐνες που παίζουν καίριο ρόλο στη διαδικασία της συναρμογής (π.χ. SF3B1, U2AF1) (Yoshida & Ogawa 2014, Quesada et al. 2012, Yoshida et al. 2011, Malcovati et al. 2015). Τέλος, μια ομάδα πρωτεϊνών, οι πρωτεΐνες SR, οι οποίες συνεισφέρουν σημαντικά στη ρύθμιση της συναρμογής και της εναλλακτικής συναρμογής (Zhou & Fu 2013), έχουν συσχετιστεί με ποικίλους τύπους καρκίνων (π.χ. καρκίνο του μαστού, του παχέος εντέρου) (Sjöblom et al., 2006). Στα γονίδια τους έχουν βρεθεί σωματικές μεταλλάξεις ή έχει παρατηρηθεί υπερέκφρασή τους σε αυτούς τους τύπους καρκίνου, όπως στην περίπτωση της SRSF1, που αποτελεί ένα πρωτο-ογκογονίδιο (Karni et al. 2007).

Όπως προαναφέρθηκε, υπάρχουν αρκετά γενετικά νοσήματα που προκαλούνται ή έχουν συσχετιστεί με μεταλλάξεις γονιδίων που κωδικοποιούν πρωτεΐνες οι οποίες είναι απαραίτητες για τη συναρμολόγηση του σωματίου συναρμογής και γενικότερα με τη διαδικασία της συναρμογής. Μέχρι τώρα έχουν πραγματοποιηθεί ορισμένες ολιστικές προσεγγίσεις που αφορούν τη δημιουργία δικτύων πρωτεϊνικών και ρυθμιστικών αλληλεπιδράσεων του σωματίου συναρμογής τα οποία μπορούν να αξιοποιηθούν για την κατασκευή του συνόλου γενετικών νοσημάτων (human disease genome) του σωματίου και αναλύονται στην επόμενη ενότητα.

### 1.5.3. Δίκτυα πρωτεϊνικών και ρυθμιστικών αλληλεπιδράσεων των συστατικών του σωματίου συναρμογής

Έχουν δημοσιευτεί αρκετές μελέτες μικρής και μεγάλης κλίμακας τόσο στον άνθρωπο όσο και σε οργανισμούς-μοντέλα για τις αλληλεπιδράσεις που λαμβάνουν χώρα εντός του σωματίου συναρμογής. Ορισμένες μελέτες περιλαμβάνουν πειραματικές προσεγγίσεις που συνεισέφεραν στη διαλεύκανση πρωτεϊνικών αλληλεπιδράσεων οι οποίες αφορούν συγκεκριμένα στάδια της διαδικασίας (π.χ. τα καταλυτικά βήματα της συναρμογής, (Coltri et al. 2011) ή λαμβάνουν χώρα μεταξύ των κύριων πρωτεϊνών του σωματίου συναρμογής (Hegele et al. 2012), επιμέρους υποσυμπλόκων (Grote et al. 2010). Άλλες μελέτες εστιάζουν στην πρόβλεψη φυσικών αλληλεπιδράσεων μεταξύ των συστατικών του σωματίου συναρμογής (Akerman et al. 2015, Pires et al. 2015) ή στη δυναμική συναρμολόγησης του συμπλόκου (Bortfeldt et al. 2010). Τέλος, μια μεγάλης κλίμακας μελέτη εξετάζει την επίδραση συγκεκριμένων πρωτεϊνών στην εναλλακτική συναρμογή (Papasaïkas et al. 2015).

Αναλυτικότερα, μια μελέτη μικρής κλίμακας αφορά την ταυτοποίηση αλληλεπιδράσεων μεταξύ συστατικών των U1snRNP και U2snRNP στον οργανισμό *Schizosaccharomyces pombe*. Αυτά τα σωματίδια παίζουν ρόλο στην έναρξη της διαδικασίας της συναρμογής (pre-spliceosome) καθώς αναγνωρίζουν τις 5' και 3' θέσεις συναρμογής, αντίστοιχα, ώστε να οριστούν σωστά τα όρια του ιντρονίου. Επιπλέον, το U2snRNP συνεισφέρει στην αναγνώριση της αδενίνης της θέσης διακλάδωσης του πρόδρομου μεταγράφου (Shao et al. 2012). Ένας επιπρόσθετος ρόλος του U2snRNP είναι η συνεισφορά του στην τοποθέτηση της αδενίνης της θέσης διακλάδωσης του πρόδρομου μεταγράφου στην κατάλληλη θέση για την πυρηνόφιλη επίθεση στην 5' θέση συναρμογής, ώστε να επιτευχθεί το πρώτο καταλυτικό βήμα της συναρμογής (Wahl et al. 2009).

Στο πλαίσιο μιας άλλης εργασίας, προσδιορίστηκαν οι αλληλεπιδράσεις μεταξύ πρωτεϊνών ή πρωτεϊνών και U2snRNA εντός του U2snRNP καθώς και αυτού του σωματίου με το πρόδρομο μετάγραφο (Dybkon et al. 2006).

Στον οργανισμό *Saccharomyces cerevisiae* προσδιορίστηκαν αλληλεπιδράσεις που λαμβάνουν χώρα εντός δυο συμπλόκων, των SF3a και SF3b, τα οποία αποτελούν συστατικά του U2snRNP. Αυτές οι αλληλεπιδράσεις συνεισφέρουν στο σχηματισμό του υποσυμπλόκου A του σωματίου συναρμογής. Οι αντίστοιχες μελέτες προσέφεραν χρήσιμες πληροφορίες για την οργάνωση των SF3a και SF3b και για τον τρόπο σύνδεσής τους με το U2snRNP, το U2snRNA και το πρόδρομο μετάγραφο (Lin & Xu 2012, Wang et al. 2005, Golas et al. 2003).

Όσον αφορά τα τελικά στάδια της συναρμογής, έχουν δημοσιευτεί δυο σχετικές μελέτες. Στην πρώτη ταυτοποιήθηκαν άμεσες και έμμεσες πρωτεϊνικές αλληλεπιδράσεις στο σύμπλοκο Prg19/CDC5L, το οποίο σταθεροποιεί τη σύνδεση των U5snRNP και U6snRNP με το ήδη συναρμολογημένο σωματίο συναρμογής κατά τη διάρκεια των καταλυτικών βημάτων της συναρμογής (Grote et al. 2010). Αυτό επιτυγχάνεται μέσω της σύνδεσης των συστατικών του Prg19/CDC5L με τα U5snRNA και U6snRNA και των τελευταίων με το πρόδρομο μετάγραφο (Grote et al. 2010). Η δεύτερη μελέτη συνεισφέρει στον προσδιορισμό των αναδιατάξεων που συμβαίνουν στις αλληλεπιδράσεις μεταξύ μορίων RNA, μεταξύ πρωτεϊνών αλλά και RNA-πρωτεϊνών έτσι ώστε να πραγματοποιηθεί το δεύτερο καταλυτικό βήμα της συναρμογής (Coltri et al. 2011). Συγκεκριμένα ταυτοποιήθηκαν 54 πρωτεΐνες οι οποίες είτε αλληλεπιδρούν με τη θηλιά του ιντρονίου που σχηματίζεται μεταξύ των δυο καταλυτικών βημάτων της συναρμογής, δηλαδή στο υποσύμπλοκο C, είτε συνδέονται με το εξόνιο στο 5' άκρο του mRNA (Coltri et al. 2011).

Μέχρι τώρα υπάρχουν πέντε δημοσιεύσεις που αφορούν μεγάλης κλίμακας προσεγγίσεις οι οποίες περιλαμβάνουν δίκτυα πρωτεϊνικών αλληλεπιδράσεων (φυσικών, συν-έκφρασης ή προβλεπόμενων) του σωματίου συναρμογής.

Μία ερευνητική εργασία αφορά την ταυτοποίηση πρωτεϊνικών αλληλεπιδράσεων που συμμετέχουν καθ' όλη τη διάρκεια των σταδίων συναρμολόγησης του σωματίου συναρμογής και τον έλεγχο της δυναμικής αυτών, μελετώντας τις μεταβολές τους σε διακριτά στάδια της συναρμολόγησης (Hegele et al. 2012). Η εργασία παρουσιάζει τη δημιουργία δέκα ιδιαίτερα αξιόπιστων δικτύων αλληλεπιδράσεων μεταξύ κύριων (core) ή βοηθητικών (non-core) πρωτεϊνών του σωματίου συναρμογής με βάση

συγκεκριμένα κριτήρια που θα αναφερθούν παρακάτω. Τα δίκτυα αυτά περιλάμβαναν από 3 έως 80 αλληλεπιδράσεις και από 2 έως 50 πρωτεΐνες. Στα πλαίσια της δυναμικής του συμπλόκου, όλες οι αλληλεπιδράσεις που ταυτοποιήθηκαν πειραματικά αξιολογήθηκαν για τον εντοπισμό συμπλεγμάτων, δηλαδή στενά συνδεδεμένων μεταξύ τους πρωτεϊνών (modules). Για ορισμένες πρωτεΐνες που συμμετέχουν σε διαφορετικά modules και πιθανώς παίζουν ρόλο στην εξέλιξη της διαδικασίας της συναρμογής, προσδιορίστηκαν οι συναγωνιστικές αλληλεπιδράσεις τους.

Αναλυτικότερα, η μελέτη αφορούσε 244 πρωτεΐνες που έχουν χαρακτηριστεί ως συστατικά του σωματίου συναρμογής σε δυο προηγούμενες ερευνητικές εργασίες (Wahl et al. 2009, Jurica & Moore 2003). Αυτές κατατάχθηκαν σε κύριες (core) και βοηθητικές (non-core) πρωτεΐνες βάσει της αφθονίας τους (αριθμός πεπτιδίων που ταυτοποιήθηκαν στα πειράματα απομόνωσης των συστατικών του σωματίου συναρμογής) και της λειτουργίας τους σύμφωνα με τους Agafonov et al. (2011). Έτσι, στις «core» πρωτεΐνες ανήκουν συστατικά που έχουν γνωστή λειτουργία στη διαδικασία της συναρμογής, αποτελούν υπομονάδες των snRNPs και παρουσιάζουν υψηλή αφθονία. Αντίστοιχα, στην κατηγορία των πρωτεϊνών «non-core» οι οποίες εμφανίζουν χαμηλή αφθονία, περιλαμβάνονται πρωτεΐνες που συνδέονται στο mRNA και παρουσιάζουν ρυθμιστικό ρόλο (π.χ. πρωτεΐνες SR και hnRNP), παράγοντες για τους οποίους υπάρχουν ενδείξεις ότι συνδέουν λειτουργικά τη συναρμογή με άλλες διαδικασίες (π.χ. μεταγραφικοί παράγοντες) και πρωτεΐνες άγνωστης λειτουργίας.

Για τις 196 από τις 244 πρωτεΐνες ταυτοποιήθηκαν 632 αλληλεπιδράσεις με το σύστημα δυο υβριδίων στο σακχαρομύκητα. Από αυτές, οι 390 αφορούσαν τις «core» πρωτεΐνες και οι 242 τις «non-core». Η σύγκριση των αποτελεσμάτων με τα δεδομένα της μετα-βάσης ConsensusPathDB (<http://consensuspathdb.org/>) έδειξε αλληλεπικάλυψη σε ποσοστό 26% για το σύνολο των πρωτεϊνικών αλληλεπιδράσεων και 41% για τις αλληλεπιδράσεις μεταξύ των «core» πρωτεϊνών. Αυτή η μετα-βάση, μεταξύ άλλων, περιέχει και δεδομένα άμεσων πρωτεϊνικών αλληλεπιδράσεων τα οποία προέρχονται από διάφορες πρωτογενείς δημόσιες βάσεις δεδομένων που στηρίζονται στη βιβλιογραφία (Kamburov et al. 2013). Ένα υποσύνολο 109 αλληλεπιδράσεων μεταξύ «core» πρωτεϊνών επιβεβαιώθηκε και με συν-ανοσοκατακρήμιση των αντίστοιχων ανασυνδυασμένων μορίων.

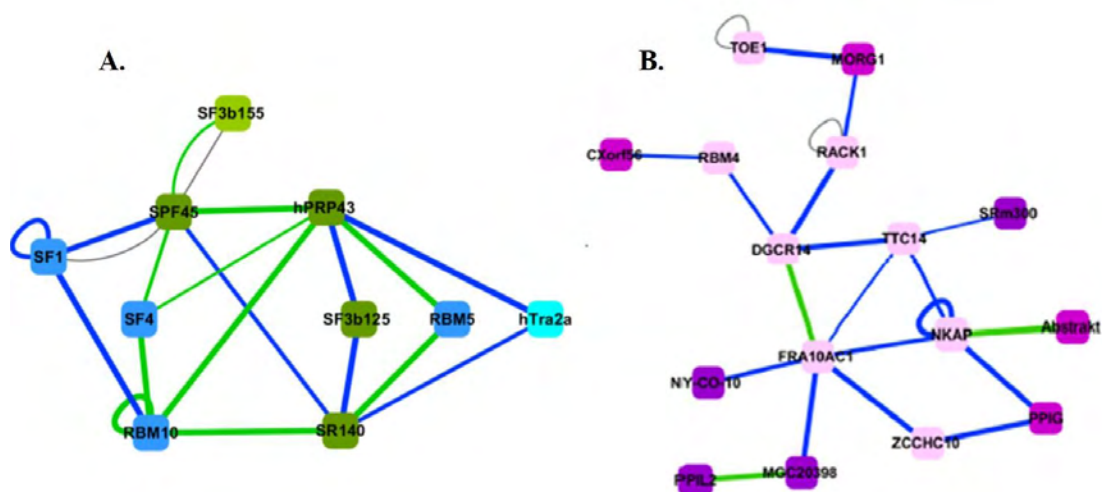
Η κατασκευή των επιμέρους δικτύων πρωτεϊνικών αλληλεπιδράσεων βασίστηκε στα ακόλουθα κριτήρια: α) στην αλληλεπίδραση των πρωτεϊνών που ταυτοποιήθηκε πειραματικά, β) στον χαρακτηρισμό τους ως «core» ή «non-core» και γ) στην αυξημένη τάση τους να αλληλεπιδρούν με πρωτεΐνες εντός της λειτουργικής ομάδας στην οποία ανήκουν. Ως λειτουργική ομάδα ορίστηκε ένα μικρό ή μεγάλο υποσύμπλοκο πρωτεϊνών που έχει συγκεκριμένο λειτουργικό ρόλο σε ένα στάδιο της συναρμογής (π.χ. exon-junction complex, EJC, Le Hir et al. 2016) ή συμμετέχει στο ίδιο στάδιο συναρμολόγησης του σωματίου συναρμογής (π.χ. όλες οι πρωτεΐνες του συμπλόκου C) βάσει των Agafonov et al. (2011).

Το κριτήριο (γ) προσδιορίστηκε με τη σύγκριση των αλληλεπιδράσεων κάθε πρωτεΐνης, οι οποίες προσδιορίστηκαν πειραματικά, με τις αντίστοιχες αλληλεπιδράσεις της που προέκυψαν από τη δημιουργία 1000 τυχαίων (random) δικτύων, διατηρώντας σταθερό τον αριθμό των αλληλεπιδράσεων της αλλά

εφαρμόζοντας τυχαία αναδιάταξη αυτών. Με αυτή τη σύγκριση υπολογίστηκε ο βαθμός της τυπικής απόκλισης (z-score) για πρωτεΐνες που ανήκουν σε μία λειτουργική ομάδα (π.χ. πρωτεΐνες «core» ή πρωτεΐνες Sm). Ως z-score ορίζεται η διαφορά του αριθμού των πρωτεϊνικών αλληλεπιδράσεων εντός μιας λειτουργικής ομάδας, οι οποίες έχουν ταυτοποιηθεί πειραματικά και του μέσου όρου του αντίστοιχου αριθμού των αλληλεπιδράσεων που προέκυψαν από 1000 τυχαία δίκτυα ως προς την τιμή της τυπικής απόκλισης από το παραπάνω μέσο όρο. Επιπλέον, υπολογίστηκε το z-score και για πρωτεΐνες διαφορετικών λειτουργικών ομάδων (π.χ. «core»-«non-core» ή Sm-LSm).

Έτσι, βρέθηκε ότι οι «core» και οι «non-core» πρωτεΐνες τείνουν να αλληλεπιδρούν κατά κύριο λόγο με τις πρωτεΐνες της κατηγορίας τους (θετικό z-score), ενώ οι «core» πρωτεΐνες δεν συνδέονται σημαντικά με τις «non-core» (αρνητικό z-score). Επιπλέον, με βάση την παραπάνω ανάλυση, δείχθηκε ότι οι πρωτεΐνες που εμφάνιζαν κοινό λειτουργικό ρόλο ή συμμετείχαν στο ίδιο στάδιο συναρμολόγησης του σωματίου συναρμογής (π.χ. οι πρωτεΐνες του υπο-συμπλόκου A του σωματίου συναρμογής, οι πρωτεΐνες Sm και Lsm) έτειναν να αλληλεπιδρούν σε μεγάλο βαθμό μεταξύ τους (θετικό z-score).

Λαμβάνοντας υπόψη τα παραπάνω κριτήρια, κατασκευάστηκαν δέκα επιμέρους δίκτυα αλληλεπιδράσεων των «core» και των «non-core» πρωτεϊνών, πέντε για κάθε κατηγορία. Ο αριθμός των αλληλεπιδράσεων κυμαίνονταν από 3 έως 20 αλληλεπιδράσεις για 4 έως 11 «core» πρωτεΐνες και από 3 έως 80 αλληλεπιδράσεις για 2 έως 50 «non-core» (Εικ. 17). Στην Εικόνα 20 φαίνεται ένα από τα πέντε δίκτυα των πρωτεϊνών «non-core» στο οποίο συμμετέχει η πρωτεΐνη FRA10AC1 (βλ. αναλυτικότερα παρακάτω). Παρόλο που κατά κανόνα πρωτεΐνες της ίδιας λειτουργικής ομάδας αλληλεπιδρούν κυρίως μεταξύ τους, σε ορισμένες περιπτώσεις παρατηρήθηκε το αντίστροφο, γεγονός που μπορεί να υποδεικνύει αλληλεπιδράσεις οι οποίες συνεισφέρουν στη στρατολόγηση πρωτεϊνών ώστε να επιτευχθούν συγκεκριμένα βήματα που είναι απαραίτητα για την εξέλιξη της συναρμολόγησης του σωματίου συναρμογής.



**Εικ. 17: Ενδεικτική απεικόνιση δικτύων πρωτεϊνικών αλληλεπιδράσεων πρωτεϊνών core (A) και non-core (B).** Τα τετράγωνα αντιπροσωπεύουν τις πρωτεΐνες (nodes) και οι γραμμές τις φυσικές αλληλεπιδράσεις (edges). Το χρώμα των γραμμών αντιστοιχεί στον τύπο του πειράματος με το οποίο

ταυτοποιήθηκαν οι αντίστοιχες αλληλεπιδράσεις (συν-ανοσοκατακρήμνιση: πράσινο, σύστημα δυο υβριδίων: μπλε) και το πάχος των γραμμών αντιπροσωπεύει το σκορ της αλληλεπίδρασης στο σύστημα των δύο υβριδίων. Οι αλληλεπιδράσεις που ανακτήθηκαν από τη βιβλιογραφία απεικονίζονται με γκρι γραμμή. Το χρώμα των κόμβων αντικατοπτρίζει το στάδιο συναρμολόγησης στο οποίο συμμετέχουν ή το σύμπλοκο στο οποίο ανήκουν {μπλε: σύμπλοκο U1snRNP ή A, γαλάζιο: πρωτεΐνες SR, πράσινο: πρωτεΐνες που σχετίζονται με το U2snRNP (U2 related), μωβ σκούρο: σύμπλοκο B *act*, ροζ σκούρο: σύμπλοκο C, ροζ: πρωτεΐνες του συμπλόκου C που εμφανίζονται σε χαμηλή αφθονία}. Φαίνονται κάποιες από τις αλληλεπιδράσεις (6 από τις 15 που ταυτοποιήθηκαν) της πρωτεΐνης FRA10AC1 με πρωτεΐνες «non-core» (B) (Hegele et al. 2012).

Στο πλαίσιο της μελέτης της δυναμικής του σωματίου συναρμογής, δημιουργήθηκαν 35 συμπλέγματα πρωτεϊνών (modules), δηλαδή ομάδες πρωτεϊνών που είναι στενά συνδεδεμένες μεταξύ τους και εμφανίζουν αραιή συνδεσιμότητα με τις υπόλοιπες, αποκλειστικά για τις core πρωτεΐνες. Η ανάλυση αυτή στηρίχτηκε σε δεδομένα πρωτεϊνικών αλληλεπιδράσεων με το σύστημα δυο υβριδίων του σακχαρομύκητα καθώς και σε δημοσιευμένα και αδημοσίευτα αποτελέσματα των συγγραφέων από 76 πειράματα που αφορούσαν την βιοχημική απομόνωση και ταυτοποίηση με φασματομετρία μάζας των συστατικών των υποσυμπλόκων του σωματίου συναρμογής. Η κατασκευή των modules πραγματοποιήθηκε μέσω του αλγορίθμου Markov clustering. Παρατηρήθηκε ότι ορισμένες πρωτεΐνες συμμετείχαν σε περισσότερα από ένα module και εμφάνιζαν διαφορετικούς «γείτονες», γεγονός που πιθανώς υποδεικνύει την απομάκρυνση πρωτεϊνών σε συγκεκριμένα στάδια της διαδικασίας και την αντικατάστασή τους από άλλες ώστε να συνεχιστεί η πορεία της συναρμολόγησης του σωματίου συναρμογής. Για κάποιες από αυτές, πιστοποιήθηκε η παραπάνω υπόθεση μέσω συναγωνιστικών δοκιμών συν-ανοσοκατακρήμνισης. Σε αυτές εξετάστηκε η ενίσχυση ή η αναστολή της αλληλεπίδρασης πρωτεϊνών (που ήταν παρούσες σε δυο modules) με τον «γείτονά τους» στο ένα module, παρουσία ή απουσία του εναλλακτικού «γείτονά» τους στο δεύτερο module. Η πειραματική αυτή προσέγγιση έδειξε ότι κάποιες αλληλεπιδράσεις είναι παροδικές και πιθανώς μπορεί να αντικατασταθούν με άλλες κατά τον *in vivo* σχηματισμό των διαφορετικών υποσυμπλόκων του σωματίου συναρμογής. Επιπλέον, σε ένα ζεύγος πρωτεϊνών, η παρουσία της μιας πρωτεΐνης μπορεί να καθίσταται απαραίτητη για τη δημιουργία μιας νέας σύνδεσης μεταξύ μιας τρίτης πρωτεΐνης με μία από τις δυο του ζεύγους.

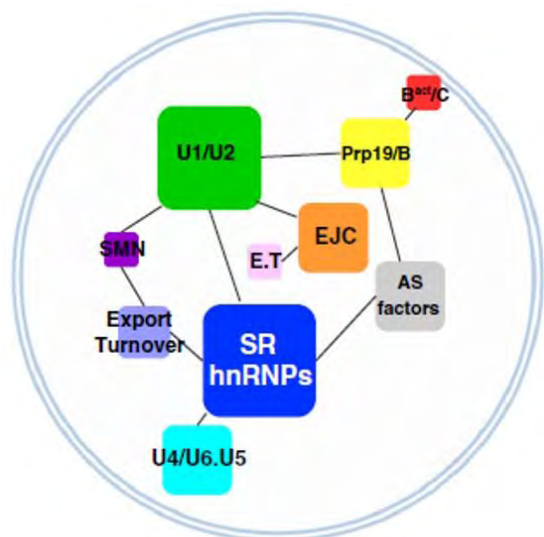
Μια άλλη ερευνητική εργασία παρουσιάζει την δημιουργία ενός δικτύου αλληλεπιδράσεων (φυσικών και συν-έκφρασης) 630 πρωτεϊνών του σωματίου συναρμογής με βάση την πρόβλεψη της πιθανότητας αλληλεπίδρασής τους, με υπολογιστικές μεθόδους. Με βάση δεδομένα πρωτεϊνικών αλληλεπιδράσεων και έκφρασης των αντίστοιχων γονιδίων από βάσεις δεδομένων, προβλέφθηκαν 198.135 αλληλεπιδράσεις μεταξύ τους. Επιπλέον, εξετάστηκαν πειραματικά, μέσω ανοσοκατακρήμνισης σε συνδυασμό με φασματομετρία μάζας, οι αλληλεπιδράσεις επιλεγμένων ρυθμιστών της συναρμογής και της εναλλακτικής συναρμογής, εστιάζοντας σε ενεργοποιητές και καταστολείς (π.χ. πρωτεΐνες SR, hnRNP) (Akerman et al. 2015).

Συγκεκριμένα, η μελέτη περιλάμβανε την πρόβλεψη των πρωτεϊνικών αλληλεπιδράσεων που λαμβάνουν χώρα εντός του σωματίου συναρμογής μέσω της

δημιουργίας ενός μοντέλου (Bayesian model) το οποίο ενσωμάτωνε: i) δεδομένα πρωτεϊνικών αλληλεπιδράσεων από τη βάση HPRD (<http://www.hprd.org/>) και ii) πρότυπα έκφρασης των αντίστοιχων γονιδίων από τη βάση δεδομένων BioGPS (<http://biogps.org/dataset/>) (U133A1/GNF1H dataset), όπως προέκυψαν από δυο αναλύσεις μικροσυστοιχιών για 44.775 γονίδια του ανθρώπου και 36.182 του ποντικού (Su et al. 2004). Τα δεδομένα αυτά αφορούσαν ένα σύνολο 630 πρωτεϊνών του σωματίου συναρμογής οι οποίες είχαν συλλεχθεί: α) από τη Spliceosome Database (<http://spliceosomedb.ucsc.edu/>) η οποία αποτελεί πηγή πληροφοριών για τις πρωτεΐνες που έχουν απομονωθεί ως συστατικά του σωματίου συναρμογής σε βιοχημικές μελέτες, β) από τη KEGG PATHWAY Database (<http://www.genome.jp/kegg/pathway.html>) η οποία περιέχει πληροφορίες για τα μονοπάτια που λαμβάνουν χώρα σε ποικίλες κυτταρικές διεργασίες και τις πρωτεΐνες που συμμετέχουν σε αυτά και γ) από δυο ερευνητικές εργασίες απομόνωσης και ταυτοποίησης των συστατικών του σωματίου συναρμογής (Hegele et al. 2012, Chen et al. 2007).

Η κατασκευή του μοντέλου αυτού (Bayesian model) στηρίχτηκε στον υπολογισμό των πιθανοτήτων αλληλεπίδρασης μεταξύ των πρωτεϊνών λαμβάνοντας υπόψη δυο παραμέτρους: πρώτον, τον αριθμό των κοινών αλληλεπιδρώντων μορίων που έχουν δυο πρωτεΐνες, μια παράμετρος γνωστή ως «transitivity» στη θεωρία των γράφων, και δεύτερον, τα κοινά πρότυπα έκφρασης κάθε ζεύγους πρωτεϊνών. Έτσι, για τις 630 πρωτεΐνες υπολογίστηκαν 198.135 αλληλεπιδράσεις που στηρίχτηκαν σε 37.231 δεδομένα άμεσων και έμμεσων πρωτεϊνικών αλληλεπιδράσεων και 31.363 κοινά πρότυπα έκφρασης. Παρατηρήθηκε ότι το 85% των πρωτεϊνικών ζευγών στη βάση δεδομένων HPRD παρουσίαζε ομοιότητα και στο πρότυπο έκφρασης.

Στη συνέχεια, υπολογίστηκε ο βαθμός ομοιότητας μεταξύ πρωτεϊνών ως προς την τάση τους να αλληλεπιδρούν με πρωτεΐνες της ίδιας λειτουργικής κατηγορίας (π.χ. παράγοντες εναλλακτικής συναρμογής) ή με πρωτεΐνες που ανήκαν στο ίδιο σύμπλοκο ή συμμετείχαν στο ίδιο στάδιο συναρμολόγησης του σωματίου συναρμογής (π.χ. συστατικά του U4/U6.U5 tri-snRNP ή του υποσυμπλόκου C). Έτσι δημιουργήθηκαν δέκα συμπλέγματα πρωτεϊνών (modules)-οι ιδιότητες των οποίων έχουν αναφερθεί προηγουμένως-που διαχωρίστηκαν σε λειτουργικά και τοπολογικά συμπλέγματα. Η διάκριση αυτή έγινε σύμφωνα με τη λειτουργία τους και το στάδιο συναρμολόγησης του σωματίου συναρμογής που συμμετείχαν, στοιχεία τα οποία προέρχονται από τις βάσεις Spliceosome Database και KEGG PATHWAY. Η παραπάνω ανάλυση έδειξε ότι οι πρωτεΐνες που χαρακτηρίζονται ως ρυθμιστές της συναρμογής-σύμφωνα με τη βάση δεδομένων RegRNA (<http://regrna2.mbc.nctu.edu.tw/>) που στηρίζεται σε βιβλιογραφικές αναφορές-δηλαδή οι πρωτεΐνες SR και hnRNP, ομαδοποιήθηκαν σε δυο διακριτά συμπλέγματα (Εικ. 18). Επιπλέον, οι πρωτεΐνες των αρχικών υποσυμπλόκων του σωματίου συναρμογής, αυτές των τελικών υποσυμπλόκων και κάποιες που συνέβαλλαν στη σύνδεση της συναρμογής με τις μετέπειτα διαδικασίες της επεξεργασίας του RNA (π.χ. έξοδος του mRNA από τον πυρήνα) κατηγοριοποιήθηκαν σε διαφορετικά συμπλέγματα (Εικ. 18).



**Εικ. 18: Κατηγοριοποίηση πρωτεϊνών του σωματίου συναρμογής σε modules.** Φαίνονται 10 λειτουργικά και τοπολογικά modules και οι συνδέσεις μεταξύ τους: πρωτεΐνες SR και hnRNP (μπλε), συστατικά των U1/U2 snRNP (πράσινο), των U4/U6.U5 snRNP (γαλάζιο), των συμπλόκων B act/C (κόκκινο), B/Prp19 (κίτρινο) και του συμπλόκου σύνδεσης στα όρια εξονίων EJC (πορτοκαλί), παράγοντες εναλλακτικής συναρμογής και πρωτεΐνες που συνδέονται στο RNA (γκρι), και τέλος πρωτεΐνες που συνεισφέρουν στη συναρμολόγηση των snRNPs (σύμπλοκο SMN, μωβ σκούρο) και στην έξοδο ή αποικοδόμηση του mRNA (export-turnover ή E.T, μωβ ανοιχτό και ροζ, αντίστοιχα). Στο module των U4/U6.U5 snRNP ταυτοποιήθηκε η πρωτεΐνη FRA10AC1 (Akerman et al. 2015).

Ακολούθησε η πιστοποίηση συγκεκριμένων αποτελεσμάτων της *in silico* ανάλυσης, μέσω ανοσοκατακρήμνισης σε συνδυασμό με φασματομετρία μάζας για δυο παράγοντες ρυθμιστές της συναρμογής (SRSF1 και hnRNPA1) παρουσία ή απουσία νουκλεάσης έτσι ώστε να εξεταστούν οι αλληλεπιδράσεις αυτών σε επίπεδο συμπλόκου, παρουσία ή απουσία RNA. Έτσι, ταυτοποίησαν 203 και 152 πρωτεΐνες που συν-κατακρημνίστηκαν με τους παράγοντες SRSF1 και hnRNPA1, αντίστοιχα, εκ των οποίων οι 114 και 60 δεν απαιτούσαν την παρουσία RNA. Παρατηρήθηκε ότι οι πρωτεϊνικές αλληλεπιδράσεις που είχαν προβλεφθεί με μεγαλύτερη ακρίβεια στο πιθανολογικό μοντέλο δικτύου της μελέτης, ήταν αυτές που δεν απαιτούσαν RNA. Μέσω του συνδυασμού των πειραματικών και *in silico* προσεγγίσεων και λαμβάνοντας υπόψη την λειτουργική και τοπολογική κατηγοριοποίηση που αναφέρθηκε προηγουμένως, εξήχθησαν συμπεράσματα για τη συνδεσιμότητα των SRSF1 και hnRNPA1. Δηλαδή, αν αλληλεπιδρούν με παράγοντες που εμφανίζουν κοινό λειτουργικό ρόλο μεταξύ τους ή/και με τις δυο πρωτεΐνες που εξετάστηκαν, και με άλλους παράγοντες που συμμετέχουν σε διαφορετικά στάδια της συναρμολόγησης του σωματίου συναρμογής καθώς και αν αυτές οι αλληλεπιδράσεις απαιτούν την παρουσία RNA. Στην περίπτωση του SRSF1, παρατηρήθηκαν αλληλεπιδράσεις με πρωτεΐνες που συμμετέχουν σε ποικίλα στάδια της διαδικασίας και εμφανίζουν κοινό λειτουργικό ρόλο μεταξύ τους, και στην πλειοψηφία τους δεν απαιτούσαν RNA. Ορισμένες από αυτές επιβεβαιώθηκαν και με δοκιμή συν-κατακρήμνισης (GST-pull down). Στην περίπτωση της hnRNPA1, οι αλληλεπιδράσεις περιορίζονταν μεταξύ πρωτεϊνών που αποτελούσαν ρυθμιστές της συναρμογής, με την πλειοψηφία αυτών να απαιτούν RNA.

Οι δυο προαναφερθείσες εργασίες (Hegele et al. 2012, Akerman et al. 2015) περιλαμβάνουν δεδομένα και για την πρωτεΐνη FRA10AC1, η οποία αναλύεται στην παρούσα διατριβή. Αναλυτικότερα, η πρώτη εργασία την κατατάσσει στις πρωτεΐνες «non-core» και συγκεκριμένα σε αυτές του υποσυμπλόκου C του σωματίου συναρμογής που εμφανίζουν χαμηλή αφθονία. Όσον αφορά τις αλληλεπιδράσεις της, δεκαπέντε ταυτοποιήθηκαν με το σύστημα δυο υβριδίων στο σακχαρομύκητα και τέσσερις από αυτές επιβεβαιώθηκαν και με συν-ανοσοκατακρήμνιση. Ένα από τα πέντε δίκτυα πρωτεϊνών «non-core» που κατασκευάστηκαν με βάση τα κριτήρια της συγκεκριμένης εργασίας (βλ. παραπάνω), περιλαμβάνει την FRA10AC1 (Εικ. 17). Το δίκτυο περιλαμβάνει 19 αλληλεπιδράσεις μεταξύ 16 πρωτεϊνών (Εικ. 17). Σε αυτό απεικονίζονται τα 6 από τα 15 άμεσα αλληλεπιδρώντα μόρια της FRA10AC1 (τα οποία ταυτοποιήθηκαν με τα δυο είδη πειραμάτων), τέσσερα από αυτά (DGCR14, TTC14, NKAP, ZCCH10) ανήκουν στις πρωτεΐνες του υποσυμπλόκου C που παρουσιάζουν χαμηλή αφθονία και τα άλλα δύο (NY-CO-10, MGC20398) στις πρωτεΐνες υψηλής αφθονίας του υποσυμπλόκου B *act* (Εικ. 20). Από τους δεύτερους «γείτονες» της FRA10AC1 που παρουσιάζονται στο δίκτυο, οι δύο (SRm300, PPIL2) ανήκουν στις πρωτεΐνες υψηλής αφθονίας του υποσυμπλόκου B *act*, οι τέσσερις (Abstrakt, PPIG, CXorf56, MORG1) στις πρωτεΐνες υψηλής αφθονίας του υποσυμπλόκου C και οι υπόλοιπες τρεις (RBM4, RACK1, TOE1) στις πρωτεΐνες χαμηλής αφθονίας του υποσυμπλόκου C (Εικ. 17).

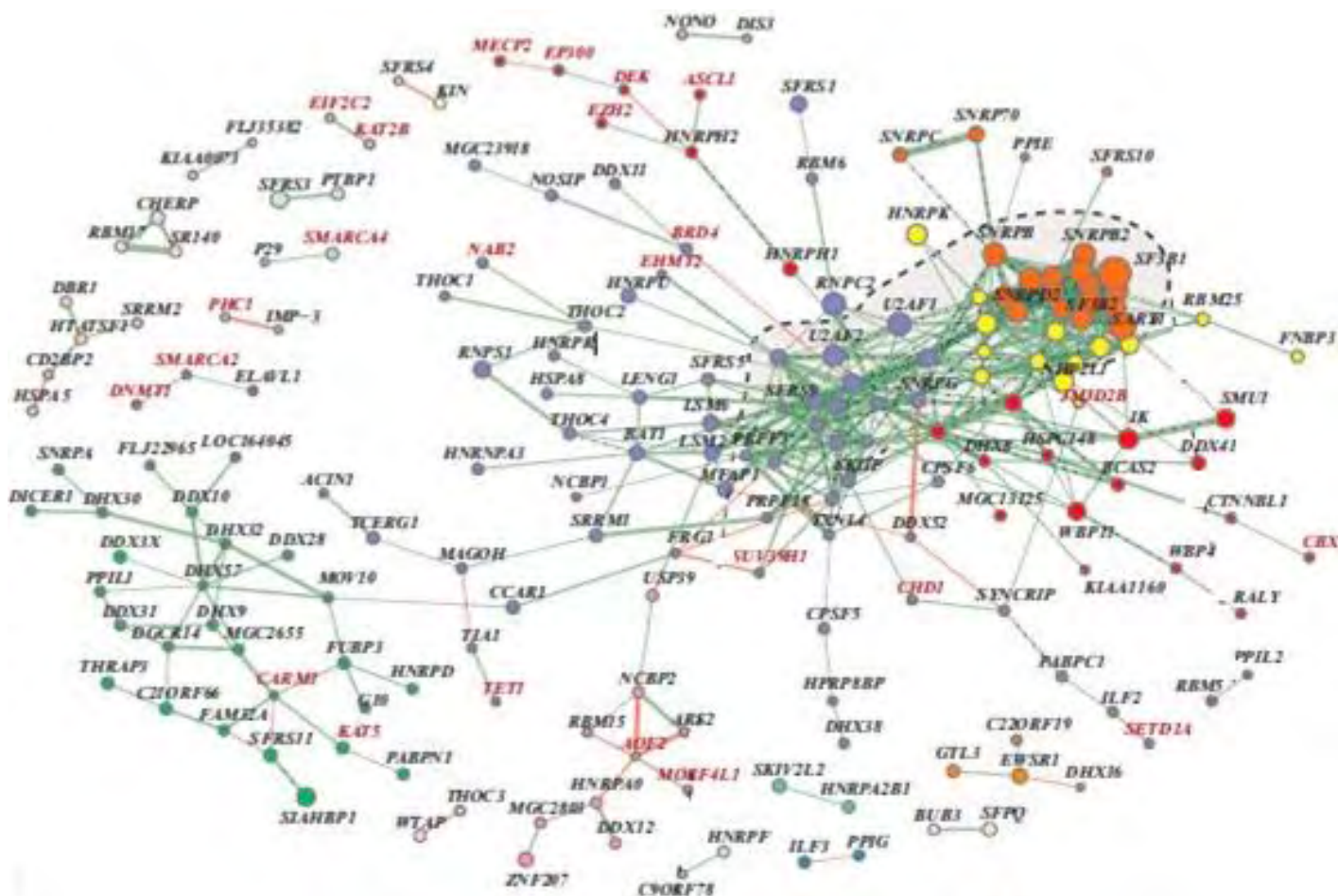
Στη δεύτερη ερευνητική εργασία υπολογίζεται η πιθανότητα αλληλεπίδρασης της πρωτεΐνης FRA10AC1 με τις υπόλοιπες 630 που εξετάστηκαν στην εργασία. Η πιθανότητα αυτή στηρίχτηκε μόνο σε δεδομένα πρωτεϊνικών αλληλεπιδράσεων και όχι σε δεδομένα συν-έκφρασης διότι δεν ήταν διαθέσιμα στη βάση δεδομένων που χρησιμοποιήθηκε. Σύμφωνα με τον τρόπο σχηματισμού των modules που χρησιμοποιήθηκε στην εργασία (βλ. παραπάνω), η πρωτεΐνη FRA10AC1 εντάχθηκε σε στο module που ονομάστηκε U4/U6.U5 στην Εικόνα 18 (Εικ. 18). Αυτό το module περιείχε 25 πρωτεΐνες κάποιες από τις οποίες (PRPF3, PRPF38A, PRPF4, SART1) αποτελούν συστατικά του τριμερούς συμπλόκου U4/U6.U5 ή συνδέονται με αυτό (SNRNP27, SART3). Από τις υπόλοιπες πρωτεΐνες, κάποιες (U2AF1, SNRNP70) σχετίζονται με την αναγνώριση των 5' και 3' θέσεων συναρμογής και με τα καταλυτικά βήματα της συναρμογής (CTNNBL1, DDX46, DHX8), τρεις συμμετέχουν στην μεταγραφή (ARGLU1, CDK11A, SNIP1), τέσσερις στην εναλλακτική συναρμογή (CHERP, RBM39, SREK1, SRSF11), μία (RNPS1) αποτελεί συστατικό του συμπλόκου σύνδεσης εξονίου-εξονίου (EJC), το οποίο συνδέει την συναρμογή με μετα-μεταγραφικές διεργασίες (π.χ. αποικοδόμηση μη κατάλληλων mRNA, έξοδος mRNA από τον πυρήνα), και οι υπόλοιπες είναι άγνωστης λειτουργίας.



Ένας άλλος τύπος δικτύου αλληλεπιδράσεων του σωματίου συναρμογής αφορά τη λειτουργική σύνδεση μεταξύ 270 πρωτεϊνών του (Wahl et al. 2009) βάσει της ομοιότητάς τους ως προς τον τρόπο ρύθμισης της εναλλακτικής συναρμογής (Papasaïkas et al. 2015). Η μελέτη περιλαμβάνει την παροδική αποσιώπηση καθενός από τα αντίστοιχα γονίδια και τη μελέτη των επιδράσεων που προκαλούν σε 35 γεγονότα εναλλακτικής συναρμογής, τα οποία, βάσει βιβλιογραφίας, συμβαίνουν φυσιολογικά σε γονίδια που σχετίζονται με την απόπτωση και τον κυτταρικό πολλαπλασιασμό. Σύμφωνα με τον υπολογισμό της ομοιότητας των αλλαγών κάθε πιθανού ζεύγους πρωτεϊνών στα πρότυπα εναλλακτικής συναρμογής και στην αφθονία των μεταγράφων που παράγονται, στο σύνολο των 35 γεγονότων εναλλακτικής συναρμογής, κατασκευάστηκε δίκτυο 541 λειτουργικών συνδέσεων μεταξύ 196 πρωτεϊνών.

Η απενεργοποίηση των γονιδίων έγινε με τη μέθοδο παρεμβολής μέσω RNA (RNAi) σε κύτταρα *HeLa* και ακολούθησε ο έλεγχος των προτύπων συναρμογής μέσω της αλυσιδωτής αντίδρασης της πολυμεράσης (PCR) με εκκινητές που αναγνώριζαν τα εναλλακτικά μετάγραφα. Ο υπολογισμός της αναλογίας των εναλλακτικών μεταγράφων που παρήχθησαν έγινε βάσει μεγέθους. Τα πρότυπα εναλλακτικής συναρμογής που παρατηρήθηκαν αφορούσαν τη χρήση εναλλακτικών θέσεων συναρμογής η οποία είχε ως αποτέλεσμα να περιλαμβάνονται κάποια εξόνια (exon inclusion) ή να παρακάμπτονται (exon skipping).

Η κατασκευή του δικτύου στηρίχτηκε στην ποσοτικοποίηση της ομοιότητας του προτύπου επίδρασης που εμφάνισε κάθε πιθανό ζεύγος πρωτεϊνών σε όλη την κλίμακα των γεγονότων εναλλακτικής συναρμογής. Με βάση αυτές τις τιμές κατασκευάστηκε ένα ενιαίο δίκτυο με πολλές συνδέσεις και επιμέρους μικρά δίκτυα (Εικ. 19). Συνολικά, τα δίκτυα περιλάμβαναν 196 πρωτεΐνες και 541 λειτουργικές συνδέσεις (Εικ. 19). Από αυτές, οι 518 αφορούσαν θετική συσχέτιση, δηλαδή οι πρωτεΐνες που συνδέονταν μεταξύ τους παρουσίαζαν κοινά πρότυπα, προκαλούσαν, δηλαδή, τους ίδιους τύπους επιδράσεων στα γεγονότα εναλλακτικής συναρμογής. Αντίθετα 23 συνδέσεις αντιπροσώπευαν αρνητική συσχέτιση καθώς οι πρωτεΐνες που συμμετείχαν, εμφάνιζαν αντίθετα πρότυπα επιδράσεων στα γεγονότα αυτά (Εικ. 19). Επιπλέον, επισημάνθηκαν στα πλαίσια του δικτύου, οι συνδέσεις των πρωτεϊνών που εμφάνισαν ισχυρότερη συσχέτιση μεταξύ τους και οι πρωτεΐνες που προκάλεσαν τις εντονότερες επιδράσεις στο σύνολο των εναλλακτικών γεγονότων (Εικ. 19). Βάσει σύγκρισης με τη βάση δεδομένων STRING (<http://string-db.org/>), το 20% των λειτουργικών συνδέσεων αυτού του δικτύου αντιστοιχούν σε φυσικές αλληλεπιδράσεις (Εικ. 19).



**Εικ. 19: Δίκτυο λειτουργικών συνδέσεων του σωματίου συναρμογής.** Οι κύκλοι αντιστοιχούν στα συστατικά του σωματίου συναρμογής και οι γραμμές στις λειτουργικές συνδέσεις μεταξύ τους. Οι θετικές και οι αρνητικές συσχετίσεις απεικονίζονται με πράσινες και κόκκινες γραμμές, αντίστοιχα. Το πάχος της κάθε γραμμής δείχνει την ισχύ της συσχέτισης. Το μέγεθος των κύκλων υποδηλώνει το μέγεθος της επίδρασης της αποσιώπησης των γονιδίων στη ρύθμιση της εναλλακτικής συναρμογής. Ο χρωματικός κώδικας των κύκλων αντικατοπτρίζει την ταξινόμησή τους σε διαφορετικά modules. Οι διακεκομμένες γραμμές αντιστοιχούν στις φυσικές αλληλεπιδράσεις που ανακτήθηκαν από τη βάση δεδομένων STRING (Papasaiikas et al. 2015).

Με στόχο τη συσχέτιση των αλλαγών που προκλήθηκαν με ομάδες πρωτεϊνών που εμφάνιζαν κοινό λειτουργικό ρόλο ή συμμετείχαν σε συγκεκριμένα στάδια της συναρμολόγησης του σωματίου συναρμογής στο πρόδρομο μετάγραφο, πραγματοποιήθηκε η κατηγοριοποίηση των πρωτεϊνών που εξετάστηκαν σύμφωνα με την λειτουργία τους. Η διάκριση αυτή στηρίχθηκε στις βάσεις δεδομένων NCBI (<https://www.ncbi.nlm.nih.gov/gene/>) και Uniprot (<http://www.uniprot.org/>) και στην κατηγοριοποίηση που ακολουθήθηκε στις ερευνητικές μελέτες των Wahl et al. (2009) και Zhou et al. (2002). Έτσι, χωρίστηκαν: i) σε παράγοντες που αποτελούν κύρια συστατικά του σωματίου (core proteins), ii) σε παράγοντες συναρμογής που δεν αποτελούν «core» πρωτεΐνες (non-core proteins), iii) σε παράγοντες επεξεργασίας του RNA που δε σχετίζονται άμεσα με τη συναρμογή και iv) σε πρωτεΐνες που σχετίζονται με την οργάνωση ή αναδιαμόρφωση της χρωματίνης.

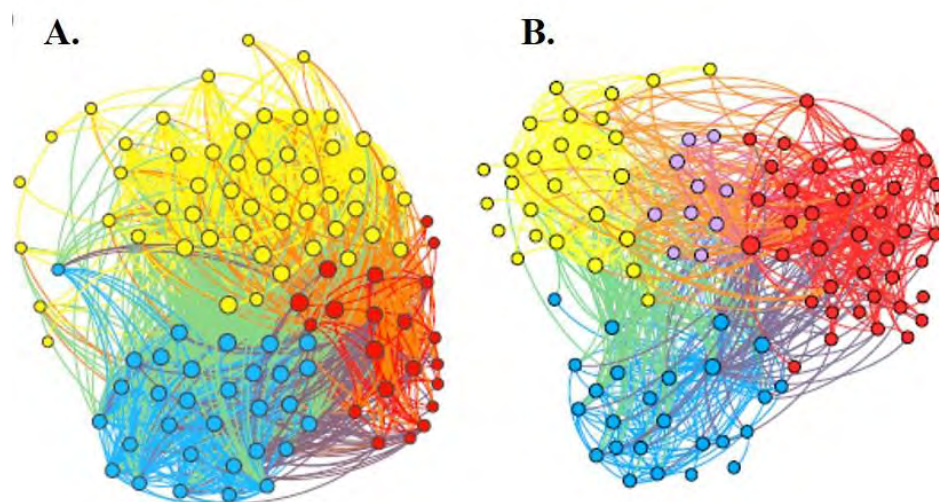
Εναλλακτικά, κατηγοριοποιήθηκαν σε τέσσερις ομάδες: α) στα συστατικά που ήταν παρόντα στα υποσύμπλοκα του σωματίου συναρμογής A, B, B *act*, C (συνεχής παρουσία σε όλα τα υποσύμπλοκα του σωματίου, persistent spliceosome), β) στις πρωτεΐνες που εμφανίζονταν στα υποσύμπλοκα A, B (παροδική παρουσία στα αρχικά υποσύμπλοκα, transient early spliceosome), γ) σε αυτές που ήταν παρούσες κατά τη διάρκεια σχηματισμού του B *act* (παροδική παρουσία στα ενδιάμεσα υποσύμπλοκα, transient middle spliceosome) και δ) στα συστατικά του υποσυμπλόκου C (παροδική παρουσία στα τελευταία υποσύμπλοκα, transient late spliceosome).

Αυτό που παρατηρήθηκε ήταν ότι οι πρωτεΐνες που αποτελούσαν «core» συστατικά του σωματίου συναρμογής και αυτές που εμφάνιζαν καθολική παρουσία στα υποσύμπλοκα A, B, B *act*, C εμφάνισαν αυξημένη θετική συσχέτιση μεταξύ τους και ισχυρότερο πρότυπο επίδρασης στα γεγονότα εναλλακτικής συναρμογής, δηλαδή, το ποσοστό των εξονίων που παρακάμπτονταν ήταν αρκετά μεγαλύτερο από το ποσοστό των εξονίων που περιλαμβάνονταν. Αντίθετα, πρωτεΐνες «non-core» και παράγοντες οργάνωσης ή αναδιαμόρφωσης της χρωματίνης και επεξεργασίας του RNA προκάλεσαν ηπιότερες επιδράσεις στα γεγονότα εναλλακτικής συναρμογής και εμφάνισαν περισσότερο ισορροπημένες αλλαγές, δηλαδή κάποια εξόνια περιλαμβάνονταν και κάποια άλλα παρακάμπτονταν με το ίδιο περίπου ποσοστό. Επιπλέον, βάσει της σύγκρισης των επιδράσεων που προκλήθηκαν από τα συστατικά του «transient early spliceosome», «transient middle spliceosome» και «transient late spliceosome», παρατηρήθηκε ότι τα πρώτα παρουσίασαν ισχυρότερες αλλαγές σε σχέση με αυτά που εμφανίζονταν στα τελικά στάδια. Το αποτέλεσμα αυτό υποδεικνύει το σημαντικό ρόλο των πρώτων βημάτων στην ακριβή αναγνώριση των κανονικών και εναλλακτικών θέσεων συναρμογής, ώστε να εξελιχθεί σωστά η διαδικασία. Στα πλαίσια του δικτύου, το 82% των λειτουργικών συνδέσεων μεταξύ των core πρωτεϊνών αντιστοιχούσε και σε φυσικές αλληλεπιδράσεις σύμφωνα και πάλι με τη βάση δεδομένων STRING (<http://string-db.org/>).

Μέχρι τώρα, έχει γίνει αναφορά σε δίκτυα πρωτεϊνικών αλληλεπιδράσεων και ρυθμιστικά δίκτυα που αφορούσαν το σωματίο συναρμογής του ανθρώπου. Μια ερευνητική εργασία περιγράφει την *in silico* κατασκευή του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής στον οργανισμό *S. cerevisiae* (Pires et al. 2015). Σύμφωνα με δυο κριτήρια που αναλύονται στη συνέχεια, κατασκευάστηκαν δυο

δίκτυα με διαφορετικό αριθμό αλληλεπιδράσεων μεταξύ 103 πρωτεϊνών. Συμπληρωματικά, έγινε η *in silico* κατασκευή του δικτύου πρωτεϊνικών αλληλεπιδράσεων του ανθρώπου χωρίς όμως να γίνεται αναφορά στον αριθμό των πρωτεϊνών και των αλληλεπιδράσεων που περιλαμβάνει.

Αναλυτικότερα, το δίκτυο πρωτεϊνικών αλληλεπιδράσεων στον οργανισμό *S. cerevisiae* περιλάμβανε 103 πρωτεΐνες του σωματίου συναρμογής, χωρίς όμως να υπάρχει συγκεκριμένη αναφορά για την πηγή προέλευσής τους. Η κατασκευή του δικτύου στηρίχτηκε σε δεδομένα πρωτεϊνικών αλληλεπιδράσεων που ανακτήθηκαν από τη βάση δεδομένων STRING (<http://string-db.org/>). Η βάση αυτή περιέχει πληροφορίες για πρωτεϊνικές αλληλεπιδράσεις και αποδίδει σε καθεμία από αυτές ένα συγκεκριμένο βαθμό αξιοπιστίας (0 έως 1) βάσει της πηγής προέλευσής της, δηλαδή αν έχει ταυτοποιηθεί με πειραματικές μεθόδους (βάσει βιβλιογραφικών αναφορών) ή αν αποτελεί προβλεπόμενη αλληλεπίδραση (predicted). Έτσι, δυο πρωτεΐνες θεωρήθηκαν ότι αλληλεπιδρούν όταν εμφάνιζαν τιμές μεγαλύτερες από τους δυο βαθμούς αξιοπιστίας της αλληλεπίδρασης (0,15 και 0,5) που ορίστηκαν ως «κατώφλι» (cut-off) για την κατασκευή δυο διαφορετικών δικτύων. Τα δίκτυα αυτά περιλάμβαναν, κατ' αντιστοιχία για τις δυο παραπάνω οριζόμενες τιμές-κατώφλι, 2538 και 881 πρωτεϊνικές αλληλεπιδράσεις (Εικ. 20).



**Εικ. 20:** Δίκτυα πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής στον οργανισμό *S. cerevisiae*. Φαίνονται τα δίκτυα πρωτεϊνικών αλληλεπιδράσεων στις δυο οριζόμενες τιμές-κατώφλι, 0,15 (A) και 0,5 (B). Ο διαφορετικός χρωματισμός των κύκλων (πρωτεϊνών) αντιστοιχεί σε διαφορετικά modules (Pires et al. 2015).

Στη συνέχεια, χαρακτηρίστηκε η «αρχιτεκτονική» των παραπάνω δικτύων σύμφωνα με δυο παραμέτρους των δικτύων, των «*nestdness*» (απουσία εξειδίκευσης) και «*modularity*». Η πρώτη παράμετρος εκτιμά την ικανότητα των πρωτεϊνών να συνδέονται εξίσου με πρωτεΐνες χαμηλής και υψηλής συνδεσιμότητας. Η ιδιότητα αυτή υποδεικνύει την ευρωστία του δικτύου ως προς την «αρχιτεκτονική» του, δηλαδή την «ανθεκτικότητά» του σε τυχόν διαταραχές (π.χ. απομάκρυνση πρωτεϊνών ή αλληλεπιδράσεων).

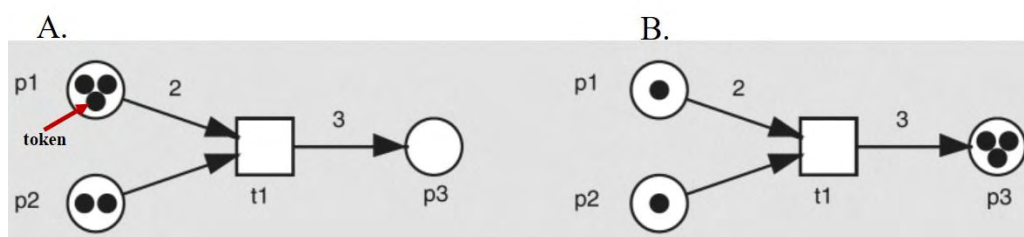
Η δεύτερη παράμετρος αναφέρεται στην ύπαρξη modules. Η σύγκριση των τιμών των δυο αυτών παραμέτρων στα δυο δίκτυα έδειξε ότι στο πιο «αυστηρό» δίκτυο η τιμή «nestedness» είναι μειωμένη, δηλαδή δεν υπήρχε μεγάλο ποσοστό πρωτεϊνών που να αλληλεπιδρούν εξίσου με πρωτεΐνες με μικρό και μεγάλο αριθμό «γειτόνων». Αντίθετα, σε αυτό το δίκτυο, παρατηρήθηκε αύξηση της τιμής «modularity», δηλαδή περιείχε περισσότερες στενά συνδεδεμένες ομάδες πρωτεϊνών που εμφάνιζαν μικρό αριθμό συνδέσεων εκτός της ομάδας (Εικ. 20).

Τέλος, κατασκευάστηκε *in silico* με τον ίδιο τρόπο το δίκτυο πρωτεϊνικών αλληλεπιδράσεων του ανθρώπου χρησιμοποιώντας ως πηγή πληροφόρησης για τα συστατικά του σωματίου συναρμογής τη βάση δεδομένων UniProt (<http://www.uniprot.org/>) και ως πηγή δεδομένων πρωτεϊνικών αλληλεπιδράσεων, εκτός από τη STRING, τις βάσεις BioGRID (<https://thebiogrid.org/>), DIP (<http://dip.doe-mbi.ucla.edu/dip/Main.cgi>), IntAct (<http://www.ebi.ac.uk/intact/>) και MINT (<http://mint.bio.uniroma2.it/>). Ο αριθμός των πρωτεϊνών και των αλληλεπιδράσεων του δικτύου του ανθρώπου δεν αναφέρονται στη δημοσίευση. Το δίκτυο αυτό, συγκρίθηκε, ως προς τις παραμέτρους της «αρχιτεκτονικής» του (οι οποίες αναφέρθηκαν παραπάνω), με τα δυο δίκτυα του σωματίου συναρμογής του οργανισμού *S. cerevisiae*, και παρατηρήθηκε ότι το δίκτυο του ανθρώπου περιείχε περισσότερες πυκνά συνδεδεμένες ομάδες πρωτεϊνών καθώς η τιμή της «modularity» ήταν αυξημένη.

Μια διαφορετική προσέγγιση ανάλυσης της διαδικασίας συναρμολόγησης του σωματίου συναρμογής στο πρόδρομο μετάγραφο-εκτός από την κατασκευή δικτύων πρωτεϊνικών αλληλεπιδράσεων και ρυθμιστικών δικτύων που αντιπροσωπεύουν την απεικόνιση στατικών αλλά και, σε κάποιες περιπτώσεις, δυναμικών αλληλεπιδράσεων-αποτελεί η μαθηματική μοντελοποίηση του σωματίου μέσω της θεωρίας κατασκευής δικτύων Petri (Bortfeldt et al. 2011). Η ερευνητική εργασία περιλαμβάνει την κατασκευή ενός μοντέλου-δικτύου (Petri network) στο οποίο γίνεται η απεικόνιση των αντιδράσεων και των αλληλεπιδράσεων που λαμβάνουν χώρα κατά τη διάρκεια των βημάτων συναρμολόγησης του σωματίου συναρμογής, ορίζοντας την κατεύθυνση αυτών, με στόχο την αναπαράσταση της δυναμικής του.

Για την κατασκευή αυτού του μοντέλου-δικτύου, χρησιμοποιήθηκαν οι πιο καλά μελετημένες αντιδράσεις (π.χ. φωσφορυλίωσης) και αλληλεπιδράσεις μεταξύ πρωτεϊνών ή RNA-πρωτεϊνών οι οποίες στηρίζονται σε πλήθος πειραματικών δεδομένων και λαμβάνουν χώρα κατά τη διάρκεια σχηματισμού των επιμέρους υποσυμπλόκων E, A, B, B *act* και C του σωματίου συναρμογής. Στο δίκτυο που κατασκευάστηκε τα υπό-εξέταση βιολογικά αντικείμενα αντιστοιχούσαν σε μόρια RNA, πρωτεΐνες και σε ενδιάμεσα σύμπλοκα και απεικονίζονται με μικρούς μαύρους κύκλους (tokens) (Εικ. 21 A & B). Η «τοποθεσία» των βιολογικών αντικειμένων (place) φαίνεται με μεγαλύτερους κύκλους (p1, p2, p3, Εικ. 21 A & B). Η διαδικασία (αντίδραση ή αλληλεπίδραση) που περιγράφεται κάθε φορά ορίζεται ως «transition» και απεικονίζεται με τετράγωνο (t1, Εικ. 21 A & B). Όταν λαμβάνει χώρα μια οποιαδήποτε διαδικασία, η πορεία της απεικονίζεται ως εξής: τα βιολογικά αντικείμενα ξεκινούν από τα αρχικά στάδια {pre-places (p1, p2), Εικ. 21 A}, περνούν από μια ενδιάμεση κατάσταση που δείχνει ποια από αυτά δεν χρησιμοποιήθηκαν {pre-places

(p1, p2), Εικ. 21 B} και καταλήγουν στην τοποθεσία {post-place (p3), Εικ. 24 B} που αποτελεί το τελικό στάδιο. Η κατεύθυνση των βιολογικών αντικειμένων απεικονίζεται με βέλη (Εικ. 21 A & B).



**Εικ. 21: Αναπαράσταση μιας διαδικασίας στο δίκτυο Petri.** A: Φαίνονται με κόκκινο βέλος τα βιολογικά αντικείμενα (μαύροι μικροί κύκλοι, tokens) στα αρχικά στάδια (p1 & p2=pre-places, μεγαλύτεροι κύκλοι) της διαδικασίας (t1). B: Απεικονίζεται το ενδιάμεσο στάδιο της διαδικασίας (t1) καθώς και ο αριθμός των βιολογικών αντικειμένων που δε χρησιμοποιήθηκαν (p1 & p2). Στο τελικό στάδιο (p3=post place) φαίνονται τα βιολογικά αντικείμενα που συμμετείχαν στη διαδικασία. Τα βέλη δείχνουν την πορεία των βιολογικών αντικειμένων (Bortfeldt et al. 2011).

Τα δυο δίκτυα που κατασκευάστηκαν περιλάμβαναν 141 βιολογικά αντικείμενα τα οποία συμμετείχαν σε 161 αντιδράσεις-μονοπάτια που λαμβάνουν χώρα κατά το σχηματισμό των συμπλόκων E, A, B, B *act* και C. Αξίζει να σημειωθεί ότι οι περισσότερες αντιδράσεις αφορούσαν το σχηματισμό των αρχικών υποσυμπλόκων E και A τα οποία είναι γενικά σε μεγαλύτερο βαθμό μελετημένα πειραματικά. Επιπλέον, αρκετές από αυτές αντιπροσώπευαν αντιδράσεις που ήταν εξαρτώμενες από το μόριο της τριφωσφορικής αδενοσίνης (ATP), όπως για παράδειγμα η φωσφορυλίωση από κινάσες, η αποφωσφορυλίωση από φωσφατάσες και σε κάποιες περιπτώσεις η αναδιάταξη των αλληλεπιδράσεων των πρωτεϊνών-RNA ή RNA-RNA από τις ελικάσες RNA (DEAD box helicases).

Συμπερασματικά, υπάρχουν αρκετές εργασίες οι οποίες εξετάζουν πειραματικά τις αλληλεπιδράσεις των κύριων συστατικών του σωματίου συναρμογής (Hegele et al. 2012) ή συστατικών επιμέρους συμπλόκων (π.χ. SF3a και SF3b, Lin & Xu 2012), προβλέπουν τις αλληλεπιδράσεις μεταξύ των συστατικών του (Akerman et al. 2012), και ελέγχουν την λειτουργική σύνδεση των συστατικών του εξετάζοντας γεγονότα εναλλακτικής συναρμογής (Papasaïkas et al. 2015).

Ωστόσο, δεν υπάρχει κάποια μελέτη που περιλαμβάνει τις αλληλεπιδράσεις μεταξύ όλων των πρωτεϊνών του σωματίου συναρμογής ή μεταξύ πρωτεϊνών-μορίων RNA, η οποία θα συνεισέφερε στην διερεύνηση του σωματίου ως ένα σύστημα. Η εξέταση του σωματίου συναρμογής ως σύστημα, θα συνέβαλλε στη διερεύνηση της λειτουργίας των συστατικών του και στην δυνατότητα ελέγχου των επιδράσεων που θα προκαλούσε η τροποποίηση-βλάβη των συστατικών του. Αυτό θα μπορούσε να επιτευχθεί μέσω της εφαρμογής ολιστικών προσεγγίσεων μεγάλης κλίμακας (ομικές αναλύσεις) σε επίπεδο μεταγράφου, πρωτεΐνης, μεταβολίτη, οι οποίες μπορούν να εξετάσουν, συνολικά ή εστιασμένα εντός του σωματίου συναρμογής, τις επιπτώσεις που προκαλούν οι διαταραχές των συστατικών του. Τέτοιες προσεγγίσεις είναι περιορισμένες, εφαρμόζονται σε μεμονωμένα επίπεδα κυτταρικής λειτουργίας (π.χ. μεταγραφικό) και η μελέτη των αλλαγών που προκαλούνται δεν περιορίζεται, στην πλειοψηφία των

περιπτώσεων, εντός του σωματίου συναρμογής αλλά επεκτείνεται σε όλες τις διαφορετικές κυτταρικές λειτουργίες που επηρεάζονται. Ενδεικτικά, σε ορισμένες μελέτες (Pacheco et al. 2006, Sen et al. 2015), πραγματοποιήθηκε η απενεργοποίηση της έκφρασης συστατικών του σωματίου συναρμογής (π.χ. κύρια πρωτεϊνική ισομορφή της U2AF1, SRSF3) μέσω της τεχνολογίας RNAi σε κύτταρα θηλαστικών, στα οποία εφαρμόστηκε μεταγραφωμική ανάλυση. Ακολούθησε ανάλυση γονιδιακής οντολογίας των μορίων που εμφάνισαν τροποποιημένα επίπεδα έκφρασης και, στη συνέχεια, είτε εξετάστηκαν ορισμένα από τα γονίδια που εμφάνισαν σημαντική αλλαγή (όχι απαραίτητα αυτά που αποτελούν συστατικά του σωματίου συναρμογής), είτε έγινε ομαδοποίηση των βιολογικών διεργασιών ώστε να εξαχθούν συμπεράσματα για τον πιθανό ρόλο της εξεταζόμενης πρωτεΐνης (π.χ. SRSF3). Εναλλακτικά, σε μία μελέτη (Pascarella et al. 2018), η μεταγραφωμική ανάλυση κυττάρων με τροποποιημένη την έκφραση του γονιδίου DNAJC17, που αποτελεί συστατικό του σωματίου συναρμογής και δεν έχει γνωστό βιολογικό ρόλο, και η ταυτοποίηση των πρωτεϊνικών αλληλεπιδράσεων με συν-ανοσοκατακρήμνιση σε συνδυασμό με φασματομετρία μάζας οδήγησαν σε στοχευμένα πειράματα κατά τα οποία ελέγχθηκαν συγκεκριμένες αλληλεπιδράσεις της DNAJC17 με συστατικά του σωματίου συναρμογής ώστε να εξαχθούν συμπεράσματα για το ρόλο της πρωτεΐνης στη συναρμογή. Καμία μελέτη, έως τώρα, δεν αφορά την μεταβολομική ανάλυση των αλλαγών που μπορεί να προκαλέσει η διαταραχή των συστατικών του σωματίου συναρμογής και καμία που να περιλαμβάνει τη συνδυαστική ανάλυση όλων των επιπέδων κυτταρικής λειτουργίας (μεταγραφικό, πρωτεϊνικό, μεταβολικό). Ο συνδυασμός και των τριών επιπέδων είναι σημαντικός διότι οι μηχανισμοί της γονιδιακής έκφρασης είναι αλληλοεξαρτώμενοι και η τροποποίηση αυτών έχει αντίκτυπο στις μεταβολικές αντιδράσεις, με αποτέλεσμα η μεμονωμένη εξέταση ενός εκ των τριών επιπέδων να μην είναι δυνατόν να προσφέρει μια ολοκληρωμένη εικόνα του συστήματος που εξετάζεται κάθε φορά.

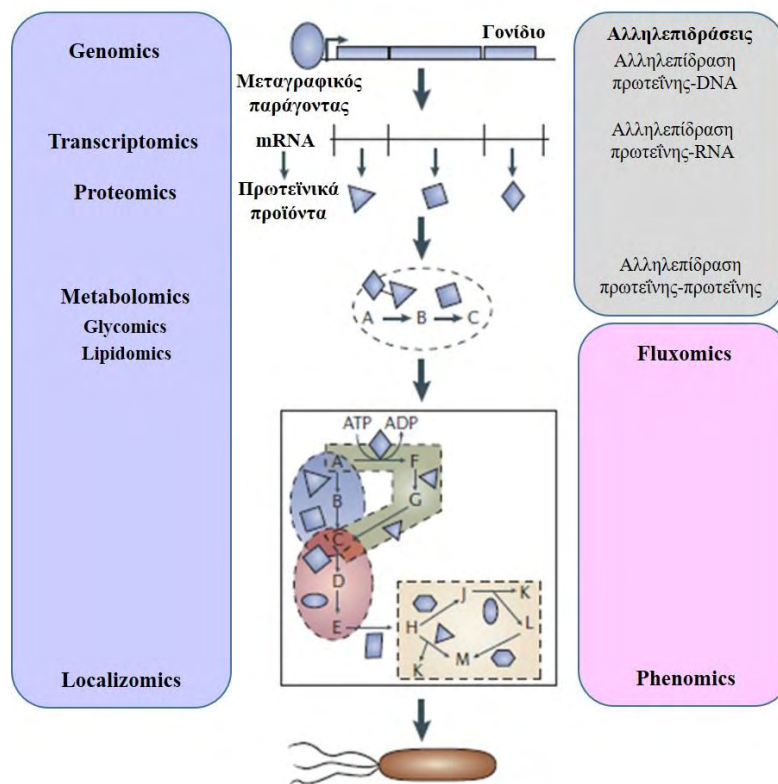
### **1.6. Αξιοποίηση των ομικών προσεγγίσεων στη Συστημική Βιολογία**

Μέχρι σήμερα, πάνω από εξακόσια γονιδιώματα ειδών έχουν αλληλουχηθεί πλήρως (Pagani et al. 2012) συνεισφέροντας σε περισσότερες από πέντε εκατομμύρια μοναδικές αλληλουχίες στις δημόσιες βάσεις δεδομένων (Benson et al. 2007, Wu et al. 2006). Ο τεράστιος όγκος αυτής της πληροφορίας έχει αποτελέσει τον ακρογωνιαίο λίθο για την κατανόηση των βιολογικών μηχανισμών και κατ' επέκταση της ζωής. Ωστόσο, γνωρίζοντας την αλληλουχία του γονιδιώματος, δεν είναι απαραίτητο ότι γνωρίζει κανείς κατευθείαν ποιες περιοχές κωδικοποιούν πρωτεΐνες, ποιες έχουν ρυθμιστικό ρόλο και ποιες είναι υπεύθυνες για τη δομή του γονιδιώματος. Με την πλήρη αλληλούχηση του ανθρώπινου γονιδιώματος εκτιμάται ότι ~20.000 γονίδια κωδικοποιούν πρωτεΐνες οι οποίες υπερβαίνουν τις 80.000 (Claverie et al. 2001, ENCODE Consortium 2012) γεγονός που υποδηλώνει ότι δεν υπάρχει σχέση ένα προς ένα μεταξύ γονιδίου και πρωτεΐνης. Ένας σημαντικός αριθμός αυτών των πρωτεϊνών δεν εμφανίζει χαρακτηρισμένη δομή και λειτουργία γεγονός που αποτελεί μία από τις μεγαλύτερες προκλήσεις για τους επιστήμονες στη μετα-γονιδιοματική εποχή.

Στη μετα-γονιδιοματική εποχή, η παρατήρηση των επιδράσεων της τροποποίησης γονιδίων, γνωστής η μη λειτουργίας, επιτυγχάνεται με την εφαρμογή βιομοριακών

αναλύσεων μεγάλης κλίμακας (ομικές τεχνολογίες) που περιλαμβάνουν τόσο πειραματικές όσο και *in silico* προσεγγίσεις. Η διερεύνηση της γονιδιακής έκφρασης και της πρωτεϊνικής ενεργότητας στο πλαίσιο των δικτύων και των συστημάτων που δημιουργούνται από τα γονίδια και τα γονιδιακά προϊόντα έρχεται να επικρατήσει και να δράσει συμπληρωματικά με την προσέγγιση της εξέτασης μικρού αριθμού γονιδίων και αντιδράσεων (Sussman et al. 2009). Οι ομικές τεχνολογίες αποτελούν ένα από τα βασικά εργαλεία της συστημικής βιολογίας διότι μέσω αυτών μπορεί να ανακτηθεί μια ολιστική εικόνα για τα μόρια που συγκροτούν ένα κύτταρο και τον τρόπο που συνεργάζονται μεταξύ τους, για τις διεργασίες που είναι απαραίτητες για την κυτταρική λειτουργία και τον τρόπο ρύθμισης αυτών καθώς και για τις αλληλεπιδράσεις του οργανισμού με το περιβάλλον του (Vidal et al. 2009). Η έρευνα μέσω ομικών αναλύσεων απαιτεί, εκτός από τον κατάλληλο εξοπλισμό, συνεργασία διεπιστημονικών ομάδων αλλά και θεμελιώδη αλλαγή στις ερευνητικές προσεγγίσεις (Hood et al. 2004). Αυτού του είδους οι προσεγγίσεις διαφέρουν από τις παραδοσιακές μελέτες οι οποίες κατά κύριο λόγο είναι προσανατολισμένες από την υπόθεση που εξετάζεται κάθε φορά (hypothesis-driven) καθώς όλα τα δεδομένα που αποκτώνται και αναλύονται καθορίζουν την υπόθεση (hypothesis-generating), η οποία πρέπει να εξεταστεί περαιτέρω (Biesecker et al. 2013, Kell & Oliver 2004). Τα προηγούμενα χρόνια, μεγαλύτερη προσοχή είχε δοθεί κυρίως στη γονιδιακή αλληλουχία και στην ανάλυση της γονιδιακής έκφρασης, μέσω κυρίως μικροσυστοιχιών, αλλά πλέον έχει αρχίσει να γίνεται όλο και πιο σαφές ότι αυτά δεν μπορούν από μόνα τους να καθορίσουν το λειτουργικό ρόλο ενός γονιδίου στο κύτταρο. Έτσι, μια προσέγγιση αποτελεί η ένταξη των γονιδίων στα πλαίσια του βιολογικού συστήματος, η ανάλυση του οποίου απαιτεί την ενσωμάτωση όλων των μοριακών προτύπων που καθορίζουν την κυτταρική λειτουργία, δηλαδή, το πρότυπο της γονιδιωματικής αλληλουχίας (genomics), το μεταγραφικό (transcriptomics), το πρωτεϊνικό (proteomics) και το μεταβολικό (metabolomics) πρότυπο καθώς και το πρότυπο της ροής των μεταβολιτών μέσω των ενζυμικών αντιδράσεων (fluxomics) (Joyce & Palsson 2006, Yao et al. 2015) (Εικ. 22).





**Εικ. 22: Ροή πληροφορίας από το γονιδίωμα μέχρι τον τελικό κυτταρικό φαινότυπο μέσω των δεδομένων των ομικών αναλύσεων.** Στο αριστερό πλαίσιο φαίνονται τα δεδομένα των ομικών αναλύσεων που προκύπτουν από το DNA (genomics), από τη μεταγραφή του DNA σε RNA (transcriptomics), από τη μετάφραση σε πρωτεΐνες (proteomics) οι οποίες μπορεί να καταλύουν αντιδράσεις παράγοντας μεταβολίτες (metabolomics), γλυκοπρωτεΐνες και ολιγοσακχαρίτες (glycomics) καθώς και ποικίλου είδους λιπίδια (lipidomics). Πολλά από αυτά τα συστατικά του κάθε μοριακού προτύπου μπορούν να σημανθούν και να εντοπιστούν στο κύτταρο (localizomics). Φαίνονται τα πιθανά είδη αλληλεπιδράσεων που μπορεί να λάβουν χώρα και να παράγουν ή να τροποποιήσουν τα συστατικά του καθενός μοριακού προτύπου (δεξί πάνω πλαίσιο). Τα διάφορα μεταβολικά μονοπάτια (κεντρικό πλαίσιο) μπορούν να σχηματίσουν δίκτυα ή χάρτες ροής (fluxomics) που αντικατοπτρίζουν την κυτταρική συμπεριφορά ή τον φαινότυπο (phenomics) (δεξί κάτω πλαίσιο) (προσαρμοσμένη από Joyce & Palsson 2006).

Παρ' όλο που καθένα από αυτά τα πρότυπα έχει σημαντική αξία όταν εξετάζεται μεμονωμένα, τα αποτελέσματα που προκύπτουν δεν αποδίδουν μια ολοκληρωμένη εικόνα για τη κυτταρική λειτουργία. Συγκεκριμένα, η μεταγραφή ενός γονιδίου είναι μια απαραίτητη αλλά όχι επαρκής κατάσταση για την υψηλή πρωτεϊνική έκφραση *in vivo*. Η ρύθμιση του RNA, της μετάφρασης, της σταθερότητας της πρωτεΐνης καθώς και των μετα-μεταφραστικών τροποποιήσεων μπορεί να μεταβάλλει τη γραμμική σχέση μεταξύ του RNA και του αντίστοιχου γονιδιακού προϊόντος. Επιπλέον, μια πρωτεΐνη μπορεί να βρίσκεται σε μεγάλες συγκεντρώσεις αλλά να μην είναι διαθέσιμες οι προαπαιτούμενες συνθήκες (π.χ. παρουσία υποστρωμάτων ή συμπαραγόντων) για την ενεργότητά της στο κυτταρικό περιβάλλον. Τέλος, στο πλαίσιο των ρυθμιστικών δικτύων, η τροποποίηση των επιπέδων έκφρασης ενός γονιδίου δεν αναμένεται να αλλάξει μόνο τη συγκέντρωση της παραγόμενης πρωτεΐνης και την ενεργότητα της αντίστοιχης βιοχημικής αντίδρασης αλλά θα μπορούσε να επηρεάσει και τα υπόλοιπα μέρη του δικτύου στο οποίο ανήκει, σύμφωνα με τη βιολογική λειτουργία που εκτελεί

μέσα στο κύτταρο. Έτσι, η εφαρμογή του συνδυασμού αυτών των μοριακών προτύπων (με προτεραιότητα του μεταγραφικού, του πρωτεϊνικού και του μεταβολικού) σε ένα κυτταρικό σύστημα θα μπορούσε να οδηγήσει στην ανακατασκευή ρυθμιστικών και μεταβολικών δικτύων, τα οποία, έπειτα από την ενσωμάτωση των δεδομένων αφθονίας των πρωτεϊνών και των φυσικών πρωτεϊνικών αλληλεπιδράσεων θα συνεισέφεραν στην εξαγωγή συμπερασμάτων σχετικά με το εκάστοτε βιολογικό ερώτημα, υπό το πρίσμα της συστημικής βιολογίας. Μία τέτοιου είδους προσέγγιση θα μπορούσε να εφαρμοστεί και σε ένα κυτταρικό σύστημα ή σε έναν οργανισμό-μοντέλο στο οποίο έχει διαταραχθεί η έκφραση ενός γονιδίου, ώστε να ανακτηθούν πληροφορίες για το λειτουργικό του ρόλο (Geschwind & Konopka 2009). Αυτή η προσέγγιση αποτελεί την αρχή της «αντίστροφης γενετικής» (reverse genetics), κατά την οποία τροποποιείται η δομή ή/και η έκφραση ενός συγκεκριμένου γονιδίου και ακολούθως ελέγχεται ο βιολογικός φαινότυπος που προκύπτει (Hardy et al. 2010, Tochitani & Hayashizaki 2007). Επειδή αυτή η προσέγγιση ακολουθήθηκε στο δεύτερο μέρος της διατριβής για την εξαγωγή συμπερασμάτων σχετικά με τη λειτουργία του γονιδίου FRA10AC1, πριν αναλυθούν τα τρία βασικά πρότυπα κυτταρικής λειτουργίας (μεταγραφικό, πρωτεϊνικό και μεταβολικό) και οι βασικότερες τεχνικές τους, θα περιγραφούν εν συντομία οι βασικές μέθοδοι τροποποίησης γονιδίων σε κυτταρικό επίπεδο και σε επίπεδο οργανισμού.

➤ *Τροποποίηση της έκφρασης γονιδίων σε κυτταρικό επίπεδο*

Εδώ και 50 περίπου χρόνια, οι κυτταρικές σειρές θηλαστικών χρησιμοποιούνται εκτενώς στις μελέτες της κυτταρικής βιολογίας. Υπάρχουν οι πρωτογενείς καλλιέργειες κυττάρων και οι διαφοροποιημένες κυτταρικές σειρές οι οποίες προέρχονται από διαφορετικούς ιστούς. Οι τελευταίες προκύπτουν έπειτα από την αθανатоποίηση τους *in vitro* και από καρκινικούς όγκους. Οι καρκινικές κυτταρικές σειρές, αν και, σε σύγκριση με τις πρωτογενείς καλλιέργειες κυττάρων, δεν διατηρούν αρκετά από τα φαινοτυπικά χαρακτηριστικά του ιστού από τον οποίο προήλθαν και φαίνεται να έχουν περιορισμένες μεταβολικές λειτουργίες, αντικατοπτρίζουν πολύ συχνά τις ιδιότητες των φυσιολογικών κυττάρων (Hornocks et al. 2003). Η αδυναμία διατήρησης των πρωτογενών καλλιεργειών *in vitro* για μεγάλο χρονικό διάστημα και χρήσης τους σε πειράματα μεγάλης κλίμακας καθώς και η υψηλή μεταβλητότητα των καλλιεργειών αυτών, η οποία οδηγεί στη δυσκολία ανάκτησης κοινού σημείου αναφοράς για τη σύγκριση των πειραματικών αποτελεσμάτων, ευνοούν τη χρήση των κυτταρικών σειρών (Grimm et al. 2004). Εξαιτίας των αποτελεσματικών πρωτοκόλλων πρόκλησης γενετικών αλλαγών, η είσοδος γενετικού υλικού στις κυτταρικές σειρές θηλαστικών καθίστανται ιδιαίτερα εύκολη. Το γεγονός αυτό σε συνδυασμό με την ομοιόμορφη γενετική αλλαγή που προκαλείται στην περίπτωση που δημιουργούνται κλώνοι κυττάρων κατά την ανάπτυξη σταθερών κυτταρικών σειρών και με τη χρήση καθορισμένων συνθηκών ανάπτυξης καθιστούν τις κυτταρικές σειρές ως ένα ευρέως χρησιμοποιούμενο βιολογικό σύστημα για μελέτες κυτταρικής ανάπτυξης, για την κατανόηση της δομής και της λειτουργίας των γονιδίων (Grimm et al. 2004). Δυο από τις ευρέως χρησιμοποιούμενες κυτταρικές σειρές θηλαστικών είναι τα κύτταρα *HeLa* που προέρχονται από αδενοκαρκίνωμα του τραχήλου της μήτρας και τα εμβρυονικά νεφρικά κύτταρα «293» τα οποία έχουν χρησιμοποιηθεί σε πολυάριθμες βιολογικές

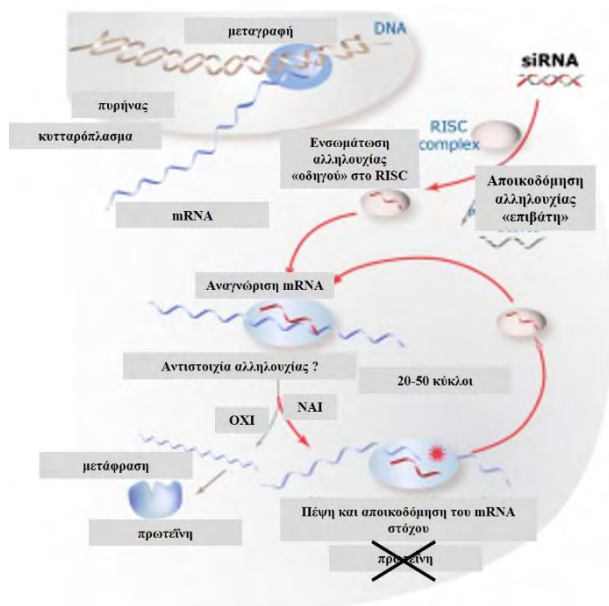
έρευνες (Grimm et al. 2004) με τα πρώτα να εμφανίζονται ως το 2013 σε περισσότερες από 60.000 δημοσιεύσεις (Landry et al. 2013).

Οι κυτταρικές σειρές θηλαστικών εμφανίζουν πολλά πλεονεκτήματα σε σχέση με τα γενετικά συστήματα ολόκληρων οργανισμών. Σε σύγκριση με τους οργανισμούς, κάθε κυτταρική σειρά περιέχει έναν ομοιόμορφο πληθυσμό, γενετικά και φαινοτυπικά, μεμονωμένων κυττάρων που προέρχονται από έναν μόνο ιστό. Οι κυτταρικές σειρές παρέχουν με αυτόν τον τρόπο μια απλουστευμένη άποψη των κυττάρων σε σύγκριση με το πολύπλοκο φυσιολογικό πολύ-κυτταρικό περιβάλλον γεγονός που αποτελεί ένα σημαντικό πλεονέκτημα, καθώς δίνεται η δυνατότητα καθορισμού συγκεκριμένων πειραματικών ρυθμίσεων οι οποίες επιτρέπουν τη διερεύνηση των συνεπειών των γενετικών αλλαγών. Επιπρόσθετα, τα κύτταρα θηλαστικών μπορούν να προέρχονται από διαφορετικούς ιστούς και η κατάσταση ανάπτυξης και διαφοροποίησής τους να διαφέρει από τη μια κυτταρική σειρά στην άλλη. Αυτό απαιτεί έναν αυστηρό έλεγχο ποιότητας έπειτα από ένα διάστημα ανακαλλιέργειών, και προσεκτική επιλογή των κυττάρων που χρησιμοποιούνται στην εκάστοτε γενετική δοκιμή ελέγχου (Grimm et al. 2004). Στα κύτταρα θηλαστικών μπορεί να γίνει στοχευμένη καταστολή ή υπερέκφραση γονιδίων μέσω διαφόρων μεθόδων, από τις οποίες, οι πιο κοινά χρησιμοποιούμενες αναλύονται παρακάτω.

- Μηχανισμός αποσιώπησης γονιδίου με παρεμβολή μέσω RNA (RNA interference, RNAi)

Στα συστήματα κυτταρικών καλλιέργειών του ανθρώπου, η πιο διαδεδομένη μέθοδος αποσιώπησης γονιδίων είναι η παρεμβολή μέσω RNA [RNA interference (RNAi)] (Carpenter & Sabatini 2004, Joga et al. 2016, Kim & Rossi 2008, Simpson et al. 2012). Η τεχνολογία RNAi εφαρμόζεται και σε οργανισμούς-μοντέλα, όπως τον *Caenorhabditis elegans* (Simpson et al. 2012, Sinha & Rae 2016) και τη *Drosophila melanogaster* (Liu et al. 2016, Zeng et al. 2015). Η διαδικασία αυτή περιλαμβάνει τη μετα-μεταγραφική αποσιώπηση του εκάστοτε γονιδίου από δίκλινα μόρια RNA (dsRNA) τα οποία είναι ομόλογα με μια συγκεκριμένη αλληλουχία του γονιδίου που πρόκειται να αποσιωπηθεί (Echeverri & Perrimon 2006, Elbashir et al. 2001). Η παρεμβολή μέσω RNA (RNAi) αξιοποιεί τα ενδογενή κυτταρικά μονοπάτια που υπάρχουν σε ποικίλα είδη από τα φυτά μέχρι και τον άνθρωπο. Αυτά τα μονοπάτια χρησιμοποιούν δυο τύπους μικρών RNA (siRNAs και miRNAs) για να κατευθύνουν την απενεργοποίηση βάσει αλληλουχίας των ενδογενών ή εξωγενών γονιδίων στόχων (Elbashir et al. 2001, Denli et al. 2004, Marques et al. 2010). Στα κύτταρα θηλαστικών εισάγονται εξωγενώς είτε μικρού μήκους δίκλινα μόρια RNA (19-23 νουκλεοτιδίων), τα οποία μιμούνται την γονιδιακή απενεργοποίηση μέσω των ενδιάμεσων (siRNAs) του ενδογενούς κυτταρικού μονοπατιού ή διαγονίδια που παράγουν μόρια RNA που είναι εν μέρει δίκλινα με δομή φουρκέτας [short hairpin RNAs (shRNAs)]. Σε κάθε περίπτωση τα δίκλινα RNA επεξεργάζονται από ένα ένζυμο που έχει ενεργότητα RNAάσης III (Dicer) και σχηματίζονται μικρά παρεμβαλλόμενα RNAs (siRNAs) τα οποία φέρουν μονόκλινα δινουκλεοτιδικά προεξέχοντα 3' άκρα και φωσφορυλιωμένα 5' άκρα. Έπειτα τα παραγόμενα siRNAs ενσωματώνονται στο σύμπλοκο RISC (RNA-Induced Silencing Complex) (Joga et al. 2016, Ohrt et al. 2006) και, βάσει της σχετικής θερμοδυναμικής αστάθειας των δύο άκρων των μικρών παρεμβαλλόμενων RNAs, ένας

κλώνος του siRNA λειτουργεί ως «οδηγός» ή ως συμπαράγοντας-στόχος και ο άλλος αποτελεί προσωρινό «επιβάτη» που γρήγορα αποικοδομείται επειδή θεωρείται ως ψευδο-στόχος. Ο κλώνος «οδηγός» χρησιμοποιείται από το συμπλοκο (RISC) για να κατευθύνει τους επαναλαμβανόμενους κύκλους αναγνώρισης του mRNA στόχου, την ενδονουκλεολυτική πέψη αυτού και κατ' επέκταση την καταστολή της έκφρασης του γονιδίου-στόχου (Joga et al. 2016, Echeverri & Perrimon 2006, Kim & Rossi 2008) (Εικ. 23).



**Εικ. 23: Διαδικασία παρεμβολής μέσω RNA (RNAi).** Με κόκκινα βέλη απεικονίζεται η διαδικασία αποσιώπησης του γονιδίου μέσω μικρών παρεμβαλλόμενων RNA (siRNAs) (προσαρμοσμένη από <http://selexel.com/en/>).

Η αποσιώπηση γονιδίων σε κύτταρα θηλαστικών μπορεί να γίνει είτε μέσω παροδικής διαμόλυνσης των κυττάρων με siRNAs ή με διαγονίδια που παράγουν δίκλιωνα μόρια RNA (shRNAs), χωρίς να πραγματοποιείται η ενσωμάτωσή τους στο χρωμοσωματικό DNA, είτε μέσω δημιουργίας σταθερών κυτταρικών σειρών κατά την οποία τα διαγονίδια (shRNAs) εισάγονται στο γονιδίωμα των κυττάρων. Σχετικά με τη δεύτερη προσέγγιση, η ενσωμάτωση του διαγονιδίου στα κύτταρα προσφέρει τη δυνατότητα δημιουργίας κλώνων κυττάρων που, θεωρητικά, εκφράζουν συνεχώς το διαγονίδιο. Ωστόσο, το διαγονίδιο ενσωματώνεται τυχαία στο γονιδίωμα των κυττάρων και όχι πάντα σε ένα αντίγραφο, γεγονός που μπορεί να έχει διάφορες επιπτώσεις (Cherbas et al. 1994, Spradling & Rubin 1983). Για παράδειγμα, η ενσωμάτωση του διαγονιδίου (ενός ή περισσότερων αντιγράφων) σε κάποιο λειτουργικό στοιχείο του γονιδιώματος μπορεί να επηρεάσει την έκφραση ενός γονιδίου (ή περισσότερων) με αποτέλεσμα οι αλλαγές που θα παρατηρηθούν να μην προέρχονται αποκλειστικά από την αποσιώπηση του εξεταζόμενου γονιδίου αλλά από τις τυχόν διαταραχές που προκαλούνται.

Το ποσοστό αποσιώπησης γονιδίων μπορεί να ελεγχθεί με την ανίχνευση και ποσοτικοποίηση είτε του RNA μέσω της αλυσιδωτής αντίδρασης της πολυμεράσης σε πραγματικό χρόνο (real time-PCR) (Deepak et al. 2007, Garibyan & Avashia 2013), είτε της πρωτεΐνης στους διαφορετικούς ιστούς με ανάλυση Western (Mahmood &

Yang 2012, Towbin et al. 1979). Για να εξεταστούν οι επιδράσεις/αλλαγές που επιφέρει η αποσιώπηση γονιδίων στην κυτταρική φυσιολογία μπορούν να εφαρμοστούν ολιστικές ομικές προσεγγίσεις στα τρία βασικά πρότυπα κυτταρικής λειτουργίας (μεταγραφικό, πρωτεϊνικό, μεταβολικό) (βλ. παρακάτω).

- Μηχανισμοί υπερέκφρασης γονιδίου

Η υπερέκφραση γονιδίων αποτελεί ένα σημαντικό εργαλείο για τη λειτουργική σύνδεση των γονιδίων με συγκεκριμένα βιολογικά μονοπάτια. Μπορεί να εφαρμοστεί σε κύτταρα θηλαστικών εισάγοντας σε αυτά εξωγενώς τα cDNAs των εξεταζόμενων γονιδίων, τα οποία είναι υποκλωνοποιημένα σε πλασμιδιακούς φορείς έκφρασης υπό τον έλεγχο ενός ισχυρού υποκινητή. Έτσι επιτυγχάνεται η υπερέκφραση των αντίστοιχων πρωτεϊνών (Tochitani & Hayashizaki 2007). Ακόμη, μπορεί να γίνει υπερέκφραση ενός cDNA το οποίο περιλαμβάνει σημειακές μεταλλάξεις σε κείρια σημεία της αλληλουχίας ή φέρει έλλειμμα σε συγκεκριμένες περιοχές. Εναλλακτικά, η υπερέκφραση μπορεί να είναι επαγόμενη με χαρακτηριστικό παράδειγμα το σύστημα του καταστολέα της τετρακυκλίνης (Gossen et al. 1995, Wanka et al. 2016). Επιπλέον, η υπερέκφραση ενός γονιδίου μπορεί να είναι επωφελής όταν εφαρμόζεται σε ένα ετερόλογο σύστημα καθώς, με αυτόν τον τρόπο, είναι δυνατόν να παραχθούν μεγάλες ποσότητες της αντίστοιχης πρωτεΐνης (Arnoldo et al. 2008). Και στην περίπτωση της υπερέκφρασης, όπως και στην αποσιώπηση γονιδίων, τα ποσοστά έκφρασης σε επίπεδο μεταγράφου ή πρωτεΐνης μπορούν να ανιχνευθούν με real time-PCR ή με ανάλυση western, αντίστοιχα, και οι αλλαγές που επιφέρει αυτή μπορούν να ελεγχθούν μέσω της εφαρμογής ολιστικών ομικών προσεγγίσεων (βλ. παρακάτω).

➤ *Τροποποίηση της έκφρασης γονιδίων σε επίπεδο οργανισμού – Διαγονιδιακά ζώα*

Ποικίλα ζωικά είδη έχουν χρησιμοποιηθεί για τη δημιουργία διαγονιδιακών μοντέλων συμπεριλαμβανομένου τα ποντίκια, τους αρουραίους, τα κουνέλια, τα γουρούνια και σπανιότερα τα πρόβατα και τις αγελάδες και επιπρόσθετα ως οργανισμοί μοντέλα τους *Caenorabditis elegans*, *Drosophila melanogaster*, *Xenopus laevis*, και το ψάρι *Oryzias latipes* (medaka), καθώς είναι διαθέσιμο ένα πλήθος από μεταλλάγματα αυτών (Houdebine et al. 2007). Ανάμεσα στα ζωικά είδη που χρησιμοποιούνται για διαγονιδιακά ζώα, ο ποντικός αποτελεί ένα ισχυρό εργαλείο για τη διαλεύκανση της γενετικής βάσης της ανθρώπινης φυσιολογίας και παθοφυσιολογίας (Peters et al. 2007). Όπως είναι γνωστό, οι ποντικοί και οι άνθρωποι εμφανίζουν ομοιότητα περίπου στο 99% των γονιδίων τους (Chinwalla et al. 2002) καθώς επίσης παρουσιάζουν κοινές κληρονομήσιμες ασθένειες τόσο μενδελικές όσο και πολυπαραγοντικές (Peters et al. 2007). Γι' αυτούς τους λόγους ο ποντικός είναι το κορυφαίο μοντέλο για τη γενετική βάση των ασθενειών του ανθρώπου. Ωστόσο, παρουσιάζει κάποια μειονεκτήματα καθώς η πρόοδος στη σεξουαλική ωρίμανση και η περίοδος κυοφορίας διαρκούν σχετικά μεγάλο χρονικό διάστημα, ~2 μήνες και 21 ημέρες, αντίστοιχα. Επίσης, η αναπτυξή τους απαιτεί ένα ειδικά διαμορφωμένο άριστα εξοπλισμένο χώρο, και ειδικά εκπαιδευμένο προσωπικό.

Οι τεχνολογίες που χρησιμοποιούνται είναι η εισαγωγή, μέσω μικροένεσης, στο γονιδίωμα των ποντικών (προπυρήνες του ζυγωτού) εξωγενών γονιδιακών αλληλουχιών ώστε να δημιουργηθούν διαγονιδιακοί ποντικοί, και η κατασκευή

μοντέλων ποντικών (knockout/knockin) που αποσκοπεί στη γονιδιακή στόχευση, δηλαδή, στην κατάργηση/αποσιώπηση ή τροποποίηση των ενδογενών γονιδίων των ποντικών, αντίστοιχα. Σχετικά με τη δεύτερη προσέγγιση η διαδικασία που ακολουθείται είναι η ίδια και στις δυο περιπτώσεις (βλ. παρακάτω) με τη διαφορά ότι στα μοντέλα ποντικών knock-in γίνεται στόχευση συγκεκριμένων αλληλουχιών (π.χ. σημειακές μεταλλάξεις, ρυθμιστικά στοιχεία) με αποτέλεσμα την τροποποίηση της λειτουργίας των γονιδίων και όχι την κατάργηση αυτής, όπως στα knock-out ποντίκια. Τα εργαλεία για την κατασκευή μοντέλων ποντικών (knock-out) έχουν εξελιχθεί καθώς πλέον είναι διαθέσιμο πλήθος κυτταρικών σειρών εμβρυονικών βλαστικών κυττάρων που μπορούν να χρησιμοποιηθούν για τη δημιουργία μοντέλων ποντικών knock-out για 20000 γονίδια, όπως είναι καταχωρημένα και εμπορικά διαθέσιμα στο International Mouse Phenotyping Consortium (<http://www.mousephenotype.org/>). Η πιο ευρέως διαδεδομένη τεχνολογία είναι η κατάργηση λειτουργίας του ενδογενούς γονιδίου (knock-out) κατά την οποία δημιουργείται κατάλληλο διαγονίδιο που διαμολύνει εμβρυονικά βλαστικά κύτταρα (ES). Σε αυτή τη μέθοδο επιτυγχάνεται ομόλογος ανασυνδυασμός μεταξύ της γενομικής αλληλουχίας και του ξένου τμήματος DNA επιτρέποντας θεωρητικά την αντικατάσταση οποιαδήποτε περιοχής του γονιδιώματος. Με τον ομόλογο ανασυνδυασμό πραγματοποιείται αντικατάσταση του ενεργού ενδογενούς γονιδίου από μια ανενεργή εκδοχή αυτού. Βασική προϋπόθεση γι' αυτήν την αντικατάσταση είναι να υπάρχουν σχετικά μεγάλους μήκους περιοχές οι οποίες είναι πανομοιότυπες στα άκρα του ενδογενούς και του ξένου τμήματος DNA (Meek et al. 2010, Tong et al. 2010). Τα τροποποιημένα εμβρυονικά βλαστικά κύτταρα εισάγονται σε βλαστοκύστες και ακολούθως σε παρένθετες μητέρες ώστε να οδηγήσουν στη γέννηση αρχικά χμιαϊκών διαγονιδιακών ζώων τα οποία με κατάλληλες διασταυρώσεις είναι ικανά να μεταφέρουν τα διαγονίδια που φέρουν στους απογόνους (Hall et al. 2009). Αυτός ο τρόπος αποτελεί την κύρια μέθοδο για την αναστολή της έκφρασης ενός ενδογενούς γονιδίου, γνωστή ως μέθοδος “knock-out” (Hall et al. 2009). Με την παραπάνω διαδικασία πραγματοποιείται απενεργοποίηση του επιθυμητού γονιδίου σε όλους τους ιστούς. Επειδή η ιδιοσυστατική καταστολή του γονιδίου μπορεί να μην είναι συμβατή με τη ζωή ακολουθείται μια εναλλακτική διαδικασία, η δημιουργία υπό συνθήκη knock-out ποντικών (conditional knock-out), κατά την οποία πραγματοποιείται καταστροφή/απαλοιφή του γονιδίου μόνο σε συγκεκριμένους ιστούς (Su et al. 2010). Ειδικότερα, σχεδιάζεται και κατασκευάζεται ένα διαγονίδιο, στο οποίο, στην κάθε πλευρά του γονιδίου-στόχου τοποθετείται μια ειδική αλληλουχία (θέση *loxP*) που αναγνωρίζεται από την Cre ρεκομπινάση, και το γονίδιο συνεχίζει να λειτουργεί. Οι διασταυρώσεις μεταξύ ποντικών που έχουν ενσωματώσει αυτή τη γονιδιακή κασέτα και διαγονιδιακών ποντικών που εκφράζουν την Cre ρεκομπινάση υπό τον έλεγχο ενός ιστοειδικού υποκινητή μπορεί να δώσουν απογόνους στους οποίους καταργείται η έκφραση του εξεταζόμενου γονιδίου σε συγκεκριμένους ιστούς.

### 1.6.1. Μεταγραφωμική (Transcriptomics)

Το πεδίο της μεταγραφωμικής παρέχει πληροφορίες για το πλήρες σύνολο των μεταγράφων και για την αφθονία αυτών, σε ένα συγκεκριμένο κυτταρικό τύπο ή ιστό σε ένα ορισμένο αναπτυξιακό στάδιο υπό μια φυσιολογική ή «τροποποιημένη» κατάσταση. Βασικοί στόχοι της μεταγραφωμικής είναι να καταγράψει όλα τα είδη των μεταγράφων συμπεριλαμβανομένων των mRNAs, των μορίων RNA που δεν κωδικοποιούν πρωτεϊνικό προϊόν (non-coding RNAs) και των μικρού μήκους RNA (siRNAs, miRNAs), και να ποσοτικοποιήσει τις αλλαγές των επιπέδων έκφρασης κάθε μεταγράφου υπό συγκεκριμένες συνθήκες (π.χ. κατά την τροποποίηση της έκφρασης ενός γονιδίου). Επιπλέον υπάρχει η δυνατότητα του ακριβή καθορισμού των θέσεων έναρξης και λήξης της μεταγραφής, των 5' και 3' άκρων των μεταγράφων, των προτύπων συναρμογής και των μετα-μεταγραφικών τροποποιήσεων (Dong & Chen 2013).

Ποικίλου τύπου τεχνολογίες έχουν αναπτυχθεί για την ανάλυση και την ποσοτικοποίηση του μεταγραφώματος περιλαμβάνοντας τεχνικές που βασίζονται στην υβριδοποίηση των μεταγράφων και στην ανάλυση της αλληλουχίας τους. Οι πρώτες βασίζονται στην ευρέως διαδεδομένη μέθοδος των μικροσυστοιχιών (microarrays). Η ανάλυση των μικροσυστοιχιών περιλαμβάνει την παραγωγή διαφορετικών συνόλων ολιγονουκλεοτιδικών ανιχνευτών που αντιστοιχούν σε κάθε μετάγραφο (τα οποία σταδιακά γίνονται όλο και περισσότερα, ~45000 σύνολα ανιχνευτών), τη μονιμοποίησή τους σε μια στερεή επιφάνεια (microchip) και την επώασή τους με τα σημασμένα συμπληρωματικά μετάγραφα (cDNAs). Έπειτα από την υβριδοποίηση των μορίων-στόχων με τους ανιχνευτές ακολουθεί η σάρωση της μικροσυστοιχίας και η ταυτοποίηση των μορίων RNA σε ποσοτικό επίπεδο στην εκάστοτε εξεταζόμενη κατάσταση. Η μέθοδος των μικροσυστοιχιών επιτρέπει τον παράλληλο προσδιορισμό των επιπέδων έκφρασης πλήθους γονιδίων αποτελώντας ένα δυναμικό εργαλείο στον τομέα της έρευνας (Govindarajan et al. 2012). Επιπλέον, έχουν σχεδιαστεί εξειδικευμένες μικροσυστοιχίες που διαθέτουν ανιχνευτές για τα σημεία σύνδεσης των εξονίων των γονιδίων (exons junctions), οι οποίες μπορούν να χρησιμοποιηθούν για την ανίχνευση και την ποσοτικοποίηση των εναλλακτικών μεταγράφων (Clark et al. 2002, Shen et al. 2010). Ένας άλλος τρόπος προσδιορισμού και ποσοτικοποίησης του μεταγραφώματος που έχει αναπτυχθεί πολύ τα τελευταία χρόνια είναι η αλληλούχηση RNA (RNA-seq) η οποία χρησιμοποιεί πρόσφατα αναπτυσσόμενες τεχνολογίες αλληλούχησης (deep-sequencing). Στη τεχνική αυτή απομονώνεται το ολικό RNA από συγκεκριμένο κυτταρικό τύπο ή ιστό, γίνεται η κατάτμησή του, μετατρέπεται σε cDNA και έπειτα προστίθενται στα άκρα των cDNAs αλληλουχίες-«προσαρμογείς» (adaptors) οι οποίες ενισχύονται μαζί με τα cDNAs με αλυσιδωτή αντίδραση πολυμεράσης. Στη συνέχεια, κάθε μόριο αλληλουχείται με αλληλούχηση νέας γενιάς (next generation sequencing) ώστε να ανακτηθούν αλληλουχίες (30bp -100kb) από το ένα ή και από τα δυο άκρα των συμπληρωματικών μεταγράφων (cDNAs). Ο αριθμός των βάσεων που αλληλουχείται κάθε φορά (reads) εξαρτάται από τον τύπο της τεχνολογίας που χρησιμοποιείται. Με τον τρόπο αυτό μπορεί να οριστούν η έναρξη και η λήξη της μεταγραφής ακόμα και σε επίπεδο ενός νουκλεοτιδίου, οι συνδέσεις μεταξύ πολλαπλών εξονίων καθώς και οι παραλλαγές στην αλληλουχία των μεταγράφων (Single

Nucleotide Polymorphisms, SNPs) (Conesa et al. 2016, Wang et al. 2009). Η αλληλούχηση RNA μπορεί να συνδυαστεί με μεθόδους *in situ* υβριδοποίησης λειτουργώντας ως ένα μεγάλης κλίμακας μοριακό μικροσκόπιο, μέσω ελέγχου της έκφρασης των μεταγράφων σε συγκεκριμένη χρονική στιγμή, σε δεδομένο χώρο, ακόμα και σε επίπεδο ενός μόνο κυττάρου (Burgess et al. 2016, Crosetto et al. 2015). Σε όλες τις παραπάνω τεχνικές ο όγκος των δεδομένων είναι πολύ μεγάλος και η επεξεργασία αυτών γίνεται με κατάλληλα λογισμικά που είναι συμβατά με την εκάστοτε τεχνολογία. Έτσι τα μεταγράφα που ταυτοποιούνται σε κάθε τρόπο ανάλυσης του μεταγραφώματος αντιστοιχίζονται με τις αλληλουχίες αναφοράς των γονιδίων (reference sequence) οι οποίες είναι άμεσα διαθέσιμες στις δημόσιες βάσεις δεδομένων. Επίσης αξιολογείται η στατιστική σημαντικότητα αυτών των μεταγράφων με κατάλληλα εργαλεία στατιστικής (Conesa et al. 2016, Selvaraj & Natarajan 2011).

Στο πλαίσιο των δικτύων πρωτεϊνικών αλληλεπιδράσεων που προαναφέρθηκαν (βλ. Ενότητα 1.5.1.) τα δεδομένα της μεταγραφωμικής ανάλυσης μπορούν να αξιοποιηθούν για την πρόβλεψη πρωτεϊνικών αλληλεπιδράσεων και κατ' επέκταση την κατασκευή πρωτεϊνικών δικτύων, δεδομένης της άποψης ότι δυο γονίδια τα οποία συν-εκφράζονται, αλληλεπιδρούν και σε επίπεδο πρωτεΐνης (De Bodt et al. 2009, Jansen et al. 2002). Τα δεδομένα έκφρασης σε συνδυασμό με ένα πρωτεϊνικό δίκτυο μπορούν να χρησιμοποιηθούν ως εργαλεία για την εύρεση συμπλεγμάτων που περιέχουν πρωτεΐνες οι οποίες αλληλεπιδρούν μεταξύ τους και έχουν κοινά πρότυπα έκφρασης, ώστε να τις εντάξουν σε κοινά βιολογικά μονοπάτια (Li et al. 2012, Padi & Quackenbush 2015, Santoni et al. 2014). Τέλος, τα δεδομένα έκφρασης των γονιδίων σε συνδυασμό με τα πρωτεϊνικά δίκτυα μπορούν να συνεισφέρουν στο καθορισμό του είδους (μόνιμες ή παροδικές) των πρωτεϊνικών αλληλεπιδράσεων (Ou-Yang et al. 2014, Tang et al. 2011, Wang et al. 2013, Wang et al. 2014, Zhang et al. 2014). Ειδικότερα αυτό επιτυγχάνεται μέσω του ελέγχου της έκφρασης των μεταγράφων σε διαφορετικές χρονικές στιγμές (π.χ. στάδια κυτταρικού κύκλου). Ανάλογα με τα επίπεδα έκφρασης των μεταγράφων κάθε δεδομένη χρονική στιγμή καθορίζεται αν, στο συγκεκριμένο χρόνο, υφίσταται αλληλεπίδραση μεταξύ των πρωτεϊνών.

### 1.6.2. Πρωτεϊνωμική (Proteomics)

Ο τομέας της πρωτεϊνωμικής παρέχει πληροφορίες για το σύνολο των πρωτεϊνών που εκφράζονται στα κύτταρα και στους ιστούς σε μια δεδομένη στιγμή υπό ορισμένες συνθήκες (Humphery-Smith et al. 2015, Wilkins et al. 1996). Η ανάλυση του πρωτεϊνώματος στοχεύει στη συστηματική ταυτοποίηση των ποικίλων ιδιοτήτων των πρωτεϊνών συμπεριλαμβανομένων της αλληλουχίας τους, της ποσότητας τους, της κατάστασης των μετα-μεταφραστικών τροποποιήσεων αυτών, και, σε συνδυασμό με άλλες μεθόδους (π.χ. φασματομετρία μάζας), των αλληλεπιδράσεών τους με άλλες πρωτεΐνες ή μόρια (DNA, RNA) (Patterson & Aebersold 2003, Picotti et al. 2013).

Πολλές τεχνολογίες έχουν αναπτυχθεί και αναπτύσσονται συνέχεια για τη συλλογή πληροφοριών που αφορούν τις παραπάνω ιδιότητες των πρωτεϊνών. Μια από τις καθιερωμένες και πιο παλιές τεχνικές που εφαρμόστηκε για την ποσοτική ανάλυση των πρωτεϊνικών προτύπων είναι η υψηλής ανάλυσης ηλεκτροφόρηση δυο διαστάσεων (2D electrophoresis). Αυτή η μέθοδος βασίζεται στο διαχωρισμό των πρωτεϊνών



χρησιμοποιώντας δυο παραμέτρους, το ισοηλεκτρικό σημείο και τη μοριακή μάζα, ο συνδυασμός των οποίων καθορίζει την κινητικότητα των πρωτεϊνών στο πήκτωμα ακρυλαμίδης (Anderson & Anderson 1996, Smith 2009). Μια εξίσου διαδεδομένη τεχνική στο πεδίο της πρωτεομικής είναι η φασματομετρία μάζας (mass spectrometry) η οποία χρησιμοποιείται για την ταυτοποίηση και ποσοτικοποίηση των πεπτιδίων των πρωτεϊνών που έχουν προέλθει είτε από ηλεκτροφόρηση δυο διαστάσεων, είτε από πρωτεϊνικό εκχύλισμα κυττάρων ή ιστών σε συνδυασμό με τεχνικές χρωματογραφίας (Di Girolamo et al. 2013).

Ειδικότερα, η διαδικασία που ακολουθείται περιλαμβάνει την αποδιάταξη του πρωτεϊνικού δείγματος, την ενζυμική πέψη σε πεπτίδια και τη σήμανση αυτών με ειδικές ετικέτες που στοχεύουν στις χημικές ομάδες συγκεκριμένων αμινοξέων. Έτσι αυξάνεται το ποσοστό ταυτοποίησης όλων των πιθανών πεπτιδίων που αντιπροσωπεύουν μια συγκεκριμένη πρωτεΐνη, διατηρείται ακέραιη η δομή τους συμπεριλαμβανομένων των μετα-μεταφραστικών τροποποιήσεων, και δίνεται η δυνατότητα της ταυτόχρονης εξέτασης πολλαπλών βιολογικών δειγμάτων (Nilse et al. 2015, Zieske et al. 2006). Στη συνέχεια ακολουθεί η τοποθέτηση των σημασμένων πεπτιδίων στη χρωματογραφική στήλη κατά την οποία τα συστατικά του μίγματος κατανέμονται μεταξύ δυο φάσεων, της ακίνητης και της κινητής. Στη περίπτωση της υγρής χρωματογραφίας η κινητή φάση είναι σε υγρή μορφή ενώ η ακίνητη μπορεί να είναι στερεό υλικό (χρωματογραφία προσρόφησης) ή υγρό προσροφημένο σε κατάλληλο στερεό υλικό (χρωματογραφία κατανομής). Στο τέλος, τα πεπτίδια διαχωρίζονται από τη στήλη με βάση τη διαλυτότητα (πολικότητα) τους και τη χημική τους δομή (Bird et al. 1989, Gosetti et al. 2013, Pitt et al. 2009).

Το επόμενο στάδιο περιλαμβάνει τρία βήματα: i) τον ιονισμό των μορίων μέσω μιας πηγής ιονισμού, δηλαδή, τη μετατροπή τους σε θετικά ή αρνητικά ιόντα μέσω της πρόσληψης ή αποβολής ηλεκτρονίων, αντίστοιχα. Ο ιονισμός των βιολογικών μακρομορίων επιτυγχάνεται με την μέθοδο «electrospray ionization» (ESI) που λειτουργεί σε αυξημένη ατμοσφαιρική πίεση και το δείγμα ψεκάζεται μέσω μιας πολύ λεπτής βελόνας σε ισχυρό ηλεκτρικό πεδίο (Kearle 2000, Wilm 2011) και την τεχνική «Matrix-assisted laser desorption /ionization» (MALDI) κατά την οποία τα ιόντα εκροφώνται από μια στερεή φάση (Calderaro et al. 2015, Hillenkamp et al. 1991), ii) το διαχωρισμό των ιονισμένων μορίων με βάση τις τιμές της αναλογίας μάζας προς φορτίο μέσω μαγνητικών ή ηλεκτρικών πεδίων, και iii) την μεταφορά τους σε αναλυτή μάζας, και την επακόλουθη ανίχνευση των διαχωρισμένων ιονισμένων μορίων, που δρούν σαν ηλεκτρικά φορτισμένα σήματα, βάσει της αφθονίας του κάθε είδους (Aebersold & Mann 2003, Chavez et al. 2016, Han et al. 2008). Η ανάλυση της μοριακής μάζας γίνεται μέσω φασματομέτρων μάζας (mass spectrometers) η οποία καθορίζεται από πολλές παραμέτρους όπως την ικανότητα διαχωρισμού των ιόντων βάσει της αναλογίας μάζας προς φορτίο που έχει προηγηθεί, τον ακριβή προσδιορισμό της μοριακής μάζας (βάσει αξιόπιστων τιμών αναλογίας μάζας προς φορτίο), το εύρος της μοριακής μάζας και τέλος την ακριβή ανάκτηση μονών (MS) ή διαδοχικών φασμάτων (MS/MS) (Mann et al. 2001, Wilhelm et al. 2014). Οι ευρέως χρησιμοποιούμενοι αναλυτές μάζας είναι το τριπλό τετράπολο (Quadrupole, Q) το οποίο χρησιμοποιεί δύο τετράπολα (Q1, Q3) ως αναλυτές μάζας και το ενδιάμεσο

τετράπολο (Q2) που χρησιμοποιείται ως χώρος θραύσης των ιονισμένων μορίων (Wilhelm et al. 2014, Schwartz & Jardine 1996). Εξειδικευμένα τετράπολα είναι τα τετράπολα παγίδευσης ιόντων (ion trap quadrupole) τα οποία είναι ανάλογα τριών διαστάσεων των γραμμικών τετραπόλων (Londry & Hager 2003, Zhu et al. 2015). Ένας άλλου είδους αναλυτής μάζας είναι αυτός κατά τον οποίο τα ηλεκτρόνια που εμφανίζουν το ίδιο φορτίο αποκτούν την ίδια κινητική ενέργεια μέσω εφαρμογής ηλεκτρικού πεδίου και σε αναλογία με το λόγο μάζας προς φορτίο κινούνται πιο γρήγορα ή πιο αργά μέχρι την τελική ανίχνευσή τους (time-of-flight mass spectrometer) (Chernushevich et al. 2001, Vestal & Campbell 2005).

Μια διαφορετικού είδους τεχνολογία της πρωτεϊνικής είναι η χρήση πρωτεϊνικών μικροσυστοιχιών μεγάλης κλίμακας που χωρίζονται σε τρεις κατηγορίες, τις αναλυτικές, τις λειτουργικές και τις αντίστροφης φάσης. Οι πρώτες βασίζονται στη χρήση ενός ή δυο αντισωμάτων για την ανίχνευση σημασμένων πρωτεϊνών. Οι δεύτερες στηρίζονται στη χρήση καθαρισμένων πρωτεϊνών εξετάζοντας τις βιοχημικές ιδιότητες αυτών, όπως τις ικανότητες σύνδεσης σε μόρια-συνδέτες (πρωτεΐνες-πεπτίδια, DNA, λιπίδια, φάρμακα, σχέση υποστρώματος ενζύμου) και τις μετα-μεταφραστικές τροποποιήσεις (φωσφορυλίωση, ακετυλίωση κ.α). Οι τελευταίες βασίζονται στη χρήση ανιχνευτών σε ιστούς, κυτταρικά εκχυλίσματα που είναι μονιμοποιημένα σε γυάλινες πλάκες για την επακόλουθη υβριδοποίηση και ταυτοποίηση συγκεκριμένων πρωτεϊνών (Sutandy et al. 2013).

Ένας μεγάλος κλάδος της πρωτεϊνικής είναι η ταυτοποίηση πρωτεϊνικών αλληλεπιδράσεων και η κατασκευή πρωτεϊνικών δικτύων (βλ. Ενότητα 1.5.1.). Στα πρωτεϊνικά δίκτυα μπορούν να ενσωματωθούν δεδομένα που αφορούν την αφθονία των πρωτεϊνών και τις μετα-μεταφραστικές τροποποιήσεις που υφίστανται αυτές, ώστε να ανακτηθούν πληροφορίες για τη δυναμική των αλληλεπιδράσεων κατά τη διάρκεια του χρόνου που λαμβάνει χώρα μία βιολογική διαδικασία (Drissi et al. 2015, Schaberg et al. 2017, Supper et al. 2017, Turriziani et al. 2014). Η ενσωμάτωση δεδομένων αφθονίας στα πρωτεϊνικά δίκτυα θα μπορούσε να συνεισφέρει στον εντοπισμό συμπλεγμάτων πρωτεϊνών που παρουσιάζουν διαφορετικό πρότυπο έκφρασης σε δυο διαφορετικές καταστάσεις. Αυτά τα συμπλέγματα θα μπορούσαν να αναλυθούν περαιτέρω, βάσει του λειτουργικού ρόλου των πρωτεϊνών που συμμετέχουν σε αυτά, ώστε να εξαχθούν συμπεράσματα για τις βιολογικές διεργασίες που επηρεάζονται σε περίπτωση ασθενειών (Zanzoni & Brun 2016).

### 1.6.3. Μεταβολομική (Metabolomics)

Ο τομέας της μεταβολομικής επιδιώκει την ταυτοποίηση του συνόλου των μεταβολιτών του κυττάρου. Οι μεταβολίτες είναι ενώσεις μικρού μοριακού βάρους και συμμετέχουν σε μεταβολικές διεργασίες (Dunn et al. 2005, Patel & Ahmed 2015). Σε αντίθεση με τα γονίδια και τις πρωτεΐνες, οι λειτουργίες των οποίων υπόκεινται σε επιγενετικές και μετα-μεταφραστικές τροποποιήσεις, αντίστοιχα, οι μεταβολίτες λειτουργούν ως άμεσες «υπογραφές» της βιοχημικής ενεργότητας με αποτέλεσμα να μπορούν να συσχετιστούν ευκολότερα με το φαινότυπο (Patti et al. 2012). Η μελέτη του μεταβολώματος περιλαμβάνει την ανάλυση μιας ευρείας κλίμακας χημικών ενώσεων, από μικρού μοριακού βάρους πολικών πτητικών ουσιών μέχρι μεγάλου

μοριακού βάρους πολικών γλυκοζιτών, μη πολικών λιπιδίων και ανόργανων ουσιών (Patel & Ahmed 2015, Lahner et al. 2003). Οι μέθοδοι που χρησιμοποιούνται εξελίσσονται συνεχώς και βασίζονται κυρίως στις τεχνικές της φασματομετρίας μάζας για την ανάλυση του μεταβολικού περιεχομένου που προέρχονται από κύτταρα ή ιστούς υπό φυσιολογικές ή «τροποποιημένες» καταστάσεις. Μέσω αυτών είναι πιθανό να μετρηθούν γρήγορα και ταυτόχρονα εκατοντάδες μεταβολίτες από ελάχιστες ποσότητες δείγματος (Patel & Ahmed 2015, Smith et al. 2006). Οι τεχνικές φασματομετρίας μάζας σε συνδυασμό με τις χρωματογραφικές τεχνικές αποτελούν τον πλέον καθιερωμένο τρόπο προσδιορισμού του μεταβολώματος. Οι πιο ευρέως χρησιμοποιούμενες είναι η αέρια χρωματογραφία-φασματομετρία μάζας (GC-MS) και η υψηλής απόδοσης υγρή χρωματογραφία-φασματομετρία μάζας (LC-MS). Οι αρχές των μεθόδων της χρωματογραφίας και της φασματομετρίας μάζας έχουν ήδη περιγραφεί στον τομέα της πρωτεϊνικής με τη διαφορά ότι στην περίπτωση της αέριας χρωματογραφίας η κινητή φάση είναι σε αέρια μορφή. Όπως αναλύθηκε παραπάνω, και στην περίπτωση του μεταβολώματος, μπορούν να χρησιμοποιηθούν διαφόρου είδους αναλυτές μάζας, με τα τριπλά τετράπολα να παρουσιάζουν μεγαλύτερη ευαισθησία (Dunn et al. 2005). Μια εναλλακτική τεχνική αποτελεί η φασματοσκοπία πυρηνικού μαγνητικού συντονισμού (NMR) κατά την οποία ένα φάσμα πυρηνικού μαγνητικού συντονισμού είναι ειδικό για μια συγκεκριμένη ένωση. Με αυτόν τον τρόπο παρέχεται πλήθος πληροφοριών σχετικά με τα συστατικά του βιολογικού δείγματος χωρίς να απαιτείται προεπιλογή των συνθηκών της ανάλυσης όπως την πηγή ιονισμού ή το είδος της χρωματογραφικής τεχνικής (Dunn et al. 2005). Όσον αφορά το σχεδιασμό ενός πειράματος μεταβολομικής μπορεί να υπάρχουν δυο προσεγγίσεις: α) η «στοχευμένη» μεταβολομική (targeted metabolomics) κατά την οποία συγκεκριμένος κατάλογος μεταβολιτών εξετάζεται δίνοντας βάση σε ένα συγκεκριμένο μονοπάτι ενδιαφέροντος ή σε σχετιζόμενα μονοπάτια (Dudley et al. 2010), και β) η μη «στοχευμένη» μεταβολομική, οι μέθοδοι της οποίας στοχεύουν στην ταυτόχρονη μέτρηση όσο το δυνατόν περισσότερων μεταβολιτών από τα βιολογικά δείγματα με αποτέλεσμα την εξαγωγή βιολογικών συμπερασμάτων σχετικά με το ερώτημα που τίθεται σε κάθε πειραματική προσέγγιση (Patti et al. 2012). Οι υπόλοιπες παράμετροι του πειράματος όπως η ανάλυση του δείγματος, και η επεξεργασία των δεδομένων εξαρτώνται κάθε φορά από τον τύπο της μεθόδου που χρησιμοποιείται η οποία καθορίζει ποιο είναι το κατάλληλο λογισμικό και ποιες βάσεις δεδομένων πρέπει να χρησιμοποιηθούν (Nicholson & Wilson 2003). Τα αποτελέσματα μιας μεταβολομικής ανάλυσης μπορούν να παρουσιαστούν σε μορφή δικτύου και σε συνδυασμό με πρωτεϊνικά δίκτυα, τα οποία απεικονίζουν τις φυσικές αλληλεπιδράσεις των ενζύμων που καταλύουν τις μεταβολικές αντιδράσεις, μπορούν να συνεισφέρουν στην πρόβλεψη νέων αλληλεπιδράσεων μεταξύ των ενζύμων ή εναλλακτικών μεταβολικών αντιδράσεων που μπορούν τα ένζυμα να συμμετέχουν (Durek & Walther 2008, Huthmacher et al. 2008).

## ΑΝΤΙΚΕΙΜΕΝΟ ΜΕΛΕΤΗΣ

Το γονίδιο *FRA10AC1* χαρτογραφείται στην εύθραυστη θέση *FRA10A* του ανθρώπου στην χρωμοσωματική περιοχή 10q23.3.. Οι φορείς της *FRA10A* εμφανίζουν φαινότυπο ήπιας νοητικής υστέρησης, νευρολογικές διαταραχές και αναπτυξιακές ανωμαλίες στα άτομα-φορείς. Η μοριακή βάση της κυτταρογενετικής εμφάνισης της *FRA10A* είναι η επέκταση (~200 φορές) τρινοκυκλοπιδικών επαναλήψεων του τύπου (CGG)<sub>n</sub> που εντοπίζονται στην 5' μη μεταφραζόμενη περιοχή του γονιδίου *FRA10AC1*. Αυτή η επέκταση οδηγεί στην υπερμεθυλίωση της περιοχής αυτής και στη μεταγραφική καταστολή του αντίστοιχου αλληλομόρφου. Το γονίδιο *FRA10AC1* είναι μεταγραφικά ενεργό σε όλους τους ιστούς και η αντίστοιχη πρωτεΐνη είναι συντηρημένη σε πλήθος ευκαρυωτικών, πολυκύτταρων ή μονοκύτταρων οργανισμών, αλλά όχι σε προκαρυωτικούς οργανισμούς. Οι ενδείξεις από προηγούμενες μελέτες υποδεικνύουν τη συσχέτιση της *FRA10AC1* με τη διαδικασία της συναρμογής ή/και με άλλες στενά συσχετιζόμενες διαδικασίες του επεξεργασίας του RNA.

Η διατριβή στοχεύει στη λειτουργική ανάλυση της πρωτεΐνης *FRA10AC1* τόσο με *in silico* όσο και πειραματικές προσεγγίσεις. Η *in silico* προσέγγιση αφορά την ανασύσταση του δικτύου άμεσων πρωτεϊνικών αλληλεπιδράσεων της *FRA10AC1*, το οποίο την εντάσσει σαφώς, βάσει και των παραπάνω ενδείξεων, στο σωματίο συναρμογής. Ακολουθώντας, με στόχο την διερεύνηση της τοπολογίας της και των ιδιοτήτων της στο παραπάνω ριβονουκλεοπρωτεϊνικό σύμπλοκο, έγινε η ανασύσταση του ευρύτερου δικτύου πρωτεϊνικών αλληλεπιδράσεων ολόκληρου του σωματίου συναρμογής και των υποσυμπλόκων του. Η ανάλυση του δικτύου με υπολογιστικές μεθόδους συνεισφέρει στην αποσαφήνιση της «αρχιτεκτονικής» του μέσω της εξέτασης των τοπολογικών παραμέτρων του δικτύου και στην διερεύνηση της τοπολογίας της *FRA10AC1* εντός του δικτύου του σωματίου συναρμογής και των υποσυμπλόκων του. Η εύρεση του λειτουργικού ρόλου της διευρυμένης «γειτονιάς» της *FRA10AC1* στα δίκτυα αυτά προσφέρει χρήσιμες πληροφορίες για το βιολογικό ρόλο της στη δυναμική και λειτουργία του σωματίου.

Όσον αφορά την πειραματική προσέγγιση, αυτή στοχεύει στην διερεύνηση του βιολογικού ρόλου της *FRA10AC1* στην κυτταρική φυσιολογία. Ως εργαλείο χρησιμοποιήθηκε ένα κυτταρικό μοντέλο με μόνιμα κατεσταλμένη την έκφραση του ενδογενούς γονιδίου *FRA10AC1*, μέσω της τεχνολογίας RNAi, έπειτα από την εισαγωγή κατάλληλου διαγονιδίου. Για την μελέτη των συνεπειών της καταστολής της έκφρασης του *FRA10AC1* εφαρμόστηκαν ομικές αναλύσεις σε μεταγραφικό, πρωτεϊνικό και μεταβολικό επίπεδο. Η ανάλυση αυτή συνεισφέρει στη διερεύνηση των μοριακών μηχανισμών στους οποίους συμμετέχει η *FRA10AC1* και κατ' επέκταση στην κατανόηση του λειτουργικού της ρόλου, κατά κύριο λόγο εντός του σωματίου συναρμογής.

## **2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ**

### **2.1. Βάσεις δεδομένων**

Οι δημόσιες βάσεις δεδομένων αποτελούν ένα σημαντικό εργαλείο στη μεταγονιδιωματική εποχή για τους βιοεπιστήμονες καθώς είναι άμεσα προσβάσιμες και περιέχουν καταχωρημένες πληροφορίες σε όλα τα επίπεδα της βιολογικής πληροφορίας, παρέχοντας τη δυνατότητα αυτόματης ανάκτησης των δεδομένων για μια συγκεκριμένη αναζήτηση. Στις περισσότερες περιπτώσεις προσφέρουν την επιλογή σύνδεσης με ποικίλου τύπου δεδομένα τα οποία είναι διαθέσιμα σε άλλες πηγές.

#### 2.1.1. Βάσεις δεδομένων για χαρακτηριστικά των γονιδίων

##### 2.1.1.1. Πηγή πληροφοριών σε επίπεδο γονιδίου (Gene) (<https://www.ncbi.nlm.nih.gov/gene>)

Η βάση αυτή αποτελεί μια κατηγορία της βάσης δεδομένων NCBI (National Center for Biotechnology Information, <https://www.ncbi.nlm.nih.gov/>) και επικεντρώνεται σε δεδομένα που αφορούν το γονίδιο (Brown et al. 2015). Συγκεκριμένα παρέχει περιγραφή της γενωμικής οργάνωσης, ανάλυση της γονιδιωματικής αλληλουχίας, των κύριων και εναλλακτικών μεταγραφών, των γονιδιακών προϊόντων και των χαρακτηριστικών τους. Επιπλέον, διαθέτει πολλαπλές παραπομπές σε άλλες βάσεις δεδομένων προσφέροντας τη δυνατότητα ανάκτησης δεδομένων σχετικά με το λειτουργικό ρόλο και την υποκυτταρική τοπολογία των προϊόντων τους, τη συσχέτισή τους με κάποιο φαινότυπο καθώς και τα ορθόλογα μόρια και τις αλληλεπιδράσεις των πρωτεϊνών που παράγονται από αυτά. Η αναζήτηση πραγματοποιείται μέσω του ονόματος του γονιδίου ή του χαρακτηριστικού κωδικού γονιδίου (Gene ID) προσθέτοντας προαιρετικά το είδος του οργανισμού. Σύμφωνα με τα υπάρχοντα στατιστικά στοιχεία που είναι καταχωρημένα στη βάση, υπάρχουν πληροφορίες για περισσότερα από 20.000.000 γονίδια για 20.617 ταξινομικές βαθμίδες, με τα 16.913.525 να αφορούν διαφορετικά είδη ευκαρυωτών. Η διερεύνηση της βάσης σε επίπεδο γονιδίου χρησιμοποιήθηκε για το σχεδιασμό των πειραμάτων υποκλωνοποίησης των αλληλεπιδρώντων μορίων της πρωτεΐνης FRA10AC1, για την ανάκτηση των Gene ID για τη δημιουργία του συνόλου δεδομένων των συστατικών του σωματίου συναρμογής.

##### 2.1.1.2. Διερεύνηση ορθόλογων μορίων (HomoloGene) (<https://www.ncbi.nlm.nih.gov/homologene/>)

Η βάση δεδομένων HomoloGene ανήκει στην NCBI και αποτελεί ένα εργαλείο γονιδιακής ομολογίας το οποίο παρέχει τη δυνατότητα διερεύνησης της ύπαρξης πιθανών ορθόλογων μορίων σε διαφορετικά είδη οργανισμών. Η αναζήτηση πραγματοποιείται μέσω της χρήσης του ονόματος του γονιδίου και οδηγεί στη δυνατότητα εύρεσης των αντίστοιχων ονομάτων των ορθόλογων μορίων σε όσους οργανισμούς είναι διαθέσιμα. Όλες οι πληροφορίες για το εκάστοτε ορθόλογο μόριο είναι διαθέσιμες μέσω παραπομπής στη βάση δεδομένων Gene του NCBI. Οι πληροφορίες για τα ορθόλογα μόρια βασίζονται στη σύγκριση νουκλεοτιδικών αλληλουχιών ενώ η προσθήκη ενός νέου ορθολόγου προϋποθέτει την πλήρη

αλληλούχηση του γονιδιώματος του. Επιπλέον παρέχεται η δυνατότητα σύγκρισης και σε επίπεδο πρωτεϊνικών ακολουθιών. Η βάση αυτή χρησιμοποιήθηκε για την εύρεση ορθόλογων γονιδίων σε οργανισμούς-μοντέλα στο πλαίσιο αναζήτησης αλληλεπιδράσεων σε αντίστοιχες βάσεις δεδομένων.

## 2.1.2. Βάσεις δεδομένων για χαρακτηριστικά, λειτουργία και οντολογία των πρωτεϊνών

### 2.1.2.1. Βάση δεδομένων Γονιδιακής Οντολογίας (Gene Ontology Consortium) (<http://www.geneontology.org/>)

Ο όρος οντολογία στον τομέα της επιστήμης των υπολογιστών υποδηλώνει μία κατηγοριοποίηση όλων των εννοιών σε κάποιο πεδίο, συμπεριλαμβανομένων των αντικειμένων, των ιδιοτήτων και των λειτουργιών που απαιτούνται για να καθορίσουν πλήρως τα αντικείμενα του πεδίου. Είναι, δηλαδή, ένα καλά δομημένο και ορισμένο λεξιλόγιο το οποίο περιγράφει μία περιοχή έρευνας, τα αντικείμενά της και τις σχέσεις μεταξύ τους. Υπάρχουν διαφορετικοί τύποι σχέσεων ανάμεσα στους όρους μιας οντολογίας. Συγκεκριμένα, υπάρχουν σχέσεις που χρησιμοποιούνται σχεδόν σε όλες τις οντολογίες (foundational relations) (π.χ. is-a, part-of) και επιπλέον μπορεί να υπάρχουν σχέσεις χωρικές (spatial relations) (π.χ. located\_in, contained\_in, adjacent\_to), χρονικές (temporal relations) (π.χ. transformation\_of, derives\_from, preceded\_by) και σχέσεις συμμετοχής των όρων μιας οντολογίας (participation relations) (has\_participant, has\_agent) (Smith & Brochhausen 2010, Bodenreider 2008, Smith et al. 2007). Δομές οντολογίας χρησιμοποιούνται ευρέως και στο πεδίο της Βιολογίας. Ιδιαίτερα στη μετα-γονιδιωματική εποχή, ο συνεχής αυξανόμενος ρυθμός αλληλούχησης, ιδιαίτερα ολόκληρων γονιδιωμάτων, άλλαξε την οπτική υπό την οποία εξετάζεται η θεωρία και η πράξη στην πειραματική βιολογία. Τα προηγούμενα χρόνια οι βιοχημικοί χαρακτήριζαν τις πρωτεΐνες βάσει των διαφορών στην ενεργότητά τους και οι γενετιστές τα γονίδια από το φαινότυπο των μεταλλάξεων τους. Πλέον, είναι γνωστό ότι υπάρχουν ομάδες γονιδίων και πρωτεϊνών που παρουσιάζουν ομολογία (σε επίπεδο αλληλουχίας) σε πληθώρα οργανισμών. Έτσι, έχει δημιουργηθεί η ανάγκη για την περιγραφή κοινών βιολογικών χαρακτηριστικών (π.χ. η απόδοση ονοματολογίας και λεπτομερούς χαρακτηρισμού των γονιδίων και των προϊόντων τους καθώς και των ιδιοτήτων αυτών) υπό έναν κοινώς αποδεκτό και τυποποιημένο τρόπο. Έτσι δημιουργήθηκε η σύμπραξη της Γονιδιακής Οντολογίας (Gene Ontology Consortium 2004, 2015, Ashburner et al. 2000), το έργο της οποίας ξεκίνησε το 1998 στο πλαίσιο της συνεργασίας που αναπτύχθηκε μεταξύ βάσεων δεδομένων που μελετούν και αναλύουν τρεις οργανισμούς, της FlyBase (*D. melanogaster*), της Saccharomyces Genome Database (SGD) και της Mouse Genome Database (MGD). Από τότε στην κοινοπραξία της Γονιδιακής Οντολογίας έχουν συμμετάσχει πολλές άλλες βάσεις δεδομένων, εκ των οποίων και μερικές από τις πιο σημαντικές πηγές πληροφοριών φυτικών, ζωικών και μικροβιακών γονιδιωμάτων.

Η Γονιδιακή Οντολογία αποτελεί ένα δομημένο και ορισμένο λεξιλόγιο που επιτρέπει την υποβολή ερωτημάτων σε διαφορετικά επίπεδα καθιστώντας την έτσι εξαιρετικά ευέλικτη. Χάρη σε αυτό το λεξιλόγιο, οι επιστήμονες που επισημειώνουν όρους Γονιδιακής Οντολογίας (Gene Ontology ή GO terms) σε γονίδια (ή γονιδιακά προϊόντα), επιλέγουν κάθε φορά τον όρο εκείνου του επιπέδου ο οποίος αναλογεί στην

διαθέσιμη πληροφορία (Gene Ontology Consortium 2015, Harris et al. 2004, Ashburner et al. 2000). Η Γονιδιακή Οντολογία διαιρείται σε τρεις κατηγορίες οντολογιών οι οποίες παρέχουν πληροφορίες για όλα τα είδη οργανισμών. Αυτές είναι: α) η μοριακή λειτουργία (molecular function) η οποία περιγράφει τη βιοχημική ενεργότητα ενός γονιδιακού προϊόντος χωρίς να καθορίζει τη χρονική στιγμή ή το χώρο που αυτή λαμβάνει χώρα, β) η βιολογική διαδικασία (biological process) η οποία περιγράφει μια σειρά γεγονότων που λαμβάνουν χώρα από το συνδυασμό μιας ή περισσότερων μοριακών λειτουργιών, χωρίς να ισοδυναμεί απαραίτητα με ένα βιολογικό μονοπάτι και γ) το κυτταρικό διαμέρισμα (cellular component) το οποίο περιγράφει τα μέρη του κυττάρου στα οποία εντοπίζονται τα προϊόντα των γονιδίων. Το σημαντικότερο στοιχείο αυτών των κατηγοριών είναι η διασφάλιση της ύπαρξης κάθε όρου (GO term) μόνο σε μία από τις τρεις κατηγορίες (Gene Ontology Consortium 2015, Harris et al. 2004, Ashburner et al. 2000). Στην παρούσα εργασία διερευνήθηκε η Γονιδιακή Οντολογία μέσω των βάσεων δεδομένων Uniprot και DAVID που περιγράφονται αμέσως μετά.

#### 2.1.2.2. Βάση δεδομένων για την περιγραφή των πρωτεϊνών (UniProt, Universal Protein Resource) (<http://www.uniprot.org/>)

Η UniProt αποτελεί την εγκυρότερη πηγή αναφοράς για την περιγραφή των πρωτεϊνών καθώς διασταυρώνει τις πληροφορίες της με έναν μεγάλο αριθμό άλλων βάσεων δεδομένων. Έτσι, ενσωματώνει όλες τις διαθέσιμες πληροφορίες που υπάρχουν για τις πρωτεΐνες, τις ισομορφές και τις ακολουθίες τους. Προκειμένου να περιορίσει τον πλεονασμό που υπάρχει σε κωδικούς αναγνώρισης, οι οποίοι αναφέρονται στην ίδια πρωτεΐνη, έχει εντάξει όλες τις πρωτεϊνικές ισομορφές που παρουσιάζουν μικρές διαφορές στην ακολουθία τους (π.χ. λόγω της ύπαρξης πολυμορφισμών) στον ίδιο κωδικό αναγνώρισης UniProt (UniProt ID). Σε περίπτωση που οι ισομορφές αυτές διαφέρουν, δηλαδή, όταν η μετάφραση εναλλακτικών εξονίων οδηγεί σε μεγάλες διαφορές στην πρωτεϊνική ακολουθία, τότε αποδίδονται σε αυτές τις ισομορφές διαφορετικά UniProt IDs. Αντιθέτως, όταν υπάρχουν πολλαπλά αντίγραφα ενός γονιδίου λόγω γονιδιακού διπλασιασμού ή ρετρομετάθεσης, τα οποία δίνουν ταυτόσημες πρωτεϊνικές ακολουθίες τότε ορίζεται ένα μοναδικό UniProt ID παρόλο που αυτό αντιστοιχεί σε διαφορετικά γονίδια και Gene IDs. Η βάση αποτελείται από τέσσερα μέρη (Consortium 2008, The UniProt Consortium 2011, Bateman et al. 2017):

1) Τη βάση γνώσεων (UniProt Knowledgebase, UniProtKB) που αποτελεί το κεντρικό σημείο πρόσβασης για πρωτεϊνικά δεδομένα και διαθέτει παραπομπές σε πολλαπλές πηγές. Παρέχει λεπτομερή περιγραφή της πρωτεϊνικής ακολουθίας (ακολουθία αμινοξέων, ύπαρξη χαρακτηριστικών μοτίβων, θέσεις μεταμεταφραστικών τροποποιήσεων, θέσεις πρόσδεσης DNA ή RNA, θέση ενεργού κέντρου), καθώς και πληροφορίες για τις πρωτεϊνικές ισομορφές και την ακολουθία αυτών. Επίσης περιλαμβάνει δεδομένα για τη δομή της πρωτεΐνης, τον υποκυτταρικό της εντοπισμό, το λειτουργικό της ρόλο μέσω του αντίστοιχου κωδικού γονιδιακής οντολογίας (Gene ontology) (βλ. παρακάτω) και δεδομένα φυλογενετικής ανάλυσης. Τα δεδομένα της βάσης γνώσεων (UniProtKB) χωρίζονται σε δυο κατηγορίες: α) στις αναθεωρημένες καταχωρήσεις των πρωτεϊνών (reviewed Uniprot IDs) οι οποίες

προέρχονται από τη βάση Swiss-Prot και έχουν επεξεργαστεί από επιμελητές της βάσης UniProt (manually annotated) και β) στις καταχωρήσεις που έχουν ανακτηθεί αυτόματα (computationally analyzed) από τη βάση TrEMBL χωρίς περαιτέρω αναθεώρηση (unreviewed Uniprot IDs). Σύμφωνα με τα στατιστικά στοιχεία της βάσης (15<sup>η</sup> Φεβρουαρίου 2017) υπάρχουν 553.655 αναθεωρημένοι κωδικοί (reviewed) και 77.483.538 μη αναθεωρημένοι (unreviewed), εκ των οποίων (και από τις δυο κατηγορίες) οι 156.851 αφορούν τον άνθρωπο.

2) Το αρχείο UniProt (UniProt Archive, UniParc) το οποίο είναι ένα κέντρο αποθήκευσης πρωτεϊνικών ακολουθιών. Αυτή η κατηγορία δημιουργήθηκε διότι είχε παρατηρηθεί διαφορετική ονοματολογία για τις ίδιες πρωτεΐνες είτε στην ίδια είτε σε διαφορετικές βάσεις δεδομένων. Έτσι, το δεύτερο αυτό μέρος της UniProt απέδωσε ένα χαρακτηριστικό κωδικό σε κάθε πρωτεϊνική ακολουθία η οποία ήταν ίδια μεταξύ των διαφορετικών ειδών ή εντός ενός είδους (π.χ. σε ένα ιό και σε ένα στέλεχος του) αλλά εμφανιζόταν, και στις δυο περιπτώσεις, με διαφορετική ονομασία.

3) Την UniProt Reference Clusters (UniRef) η οποία περιέχει συμπλέγματα πρωτεϊνών (clusters) που παρουσιάζουν ομοιότητα στην πρωτεϊνική τους ακολουθία (συμπεριλαμβανομένων των ισομορφών) αντλώντας δεδομένα από τη UniProtKB και επιλεγμένες καταχωρήσεις από τη UniParc. Συγκεκριμένα, αν κάποιες ακολουθίες από οποιοδήποτε οργανισμό (ίδιο ή διαφορετικό) είναι ίδιες σε περισσότερα από έντεκα συνεχόμενα κατάλοιπα, τότε σε αυτό το σύμπλεγμα αποδίδεται ένας κωδικός UniRef100. Στη συνέχεια οι ακολουθίες του ίδιου συμπλέγματος UniRef100 συγκρίνονται μεταξύ τους και όσες παρουσιάζουν 90% ταυτότητα ή 80% αλληλοεπικάλυψη με τη μεγαλύτερη ακολουθία του συμπλέγματος δημιουργούν ένα επιπλέον σύμπλεγμα με κωδικό UniRef90. Τέλος, οι ακολουθίες που ανήκουν σε έναν συγκεκριμένο κωδικό UniRef90 συγκρίνονται μεταξύ τους και όσες παρουσιάζουν 50% ταυτότητα ή 80% αλληλοεπικάλυψη με τη μεγαλύτερη ακολουθία του συμπλέγματος, δημιουργούν ένα καινούργιο σύμπλεγμα με κωδικό UniRef50.

4) Την Proteomes που περιέχει πληροφορίες για τη σύσταση του πρωτεϊνώματος 147.053 ειδών οργανισμών.

Στην παρούσα εργασία χρησιμοποιήθηκε η πρώτη κατηγορία (UniProtKB), τόσο για την αναζήτηση των αναθεωρημένων κωδικών πρωτεΐνης (reviewed Uniprot IDs) όσο και για την αναζήτηση της γονιδιακής οντολογίας (Gene ontology) για τις υπό μελέτη πρωτεΐνες. Σε κάθε περίπτωση, η αναζήτηση έγινε με το επίσημο σύμβολο του γονιδίου.

### 2.1.2.3. Βάση δεδομένων DAVID (Database for Annotation, Visualization and Integrated Discovery) (<http://david.abcc.ncifcrf.gov/>)

Η βάση δεδομένων DAVID παρέχει τη δυνατότητα αναζήτησης όρων γονιδιακής οντολογίας (GO terms) για καταλόγους γονιδίων πολλών διαφορετικών ειδών οργανισμών, συμπεριλαμβανομένου και του ανθρώπου. Η πληροφορία για τους όρους γονιδιακής οντολογίας προέρχεται από τις βάσεις δεδομένων GO (<http://www.geneontology.org/page/go-database>) και PANTHER (<http://www.pantherdb.org/>) οι οποίες ανήκουν στο Gene Ontology Consortium (<http://www.geneontology.org/>) και τη βάση δεδομένων BioPortal (<https://bioportal.bioontology.org/>) η



οποία έχει ως σημείο αναφοράς το Gene Ontology Consortium για τον προσδιορισμό των όρων γονιδιακής οντολογίας. Η σύμπραξη γονιδιακής οντολογίας (Gene Ontology Consortium) χρησιμοποιεί τρεις μεθόδους απόδοσης GO terms στα γονίδια. Η πρώτη περιλαμβάνει τον προσδιορισμό GO term από ειδικούς επιμελητές (curators) οι οποίοι αξιολογούν τις ερευνητικές εργασίες που αφορούν πειραματικές προσεγγίσεις (manually annotated experimental gene products). Η δεύτερη περιλαμβάνει τον προσδιορισμό των οντολογικών κατηγοριών των γονιδίων για τα οποία δεν υπάρχουν βιβλιογραφικές αναφορές, από επιμελητές οι οποίοι χρησιμοποιούν αυτοματοποιημένους τρόπους για να εξετάσουν τα χαρακτηριστικά της αλληλουχίας των γονιδίων καθώς και τις πληροφορίες που υπάρχουν είτε σε βάσεις δεδομένων είτε στη βιβλιογραφία και αφορούν όμως το λειτουργικό ρόλο των ορθόλογων μορίων (computational curated annotated gene products). Η τρίτη μέθοδος είναι τελείως αυτοματοποιημένη, χωρίς την εποπτεία επιμελητή, και κατά αυτήν προσδιορίζονται οι λειτουργίες των γονιδίων, κατά κύριο λόγο βάσει της σύγκρισης των αλληλουχιών τους με γονίδια που έχουν παρόμοια αλληλουχία και γνωστό λειτουργικό ρόλο (computational electronic annotated gene products) (The Gene Ontology Consortium 2013, 2015). Σύμφωνα με τα στατιστικά στοιχεία της βάσης (15<sup>η</sup> Φεβρουαρίου 2017), υπάρχουν καταχωρημένοι πάνω από 40 εκατομμύρια όροι γονιδιακής οντολογίας για 45.000 είδη οργανισμών. Όλοι οι παραπάνω όροι αναφέρονται και στις τρεις κατηγορίες γονιδιακής οντολογίας αλλά στη διατριβή χρησιμοποιήθηκε μόνο η κατηγορία της βιολογικής διαδικασίας (biological process). Σε κάθε οντολογικό όρο περιγραφής αποδίδεται ένας βαθμός σημαντικότητας (Ease score) ο οποίος δείχνει τον εμπλουτισμό των οντολογικών κατηγοριών με γονίδια, για τα οποία έγινε η αναζήτηση της οντολογίας τους. Επιπλέον, η βάση παρέχει τη δυνατότητα ομαδοποίησης παρόμοιων όρων γονιδιακής οντολογίας συγκεκριμένης κατηγορίας (π.χ. biological process) (Functional Annotation Clustering) οι οποίοι περιλαμβάνουν κοινά γονίδια, κατασκευάζοντας με αυτόν τον τρόπο συμπλέγματα επισημείωσης όρων (annotation clusters). Τέλος, στη βάση αυτή δίνεται η δυνατότητα ανάκτησης δεδομένων τα οποία αφορούν βιολογικά μονοπάτια, πρωτεϊνικές αλληλεπιδράσεις, πρωτεϊνικά μοτίβα και οικογένειες, σχέσεις γονιδίων-φαινοτύπου και πρότυπα έκφρασης γονιδίων σε συγκεκριμένους ιστούς (Huang et al. 2007).

#### 2.1.2.4. Βάση δεδομένων για την υποκυτταρική τοπολογία πρωτεϊνών και τα επίπεδα έκφρασης γονιδίων (Human Protein Atlas) (<http://www.proteinatlas.org/>)

Η πηγή αναφοράς Human Protein Atlas περιέχει πληροφορίες για την υποκυτταρική τοπολογία των πρωτεϊνών και για τα επίπεδα έκφρασης των γονιδίων σε μεγάλο εύρος ιστών και οργάνων του ανθρώπου. Αποτελείται από τρία μέρη: την κατηγορία του κυττάρου (cell), του φυσιολογικού ιστού (normal tissue) και του καρκινικού ιστού (cancer tissue) (Pontén et al. 2011, Asplund et al. 2012, Uhlen et al. 2010).

Η πρώτη κατηγορία περιλαμβάνει πληροφορίες: α) για το πρότυπο υποκυτταρικής κατανομής των πρωτεϊνών σε 32 κυτταρικές δομές και οργανίδια μέσω πειραμάτων ανοσοφθορισμού και β) για τα πρότυπα έκφρασης των γονιδίων σε 56 κυτταρικές σειρές του ανθρώπου μέσω της μέτρησης των επιπέδων mRNA με αλληλούχηση RNA (RNA seq). Επιπρόσθετα, διαθέτει αποτελέσματα για το ποσοστό αποσιώπησης των

γονιδίων στην κυτταρική σειρά U-2OS (Human Bone Osteosarcoma Epithelial Cells) μέσω της μεθόδου παρεμβολής μέσω RNA (RNAi) ώστε να γίνει η επιβεβαίωση της υποκυτταρικής τοπολογίας των πρωτεϊνών.

Η δεύτερη κατηγορία περιλαμβάνει δεδομένα για την έκφραση των γονιδίων σε επίπεδο RNA και πρωτεΐνης σε διάφορους ιστούς. Τα δεδομένα σε επίπεδο RNA αφορούν την κατά μέσο όρο έκφραση του εκάστοτε γονιδίου στους αντίστοιχους ιστούς με αλληλούχηση RNA (RNA seq). Όσον αφορά τα πρωτεϊνικά δεδομένα, αυτά προέρχονται από πειράματα ανοσοϊστοχημείας (immunohistochemical staining) σε 76 διαφορετικούς κυτταρικούς τύπους που αντιστοιχούν σε 44 διαφορετικούς ιστούς.

Η τελευταία κατηγορία περιέχει 216 διαφορετικά καρκινικά δείγματα τα οποία αντιπροσωπεύουν τους 20 πιο κοινούς τύπους καρκίνου. Αυτά τα δείγματα χρησιμοποιούνται για την ανάκτηση των προτύπων έκφρασης των πρωτεϊνών μέσω ανοσοϊστοχημείας.

Σύμφωνα με τα στατιστικά στοιχεία της τελευταίας έκδοσης της βάσης (version 16, Release date: 04.12.2016) όλα τα παραπάνω δεδομένα αναφέρονται σε περισσότερα από 17.000 γονίδια του ανθρώπου (86% των γονιδίων που κωδικοποιούν πρωτεΐνες). Συγκεκριμένα, τα δεδομένα υποκυτταρικού εντοπισμού αφορούν περισσότερα από 12.000 πρωτεΐνες και από αυτά αντλήθηκαν πληροφορίες που χρησιμοποιήθηκαν στο πλαίσιο αυτής της εργασίας.

#### 2.1.2.5. Βάση δεδομένων για τις πρωτεΐνες που αποτελούν συστατικά του σωματίου συναρμογής (Spliceosome Database) (<http://spliceosomedb.ucsc.edu/>)

Η Spliceosome Database αποτελεί ένα εργαλείο αναζήτησης συστατικών του σωματίου συναρμογής (spliceosome) με βάση την ονομασία τους και το σύμπλοκο του σωματίου συναρμογής στο οποίο ανήκουν. Επιπλέον, μέσω της βάσης, μπορούν να αναζητηθούν οι ερευνητικές εργασίες (και μέσω του ονόματος του συγγραφέα των εργασιών αυτών) οι οποίες περιέχουν πειράματα απομόνωσης και ταυτοποίησης των συστατικών του σωματίου συναρμογής. Οι αναζητήσεις αυτές μπορεί να γίνουν σε εννέα διαφορετικά είδη οργανισμών συμπεριλαμβανομένου και του ανθρώπου. Επίσης παρέχεται η δυνατότητα εύρεσης πιθανών ορθόλογων μορίων και πληροφοριών σχετικών με την αλληλουχία των πρωτεϊνών (π.χ. συντηρημένα μοτίβα) μέσω σύνδεσης με άλλες βάσεις δεδομένων (Cvitkovic & Jurica 2013). Για να συμπεριληφθεί μια πρωτεΐνη ως συστατικό του σωματίου συναρμογής προϋποθέτει να πληροί ένα από τα παρακάτω κριτήρια: α) να συμμετέχει στη λειτουργία του σωματίου συναρμογής βάσει πειραματικών μεθόδων, β) να παρουσιάζει ομολογία με γνωστό παράγοντα συναρμογής και γ) να έχει ταυτοποιηθεί ως συστατικό του σωματίου συναρμογής μέσω φασματομετρίας μάζας (Cvitkovic & Jurica 2013). Στην παρούσα διατριβή, η βάση χρησιμοποιήθηκε συνδυαστικά με την PubMed για την εύρεση των ερευνητικών εργασιών που αφορούν την απομόνωση συστατικών του σωματίου συναρμογής και των επιμέρους συμπλόκων του. Αυτό επιτεύχθηκε με αναζήτηση στην κατηγορία των πειραμάτων φασματομετρίας μάζας (<http://spliceosomedb.ucsc.edu/experiments>) η οποία δίνει τη δυνατότητα εύρεσης των ερευνητικών εργασιών που αναφέρονται στα επιμέρους υποσύμπλοκα (E, A, B, B act, C και P). Σύμφωνα με την τελευταία δημοσίευση της βάσης δεδομένων-καθώς δεν είναι διαθέσιμα στατιστικά στοιχεία στην

δημόσια πλατφόρμα- μέχρι το 2013 υπήρχαν περισσότερες από 3600 καταχωρήσεις γονιδίων-πρωτεϊνών που προέρχονται από οκτώ είδη οργανισμών (Cvitkovic & Jurica 2013).

### 2.1.3. Βάσεις δεδομένων για εύρεση βιβλιογραφικών αναφορών και πληροφοριών για γενετικά νοσήματα

#### 2.1.3.1. Βάση δεδομένων βιβλιογραφικών αναφορών (PubMed) (<https://www.ncbi.nlm.nih.gov/pubmed/>)

Η PubMed αποτελεί κατηγορία της βάσης NCBI και περιέχει πάνω από είκοσι έξι εκατομμύρια βιβλιογραφικές αναφορές στα πεδία της βιοϊατρικής και της υγείας που προέρχονται από τη βάση δεδομένων MEDLINE ([https://www.nlm.nih.gov/pubs/fact\\_sheets/medline.html](https://www.nlm.nih.gov/pubs/fact_sheets/medline.html)), από διεθνή επιστημονικά περιοδικά και ηλεκτρονικά διαθέσιμα βιβλία. Η αναζήτηση μιας συγκεκριμένης ερευνητικής μελέτης μπορεί να γίνει βάσει λέξεων-κλειδιών, της χρονολογίας δημοσίευσης και των συγγραφέων.

#### 2.1.3.2. Βάση δεδομένων για γενετικές διαταραχές του ανθρώπου (OMIM) (<http://www.ncbi.nlm.nih.gov/omim/>)

Η OMIM (Online Mendelian Inheritance in Man) αποτελεί κατηγορία της βάσης NCBI και είναι μια ολοκληρωμένη και έγκυρη βάση γνώσεων των γενετικών διαταραχών του ανθρώπου και των γονιδίων που σχετίζονται με αυτά. Συνεισφέρει στην υποστήριξη της έρευνας της γενετικής του ανθρώπου και στην πρακτική της κλινικής γενετικής (Amberger & Hamosh 2017). Σύμφωνα με τα στατιστικά στοιχεία της βάσης κατά την τελευταία ενημέρωσή της (10<sup>η</sup> Φεβρουαρίου 2017), έχει καταχωρημένα 3.708 γονίδια, μεταλλάξεις των οποίων προκαλούν ή συσχετίζονται με νοσήματα. Από αυτά, τα 3.361 γονίδια είναι υπεύθυνα για 4950 μενδελικές διαταραχές, τα 501 έχουν συσχετιστεί με 702 πολυπαραγοντικά νοσήματα και τα 117 εμφανίζουν μεταλλάξεις σε σωματικά κύτταρα σε 211 τύπους καρκίνου. Μερικά γονίδια προσμετρώνται περισσότερες από μία φορές καθώς κάποιες διαταραχές εμφανίζουν γενετική ετερογένεια και κάποια γονίδια συσχετίζονται με διαφορετικά πολυπαραγοντικά νοσήματα. Η βάση αυτή χρησιμοποιήθηκε για την αναζήτηση των γενετικών νοσημάτων με τα οποία σχετίζονται τα μόρια του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής.

#### 2.1.4. Βάσεις δεδομένων πρωτεϊνικών αλληλεπιδράσεων του ανθρώπου

Τα δεδομένα πρωτεϊνικών αλληλεπιδράσεων που χρησιμοποιήθηκαν στην παρούσα εργασία ανακτήθηκαν από τη μετα-βάση PICKLE (Klapa et al. 2013, Gioutlakis et al. 2017) η οποία περιλαμβάνει πληροφορίες από πέντε κύριες πρωτογενείς δημόσιες βάσεις δεδομένων πρωτεϊνικών αλληλεπιδράσεων του ανθρώπου που στηρίζονται στη βιβλιογραφία (HPRD, BioGRID, IntAct, MINT, DIP). Στη συνέχεια παρατίθενται αναλυτικά τα χαρακτηριστικά της καθεμίας.

2.1.4.1. BioGRID (Biological General Repository for Interaction Datasets) (<https://thebiogrid.org>)

Η βάση δεδομένων BioGRID παρέχει πληροφορίες για πρωτεϊνικές αλληλεπιδράσεις (physical interactions), γενετικές αλληλεπιδράσεις (genetic interactions), δηλαδή αλληλεπιδράσεις μεταξύ γονιδίων που είναι υπεύθυνες για την εκδήλωση ενός φαινοτύπου, χημικές αλληλεπιδράσεις μεταξύ γονιδίων και χημικών ουσιών καθώς και μετα-μεταφραστικές τροποποιήσεις για 61 διαφορετικά είδη οργανισμών συμπεριλαμβανομένου του ανθρώπου και όλων των κύριων οργανισμών-μοντέλων (Chatr-Aryamontri et al. 2017, Stark 2006). Η βάση αυτή αποτελεί τη μεγαλύτερη βάση δεδομένων καθώς διαθέτει πολύ μεγάλο όγκο δεδομένων πρωτεϊνικών αλληλεπιδράσεων τόσο συνολικά για όλα τα είδη (477.535) όσο και για τον άνθρωπο (278.407), και τη βάση αναφοράς για τους περισσότερους ερευνητές. Για κάθε πρωτεϊνική αλληλεπίδραση, υπάρχει ο τύπος του πειράματος με τον οποίο ταυτοποιήθηκε, ο χαρακτηρισμός του ως πείραμα μικρής ή μεγάλης κλίμακας και η αντίστοιχη βιβλιογραφική αναφορά. Ωστόσο, δεν γίνεται διαχωρισμός μεταξύ των άμεσων αλληλεπιδράσεων και των αλληλεπιδράσεων εντός ενός συμπλόκου. Η αναζήτηση των αλληλεπιδράσεων μπορεί να γίνει είτε βάσει του κωδικού Gene ID ή του επίσημου ονόματος του γονιδίου. Τα στατιστικά στοιχεία της τελευταίας έκδοσης της βάσης (3.4.145, Φεβρουάριος 2017) παρουσιάζονται στον Πίνακα 1. Φαίνεται ο αριθμός των πρωτεϊνών, των γονιδίων, το σύνολο των πρωτεϊνικών, γενετικών και χημικών αλληλεπιδράσεων και ο αριθμός των δημοσιεύσεων. Επιπλέον σημειώνεται ο αριθμός των πρωτεϊνών που υφίσταται μετα-μεταφραστικές τροποποιήσεις και οι αντίστοιχες δημοσιεύσεις. Αυτά τα στοιχεία αναφέρονται σε όλα τα είδη οργανισμών που είναι καταχωρημένα στη βάση ενώ στις παρενθέσεις παρατίθενται τα αντίστοιχα δεδομένα για τον άνθρωπο.

**Πιν. 1: Στατιστικά στοιχεία της βάσης δεδομένων BioGRID.**

Αλληλεπιδράσεις	Αριθμός αλληλεπιδράσεων	Πρωτεΐνες-Γονίδια	Δημοσιεύσεις	Χημικές ουσίες
Πρωτεϊνικές	477.535 (278.407)	58.695 (20.935)	39.999 (25.791)	-
Γενετικές	643.354 (1.768)	17.398 (1.663)	11.577 (287)	-
Χημικές	11.805 (10.646)	2.525 (2.136)	9.083 (8.808)	5.031 (4.450)
Μετα-μεταφραστικές τροποποιήσεις	-	14.999 (10.523)	4.317 (3.332)	-

2.1.4.2. HPRD (Human Protein Reference Database) (<http://www.hprd.org/>)

Η βάση δεδομένων HPRD αποτελεί πηγή πληροφοριών για το πρωτεϊνωμα αποκλειστικά του ανθρώπου. Ειδικότερα, περιλαμβάνει δεδομένα για τις πρωτεϊνικές αλληλεπιδράσεις, τις μετα-μεταφραστικές τροποποιήσεις, την έκφραση των γονιδίων σε συγκεκριμένους ιστούς, τον υποκυτταρικό εντοπισμό των πρωτεϊνών, την ύπαρξη χαρακτηριστικών περιοχών (domains) και μοτίβων (motif) καθώς και σχετιζόμενες ασθένειες (Prasad et al. 2009). Αποτελεί μια από τις παλαιότερες βάσεις δεδομένων

πρωτεϊνικών αλληλεπιδράσεων (δημιουργήθηκε το 2003) και λειτουργεί μέσω μιας ομάδας επιμελητών που διαβάζουν και επιλέγουν δεδομένα από τη βιβλιογραφία (Prasad et al. 2009). Οι αλληλεπιδράσεις κατατάσσονται σε άμεσες (direct) και αυτές που λαμβάνουν χώρα εντός ενός συμπλόκου (complex). Για την καθεμία από αυτές αναφέρεται ο τύπος του πειράματος με τον οποίο ταυτοποιήθηκε και δίνεται η δυνατότητα σύνδεσης με την αντίστοιχη ερευνητική εργασία στη βάση δεδομένων PubMed. Για τα άλλα δεδομένα υπάρχουν πολλαπλές παραπομπές σε άλλες βάσεις δεδομένων με κυρίαρχη την PubMed. Σύμφωνα με τα στατιστικά στοιχεία της βάσης (15<sup>η</sup> Φεβρουαρίου 2017), υπάρχουν καταχωρημένες 30.047 πρωτεΐνες, 41.327 πρωτεϊνικές αλληλεπιδράσεις, 112.158 μετα-μεταφραστικές τροποποιήσεις, 112.158 και 22.490 καταχωρήσεις για το πρότυπο έκφρασης των γονιδίων και για τον υποκυτταρικό εντοπισμό των πρωτεϊνών, αντίστοιχα, 470 καταχωρήσεις χαρακτηριστικών περιοχών (domains) και 453.521 συνδέσεις με τη βάση δεδομένων PubMed. Παρόλου που στην παρουσίαση των στατιστικών στοιχείων δεν διαχωρίζεται ο αριθμός των πρωτεϊνών στον οποίο αναφέρονται οι πρωτεϊνικές αλληλεπιδράσεις, αυτή η βάση περιέχει ένα σημαντικό μεγάλο όγκο δεδομένων πρωτεϊνικών αλληλεπιδράσεων που την κατατάσσει στις μεγαλύτερες βάσεις δεδομένων αυτής της κατηγορίας. Η αναζήτηση στη συγκεκριμένη βάση μπορεί να γίνει μέσω της χρήσης του συμβόλου της πρωτεΐνης ή του κωδικού Gene ID γονιδίου.

#### 2.1.4.3. IntAct (Database of protein interaction data) (<http://www.ebi.ac.uk/intact>)

Η βάση δεδομένων IntAct αποτελεί ενεργό μέλος του International Molecular Exchange (IMEx) Consortium, μίας σύμπραξης (<http://www.imexconsortium.org/>) μεταξύ 16 δημόσιων βάσεων δεδομένων πρωτεϊνικών αλληλεπιδράσεων (12 ενεργά μέλη, 2 ανενεργά και 2 υποψήφια, μεταξύ των οποίων η BioGRID). Αυτές έχουν καθιερώσει ένα κοινό αποδεκτό σύνολο κανόνων επιμέλειας, οι οποίοι προσδιορίζουν τον τρόπο ανάκτησης πληροφοριών για την εκάστοτε αλληλεπίδραση, τη διαδικασία ανάλυσης κάθε δημοσίευσης με στόχο τη σωστή καταγραφή των αλληλεπιδράσεων που περιλαμβάνει, και το πρότυπο παρουσίασης των πληροφοριών (Orchard et al. 2012). Η IntAct παρέχει πληροφορίες κυρίως για δεδομένα πρωτεϊνικών αλληλεπιδράσεων και σε μικρότερο βαθμό δεδομένα που αφορούν αλληλεπιδράσεις πρωτεϊνών-μικρών μορίων (π.χ. φωσφολιπιδίων) και πρωτεϊνών-νουκλεϊκών οξέων (Kerrien et al., 2012). Οι πρωτεϊνικές αλληλεπιδράσεις διακρίνονται σε άμεσες («non spoke») και σε αυτές που λαμβάνουν χώρα εντός ενός συμπλόκου («spoke») παρέχοντας τη δυνατότητα διάκρισής τους κατά την ανάκτηση. Σύμφωνα με την τελευταία έκδοση της βάσης (έκδοση 4.2.6, Φεβρουάριος 2017), υπάρχουν καταχωρημένες 454.760 αλληλεπιδράσεις για 95.487 αλληλεπιδρώντα μόρια που στηρίζονται σε 14.495 δημοσιεύσεις και 38.781 πειραματικές μεθόδους. Από τα παραπάνω, το 52,2% των αλληλεπιδράσεων αφορούν τον άνθρωπο, 19,6% τον σακχαρομύκητα, 9,9% τον ποντικό, 9,6% τους οργανισμούς μοντέλα *D. melanogaster* και *C. elegans* και το υπόλοιπο 8,7% άλλα είδη οργανισμών συμπεριλαμβανομένου του *E. coli* και του *A. thaliana*. Όσον αφορά τις πειραματικές μεθόδους, το 35,6% των αλληλεπιδράσεων ταυτοποιήθηκαν με το σύστημα δυο υβριδίων στο σακχαρομύκητα, το 15,2% με καθαρισμό συγγένειας (tandem affinity purification), το 27,4 % με

μεθόδους συν-ανοσοκατακρήμνισης και συν-κατακρήμνισης (13,2% anti-tag co-immunoprecipitation, 4,6% anti bait co-immunoprecipitation, 9,6% pull down) και το υπόλοιπο 21,8% με άλλες μεθόδους. Από το 2013 η βάση αυτή έχει ενσωματώσει και τα δεδομένα της βάσης MINT (η οποία αναλύεται αμέσως μετά) (Orchard et al. 2014) γεγονός που πρέπει να ληφθεί υπόψιν στη σύγκριση των δεδομένων της σε σχέση με άλλες βάσεις δεδομένων. Η αναζήτηση πραγματοποιήθηκε μέσω του χαρακτηριστικού κωδικού πρωτεΐνης (Uniprot ID) ενώ εναλλακτικά μπορεί να γίνει με το επίσημο όνομα του γονιδίου.

#### 2.1.4.4. MINT (Molecular INTeraction database) (<http://mint.bio.uniroma2.it/mint/>)

Η βάση δεδομένων MINT αποτελεί και αυτή ενεργό μέλος του IMEx Consortium (Orchard et al. 2012) και από το 2013 τα δεδομένα της έχουν ενσωματωθεί στην IntAct (Orchard et al. 2014, Licata & Orchard 2016). Περιέχει δεδομένα πρωτεϊνικών αλληλεπιδράσεων για τον άνθρωπο και άλλα 610 είδη οργανισμών. Για την καθεμία από αυτές αναφέρει τον αντίστοιχο τύπο του πειράματος με τον οποίο έχει ταυτοποιηθεί η εκάστοτε αλληλεπίδραση, χωρίς να κάνει διάκριση μεταξύ των άμεσων αλληλεπιδράσεων και αυτών που λαμβάνουν χώρα εντός ενός συμπλόκου. Σύμφωνα με τα στατιστικά στοιχεία της βάσης (15<sup>η</sup> Φεβρουαρίου 2017), περιέχει 125.464 πρωτεϊνικές αλληλεπιδράσεις για 25.530 πρωτεΐνες που στηρίζονται σε 5.941 βιβλιογραφικές αναφορές. Συγκεκριμένα για τον άνθρωπο, διαθέτει 10.807 δεδομένα πρωτεϊνικών αλληλεπιδράσεων τα οποία αναφέρονται σε 8.644 δημοσιεύσεις. Ο όγκος των δεδομένων που αφορούν τον άνθρωπο δεν είναι ιδιαίτερα μεγάλος γεγονός που την κατατάσσει στις μικρότερες βάσεις δεδομένων. Ο μεγαλύτερος όγκος δεδομένων αναφέρεται στο σακχαρομύκητα (61.071) και στον ποντικό (23.148). Η αναζήτηση στη συγκεκριμένη βάση μπορεί να γίνει είτε μέσω του χαρακτηριστικού κωδικού πρωτεΐνης (Uniprot ID) ή του επίσημου ονόματος του γονιδίου.

#### 2.1.4.5. DIP (Database of Interacting Proteins) (<http://dip.doe-mbi.ucla.edu>)

Η βάση δεδομένων DIP αποτελεί ενεργό μέλος IMEx Consortium (Orchard et al. 2012). Περιλαμβάνει δεδομένα πρωτεϊνικών αλληλεπιδράσεων για δέκα διαφορετικά είδη οργανισμών συμπεριλαμβανομένου και του ανθρώπου. Διαχωρίζει τις αλληλεπιδράσεις σε άμεσες («direct») και σε αυτές που λαμβάνουν χώρα εντός ενός συμπλόκου («complex») (Salwinski 2004). Σε καθεμία από αυτές δίνει την πληροφορία της κλίμακας του πειράματος που ταυτοποιήθηκε η εκάστοτε αλληλεπίδραση δηλαδή αν αυτό είναι μικρής (low-throughput) ή μεγάλης κλίμακας (high-throughput). Σύμφωνα με τα στατιστικά στοιχεία της βάσης (15<sup>η</sup> Φεβρουαρίου 2017), περιλαμβάνει 80.820 δεδομένα πρωτεϊνικών αλληλεπιδράσεων για 28.293 πρωτεΐνες που στηρίζονται σε 79.337 πειραματικές μεθόδους. Συγκεκριμένα για τον άνθρωπο, περιέχει 9078 αλληλεπιδράσεις για 5046 πρωτεΐνες που στηρίζονται σε 13.817 πειραματικές μεθόδους. Παρατηρώντας τα στοιχεία που αφορούν τον άνθρωπο αλλά και συνολικά όλα τα είδη, προκύπτει ότι η συγκεκριμένη βάση ανήκει στις μικρότερες βάσεις δεδομένων αυτής της κατηγορίας.

#### 2.1.4.6. Μετα-βάση πρωτεϊνικών αλληλεπιδράσεων του ανθρώπου PICKLE (<http://www.pickle.gr/>)

Η μετα-βάση PICKLE (Klapa et al. 2013, Gioutlakis et al. 2017) ενσωματώνει, όπως προαναφέρθηκε, δεδομένα πρωτεϊνικών αλληλεπιδράσεων από 5 βάσεις δεδομένων πρωτεϊνικών αλληλεπιδράσεων χρησιμοποιώντας ως σύνολο αναφοράς το συνολικό πρωτεϊνωμα του ανθρώπου όπως αυτό προκύπτει με τη συλλογή όλων των αναθεωρημένων κωδικών πρωτεΐνης (UniProt IDs) που βρίσκονται καταχωρημένοι στη UniProt (UniProtKB). Για τη συμβατότητα αυτών των δεδομένων, έχει γίνει αυτόματη μετατροπή των διαφόρων κωδικών πρωτεΐνης ή γονιδίου (με τους οποίους είναι καταχωρημένες οι πληροφορίες στην εκάστοτε βάση δεδομένων) στον αντίστοιχο κωδικό UniProt ID. Συγκεκριμένα, η PICKLE ενσωματώνει όλες τις αλληλεπιδράσεις από τις πέντε βάσεις δεδομένων και τις κατατάσσει σε τρεις κλάσεις (class), φιλτράροντας όλες τις γενετικές αλληλεπιδράσεις και τις αλληλεπιδράσεις ανάμεσα σε πρωτεΐνη και νουκλεϊκό οξύ. Στην 1<sup>η</sup> κατηγορία («first» class) περιλαμβάνονται οι αλληλεπιδράσεις που φέρουν το χαρακτηρισμό «άμεσες» (direct) στις βάσεις δεδομένων ή έχουν ταυτοποιηθεί με πειραματικές μεθόδους, μεγάλης ή μικρής κλίμακας, που περιγράφουν άμεσες αλληλεπιδράσεις, για παράδειγμα, το σύστημα δυο υβριδίων, μέθοδοι κρυσταλλογραφίας, δοκιμή φωσφορυλίωσης και μεθυλίωσης (phosphorylation & methylation reaction), τη μέθοδο μεταφοράς φθορισμού μέσω συντονισμού (FRET), δοκιμή συμπληρωματικότητας πρωτεϊνών (protein complementation assay, PCA), ανασύσταση συμπλόκου (reconstituted complex) κ.α., οι οποίες περιγράφονται αναλυτικά στους Gioutlakis et al. (2017). Οι αλληλεπιδράσεις αυτής της κατηγορίας έχουν τον μεγαλύτερο βαθμό αξιοπιστίας (confidence score) ο οποίος έχει οριστεί στην τιμή 1. Στην 2<sup>η</sup> κατηγορία περιλαμβάνονται οι αλληλεπιδράσεις που έχουν ταυτοποιηθεί με οποιαδήποτε μέθοδο (εκτός από αυτές που αναφέρονται στην 1<sup>η</sup> κατηγορία) σε τουλάχιστον μία μελέτη μικρής κλίμακας, ενώ στην 3<sup>η</sup> κατηγορία αυτές που έχουν ταυτοποιηθεί μόνο σε μελέτες μεγάλης κλίμακας. Ο βαθμός αξιοπιστίας (confidence score) αυτών των δυο κατηγοριών έχει οριστεί στις τιμές 2 και 3, αντίστοιχα. Οι αλληλεπιδράσεις της 1<sup>ης</sup> και της 2<sup>ης</sup> κατηγορίας είναι καταχωρημένες στα σύνολα δεδομένων της μεταβάσης που χαρακτηρίζονται ως «βασικό» (standard) και ως «ελεγμένο» (cross-checked). Αυτά τα δυο σύνολα περιέχουν τις ίδιες αλληλεπιδράσεις με τη διαφορά ότι στο σύνολο δεδομένων cross-checked γίνεται περαιτέρω έλεγχος του βαθμού αξιοπιστίας των αλληλεπιδράσεων διασταυρώνοντας πληροφορίες που προέρχονται από διαφορετικές βάσεις δεδομένων και έτσι η αξιοπιστία κάποιων αλληλεπιδράσεων μπορεί να αναβαθμιστεί ή και να υποβιβαστεί (Gioutlakis et al. 2017). Οι αλληλεπιδράσεις όλων των κατηγοριών είναι καταχωρημένες στο «μη φιλτραρισμένο» (unfiltered) σύνολο δεδομένων (Gioutlakis et al. 2017). Στην παρούσα εργασία χρησιμοποιήθηκε το σύνολο δεδομένων cross-checked.

Σύμφωνα με τα στατιστικά στοιχεία, η PICKLE (PICKLE Release 2.2) περιέχει 161.007 αλληλεπιδράσεις για 15.434 UniProt IDs που στηρίζονται σε 38.268 δημοσιευμένες εργασίες. Τα δεδομένα πρωτεϊνικών αλληλεπιδράσεων που χρησιμοποιήθηκαν στη παρούσα διατριβή προήλθαν από τα συνεργαζόμενα

εργαστήρια Γενικής Βιολογίας (Ιατρική Σχολή, Πανεπιστήμιο Πατρών) και Μεταβολικής Μηχανικής και Βιολογίας Συστημάτων, Ινστιτούτο Χημικών Μηχανικών Επιστημών (Ίδρυμα Έρευνας και Τεχνολογίας, FORTH/ICE-HT, Πάτρα) (Επικεφαλής: Καθ. Νικόλαος Μοσχονάς, Δρ. Κλάπα Μαρία, αντίστοιχα).

#### 2.1.5. Βάσεις δεδομένων πρωτεϊνικών αλληλεπιδράσεων σε οργανισμούς-μοντέλα

Τα δεδομένα πρωτεϊνικών αλληλεπιδράσεων που χρησιμοποιήθηκαν στην παρούσα εργασία προήλθαν, επιπλέον, και από πρωτογενείς βάσεις δεδομένων στους οργανισμούς μοντέλα *D. melanogaster* και *C. elegans*, οι οποίες αναλύονται παρακάτω.

##### 2.1.5.1. Droid (Drosophila Interactions Database) (<http://droidb.org/>)

Η βάση δεδομένων Droid είναι μια δημόσια πηγή πληροφοριών που στηρίζεται στη βιβλιογραφία για πρωτεϊνικές και γενετικές αλληλεπιδράσεις καθώς και αλληλεπιδράσεις μεταξύ γονιδίων-μεταγραφικών παραγόντων και γονιδίων-microRNA (Murali et al. 2011, Yu et al. 2008). Τα δεδομένα πρωτεϊνικών αλληλεπιδράσεων στη *D. melanogaster* έχουν προκύψει κυρίως με το σύστημα δυο υβριδίων στο σακχαρομύκητα και με τη μέθοδο της ανοσοκατακρήμνισης σε συνδυασμό με φασματομετρία μάζας. Επίσης, περιλαμβάνει δεδομένα πρωτεϊνικών αλληλεπιδράσεων από άλλες βάσεις δεδομένων όπως είναι η BioGRID, η IntAct και η MINT. Ακόμη, διαθέτει ένα σύνολο προβλεπόμενων (predicted) αλληλεπιδράσεων μεταξύ πρωτεϊνών στη *D. melanogaster* βάσει διαθέσιμων πειραματικών δεδομένων για την αλληλεπίδραση των ορθόλογων πρωτεϊνών τους σε άλλα είδη οργανισμών, στηριζόμενη στην ιδέα των ιντερόλογων (Interolog data), δηλαδή των συντηρημένων αλληλεπιδράσεων μεταξύ διαφορετικών ειδών. Τέλος, περιέχει ένα σύνολο αδημοσίευτων δεδομένων από ένα πείραμα μεγάλης κλίμακας με το σύστημα δυο υβριδίων στο σακχαρομύκητα, που πραγματοποιήθηκε από την ερευνητική ομάδα του Dr. R. Finley ([www.proteome.wayne.edu](http://www.proteome.wayne.edu)) (Murali et al. 2012). Στη παρούσα διατριβή, η εύρεση των αλληλεπιδρώντων μορίων πραγματοποιήθηκε με τη χρήση του ονόματος του γονιδίου του ανθρώπου, και επιλέχθηκαν μόνο εκείνες οι αλληλεπιδράσεις που προέρχονταν από το σύστημα των δυο υβριδίων στο σακχαρομύκητα. Σύμφωνα με τα στατιστικά στοιχεία της βάσης (16 Δεκεμβρίου 2015), υπάρχουν 537.562 καταχωρημένες αλληλεπιδράσεις για 15.188 πρωτεΐνες. Συγκεκριμένα, τα δεδομένα που προέρχονται από πειράματα με το σύστημα δύο υβριδίων αφορούν 30.355 αλληλεπιδράσεις για 11.557 πρωτεΐνες, ενώ αυτά που προέρχονται από μεθόδους ανοσοκατακρήμνισης σε συνδυασμό με φασματομετρία μάζας είναι 61.643 αλληλεπιδράσεις για 5231 πρωτεΐνες.

##### 2.1.5.2. Worm Interactome Database ([http://interactome.dfci.harvard.edu/C\\_elegans/](http://interactome.dfci.harvard.edu/C_elegans/))

Η βάση δεδομένων Worm Interactome Database περιέχει δεδομένα πρωτεϊνικών αλληλεπιδράσεων στον *C. elegans* τα οποία έχουν προκύψει από το συνδυασμό τριών πειραμάτων μεγάλης κλίμακας με το σύστημα δυο υβριδίων στο σακχαρομύκητα. Σύμφωνα με την τελευταία ενημέρωση του 2009 (δεν είναι διαθέσιμη πιο πρόσφατη ενημέρωση) περιέχει 3.864 άμεσες πρωτεϊνικές αλληλεπιδράσεις (Simonis et al. 2009).



Για την αναζήτηση στη συγκεκριμένη βάση δεδομένων, πρώτα προηγήθηκε η διερεύνηση στην βάση Homologene για τον έλεγχο της ύπαρξης ορθόλογων μορίων των πρωτεϊνών του ανθρώπου στον *C. elegans*, και έπειτα πραγματοποιήθηκε η αναζήτηση των αλληλεπιδράσεων των πρωτεϊνών μέσω της χρήσης των κωδικών του *C. elegans*.

## 2.2. Υπολογιστικά εργαλεία οπτικοποίησης και ανάλυσης δικτύων

### 2.2.1. Πρόγραμμα οπτικοποίησης δικτύων, Cytoscape ([www.cytoscape.org](http://www.cytoscape.org), version 3.2.1)

Η οπτικοποίηση του ανακατασκευασμένου δικτύου έγινε με τη χρήση του προγράμματος Cytoscape (Shannon et al. 2003) σε διάταξη «Edge-weighted Spring-Embedded». Σε αυτή τη διάταξη οι πρωτεΐνες-κόμβοι του δικτύου ορίζονται ως φυσικά αντικείμενα που απωθούνται μεταξύ τους και οι συνδέσεις ανάμεσα στις πρωτεΐνες-κόμβους ορίζονται σαν μεταλλικά «ελατήρια» που συνδέονται με το ζεύγος των κόμβων. Αυτά τα «ελατήρια» απωθούν ή έλκουν τα καταληκτικά τους σημεία (end points) βάσει των δυνάμεων των συνδέσεων ανάμεσα στους κόμβους με στόχο να οριστούν οι θέσεις των πρωτεϊνών-κόμβων με τέτοιο τρόπο ώστε να ελαχιστοποιείται το σύνολο των δυνάμεων του δικτύου (Kamada & Kawai 1989). Ο προσδιορισμός των απλών (simple) και των σύνθετων (complex) τοπολογικών παραμέτρων του δικτύου έγινε με την εφαρμογή Network Analyzer plugin (Assenov et al. 2008) του προγράμματος Cytoscape.

#### ➤ Δημιουργία συμπλεγμάτων πρωτεϊνών με τον αλγόριθμο Cluster one

Ο αλγόριθμος «Cluster one» λειτουργεί αναπτύσσοντας περιοχές που περιέχουν πρωτεΐνες στενά συνδεδεμένες μεταξύ τους ξεκινώντας από μία ή περισσότερες πρωτεΐνες-κόμβους και χρησιμοποιώντας ένα μέτρο ποιότητας της συνδεσιμότητας που ονομάζεται συνεκτικότητα (cohesiveness) (Nepusz et al. 2012). Το μέτρο αυτό υπολογίζει πόσο πιθανό είναι για μια ομάδα πρωτεϊνών να σχηματίσει ένα σύμπλεγμα (cluster) με βάση δυο ιδιότητες: α) την ύπαρξη πολλών συνδέσεων μεταξύ των πρωτεϊνών-κόμβων αποκλειστικά εντός του συμπλέγματος (υψηλή τιμή «in-weight») και β) τον καλό διαχωρισμό του συμπλέγματος από το υπόλοιπο δίκτυο, δηλαδή την ύπαρξη μικρού αριθμού συνδέσεων των πρωτεϊνών του συμπλέγματος με άλλες εκτός αυτού (Nepusz et al. 2012). Το κατώφλι για τον αριθμό των κόμβων ορίστηκε στο 5, δηλαδή δεν υπολογίστηκαν τα συμπλέγματα που περιείχαν λιγότερους από πέντε κόμβους, και για το βαθμό πυκνότητας {συνολικός αριθμός συνδέσεων των πρωτεϊνών-κόμβων που βρίσκονται αποκλειστικά εντός του συμπλέγματος (in-weight) διαιρούμενος με όλες τις θεωρητικά πιθανές πρωτεϊνικές αλληλεπιδράσεις εντός αυτού} ορίστηκε το 0,2. Οι προκαθορισμένες τιμές του αλγορίθμου ήταν 3 για την 1<sup>η</sup> παράμετρο και ρυθμισμένη στην αυτόματη τιμή (auto) από το σύστημα για τη 2<sup>η</sup> παράμετρο. Όσο αυξάνεται ο αριθμός των κόμβων των συμπλεγμάτων και ο βαθμός πυκνότητας τόσο μειώνεται ο αριθμός των συμπλεγμάτων που προκύπτουν διότι το μέτρο αυστηρότητας αυξάνει. Σε αυτόν τον αλγόριθμο η στατιστική σημαντικότητα (p-value) εξετάζεται μέσω της στατιστικής ανάλυσης «one-sided Mann-Whitney U test» (Rosner & Grove 1999, Galbraith et al. 2010). Αυτή η ανάλυση εξετάζει τη στατιστική σημαντικότητα ενός συμπλέγματος συγκρίνοντας το σύνολο των συνδέσεων των

πρωτεϊνών-κόμβων που τοποθετούνται αποκλειστικά εντός του συμπλέγματος (in-weight) με το σύνολο των συνδέσεων των πρωτεϊνών-κόμβων, των οποίων το ένα καταληκτικό σημείο τοποθετείται μέσα στο σύμπλεγμα ενώ το άλλο βρίσκεται έξω από αυτό (out-weight). Η αξιοπιστία ενός «cluster» αυξάνεται ( $p\text{-value} < 0,05$ ) όταν ο αριθμός των «in-weights» είναι μεγαλύτερος από τον αριθμό των «out-weights».

## 2.3. Καλλιέργεια και μετασχηματισμός βακτηρίων

### 2.3.1. Βακτηριακά στελέχη

Το στέλεχος βακτηρίων *E. coli* που χρησιμοποιήθηκε ήταν το *DH5a*. Το γονιδίωμα αυτού του στελέχους είναι κυκλικό, απλοειδές, μεγέθους 4,6 Mb και περιλαμβάνει 4359 γονίδια από τα οποία τα 4128 κωδικοποιούν πρωτεΐνες (Taylor et al. 1993). Το συγκεκριμένο στέλεχος δεν είναι παθογόνο, χρησιμοποιείται ευρέως μόνο για εργαστηριακή χρήση και διαθέτει πολλαπλές μεταλλάξεις οι οποίες επιτρέπουν υψηλής απόδοσης μετασχηματισμούς ( $> 2 \times 10^6$  αποικίες/μg πλασμιδιακού DNA) με πλασμιδιακούς φορείς. Οι μεταλλάξεις που διαθέτουν είναι: α) η *lacZ* Delta M15 στο αμινοτελικό τμήμα της β-γαλακτοσιδάσης, η οποία επιτρέπει τη δυνατότητα επιλογής των βακτηριακών αποικιών που περιέχουν ορισμένα ανασυνδυασμένα πλασμίδια, παρουσία ενός αδρανούς χρωμογόνου υποστρώματος (X-gal) για τη β-γαλακτοσιδάση. Συγκεκριμένα, αυτά τα πλασμίδια πρέπει να διαθέτουν τμήμα του γονιδίου της β-γαλακτοσιδάσης που κωδικοποιεί τα 146 αμινοτελικά αμινοξέα της πρωτεΐνης (α-πολυπεπτίδιο). Μέσα στην αλληλουχία αυτή είναι ενσωματωμένη η κασέτα κλωνοποίησης του φορέα (polylinker) με αποτέλεσμα η εισαγωγή του ενθέματος να εμποδίζει την παραγωγή λειτουργικής β-γαλακτοσιδάσης και κατ' επέκταση να αναπτύσσονται στο τρυβλίο καλλιέργειας λευκές αποικίες. Στην αντίθετη περίπτωση παράγεται το α-πολυπεπτίδιο της πρωτεΐνης από το πλασμίδιο (καθώς δεν μπορεί να παραχθεί από το βακτήριο λόγω μετάλλαξης) και το ω-πολυπεπτίδιο από το βακτήριο μέσω μιας διαδικασίας, γνωστής ως α-συμπληρωματικότητα (α-complementation) με αποτέλεσμα να εμφανίζονται μπλε αποικίες, β) η μετάλλαξη στο γονίδιο *endA1* (κωδικοποιεί μια ενδονουκλεάση), μέσω της οποίας αποτρέπεται η αποικοδόμηση του DNA των πλασμιδίων που εισέρχονται στα βακτήρια, διασφαλίζοντας με αυτόν τον τρόπο υψηλά επίπεδα μετασχηματισμού, και γ) η μετάλλαξη *recA1* η οποία μειώνει τα επίπεδα ανασυνδυασμού των πλασμιδίων με το χρωμοσωματικό βακτηριακό DNA, αυξάνοντας τη σταθερότητα του ενθέματος (Taylor et al. 1993, Adikesavan et al. 2011). Χρησιμοποιήθηκαν είτε βακτήρια *DH5a* που ήταν ικανά να εισάγουν πλασμίδια, γνωστά ως δεκτικά κύτταρα (competent cells), και προέρχονταν από εταιρία, είτε βακτήρια *DH5a* που επεξεργάστηκαν κατάλληλα ώστε να γίνουν δεκτικά σε πλασμίδια μέσω κατάλληλου πρωτοκόλλου (βλ. ενότητα παρακάτω).

### 2.3.2. Καλλιέργεια βακτηρίων

Η καλλιέργεια βακτηρίων πραγματοποιείται είτε σε υγρό θρεπτικό μέσο ανάπτυξης σε ειδικά σωληνάρια και κωνικές φιάλες καλλιέργειας ή σε στερεό θρεπτικό υπόστρωμα σε τρυβλία καλλιέργειας βακτηρίων.

Το υγρό και στερεό θρεπτικό μέσο που χρησιμοποιήθηκε ήταν της εταιρίας Sharlau Microbiology {LB broth (Luria Broth), Cat. No 02-384, LB agar, Cat. No 01-385,

αντίστοιχα}. Τα θρεπτικά μέσα είναι σε μορφή σκόνης η οποία διαλύεται χρησιμοποιώντας 15g/L δις-απιονισμένου νερού στην περίπτωση του υγρού θρεπτικού μέσου και 40g/L νερού στην περίπτωση του στερεού. Σε κάθε διάλυμα που παρασκευάστηκε στην εργασία και σε οποιαδήποτε επαναδιάλυση χρησιμοποιήθηκε δις-απιονισμένο νερό. Η σύσταση και των δυο ειδών θρεπτικών μέσων είναι: 10 gr/L νερού καζεΐνη-τρουπτόνη (casein-tryptone), 5 gr/L εκχύλισμα ζύμης, 10 gr/L NaCl, και το στερεό θρεπτικό μέσο περιέχει επιπλέον 15 gr άγαρ/L. Και στις δυο περιπτώσεις, η επαναδιάλυση της σκόνης έγινε σε κωνική φιάλη, το pH ρυθμίστηκε στην τιμή 7-7,5 με διάλυμα 5M NaOH, και ακολούθησε αποστείρωση. Έπειτα, και τα δυο θρεπτικά μέσα παρέμειναν σε θερμοκρασία δωματίου ώστε να πέσει η θερμοκρασία τους και να προστεθεί το αντιβιοτικό. Στην περίπτωση του στερεού μέσου, μετά την προσθήκη αντιβιοτικού το μίγμα τοποθετήθηκε σε τρυβλία καλλιέργειας βακτηρίων (Cat. No G090, Kisker), όπου στερεοποιήθηκε σε θερμοκρασία δωματίου. Τα διαλύματα αντιβιοτικών που χρησιμοποιήθηκαν ανάλογα με τα γονίδια ανθεκτικότητας που διέθεταν οι φορείς κλωνοποίησης ήταν η αμπικιλίνη (Cat. No A9393, Sigma) και η καναμυκίνη (Cat. No 60615, Sigma). Τα αντιβιοτικά ήταν σε μορφή σκόνης και πραγματοποιήθηκε η επαναδιάλυση κατάλληλης ποσότητας σε νερό έτσι ώστε η συγκέντρωση των διαλυμάτων να είναι 100mg/ml. Ακολούθησε το φιλτράρισμα του διαλύματος με φίλτρο 0,22μm (Cat. No SLGP033RS, Merck Millipore), ο διαμερισμός του σε σωληνάρια (ependorf tubes) του 1,5ml (aliquots), τοποθετώντας 100μl στο καθένα, και η αποθήκευση των «aliquots» στους -20°C. Η τελική συγκέντρωση αμπικιλίνης που χρησιμοποιήθηκε στα πειράματα ήταν 100μg/ml για τους φορείς κλωνοποίησης pcDNA<sup>TM</sup>6/TR, pLKO.1-puro, pcDNA 4/TO και αυτή της καναμυκίνης ήταν 30μg/ml για τον φορέα κλωνοποίησης pmCherry-C1 (βλ. παρακάτω).

### 2.3.3. Δεκτικά βακτηριακά κύτταρα *DH5a*

Τα δεκτικά βακτηριακά κύτταρα (competent cells) *DH5a* που χρησιμοποιήθηκαν προήλθαν από την εταιρία Takara (Cat. No 636763) ή παρασκευάστηκαν στο εργαστήριο. Τα πρώτα χρησιμοποιήθηκαν αποκλειστικά για αντιδράσεις σύνδεσης διότι εμφάνιζαν καλύτερη απόδοση μετασχηματισμού και τα δεύτερα για όλους τους υπόλοιπους μετασχηματισμούς για την παραγωγή μεγάλων ποσοτήτων DNA ανασυνδυασμένων ή μη πλασμιδιακών φορέων.

#### 2.3.3.1. Διαδικασία παρασκευής δεκτικών βακτηριακών κυττάρων *DH5a*

Το πρώτο στάδιο της διαδικασίας περιλάμβανε την επίστρωση βακτηριακών κυττάρων *DH5a* σε στερεό μέσο καλλιέργειας (LB agar), την ολονύκτια επώαση αυτών στους 37°C για την ανάπτυξη των βακτηριακών αποικιών, και την επιλογή και ανάπτυξη μιας μοναδιαίας αποικίας σε υγρό θρεπτικό μέσο (LB broth) στους 37°C για 18 ώρες. Έπειτα, 1ml από την υγρή καλλιέργεια μεταφέρθηκε σε 100ml υγρού θρεπτικού υλικού σε κωνική φιάλη του 1lt και η καλλιέργεια αυτή επώαστηκε στους 37°C υπό ανάδευση στις 200rpm μέχρι τα κύτταρα να φτάσουν σε εκθετική φάση ανάπτυξης, δηλαδή, η τιμή της οπτικής πυκνότητας να είναι 0,7-0,8 στα 600nm (OD<sub>600</sub>). Ακολούθησε μεταφορά της υγρής καλλιέργειας σε σωληνάρια (falcon) των 50ml και επώαση στον πάγο για 10min. Το επόμενο στάδιο περιλάμβανε συλλογή με φυγοκέντρηση των κυττάρων στις 4000rpm για 20min στους 4°C, απομάκρυνση του

υπερκειμένου διαλύματος και επαναιώρηση της πελλέτας των βακτηρίων σε διάλυμα 80mM MgCl<sub>2</sub>-20mM CaCl<sub>2</sub> (φιλτραρισμένο με φίλτρο 0,22μm, διατήρηση στους 4°C), επώαση σε πάγο για 20min, επαναφυγοκέντρηση και απομάκρυνση του υπερκειμένου. Στο τελικό στάδιο της διαδικασίας, η κάθε πελλέτα των βακτηρίων επαναιωρήθηκε σε 2ml διαλύματος 0,1mM CaCl<sub>2</sub> (φιλτραρισμένο με φίλτρο 0,22μm, διατήρηση στους 4°C). Ακολούθησε η προσθήκη στάγδην 70μl κρυοπροστατευτικού μέσου, DMSO (Dimethyl sulfoxide, Cat. No D8418, Sigma), ήπια ανάδευση και προσθήκη ίδιας ποσότητας DMSO. Το τελικό εναιώρημα των βακτηριακών κυττάρων μεταφέρθηκε σε αποστειρωμένα σωληνάρια 1,5ml (ependorf tubes) σε πάγο. Συνολικά παρασκευάστηκαν ~20 «aliquots» (200μl το καθένα) δεκτικών κυττάρων *DH5a* τα οποία αποθηκεύτηκαν στους -80°C.

#### 2.3.4. Μετασχηματισμός βακτηρίων με θερμικό σοκ

Μετασχηματισμός είναι ο μηχανισμός κατά τον οποίον εισάγονται τμήματα ελεύθερου DNA που προέρχονται από άλλη πηγή (συνήθως πλασμίδια) σε βακτηριακά κύτταρα. Τα περισσότερα είδη βακτηρίων έχουν την ιδιότητα να προσλαμβάνουν και να ενσωματώνουν «ξένο» DNA από το θρεπτικό τους υλικό, το οποίο όμως αποικοδομείται κατά την ενσωμάτωσή του στο βακτηριακό γονιδίωμα. Σε άλλες περιπτώσεις και υπό συγκεκριμένες συνθήκες, στο βακτήριο μπορεί να εισαχθεί «ξένο» DNA υπό τη μορφή πλασμιδίων (Froger & Hall 2007). Μία παράμετρος που βελτιώνει την ικανότητα των κυττάρων *E. coli* να προσλαμβάνουν πλασμιδιακό DNA είναι τα ιόντα Ca<sup>2+</sup> λόγω της κατακρήμνισης και συσσώρευσης του DNA στην εξωτερική επιφάνεια του βακτηρίου (Cohen et al. 1973, Lorenz & Wackernagel 1994).

Τα αρχικά βήματα του πρωτοκόλλου μετασχηματισμού (Sambrook et al. 1989) που χρησιμοποιήθηκε στην περίπτωση των δεκτικών κυττάρων που παρασκευάστηκαν στο εργαστήριο περιλάμβαναν τη μεταφορά και την επώαση ενός δείγματος (aliquot) δεκτικών κυττάρων από τους -80°C για 10min σε πάγο (για να πραγματοποιηθεί η τήξη του δείγματος), την προσθήκη κατάλληλης ποσότητας πλασμιδιακού DNA (~1-5ng πλασμιδιακού DNA) και την επώαση σε πάγο για 20min κατά την οποία το DNA έρχεται σε επαφή με τα βακτήρια. Έπειτα, το δείγμα τοποθετήθηκε σε υδατόλουτρο στους 42°C για 1,5min ώστε τα βακτήρια να υποστούν θερμικό σοκ για να γίνουν διαπερατά τα τοιχώματά τους, και ακολούθησε άμεση μεταφορά στον πάγο όπου παρέμεινε για 1min. Στο δείγμα προστέθηκε 1ml υγρού θρεπτικού μέσου LB και ακολούθησε επώαση για 1h στους 37°C ώστε να εκφραστεί το γονίδιο ανθεκτικότητας του πλασμιδίου με αποτέλεσμα να είναι δυνατή η επιλογή των μετασχηματισμένων βακτηρίων. Τέλος, τα κύτταρα συλλέχθηκαν με φυγοκέντρηση στις 2500rpm για 10min σε θερμοκρασία δωματίου και η μεγαλύτερη ποσότητα του υπερκειμένου θρεπτικού μέσου απομακρύνθηκε αφήνοντας ελάχιστη ποσότητα για την επαναιώρηση της πελλέτας των κυττάρων έτσι ώστε να γίνει η επίστρωσή τους σε αποστειρωμένο και στεγνό τρυβλίο καλλιέργειας με το κατάλληλο αντιβιοτικό (σύμφωνα με τον πλασμιδιακό φορέα που χρησιμοποιήθηκε κάθε φορά). Η ανάπτυξη των βακτηριακών κυττάρων πραγματοποιήθηκε έπειτα από ολονύκτια επώαση στους 37°C.

Ο μετασχηματισμός των δεκτικών κυττάρων *DH5a* που προήλθαν από εταιρία έγινε βάσει του πρωτοκόλλου της εταιρίας Takara, σύμφωνα με το οποίο η εισαγωγή του

πλασμιδιακού DNA (>200ng) στο βακτήριο επιτεύχθηκε με την προσωρινή αλλαγή της θερμοκρασίας (θερμικό σοκ) από τον πάγο στους 42°C για 45 δευτερόλεπτα. Τα επόμενα βήματα του πρωτοκόλλου ήταν ίδια με αυτά που περιγράφηκαν παραπάνω.

## **2.4. Χειρισμός και τροποποίηση κυττάρων θηλαστικών *HeLa***

### 2.4.1. Κυτταρικές σειρές θηλαστικών

Οι διαφοροποιημένες κυτταρικές σειρές προέρχονται από διαφορετικούς ιστούς και προκύπτουν έπειτα από την αθανатоποίηση τους *in vitro* καθώς και από καρκινικούς όγκους οι οποίοι, παρά την καρκινική τους προέλευση, αντικατοπτρίζουν εν μέρει τις ιδιότητες των φυσιολογικών κυττάρων (Hogrocks et al. 2003). Χρησιμοποιήθηκε η καρκινική κυτταρική σειρά *HeLa* η οποία προέρχεται από επιθηλιακά κύτταρα ανθρώπου από αδενοκαρκίνωμα τραχήλου της μήτρας. Τα κύτταρα αυτά, όπως και των υπόλοιπων καρκινικών κυτταρικών σειρών, έχουν την ικανότητα να πολλαπλασιάζονται πολύ γρήγορα διότι κατά τη διάρκεια της κυτταρικής διαίρεσης το ένζυμο τελομεράση παραμένει σε ενεργή μορφή εμποδίζοντας τη βράχυνση των τελομερών. Έτσι, ο αριθμός των κυτταρικών διαιρέσεων ξεπερνάει αυτών των φυσιολογικών κυττάρων πριν τα κύτταρα φτάσουν στο στάδιο της γήρανσης (Lucey et al. 2009). Τα κύτταρα *HeLa* είναι τριπλοειδή (3n) (Ghosh S. & Ghosh I. 1975) και αναπτύσσονται προσκολλημένα σε στερεό υπόστρωμα (adherent cells), σχηματίζοντας μονοστοιβάδες.

### 2.4.2. Καλλιέργεια κυττάρων θηλαστικών

#### 2.4.2.1. Συνθήκες και βασικές αρχές κυτταρικής καλλιέργειας

Ο χειρισμός και η ανάπτυξη όλων των κυτταρικών σειρών πραγματοποιείται σε ειδικά διαμορφωμένο χώρο (cell culture room) ο οποίος περιλαμβάνει τα μηχανήματα που είναι απαραίτητα για τις κυτταροκαλλιέργειες και οι συνθήκες σε αυτόν είναι άσηπτες για την αποφυγή μολύνσεων. Η ανάπτυξη των κυττάρων γίνεται σε επωαστικό κλίβανο με σταθερή θερμοκρασία (37°C), κατάλληλα επίπεδα CO<sub>2</sub> (5%) για την διατήρηση του pH στο θρεπτικό μέσο καλλιέργειας και κατάλληλες συνθήκες υγρασίας. Η παρατήρηση των κυττάρων γίνεται σε οπτικό μικροσκόπιο ανάστροφης φάσης. Καθώς το φως διέρχεται από τα κύτταρα, επέρχεται απορρόφηση μέρος αυτού που σημαίνει ότι μειώνεται το πλάτος ταλάντωσης του φωτεινού κύματος (η μέγιστη απομάκρυνση ενός σημείου από τη θέση ισορροπίας του κατά την αρμονική ταλάντωση που εκτελεί λόγω της διέλευσης του κύματος). Ταυτόχρονα, το φως διαθλάται έπειτα από την αλληλεπίδρασή του με το παρασκεύασμα και προκαλείται αλλαγή της φάσης (επιβράδυνση) του φωτεινού κύματος. Ο οφθαλμός δεν αντιλαμβάνεται τη διαφορά φάσης αλλά η μικροσκοπία αντίθεσης φάσεως μετατρέπει τις διαφορές φάσης σε διαφορές έντασης του φωτός οι οποίες γίνονται ορατές από τον οφθαλμό. Δηλαδή, η μικροσκοπία αντίθεσης φάσης αξιοποιεί τα φαινόμενα της διάθλασης και επιβράδυνσης του διερχόμενου φωτός για τη δημιουργία αντίθεσης (Zernike 1955). Η μακροχρόνια συντήρηση των κυττάρων πραγματοποιείται σε δεξαμενή υγρού αζώτου (-196 °C ως -156 °C) (Werner & Noé 1993). Κατά την καλλιέργεια των κυττάρων υπάρχει κίνδυνος μόλυνσης με βακτήρια, μύκητες ή ζύμες, εάν δεν είναι σωστός ο χειρισμός τους, τα οποία μπορούν να εντοπιστούν εύκολα με

γυμνό μάτι μέσω της παρατήρησης των αλλοιώσεων που προκαλούν στις κυτταροκαλλιέργειες. Ωστόσο, τυχόν μόλυνση από ιούς ή από το βακτήριο μυκόπλασμα δεν είναι ορατή και για αυτό απαιτούνται ειδικές μέθοδοι ανίχνευσης, όπως η αλυσιδωτή αντίδραση πολυμεράσης (PCR). Έτσι, ο χειρισμός των κυττάρων πρέπει να γίνεται σε άσηπτες συνθήκες. Για την επίτευξη άσηπτων συνθηκών χρησιμοποιείται εστία κάθετης νηματικής ροής στην οποία πραγματοποιείται η δημιουργία ενός προστατευτικού «τείχους» μέσω της ανακύκλωσης του αέρα και η χρήση υπεριώδους ακτινοβολίας έπειτα από κάθε χρήση. Επιπλέον, απαιτείται η αποστείρωση των εξωτερικών επιφανειών των υλικών και των θρεπτικών μέσων ή διαλυμάτων που χρησιμοποιούνται με τη χρήση 70% αιθανόλης.

Η κυτταρική σειρά *HeLa* καλλιεργήθηκε σε θρεπτικό μέσο DMEM (Invitrogen-Life Technologies, Cat. No 41966-029), το οποίο έχει την ακόλουθη σύσταση: i) υψηλή συγκέντρωση γλυκόζης (4500mg/L), ii) δεκαπέντε αμινοξέα σε συνολική συγκέντρωση 1.570mg/L με το αμινοξύ της γλουταμίνης σε μεγαλύτερη συγκέντρωση (584mg/L), iii) επτά είδη ανόργανων αλάτων για τη ρύθμιση του pH σε συνολική συγκέντρωση 11.105mg/L, με τη μεγαλύτερη συγκέντρωση να παρουσιάζουν το χλωριούχο νάτριο (NaCl, 6400mg/L) και το όξινο ανθρακικό νάτριο (NaHCO<sub>3</sub>, 3700mg/L), iv) οκτώ είδη βιταμινών σε συνολική συγκέντρωση 31,6 mg/L, και v) ερυθρό της φαινόλης {Phenol Red (15mg/L)} που είναι ένας δείκτης pH και πυροσταφυλικό νάτριο {(Sodium Pyruvate, 110mg/L)} που αποτελεί πηγή άνθρακα επιπρόσθετη της γλυκόζης. Για την παρασκευή πλήρους θρεπτικού μέσου, προστέθηκε στο DMEM ορός εμβρύου βοοειδούς (Fetal Bovine Serum, FBS) σε συγκέντρωση 10% (Cat. No 10270-106, Invitrogen-Life Technologies) ο οποίος παρέχει στα κύτταρα αυξητικούς παράγοντες και ορμόνες που ευνοούν την ανάπτυξή τους. Πριν την προσθήκη του ορού είχε προηγηθεί η επώασή του στους 56°C για 30min, ώστε να απενεργοποιηθούν θερμοευαίσθητες πρωτεΐνες του συμπληρώματος οι οποίες μπορεί να παρεμβαίνουν στην ανοσοαπόκριση των κυττάρων (Soltis et al. 1979). Επιπλέον στο μίγμα DMEM και FBS προστέθηκε διάλυμα αντιβιοτικών στρεπτομυκίνης και πενικιλίνης (συγκέντρωσης 100x, 10,000 units/mL, Invitrogen-Life Technologies, Cat. No 15140122) για την αποφυγή μολύνσεων από μικροοργανισμούς.

#### 2.4.2.2. Διατήρηση κυτταροκαλλιεργειών

Οι κυτταροκαλλιέργειες διατηρούνται σε ειδικές φλάσκες εμβαδού 25cm<sup>2</sup> (Corning, Cat. No 430168) ή 75cm<sup>2</sup> (Corning, Cat. No 430720) ενώ οι πειραματικές διαδικασίες πραγματοποιήθηκαν σε τρυβλία καλλιέργειας διαμέτρου 10 cm (Corning, Cat. No 430167), τρυβλία έξι θέσεων (6 well plate, Corning, Cat. No 3506) ή είκοσι τεσσάρων θέσεων (24 well plate, Corning, Cat. No 3527). Τα κύτταρα, όταν καλλιεργούνται σε στερεό υπόστρωμα, διαχωρίζονται με τη διαδικασία της θρυψινοποίησης (περιγράφεται αμέσως μετά) και συλλέγονται σε εκθετική φάση ανάπτυξης, δηλαδή όταν έχουν καλύψει το ~80% της διαθέσιμης επιφάνειας πριν η ανάπτυξή τους παρεμποδιστεί από την έλλειψη χώρου. Στις κυτταρικές καλλιέργειες, όταν παρατηρείται έλλειψη χώρου, ο κυτταρικός πολλαπλασιασμός αναστέλλεται και οι μεταβολικές διαδικασίες των κυττάρων τροποποιούνται. Τα κύτταρα δεν βρίσκουν διαθέσιμη επιφάνεια για να προσκολληθούν, πεθαίνουν, αποκολλώνται και αιωρούνται

στο θρεπτικό μέσο της καλλιέργειας. Επομένως, η ανακαλλιέργειά τους θα πρέπει να πραγματοποιείται σε τακτά χρονικά διαστήματα (3-4 μέρες), έτσι ώστε να διασφαλίζεται ο κατάλληλος αριθμός των κυττάρων για πειράματα αλλά και η ομαλή ανάπτυξή τους. Η αποκόλληση των κυττάρων από το υπόστρωμα έγινε μέσω της χρήσης του παγκρεατικού ενζύμου θρυψίνη. Το ένζυμο αυτό είναι μια πρωτεάση η οποία διασπά τους πεπτιδικούς δεσμούς στο καρβοξυτελικό άκρο των αμινοξέων λυσίνης και αργινίνης. Το ένζυμο είναι διαλυμένο σε αιθυλενο-διαμινο-τετραοξικό οξύ (EDTA) το οποίο αποτελεί ένα ισχυρό χηλικό αντιδραστήριο των ιόντων ασβεστίου που είναι απαραίτητα για την προσκόλληση των κυττάρων μέσω ιντεγκρινών (Buck & Horwitz 1987). Οι πρωτεΐνες του ορού, βιτρονεκτίνη και φμπρονεκτίνη σχηματίζουν δεσμούς (μέσω των ατόμων άνθρακα) με το υλικό του στερεού υποστρώματος (πολυστυρένιο) με αποτέλεσμα να καλύπτουν την κάτω επιφάνεια των φλασκών ή των τρυβλίων. Η προσκόλληση των κυττάρων επιτυγχάνεται μέσω του σχηματισμού πεπτιδικών δεσμών μεταξύ των αμινοξέων των διαμεμβρανικών κυτταρικών υποδοχέων (ιντεγκρίνες) και των πρωτεϊνών του ορού (Steele et al. 1992). Το διάλυμα θρυψίνης-EDTA (GIBCO Cat. no 15400-054, 10x) ήταν από πάγκρεας χοίρου. Διατηρείται στους  $-20^{\circ}\text{C}$  σε aliquots συγκέντρωσης 1x με αραιώση σε ρυθμιστικό διάλυμα φωσφορικών αλάτων (1xPBS, Invitrogen-Life Technologies, Cat. No 10010-015). Η χρήση του ενζύμου στοχεύει στη διάσπαση των πεπτιδικών δεσμών των αμινοξέων των πρωτεϊνών, μέσω των οποίων προσκολλώνται στο στερεό τους υπόστρωμα, με αποτέλεσμα την αποκόλλησή τους από αυτό. Η διαδικασία της αποκόλλησης των κυττάρων με τη χρήση θρυψίνης περιλάμβανε: i) την απομάκρυνση του θρεπτικού μέσου, ii) την έκπλυση με διάλυμα 1xPBS, iii) την προσθήκη διαλύματος θρυψίνης (1x) και την επώαση για ~1-2min στους  $37^{\circ}\text{C}$  ή για ~5min σε θερμοκρασία δωματίου και iv) την προσθήκη πλήρους θρεπτικού μέσου για την αναστολή της δράσης της θρυψίνης μέσω της ταυτόχρονης δράσης της έναντι των πρωτεϊνών του ορού. Στη συνέχεια πραγματοποιήθηκε αραιώση των κυττάρων ανάλογα με τον αριθμό που χρειαζόταν για τη συνέχεια του πειράματος ή μέτρηση των κυττάρων (βλ. Ενότητα 2.4.2.5.) όταν αυτό ήταν απαραίτητο για επακόλουθο πείραμα.

#### 2.4.2.3. Κρυοδιατήρηση κυττάρων

Η κρυοδιατήρηση των κυττάρων αποτελεί μια διαδικασία αποθήκευσης σε πολύ χαμηλές θερμοκρασίες σε υπερκαταψύκτες ( $-80^{\circ}\text{C}$ ) για συντήρηση για μικρό χρονικό διάστημα ή σε υγρό άζωτο ( $-196^{\circ}\text{C}$  ως  $-156^{\circ}\text{C}$ ) για να διατηρηθούν βιώσιμα έως και πάρα πολλά χρόνια. Σε μια πρόσφατη μελέτη, στην οποία εξετάστηκαν κύτταρα που είχαν διατηρηθεί στο άζωτο 28 χρόνια, βρέθηκε ότι οι αλλαγές που επέφερε ο χρόνος στην μορφολογία και στην ανάπτυξη των κυττάρων καθώς και στην έκφραση των γονιδίων τους ήταν ελάχιστες (Fazekas et al. 2017). Η κρυοδιατήρηση των κυττάρων σε τόσο χαμηλές θερμοκρασίες επιτρέπει την αναστολή όλων των χημικών αντιδράσεων και των βιολογικών διαδικασιών των κυττάρων με την ταυτόχρονη διατήρηση της δομής των κυτταρικών οργανιδίων έτσι ώστε κατά τη διάρκεια της επαναφοράς τους σε συνθήκες καλλιέργειας, οι παραπάνω διαδικασίες να επανενεργοποιούνται και τα κύτταρα να αναπτύσσονται φυσιολογικά.

Για να επιτευχθεί η κρυοδιατήρηση των κυττάρων θα πρέπει να πραγματοποιείται η σταδιακή μετάβαση τους στην τελική θερμοκρασία ψύξης. Η απότομη μετάβαση από τους 37°C σε πολύ χαμηλές θερμοκρασίες οδηγεί στην απώλεια του ενδοκυτταρικού υγρού, στην αύξηση της συγκέντρωσης των ηλεκτρολυτών και στο σχηματισμό κρυστάλλων πάγου μέσα στα κύτταρα. Έτσι τα κύτταρα παραμορφώνονται και οι ενδοκυτταρικές δομές τους καταστρέφονται (Bakhach 2009, Yokoyama et al. 2012).

Για να επιβιώσουν τα κύτταρα κατά την κρυοσυντήρησή τους, χρησιμοποιούνται διάφορες κρυοπροστατευτικές ουσίες, όπως το διμεθυλοσουλφοξείδιο (DMSO), που χρησιμοποιήθηκε στην εργασία, το οποίο αποτελεί έναν πολικό οργανικό διαλύτη. Το DMSO αποτελεί μια εξαιρετική κρυοπροστατευτική ουσία που έχει το πλεονέκτημα να διέρχεται με μεγάλη ταχύτητα μέσω των κυτταρικών μεμβρανών και ταυτόχρονα να μην εμφανίζει τοξικότητα. Καθώς μειώνεται η θερμοκρασία, τα υδατικά διαλύματα κρυοπροστατευτικών ουσιών καθίστανται σταδιακά πιο πυκνόρρευστα μέχρι πλήρους στερεοποίησης, μια διαδικασία που ονομάζεται υαλοποίηση (Bakhach 2009). Με αυτόν τον τρόπο πραγματοποιείται η σταδιακή μετάβαση των κυττάρων στην τελική θερμοκρασία ψύξης, αποφεύγεται ο σχηματισμός κρυστάλλων στο εσωτερικό τους, και ελαχιστοποιούνται οι κυτταρικές βλάβες και η απώλεια του ενδοκυτταρικού υγρού (Bakhach 2009, Yokoyama et al. 2012).

Η κρυοσυντήρηση των κυττάρων που χρησιμοποιήθηκαν σε αυτή την εργασία έγινε όταν τα κύτταρα βρίσκονταν σε εκθετική φάση ανάπτυξης (~80% κάλυψη της διαθέσιμης επιφάνειας). Η διαδικασία περιλάμβανε την αποκόλληση των κυττάρων, όπως περιγράφηκε παραπάνω, τη συλλογή τους και την επακόλουθη φυγοκέντρωσή τους για 5min στις 1000 rpm σε θερμοκρασία δωματίου. Έπειτα, ακολούθησε επαναιώρηση των κυττάρων με εμβρυϊκό ορό βοοειδούς (FBS) που περιείχε 10% DMSO και η μεταφορά των δειγμάτων σε κατάλληλα φιαλίδια (cryovials, Cat. No CV11-2, Kisker). Τέλος, τα δείγματα τοποθετήθηκαν στους -20°C για ~30min και μεταφέρθηκαν στους -80°C για 24h. Η τελική αποθήκευσή τους έγινε σε υγρό άζωτο (-196 °C ως -156 °C).

#### 2.4.2.4. Επανακαλλιέργεια κυττάρων

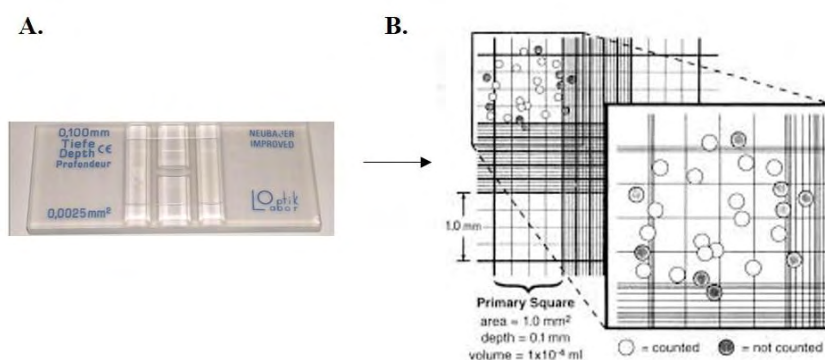
Η πιθανότητα επιβίωσης των κυττάρων μετά την κρυοσυντήρησή τους δεν καθορίζεται μόνο από τον ρυθμό ψύξης, αλλά και από τον ρυθμό απόψυξής τους. Ο ρυθμός επαναφοράς των κρυοσυντηρημένων κυττάρων πρέπει να είναι γρήγορος. Γι' αυτό το λόγο, τα κύτταρα που πρόκειται να επανακαλλιεργηθούν, μεταφέρονται σε υδατόλουτρο στους 37°C όπου και αναδεύονται ήπια μέχρι την πλήρη τήξη του θρεπτικού μέσου (1-2min). Η διαδικασία πραγματοποιείται όσο το δυνατόν ταχύτερα, προκειμένου τα κύτταρα να μην εκτεθούν στο DMSO (το οποίο μπορεί να δράσει τοξικά) καθώς και για να αποφευχθεί ο σχηματισμός ενδοκυτταρικών κρυστάλλων κατά τη διαδικασία της απόψυξής τους. Ακολουθεί η μεταφορά των κυττάρων στη φιάλη καλλιέργειας στην οποία έχει ήδη προστεθεί πλήρες θρεπτικό μέσο.

#### 2.4.2.5. Καταμέτρηση κυττάρων για επιστροφή τρυβλίων καλλιέργειας

Για την εκτίμηση του αριθμού των κυττάρων που επιστρώνονται στα τρυβλία καλλιέργειας ανάλογα με το εκάστοτε πείραμα, είναι απαραίτητη η καταμέτρησή τους.



Αυτή γίνεται μέσω του αιμοκυτταρόμετρου, η χρήση του οποίου είναι άμεση, απλή και οικονομική. Το αιμοκυτταρόμετρο (πλάκα Neubauer) είναι μια τροποποιημένη αντικειμενοφόρος πλάκα που έχει δύο κατάλληλα επεξεργασμένες λείες επιφάνειες. Κάθε μια από αυτές διαθέτει ένα τετραγωνισμένο πλέγμα, το οποίο αποτελείται από εννιά κύρια τετράγωνα με μήκος πλευράς 1mm (εμβαδόν  $1\text{mm}^2$ ). Το κάθε ένα από αυτά τα τετράγωνα ορίζεται από τρεις παράλληλες ισαπέχουσες γραμμές (απέχουν μεταξύ τους  $2,5\mu\text{m}$ ), με βάση τις οποίες καθορίζεται εάν τα κύτταρα βρίσκονται εντός ή εκτός του τετραγώνου. Κάθε ένα από τα κύρια τετράγωνα διαιρείται σε δεκαέξι μικρότερα τετράγωνα ώστε να διευκολύνεται η μέτρηση των κυττάρων κατά την παρατήρηση στο οπτικό μικροσκόπιο. Το επίπεδο του διαβαθμισμένου πλέγματος βρίσκεται  $0,1\text{mm}$  χαμηλότερα από δύο παράλληλες προεξοχές, στις οποίες στηρίζεται η καλυπτρίδα (Εικ. 24). Ανάμεσα στην εξωτερική πλευρά της τετραγωνισμένης λείας επιφάνειας και στις προεξοχές αυτές υπάρχει μια κοίλη επιφάνεια στην οποία εισάγεται το κυτταρικό εναιώρημα το οποίο απλώνεται με τριχοειδικά φαινόμενα στην διαβαθμισμένη επιφάνεια του αιμοκυτταρόμετρου. Επειδή ο όγκος που καταλαμβάνει το κυτταρικό εναιώρημα σε καθένα από τα εννέα τετράγωνα είναι  $0,1\text{mm}^3$  που ισούται με  $10^{-4}\text{cm}^3$  ( $1\text{cm}^3$  ισοδυναμεί με  $1\text{ml}$ ) (Εικ. 24), ο αριθμός των κυττάρων ανά  $\text{ml}$  καλλιέργειας υπολογίζεται από το μέσο αριθμό κυττάρων ανά τετράγωνο επί  $1 \times 10^4$  (Phelan & Lawler 2001).



**Εικ. 24: A: Πλάκα Neubauer B: Τετραγωνισμένο πλέγμα στο αιμοκυτταρόμετρο Neubauer για την μέτρηση των κυττάρων** (Hemocytometer Cell Counting Protocol, Provost & Wallert research).

Η διαδικασία που ακολουθήθηκε περιλάμβανε τη θρυψινοποίηση και επαναιώρηση των κυττάρων σε κατάλληλο όγκο πλήρους θρεπτικού μέσου, τη μεταφορά  $10\mu\text{l}$  κυτταρικού εναιωρήματος στην κοίλη επιφάνεια του αιμοκυτταρόμετρου, τη μέτρηση των κυττάρων με παρατήρηση στο οπτικό μικροσκόπιο και τον υπολογισμό του αριθμού τους στο εναιώρημα. Στην παρούσα διατριβή, ο επιθυμητός αριθμός κυττάρων προς ανάπτυξη σε τρυβλίο καλλιέργειας διαμέτρου  $10\text{cm}$  ήταν  $1,2 \times 10^6$  για τη διαμόλυνση. Όσον αφορά την εύρεση του κατάλληλου πρωτοκόλλου συλλογής των κυττάρων που προορίζονταν για ομικές αναλύσεις και την ανάλυση του μεταγραφικού, πρωτεϊνικού και μεταβολικού προτύπου μέσω των ομικών τεχνολογιών ο αριθμός των κυττάρων που χρησιμοποιήθηκε σε τρυβλίο  $10\text{cm}$  ήταν  $10^6$ .

### 2.4.3. Δημιουργία μόνιμα διαμολυσμένης κυτταρικής σειράς

Υπό κατάλληλες συνθήκες, μόρια ανασυνδυασμένου πλασμιδιακού DNA μπορεί να εισαχθούν και να ενσωματωθούν στο DNA ευκαρυωτικών κυττάρων όπου είναι δυνατόν να εκφραστούν. Η έκφραση του πλασμιδιακού DNA μπορεί να είναι μόνιμη (stable) εφόσον το ξένο DNA ενσωματώνεται στο χρωμόσωμα ή παροδική (transient) εφόσον δεν ενσωματώνεται στο χρωμόσωμα αλλά παραμένει ως πλασμιδιακό και αποικοδομείται μετά την παρέλευση ενός μιτωτικού κύκλου.

Η δημιουργία της μόνιμα διαμολυσμένης κυτταρικής σειράς *HeLa* με καταστολή της έκφρασης του γονιδίου *FRA10AC1* και της σταθερά διπλά διαμολυσμένης κυτταρικής σειράς *HeLa* για την επαγόμενη έκφραση του γονιδίου *DGCR14* πραγματοποιήθηκε με την ενσωμάτωση στο γονιδίωμα των κυττάρων των αντίστοιχων πλασμιδίων. Η επιλογή των κυτταρικών κλώνων έγινε μέσω της χρήσης αντιβιοτικών με βάση τα αντίστοιχα γονίδια ανθεκτικότητας των πλασμιδιακών φορέων

Συγκεκριμένα, στην πρώτη περίπτωση, δημιουργήθηκαν κλώνοι κυττάρων *HeLa* που εκφράζουν μόνιμα τα τέσσερα διαγονίδια τα οποία παράγουν τα shRNAs που στοχεύουν στην κωδική περιοχή του γονιδίου και κατ' επέκταση στην καταστολή της έκφρασής του (βλ. Ενότητα 2.5.2.1.), καθώς και το διαγονίδιο που χρησιμοποιήθηκε ως αρνητικός έλεγχος του συγκεκριμένου πειράματος (βλ. Ενότητα 2.5.2.1.).

Στη δεύτερη περίπτωση, το πρώτο βήμα ήταν η δημιουργία κλώνων κυττάρων *HeLa* που εκφράζουν μόνιμα τον καταστολέα της τετρακυκλίνης (tetR). Έπειτα ο κλώνος κυττάρων που επιλέχθηκε (βλ. Ενότητα 3.6.1.) διαμολύνθηκε με το διαγονίδιο pCDNA4/TO-*DGCR14* με αποτέλεσμα να δημιουργηθεί μια σταθερά διπλά διαμολυσμένη κυτταρική σειρά.

Το πρώτο βήμα για τη δημιουργία της μόνιμα διαμολυσμένης κυτταρικής σειράς με τροποποιημένη την έκφραση του γονιδίου *FRA10AC1* ήταν ο καθορισμός της ελάχιστης συγκέντρωσης πουρομυκίνης (γονίδιο ανθεκτικότητας σε κύτταρα θηλαστικών του φορέα κλωνοποίησης pLKO.1-puro) που είναι ικανή να σκοτώσει τα κύτταρα εκείνα που δεν φέρουν ενσωματωμένο στο γονιδιωματικό τους DNA το διαγονίδιο pLKO.1-puro/shRNA ή το διαγονίδιο του αρνητικού ελέγχου (βλ. Ενότητα 2.5.2.1.), σε διάστημα ~15 ημερών. Έτσι, δοκιμάστηκε ένα εύρος συγκεντρώσεων πουρομυκίνης (Sigma, Cat. No P9620, >98% καθαρότητα) στα κύτταρα *HeLa*. Ακολουθήθηκε η εξής διαδικασία: τα κύτταρα *HeLa* επιστρώθηκαν σε τρυβλία 6 θέσεων (6 well plate) με πλήρες θρεπτικό μέσο DMEM με ~25% κάλυψη επιφάνειας ώστε να μπορεί να δράσει αποτελεσματικά η πουρομυκίνη. Μετά από 24h πραγματοποιήθηκε αλλαγή του θρεπτικού μέσου και προσθήκη νέου που περιείχε εύρος συγκεντρώσεων αντιβιοτικού που προτείνεται από την εταιρία Sigma (0,01-2μg/ml). Ακολούθως, γινόταν αλλαγή του θρεπτικού μέσου με τη πουρομυκίνη κάθε τρεις μέρες για δυο εβδομάδες. Τα κύτταρα που είναι ευαίσθητα στη πουρομυκίνη πεθαίνουν, αποκολλούνται από την επιφάνεια του τρυβλίου και αιωρούνται στο θρεπτικό μέσο.

Ακολούθησε η διαμόλυνση των κυττάρων *HeLa* με τα διαγονίδια που περιλαμβάνουν τους τέσσερις διαφορετικούς κλώνους των shRNA και το διαγονίδιο του αρνητικού ελέγχου (βλ. Ενότητα 2.5.2.1.). Ειδικότερα, κύτταρα *HeLa* καλλιεργήθηκαν σε δυο φλάσκες 75 cm<sup>2</sup> με πλήρες θρεπτικό μέσο. Όταν τα κύτταρα

έφθασαν σε κάλυψη 80-90%, έγινε η θρυψινοποίηση τους, η καταμέτρησή τους σε πλάκα Neubauer και η επίστρωσή τους σε πέντε τρυβλία διαμέτρου 10cm χρησιμοποιώντας  $1,2 \times 10^6$  κύτταρα στο καθένα. Η διαμόλυνση των κυττάρων έγινε α) με τα τέσσερα διαγονίδια γραμμοποιημένα με το ένζυμο περιορισμού *KpnI* (nt 1189) που εντοπίζεται ανοδικά της 3' θέσης LTR το οποίο δεν διαταράσσει την αλληλουχία των shRNA καθώς και άλλα στοιχεία που απαιτούνται για την έκφραση σε κύτταρα θηλαστικών (βλ. Ενότητες 2.5.2.1., 2.5.4.) και β) με το γραμμοποιημένο διαγονίδιο με το ένζυμο περιορισμού *KpnI* που χρησιμοποιήθηκε ως αρνητικός έλεγχος. Η γραμμοποίηση των διαγονιδίων δεν αυξάνει την αποτελεσματικότητα της διαμόλυνσης αλλά μειώνει τις πιθανότητες να κοπεί τυχαία το πλασμίδιο κατά την ενσωμάτωσή του στο γονιδίωμα των κυττάρων. Η μέθοδος διαμόλυνσης που χρησιμοποιήθηκε βασίζεται στη συνκατακρήμνιση του DNA με  $\text{Ca}_3(\text{PO}_4)_2$ , η οποία αποτελεί μια ευρέως χρησιμοποιούμενη μέθοδο (Graham & van der Eb 1973, Kim & Eberwine 2010). Η προσθήκη του  $\text{Ca}_3(\text{PO}_4)_2$  στο διάλυμα του DNA έχει ως αποτέλεσμα τη δημιουργία συμπλόκων (με τη μορφή λεπτών κρυστάλλων) μεταξύ του DNA και των θετικά φορτισμένων μορίων-φορέων, τα οποία προσκολλώνται στην κυτταρική μεμβράνη και εισέρχονται στο κύτταρο με ενδοκύτωση (Orlantia & Chang 1990). Μία μέρα πριν τη διαδικασία της διαμόλυνσης επιστρώθηκαν  $1,2 \times 10^6$  κύτταρα σε τρυβλία καλλιέργειας διαμέτρου 10cm. Πριν ξεκινήσει η διαδικασία (1-2h πριν) έγινε αλλαγή πλήρους θρεπτικού μέσου DMEM και προσθήκη νέου. Για το μίγμα της διαμόλυνσης χρησιμοποιήθηκαν 20μg γραμμικού DNA (ανά τρυβλίο) σε 450μl  $\text{H}_2\text{O}$  στα οποία προστέθηκαν 50μl  $\text{CaCl}_2$  2.5M (μετά την παρασκευή του, φιλτραρίστηκε με φίλτρο 0,22μm, διαμερίστηκε σε σωληνάρια και αποθηκεύτηκε στους  $-80^\circ\text{C}$ ) και 500μl διαλύματος 2xHEBS (280mM NaCl, 50mM Hepes, 5mM  $\text{Na}_2\text{HPO}_4$ , pH 7.12 έπειτα από προσαρμογή μέσω της χρήσης πυκνού υδροχλωρικού οξέος, αφού παρασκευάστηκε, φιλτραρίστηκε με φίλτρο 0,22μm, διαμερίστηκε σε σωληνάρια και αποθηκεύτηκε στους  $4^\circ\text{C}$ ). Το μίγμα επώαστηκε για 20min σε θερμοκρασία δωματίου για να σχηματιστούν τα σύμπλοκα DNA- $\text{Ca}_3(\text{PO}_4)_2$ . Σε αυτό το χρονικό διάστημα ελεγχόταν αν το δείγμα γινόταν άμεσα θολό ή αν σχηματιζόταν ίζημα, το οποίο θα σήμαινε ότι το DNA που χρησιμοποιήθηκε δεν ήταν καθαρό από πρωτεΐνες, RNA ή άλλες χημικές ουσίες. Αν το μίγμα ήταν διαυγές τότε προστίθετο στάγδην στα κύτταρα. Εφόσον είχε περάσει το διάστημα των 24h μετά τη διαμόλυνση έγινε έκπλυση των κυττάρων με 1xPBS και αλλαγή πλήρους θρεπτικού μέσου DMEM (χωρίς πουρομυκίνη). Μετά από 24h πραγματοποιήθηκε θρυψινοποίηση των κυττάρων και επανακαλλιέργειά τους σε τρία τρυβλία 10cm το καθένα έτσι ώστε η κάλυψη της επιφάνειας να είναι ~25%. Μετά από 12h έγινε αλλαγή θρεπτικού μέσου και προσθήκη νέου με την κατάλληλη συγκέντρωση πουρομυκίνης. Κάθε τρεις μέρες γινόταν αλλαγή του θρεπτικού με το αντιβιοτικό ενώ συνολικά πραγματοποιήθηκαν τέσσερις αλλαγές.

Τα κύτταρα *HeLa* που είναι ανθεκτικά στη πουρομυκίνη, ως συνέπεια της ενσωμάτωσης των διαγονιδίων, προσκολλώνται στην επιφάνεια του τρυβλίου και σχηματίζουν αποικίες. Αρχικά οι αποικίες εντοπίζονται στο μικροσκόπιο και σημειώνονται με μαρκαδόρο στην εξωτερική επιφάνεια του τρυβλίου. Μετά την έκπλυση των κυττάρων με 1xPBS, ειδικοί αποστειρωμένοι κύλινδροι (cloning cylinders, Sigma, Cat. No C1059-1EA, όγκου 150μl) τοποθετήθηκαν σε κάθε

σημειωμένη αποικία με τη χρήση λαβίδας και σταθεροποιήθηκαν στο τρυβλίο με τη χρήση αποστειρωμένης βαζελίνης. Σε κάθε κύλινδρο προστέθηκαν 100μl θρυψίνης ώστε να αποκολληθούν τα κύτταρα της αποικίας από την επιφάνεια του τρυβλίου. Μετά από επώαση ~5min σε θερμοκρασία δωματίου, τα κύτταρα συλλέχθηκαν και καλλιεργήθηκαν παρουσία πουρομυκίνης σε τρυβλία 24 θέσεων. Η καλλιέργεια των αποικιών συνεχίστηκε σε σταδιακά μεγαλύτερη επιφάνεια μέχρι φλάσκες εμβαδού 25cm<sup>2</sup>.

Για τη δημιουργία κλώνων κυττάρων *HeLa* που εκφράζουν μόνιμα τον καταστολέα της τετρακυκλίνης (tetR) ακολουθήθηκε ο ίδιος τρόπος που περιγράφηκε παραπάνω με κάποιες διαφοροποιήσεις. Το πρώτο βήμα περιλάμβανε τον καθορισμό της ελάχιστης συγκέντρωσης βλαστισιδίνης (γονίδιο ανθεκτικότητας του φορέα pcDNA6/TR σε κύτταρα θηλαστικών) που είναι ικανή να σκοτώσει τα κύτταρα εκείνα που δεν θα λάβουν το πλασμιδίο pcDNA6/TR. Στη συγκεκριμένη περίπτωση χρησιμοποιήθηκαν τρεις διαφορετικές συγκεντρώσεις αντιβιοτικού, 10μg/ml, 15μg/ml και 20μg/ml. Στο δεύτερο βήμα, η διαμόλυνση έγινε με το γραμμικό πλασμιδιακό φορέα pcDNA6/TR (έπειτα από την πέψη του με το ένζυμο περιορισμού *AhdI*). Το ένζυμο αυτό (5745nt) τέμνει το γονίδιο ανθεκτικότητας σε αμπικιλίνη. Η διαδικασία διαμόλυνσης ήταν ίδια με την παραπάνω, ο αριθμός των κυττάρων που χρησιμοποιήθηκαν σε κάθε τρυβλίο ήταν και σε αυτή την περίπτωση  $1,2 \times 10^6$  και η ποσότητα του πλασμιδιακού DNA ήταν 20μg. Με την ίδια ακριβώς διαδικασία με παραπάνω, τα κύτταρα επώαστηκαν με αλλαγή θρεπτικού υλικού που περιείχε την κατάλληλη συγκέντρωση βλαστισιδίνης κάθε ~3 μέρες μέχρι την τελική επιλογή των αποικιών των κυττάρων τα οποία ήταν ανθεκτικά στο αντιβιοτικό. Ομοίως, η καλλιέργεια των αποικιών συνεχίστηκε σε σταδιακά μεγαλύτερη επιφάνεια μέχρι φλάσκες εμβαδού 25cm<sup>2</sup>.

Ένας κλώνος κυττάρων (#5) που εκφράζει τον καταστολέα της τετρακυκλίνης επιλέχθηκε (βλ. Ενότητα 3.6.1.) για να διαμολυνθεί με το διαγονίδιο pcDNA4/TO-*DGCR14* ώστε να δημιουργηθεί η σταθερή διπλά διαμολυσμένη κυτταρική σειρά. Η διαδικασία παραμένει ίδια με αυτές των άλλων σταθερών κυτταρικών σειρών που αναφέρθηκαν παραπάνω. Ειδικότερα, δυο τρυβλία 10cm επιστρώθηκαν με  $1,2 \times 10^6$  κύτταρα του κλώνου #5, τα οποία διαμολύνθηκαν με 20μg του γραμμοποιημένου διαγονιδίου (με το ένζυμο *PvuI*) pcDNA4/TO-*DGCR14* με τη μέθοδο συνκατακρήμνισης του DNA με Ca<sub>3</sub>(PO<sub>4</sub>)<sub>2</sub>. Το ένζυμο αυτό (4525nt) τέμνει το γονίδιο ανθεκτικότητας σε αμπικιλίνη σε βακτήρια. Έπειτα από τη διαμόλυνση, τα κύτταρα του κάθε τρυβλίου θρυψινοποιήθηκαν σε άλλα τρία (όπως αναφέρθηκε και παραπάνω) και προστέθηκε πλήρες θρεπτικό υλικό με βλαστισιδίνη και τέσσερις διαφορετικές συγκεντρώσεις ζεοσίνης (γονίδιο ανθεκτικότητας του φορέα pcDNA4/TO σε ευκαρυωτικά κύτταρα, Cat. No R25001, ThermoFischer Scientific) 350 μg/ml, 500 μg/ml, 650 μg/ml και 800 μg/ml. Για να γίνει η επιλογή των συγκεντρώσεων ζεοσίνης δοκιμάστηκε ένα εύρος συγκεντρώσεων από 200-1200μg/ml σε κύτταρα *HeLa*. Η βιωσιμότητα των κυττάρων ελέγχθηκε σε χρονικό διάστημα 15 ημερών. Επιλέχθηκαν οι 4 συγκεντρώσεις ζεοσίνης που προαναφέρθηκαν διότι προκαλούσαν παρόμοιες επιδράσεις στη βιωσιμότητα των κυττάρων, δηλαδή τα κύτταρα πέθαιναν σε 3-7 ημέρες. Η συλλογή και η ανάπτυξη των αποικιών έγινε με τον τρόπο που αναλύθηκε παραπάνω.

#### 2.4.4. Πρωτόκολλο επεξεργασίας και συλλογής κυττάρων για ομικές αναλύσεις

Για την ανάλυση των κυτταρικών κλώνων με τροποποιημένη την έκφραση του γονιδίου *FRA10AC1* σε μεταγραφικό, πρωτεϊνικό και μεταβολικό επίπεδο μέσω ομικών τεχνολογιών, αναπτύχθηκε ένα κατάλληλο πρωτόκολλο επεξεργασίας και συλλογής τους. Το πρωτόκολλο αυτό θα έπρεπε να είχε την ελάχιστη επίδραση στην κυτταρική φυσιολογία έτσι ώστε οι μετρήσεις που θα προέκυπταν από τις υψηλής απόδοσης βιομοριακές αναλύσεις να μπορούν να αντικατοπτρίζουν την πραγματική κυτταρική φυσιολογία του υπό μελέτη συστήματος. Χρησιμοποιήθηκαν τρία διαφορετικά πρωτόκολλα και το τελευταίο δοκιμάστηκε σε δυο διαφορετικούς χρόνους συλλογής. Ο έλεγχος των επιδράσεων του εκάστοτε πρωτοκόλλου στην κυτταρική φυσιολογία πραγματοποιήθηκε με την ανάλυση του μεταβολικού προτύπου από το συνεργαζόμενο εργαστήριο Μεταβολικής Μηχανικής και Βιολογίας Συστημάτων, Ινστιτούτο Χημικών Μηχανικών Επιστημών, Ίδρυμα Έρευνας και Τεχνολογίας, FORTH/ICE-HT, Πάτρα, (Επικεφαλής: Δρ. Μ. Κλάπα).

Στα πειράματα που πραγματοποιήθηκαν με τα διαφορετικά πρωτόκολλα χρησιμοποιήθηκαν κύτταρα *HeLa* ίδιας ή διαφορετικής ανακαλλιέργειας από τη στιγμή της απόψυξής τους, συγκεκριμένου αριθμού ( $10^6$  κύτταρα), τα οποία επιστρώθηκαν σε τρυβλία καλλιέργειας 10cm και μετά από 48 ώρες πραγματοποιήθηκε η συλλογή τους με το εκάστοτε πρωτόκολλο.

Στα βήματα έκπλυσης των κυττάρων πριν τη συλλογή τους, χρησιμοποιήθηκαν δυο διαφορετικά διαλύματα έκπλυσης: α) το ρυθμιστικό διάλυμα φωσφορικών αλάτων 1xPBS και β) ο φυσιολογικός ορός (0,9% w/v NaCl) που παρασκευάστηκε στο συνεργαζόμενο εργαστήριο Μεταβολικής Μηχανικής και Βιολογίας Συστημάτων, από υψηλής καθαρότητας NaCl (Cat. No 35275, Riedel-de Haën) διαλυμένο σε νερό HPLC (Cat. No 34877, Sigma). Στα πειράματα που πραγματοποιήθηκαν, χρησιμοποιήθηκε είτε το ίδιο διάλυμα σε όλα τα βήματα πλύσης ή, όταν υπήρξε συνδυασμός, τα αρχικά βήματα πλύσης έγιναν με PBS και ο φυσιολογικός ορός (0.9% w/v NaCl) χρησιμοποιήθηκε στα τελικά στάδια πλύσης.

Τα πρωτόκολλα 1 και 2 βασίστηκαν σε μία μέθοδο συλλογής των προσκολλημένων κυττάρων η οποία περιλαμβάνει: i) μηχανική (με τη χρήση ειδικής ξύστρας) αποκόλληση των κυττάρων από το τρυβλίο καλλιέργειας και ii) πολλαπλά στάδια φυγοκεντρήσεων και επαναιωρήσεων σε ρυθμιστικό διάλυμα μέχρι τη συλλογή ενός καθαρού από το θρεπτικό μέσο ιζήματος κυττάρων. Στην παρούσα εργασία δεν επιλέχθηκε ο ενζυματικός χειρισμός καθώς η χρήση της θρυψίνης μπορεί να επηρεάσει τη συνολική ενεργότητα των πρωτεϊνών διότι αλληλεπιδρά με τις μεμβρανικές πρωτεΐνες, για παράδειγμα τις ιντεγκρίνες, με αποτέλεσμα την πρωτεόλυσή τους (Huang et al. 2010). Επιπλέον, έχει αρνητικές επιδράσεις στη μεταβολική φυσιολογία των κυτταρικών καλλιέργειών καθώς έχει δείχθει ότι η επώαση των κυττάρων με θρυψίνη έχει ως αποτέλεσμα τη διαρροή μικρών μεταβολιτών (π.χ. αμινοξέων) μέσω της κυτταροπλασματικής μεμβράνης (Dettmer et al. 2011).

Στο πρωτόκολλο 1, το πρώτο βήμα ήταν η αφαίρεση του θρεπτικού μέσου καλλιέργειας με πιπέτα Pasteur συνδεδεμένη με αντλία κενού. Η έκπλυση των κυττάρων έγινε τρεις φορές στο τρυβλίο καλλιέργειας με 1xPBS. Μετά την 3<sup>η</sup> έκπλυση, προστέθηκε στο τρυβλίο 1xPBS ή φυσιολογικός ορός (0.9% w/v NaCl) και

ακολούθησε μηχανική αποκόλληση των κυττάρων με χρήση ειδικής ξύστρας, μήκους επιφάνειας 1,8 cm (Corning, Cat. No CLS3010-100EA). Έπειτα πραγματοποιήθηκε η συλλογή των κυττάρων σε σωληνάριο των 15ml (falcon). Ακολούθησε φυγοκέντρηση για 8min στις 1000rpm, αφαίρεση του υπερκείμενου και επαναιώρηση των κυττάρων στο διάλυμα συλλογής τους (1xPBS ή 0.9% w/v NaCl). Το κυτταρικό εναιώρημα μεταφέρθηκε σε σωληνάκι του 1,5ml (ependorf tube) και φυγοκεντρήθηκε για 5min στις 1000rpm. Οι τελικές πελλέτες κυττάρων ψύχθηκαν άμεσα στους -80°C μέχρι την περαιτέρω ανάλυσή της.

Στο πρωτόκολλο 2, αρχικά αφαιρέθηκε το θρεπτικό μέσο από το τρυβλίο καλλιέργειας, με τον τρόπο που αναφέρθηκε στο πρωτόκολλο 1. Ακολούθησε η προσθήκη πλήρους θρεπτικού μέσου DMEM και η μηχανική αποκόλληση των κυττάρων με τη χρήση της ειδικής ξύστρας. Τα κύτταρα φυγοκεντρήθηκαν για 8min στις 1000rpm, αφαιρέθηκε το υπερκείμενο και συλλέχθηκε η πελλέτα των κυττάρων. Προστέθηκε διάλυμα έκπλυσης (1xPBS ή 0.9% w/v NaCl) και πραγματοποιήθηκε φυγοκέντρηση για 8min στις 1000rpm. Το βήμα αυτό επαναλήφθηκε άλλες 2 φορές. Η τελική πελλέτα κυττάρων επαναιωρήθηκε σε φυσιολογικό ορό (0.9% w/v NaCl). Το κυτταρικό εναιώρημα μεταφέρθηκε σε σωληνάκι του 1,5ml (ependorf tube) και φυγοκεντρήθηκε για 5min στις 1000rpm. Έπειτα από την αφαίρεση του υπερκείμενου, οι πελλέτες των κυττάρων αποθηκεύτηκαν στους -80°C μέχρι την περαιτέρω ανάλυσή της. Πριν την αφαίρεση του θρεπτικού μέσου από τα τρυβλία και σε κάθε βήμα φυγοκέντρησης γινόταν συλλογή 1ml υπερκείμενου διαλύματος και προσθήκη σε αυτό ίσης ποσότητας παγωμένης μεθανόλης HPLC (Cat No 494291, Sigma). Επιλέχθηκε να γίνει η αποκόλληση των κυττάρων σε θρεπτικό μέσο DMEM αντί σε διάλυμα έκπλυσης, για να αποφευχθεί ο σημαντικός παράγοντας της έλλειψης των θρεπτικών συστατικών ενώ τα κύτταρα υπόκεινται σε συνθήκες στρες κατά τη διάρκεια της μηχανικής αποκόλλησης.

Ο συνολικός χρόνος διάρκειας των πρωτοκόλλων 1 και 2 ήταν 30 και 45min, αντίστοιχα.

Τα πρωτόκολλα 3A και 3B προσαρμόστηκαν με βάση τη διαδικασία που περιγράφεται στους Teng et al. (2009). Με αυτή τη μέθοδο, δηλαδή την προσθήκη μεθανόλης για την επακόλουθη συλλογή των κυττάρων, αποφεύγονται οι συνέπειες της μηχανικής αποκόλλησης και της παραμονής των κυττάρων σε συνθήκες έλλειψης θρεπτικών συστατικών κατά τα βήματα των συνεχών φυγοκεντρήσεων και επαναιωρήσεων.

Στα πρωτόκολλα 3A και 3B, το αρχικό βήμα ήταν η αφαίρεση του θρεπτικού μέσου από τα προσκολλημένα κύτταρα. Και στα 2 πρωτόκολλα οι εκπλύσεις ήταν τρεις. Στο πρωτόκολλο 3A τα διαλύματα έκπλυσης ήταν το θρεπτικό μέσο DMEM ή το 1xPBS ή ο φυσιολογικός ορός 0.9% w/v NaCl τα οποία χρησιμοποιήθηκαν είτε ξεχωριστά το καθένα είτε σε συνδυασμό τα 2 τελευταία. Το DMEM χρησιμοποιήθηκε ώστε το περιβάλλον που εκτίθενται τα κύτταρα κατά τη διάρκεια της έκπλυσης να προσομοιάζει κατά πολύ το περιβάλλον ανάπτυξής τους, και να μην διαταράσσει, συγκριτικά με τα άλλα διαλύματα έκπλυσης, την κυτταρική φυσιολογία. Στο πρωτόκολλο αυτό, η αναρρόφηση του θρεπτικού μέσου ή των διαλυμάτων έκπλυσης από το τρυβλίο καλλιέργειας πραγματοποιήθηκε μέσω αντλίας κενού, αφήνοντας μια μικρή ποσότητα

ώστε να μην αφυδατωθούν τα κύτταρα. Στο πρωτόκολλο 3B η έκπλυση των κυττάρων έγινε είτε με 1xPBS είτε με 0.9% w/v NaCl ξεχωριστά ή με το συνδυασμό αυτών των δυο. Το θρεπτικό μέσο ή τα διαλύματα έκπλυσης απομακρύνθηκαν γρήγορα από το τρυβλίο καλλιέργειας με απόχυση και ακολούθησε άμεση προσθήκη του νέου διαλύματος. Και στα δυο πρωτόκολλα, μετά την απομάκρυνση του τελευταίου διαλύματος έκπλυσης, προστέθηκε στα κύτταρα παγωμένη μεθανόλη HPLC και ακολούθησε αποκόλληση των κυττάρων με τη χρήση της ειδικής ξύστρας όπως και στα άλλα δυο πρωτόκολλα. Η χρήση της μεθανόλης εξυπηρετεί: α) την αδρανοποίηση των ενζυμικών διεργασιών και β) την εκχύλιση των ενδοκυτταρικών πολικών μεταβολιτών σε ικανοποιητικό βαθμό σε σύγκριση με ένα υδατικό διάλυμα διότι η μεθανόλη αποτελεί έναν οργανικό διαλύτη ο οποίος διευκολύνει την εξαγωγή των μεταβολιτών έξω από τα κύτταρα (Villas-Bôas et al. 2005). Έπειτα ακολούθησε η συλλογή και η αποθήκευση των κυττάρων στους  $-80^{\circ}\text{C}$  μέχρι την περαιτέρω ανάλυσή τους. Στο πρωτόκολλο 3A είχε συλλεχθεί 1ml υπερκείμενων διαλυμάτων σε τρία στάδια (πριν από την 1<sup>η</sup>, 2<sup>η</sup> και 3<sup>η</sup> έκπλυση) με επακόλουθη προσθήκη σε αυτά ίσου όγκου παγωμένης μεθανόλης. Στο πρωτόκολλο 3B, τα υπερκείμενα διαλύματα συλλέχθηκαν σε δυο στάδια (πριν από την 1<sup>η</sup> και την 3<sup>η</sup> έκπλυση) και προστέθηκε σε αυτά μεθανόλη. Σε όλες τις περιπτώσεις τα διαλύματα αποθηκεύτηκαν στους  $-80^{\circ}\text{C}$ . Η χρονική διάρκεια του πρωτοκόλλου 3A μέχρι και την προσθήκη μεθανόλης στα τρυβλία καλλιέργειας ήταν 10min ενώ η αντίστοιχη διάρκεια του πρωτοκόλλου 3B ήταν 1min. Αυτή η χρονική διαφορά μεταξύ των δυο πρωτοκόλλων οφείλεται στην άμεση απόχυση του θρεπτικού μέσου καλλιέργειας και των διαλυμάτων έκπλυσης από το τρυβλίο καλλιέργειας, στο πρωτόκολλο 3B.

## **2.5. Απομόνωση νουκλεϊκών οξέων & τεχνικές ανασυνδυασμένου DNA**

### **2.5.1. Απομόνωση γονιδιωματικού DNA**

Η απομόνωση γενετικού υλικού από τα κύτταρα *HeLa* που αναπτύχθηκαν και από τις σταθερές κυτταρικές σειρές *HeLa* που δημιουργήθηκαν, έγινε με σκοπό τον έλεγχο τους για την παρουσία μυκοπλάσματος το οποίο ανήκει σε ένα γένος βακτηρίων που δεν διαθέτουν κυτταρικό τοίχωμα. Το DNA που απομονώθηκε χρησιμοποιήθηκε για PCR για την ανίχνευση μυκοπλάσματος διότι, τουλάχιστον στα αρχικά στάδια της μόλυνσης από αυτό το παθογόνο, τα κύτταρα δεν επηρεάζονται στη μορφολογία και στην ανάπτυξή τους. Ειδικότερα, σε αυτά τα στάδια, τα βακτήρια επικάθονται στην κυτταρική μεμβράνη, χωρίς όμως να δρουν τοξικά και χωρίς να προκαλούν αλλαγές στο pH. Έπειτα, ανταγωνίζονται με τα κύτταρα για τις πρόδρομες βιοσυνθετικές ουσίες και τα θρεπτικά συστατικά αναστέλλοντας το μεταβολισμό τους. Τα βακτήρια αυτά μπορεί να προκαλέσουν διαταραχές στη σύνθεση του DNA, του RNA και των πρωτεϊνών, μείωση στα επίπεδα των αμινοξέων και της τριφωσφορικής αδενοσίνης (ATP) και τροποποίηση των αντιγόνων της κυτταροπλασματικής μεμβράνης (Young et al. 2010). Η απομόνωση του DNA των εκάστοτε κυττάρων πραγματοποιήθηκε με κατάλληλο κιτ απομόνωσης DNA (Cat. No. 740952.50, Nacheary Nagel, NucleoSpin Tissue). Σε αυτό το πρωτόκολλο, η λύση των κυττάρων επιτυγχάνεται με την προσθήκη ενός διαλύματος που περιέχει το ανιονικό απορρυπαντικό SDS (δωδεκακυλοθειϊκό νάτριο) σε συνδυασμό με υψηλή θερμοκρασία. Το DNA συνδέεται

σε μεμβράνη πυριτίου, που είναι ειδική για τη σύνδεση μόνο των νουκλεϊκών οξέων, μέσα σε μια στήλη, μέσω της προσθήκης ενός διαλύματος το οποίο περιέχει μεγάλες ποσότητες χαοτροπικών ιόντων (για την απελευθέρωση του DNA από τις πρωτεΐνες) και αιθανόλης (για την κατακρήμνισή του). Έτσι, πραγματοποιήθηκε η γρήγορη απομόνωση, υψηλής καθαρότητας, γενωμικού DNA ποσότητας 15-25μg από κύτταρα που προήλθαν από το 1/10 μίας φλάσκας εμβαδού 25cm<sup>2</sup>.

### 2.5.2. Απομόνωση πλασμιδιακού DNA

Η απομόνωση πλασμιδιακού DNA πραγματοποιήθηκε είτε σε μικρή κλίμακα (mini prep) κατά τα στάδια υποκλωνοποίησης των τριών υποψήφιων αλληλεπιδρώντων μορίων της πρωτεΐνης FRA10AC1 στο φορέα κλωνοποίησης pmCherry-C1 ή σε μεγάλη κλίμακα (maxi prep) για την παραγωγή μεγάλων ποσοτήτων DNA τόσο των πλασμιδιακών φορέων κλωνοποίησης όσο και των διαγονιδίων που χρησιμοποιήθηκαν. Η απομόνωση DNA σε μικρή κλίμακα πραγματοποιήθηκε με το πρωτόκολλο αλκαλικής λύσης (Sambrook et al. 1989) που περιλάμβανε τα εξής βήματα: Αρχικά, πραγματοποιήθηκε η ανάπτυξη μοναδιαίων αποικιών των μετασχηματισμένων βακτηρίων σε 2-3ml υγρού θρεπτικού υλικού (LB broth) με το κατάλληλο αντιβιοτικό στους 37°C ολονύκτια υπό ανάδευση. Ειδικότερα για τους φορείς κλωνοποίησης pcDNATM6/TR, pLKO.1-puro και pcDNA 4/TO χρησιμοποιήθηκε αμπικιλίνη (100mg/ml) σε τελική συγκέντρωση 100μg/ml και για τον φορέα pmCherry-C1 χρησιμοποιήθηκε καναμυκίνη (50mg/ml) σε τελική συγκέντρωση 30μg/ml. Ακολούθησε η μεταφορά ~1,5ml καλλιέργειας σε σωληνάριο των 2ml (eppendorf tube), η συλλογή των κυττάρων με φυγοκέντρηση στις 3000 rpm για 15min στους 4°C και η άμεση απομάκρυνση του υπερκειμένου θρεπτικού υλικού. Έπειτα, έγινε πλήρης επαναιώρηση των κυττάρων σε 300μl διαλύματος P1 (15mM Tris-Cl, pH=8, mM EDTA, pH=8), προσθήκη 300μl διαλύματος αποδιάταξης P2 (200mM NaOH, 1%SDS) σε συνδυασμό με ήπια ανάδευση, και επώαση για 5min μέχρι το διάλυμα να γίνει διαυγές. Με αυτόν τον τρόπο πραγματοποιήθηκε η λύση των βακτηριακών κυττάρων. Στη συνέχεια, προστέθηκε 300μl ρυθμιστικού διαλύματος P3 (3M οξικό κάλιο, pH 5,5) με ταυτόχρονη ισχυρή ανάδευση και επώαση στον πάγο για 15min για τη γρήγορη ουδετεροποίηση του διαλύματος. Τα δείγματα φυγοκεντρήθηκαν στις 13000rpm στους 4°C για 15min. Το υπερκείμενο διάλυμα συλλέχθηκε προσπαθώντας να μην συλλεχθούν άσπρα συσσωματώματα (άλατα). Τα δείγματα φυγοκεντρήθηκαν ξανά στις ίδιες συνθήκες και το τελικό υπερκείμενο διάλυμα μεταφέρθηκε σε καθαρά σωληνάκια. Τα τελικά βήματα περιλάμβαναν: α) την κατακρήμνιση του πλασμιδιακού DNA με την προσθήκη ισοπροπανόλης (0,8 του όγκου του υπερκειμένου), ακόλουθη ανάδευση και φυγοκέντρηση στις 12000rpm στους 4°C για 15min, β) την απομάκρυνση του υπερκειμένου και την έκπλυση του DNA με 70% αιθανόλη, γ) την φυγοκέντρηση στις 12000rpm για 10min στους 4°C, δ) την απομάκρυνση της αιθανόλης, ε) την επαναδιάλυση του DNA σε νερό. Η ποσότητα του πλασμιδιακού DNA που απομονώνεται εξαρτάται από τον αριθμό αντιγράφων του πλασμιδίου και ήταν της τάξεως των αρκετών μg (1-10μg).



Στην περίπτωση της απομόνωσης σε μεγάλη κλίμακα χρησιμοποιήθηκαν δυο τρόποι. Ο πρώτος είναι η ίδια διαδικασία που περιγράφηκε παραπάνω με τη διαφορά ότι ο όγκος της βακτηριακής καλλιέργειας αυξήθηκε στα 2lt, οι όγκοι των διαλυμάτων P1, P2, P3 αυξήθηκαν στα 3mL και οι φυγοκεντρήσεις πραγματοποιήθηκαν στις 3000rpm για 20min στους 4°C. Έπειτα από την προσθήκη του διαλύματος P3, τη φυγοκέντρωση και τη συλλογή του υπερκείμενου ακολουθήθηκε, στη συγκεκριμένη περίπτωση, περαιτέρω επεξεργασία διότι τα δείγματα έπρεπε να ήταν καθαρά για τη διαδικασία της διαμόλυνσής τους στα ευκαρυωτικά κύτταρα *HeLa*. Η μετέπειτα διαδικασία περιλάμβανε την επώαση με RNAση (Cat. No 740505, Macherey Nagel) σε τελική συγκέντρωση 100μg/ml για 1h στους 37°C, και επακόλουθη επώαση με πρωτεΐνάση K (Cat No P6556, Sigma) σε τελική συγκέντρωση 100μg/ml 3h ή ολονύκτια στους 55°C σε κατάλληλο ρυθμιστικό διάλυμα 1x (50mM Tris.Cl, pH 7.5, 100mM NaCl, 0,02% SDS, 2mM EDTA) για την αποικοδόμηση των πρωτεϊνών και για την απενεργοποίηση των ενδογενών νουκλεασών που μπορεί να καταστρέψουν το DNA. Έπειτα, πραγματοποιήθηκαν διαδοχικές εκχυλίσεις με φαινόλη-χλωροφόρμιο για την απομάκρυνση των πρωτεϊνών και των λιπιδίων. Αρχικά προστέθηκε φαινόλη και χλωροφόρμιο σε αναλογία όγκου 1:1, ακολούθησε ισχυρή ανάδευση και φυγοκέντρωση στις 11000rpm για 10min στους 4°C. Η φυγοκέντρωση είχε ως αποτέλεσμα τη δημιουργία τριών φάσεων, την υπερκείμενη φάση που περιείχε υδατικό διάλυμα νουκλεϊκών οξέων, τη μεσόφαση στην οποία βρίσκονταν οι πρωτεΐνες και τα λιπίδια και την κατώτερη φάση η οποία περιλάμβανε τη φαινόλη και το χλωροφόρμιο. Στη συνέχεια συλλέχθηκε η υπερκείμενη φάση, και το βήμα της φυγοκέντρωσης επαναλαμβανόταν μέχρι να προκύψει καθαρή μεσόφαση. Έπειτα ακολούθησε προσθήκη χλωροφορμίου σε όγκο ίσο με τον όγκο του δείγματος ώστε να απομακρυνθεί η φαινόλη, φυγοκέντρωση στις 11000rpm για 5min στους 4°C και συλλογή της υπερκείμενης φάσης. Στο υπερκείμενο διάλυμα, προστέθηκε οξικό νάτριο (CH<sub>3</sub>COONa, 1/10 του υπερκείμενου διαλύματος) συγκέντρωσης 3M και pH=5.2 και απόλυτη αιθανόλη (σε τελική συγκέντρωση 70-75%). Ακολούθησε ήπια ανάδευση για να επιτευχθεί η κατακρήμνιση του DNA. Αν σχηματιζόταν νημάτια DNA, τότε γινόταν η συλλογή τους («ψάρεμα») και η τοποθέτησή τους σε διαφορετικό σωληνάριο. Εάν όχι, τότε γινόταν επώαση στους -80°C για 20min και επακόλουθη φυγοκέντρωση στις 12000rpm για 15 min στους 4°C με αποτέλεσμα τη συλλογή του DNA. Και στις δυο περιπτώσεις το DNA εκπλύθηκε με 70% αιθανόλη και ακολούθησε φυγοκέντρωση στις 12000rpm για 5min στους 4°C. Έπειτα από την αφαίρεση της αιθανόλης, τα δείγματα παρέμειναν στους 37°C για την πλήρη εξάτμιση της αιθανόλης και επαναδιαλύθηκαν σε νερό. Η ποσότητα του DNA ήταν της τάξεως των mg (~1mg).

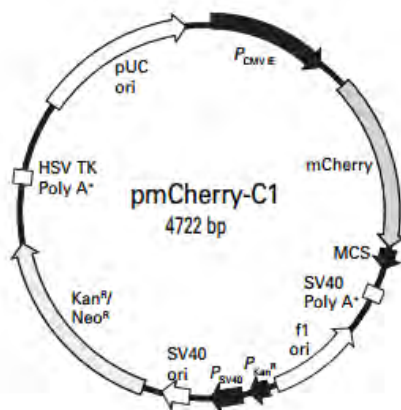
Ο δεύτερος τρόπος ήταν η χρήση κατάλληλου κιτ απομόνωσης (QIAGEN® Plasmid Maxi Kit, Cat. No. 12162). Το πρωτόκολλο αυτό βασίζεται στη διαδικασία της αλκαλικής λύσης συνοδευόμενη από τη σύνδεση του πλασμιδιακού DNA σε μια μεμβράνη ρητίνης, σε ειδικά διαμορφωμένη στήλη (QIAGEN-tip), η οποία λειτουργεί μέσω της ροής λόγω βαρύτητας υπό συνθήκες χαμηλής συγκέντρωσης άλατος και pH. Το RNA, οι πρωτεΐνες και τα μόρια μικρού μοριακού βάρους απομακρύνονται με τη χρήση διαλυμάτων πλύσης μέτριας συγκέντρωσης άλατος. Η απομόνωση του

πλασμιδιακού DNA είναι σχετικά γρήγορη, το παραγόμενο DNA είναι καλής ποιότητας και μεγάλης ποσότητας (500μg-1mg).

### 2.5.2.1. Πλασμιδιακοί φορείς

#### **1. pmCherry-C1**

Ο φορέας κλωνοποίησης pmCherry-C1 (<http://www.clontech.com/>, Cat. No 632524) χρησιμοποιήθηκε για την υποκλωνοποίηση των υποψήφιων αλληλεπιδρώντων μορίων της πρωτεΐνης FRA10AC1 με τη μέθοδο μεταφοράς ενέργειας φθορισμού μέσω συντονισμού (FRET) (Εικ. 25). Ο φορέας έχει μήκος 4722bp και διαθέτει: i) αλληλουχίες έναρξης της αντιγραφής σε κύτταρα θηλαστικών από τον ιό SV40 (simian virus 40 ori) οι οποίες επιτρέπουν την αντιγραφή των γονιδίων σε κύτταρα θηλαστικών που εκφράζουν το αντιγόνο T (μια πρωτεΐνη του ιού απαραίτητη για την έναρξη της αντιγραφής) και σε βακτήρια (pUC origin), ii) αλληλουχίες πολυαδενυλίωσης από τον ιό SV40 που είναι κατάλληλες για την επεξεργασία του 3' άκρου του mRNA που περιλαμβάνει το ένθεμα και τον επίτοπο mCherry και iii) γονίδια ανθεκτικότητας στην καναμυκίνη και στη νεομυκίνη για την ανάπτυξη σε βακτηριακά κύτταρα και σε κύτταρα θηλαστικών, αντίστοιχα (Εικ. 25). Ο φορέας αυτός κατευθύνει την έκφραση μιας χμιαρικής πρωτεΐνης που φέρει την mCherry στο αμινοτελικό της άκρο. Η πρωτεΐνη mCherry ανήκει στην κατηγορία των κόκκινων φθορίζουσών πρωτεϊνών και παράγεται από ένα γονίδιο το οποίο απομονώθηκε από το κοράλλι *Discosoma sp.* Το μέγιστο φάσμα απορρόφησης και εκπομπής της mCherry είναι 587nm και 610nm, αντίστοιχα, και χρησιμοποιείται ευρέως σε σύντηξη με πολλές πρωτεΐνες για τεχνικές απεικόνισης (Day & Davidson 2012, Shaner et al. 2004).

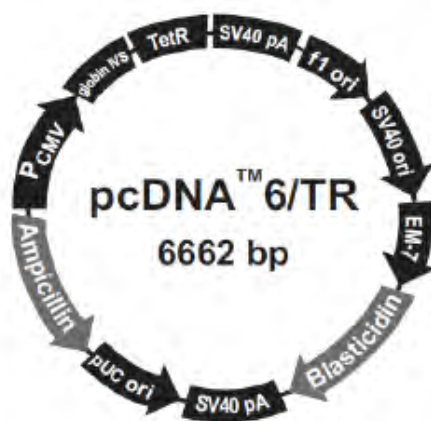


**Εικ. 25: Αδρός φυσικός χάρτης του φορέα pm-Cherry (από <http://www.clontech.com/>).** Η έκφραση του χμιαρικού mRNA ελέγχεται από τον ισχυρό υποκινητή του κυτταρομεγαλοϊού (P<sub>CMV</sub>). Επιπλέον, απεικονίζονται οι αλληλουχίες πολυαδενυλίωσης SV40, η αλληλουχία έναρξης της αντιγραφής σε κύτταρα θηλαστικών (SV40) και σε βακτήρια (pUC) και το γονίδιο που παρέχει ανθεκτικότητα σε καναμυκίνη και νεομυκίνη για την ανάπτυξη σε βακτηριακά κύτταρα και κύτταρα θηλαστικών, αντίστοιχα.

#### **2. pcDNA<sup>TM</sup>6/TR**

Ο φορέας pcDNA<sup>TM</sup>6/TR (<http://www.invitrogen.com>, Cat. No V1025-20) χρησιμοποιήθηκε για τη δημιουργία σταθερά διαμολυσμένης κυτταρικής σειράς *HeLa* στην οποία εκφράζεται ιδιοσυστατικά ο καταστολέας της τετρακυκλίνης (Tet

repressor, TetR). Ο φορέας αυτός είναι ένα «ρυθμιστικό» πλασμίδιο το οποίο χρησιμοποιείται συνδυαστικά με άλλα εργαλεία για την επαγόμενη έκφραση του υπό μελέτη γονιδίου υπό τον έλεγχο του συστήματος της τετρακυκλίνης. Έχει μέγεθος 6662bp και εκφράζει σε υψηλά επίπεδα τον καταστολέα τετρακυκλίνης (TetR) υπό τον έλεγχο του υποκινητή CMV (Εικ. 26). Στα περισσότερα κύτταρα θηλαστικών μπορεί να επιτευχθεί σταθερή και υψηλή έκφραση του καταστολέα. Με τον τρόπο αυτόν, μπορεί να επιτευχθεί η επαγόμενη έκφραση ενός γονιδίου, διαμολύνοντας τα κύτταρα και με ένα δεύτερο πλασμίδιο που φέρει το cDNA του υπό μελέτη γονιδίου υπό τον έλεγχο ενός υποκινητή που ρυθμίζεται από τον καταστολέα της τετρακυκλίνης. Το γονίδιο του καταστολέα TetR, που υπάρχει στο φορέα, προέρχεται από το τρανσποζόνιο Tn10. Η πρωτεΐνη του καταστολέα αποτελείται από 207 αμινοξέα και έχει μοριακό βάρος 23 kD. Το πλασμίδιο pcDNA6/TR περιέχει ακόμη τα γονίδια ανθεκτικότητας στην αμικικιλίνη και στη βλαστισιδίνη σε βακτηριακά κύτταρα και σε κύτταρα θηλαστικών, αντίστοιχα (Εικ. 26).



**Εικ. 26:** Χάρτης πλασμιδιακού φορέα pcDNA<sup>TM</sup>6/TR (<http://www.invitrogen.com>). Η έκφραση του γονιδίου του καταστολέα TetR ελέγχεται από τον ισχυρό υποκινητή του κυτταρομεγαλοϊού CMV. Επιπλέον, απεικονίζονται οι αλληλουχίες πολυαδενυλίωσης SV40, η αλληλουχία έναρξης της αντιγραφής σε κύτταρα θηλαστικών (SV40) και σε βακτήρια (pUC) και τα γονίδια που παρέχουν ανθεκτικότητα σε αμικικιλίνη και βλαστισιδίνη για την ανάπτυξη σε βακτηριακά κύτταρα και κύτταρα θηλαστικών, αντίστοιχα.

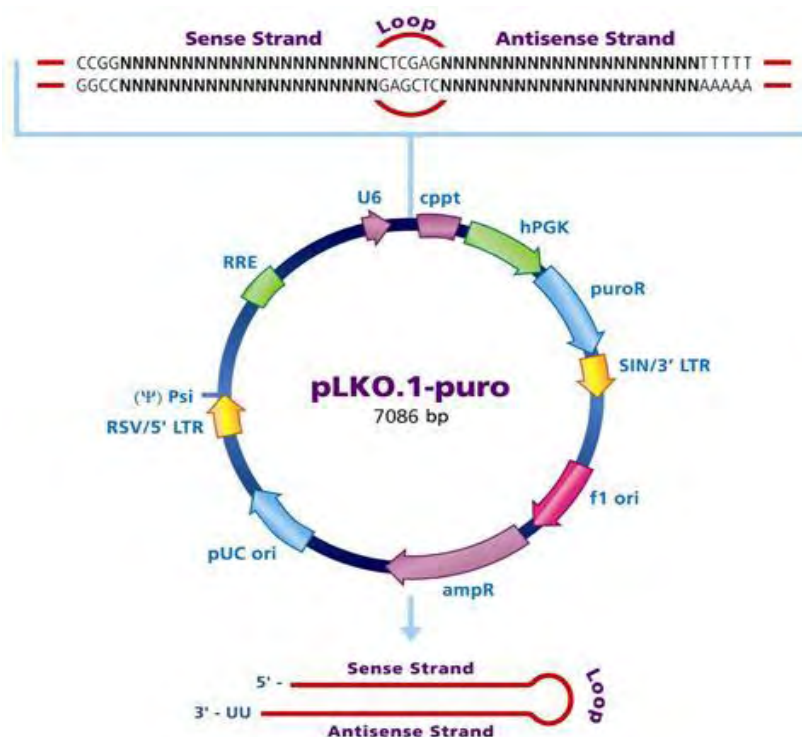
#### 2.5.2.2. Πλασμιδιακές κατασκευές

Οι πλασμιδιακές κατασκευές που χρησιμοποιήθηκαν στην παρούσα εργασία σκόπευαν: α) στην καταστολή της έκφρασης του γονιδίου *FRA10AC1* στα κύτταρα *HeLa*, β) στην υποκλωνοποίηση των υποψήφιων αλληλεπιδρώντων μορίων της πρωτεΐνης *FRA10AC1* στο φορέα κλωνοποίησης pmCherry-C1 και γ) στη δημιουργία της σταθερά διαμολυσμένης κυτταρικής σειράς *HeLa* που εκφράζει επαγόμενα το γονίδιο *DGCR14*.

##### **α) pLKO.1-puro/shRNA**

Για την καταστολή της έκφρασης του γονιδίου *FRA10AC1* χρησιμοποιήθηκε ο φορέας έκφρασης σε κύτταρα θηλαστικών pLKO.1-puro της εταιρίας Sigma-Aldrich ο οποίος περιείχε μόρια shRNA (short-hairpin RNAs) που στοχεύουν στο γονίδιο. Ως αρνητικός έλεγχος του πειράματος χρησιμοποιήθηκε μόριο shRNA που δεν στοχεύει

σε κανένα γνωστό γονίδιο των θηλαστικών (pLKO.1-puro Non-Mammalian shRNA Control, #SHC002, Sigma) (Εικ. 27). Τα χαρακτηριστικά του φορέα αυτού επιτρέπουν την παροδική ή τη σταθερή διαμόλυνση κυττάρων θηλαστικών με τα μόρια shRNA (Stewart et al. 2003). Χωρίς το shRNA, ο φορέας αυτός έχει μήκος 7052bp, διαθέτει γονίδια ανθεκτικότητας στην αμικικιλίνη για την ανάπτυξη σε βακτηριακά κύτταρα και στην πουρομυκίνη για την ανάπτυξη σε κύτταρα θηλαστικών. Διαθέτει θέση έναρξης της αντιγραφής σε βακτήρια (pUC origin) και ισχυρό υποκινητή U6 ο οποίος χρησιμοποιείται από την RNA πολυμεράση III για τη σύνθεση των shRNA, tRNA, rRNA κ.α εκτός από το mRNA (Εικ. 27). Η σύνθεση του shRNA, ενισχύει το μηχανισμό παρεμβολής μέσω RNA (RNAi) (Xia et al. 2003). Επιπλέον διαθέτει περιοχές (5' & 3' LTR- self inactivating long terminal repeat, Psi- RNA packaging signal, f1 origin) που είναι απαραίτητες για την παραγωγή ιικών μορίων RNA ενός είδους ρετροϊού (lentivirus) τα οποία εμπεριέχουν την αλληλουχία shRNA. Οι περιοχές αυτές επιτρέπουν τη μόλυνση και την αντιγραφή αυτών των ιών έτσι ώστε να ενσωματωθούν στο γονιδίωμα (μια διαδικασία γνωστή ως «transduction») μη διαφοροποιημένων ή μη διαιρούμενων κυττάρων όπως οι νευρώνες ή τα κύτταρα των δενδριτών, ξεπερνώντας με αυτόν τον τρόπο την χαμηλή αποτελεσματικότητα της διαμόλυνσής τους (Εικ. 27).

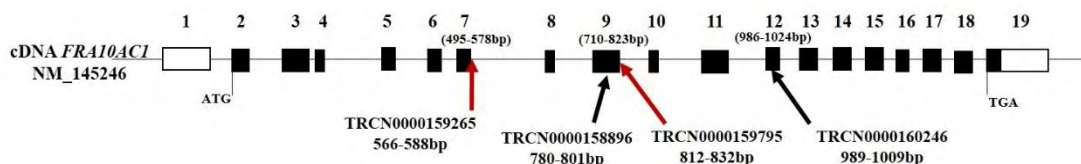


**Εικ. 27:** Αδρός φυσικός χάρτης του πλασμιδιακού φορέα pLKO.1-puro (<https://www.sigmaaldrich.com/>). Ο φορέας εκφράζει το κλωνοποιημένο shRNA υπό τον έλεγχο του υποκινητή U6 του ανθρώπου και περιέχει γονίδια που προσδίδουν ανθεκτικότητα στην αμικικιλίνη σε βακτήρια, και στην πουρομυκίνη σε κύτταρα θηλαστικών. Φαίνεται η θέση ενσωμάτωσης του shRNA καθώς και οι αλληλουχίες βάσεων (CCGG/GGCC, TTTT/AAAAA) που είναι σταθερές σε κάθε shRNA (επάνω μέρος της εικόνας).

Στην περίπτωση του γονιδίου *FRA10AC1* χρησιμοποιήθηκε ο φορέας pLKO.1-puro που έφερε τα παρακάτω shRNA (short-hairpin RNAs) μεγέθους 59 βάσεων:

- shRNA-TRCN0000160246 με κωδικό κλώνου NM\_145246.2-1007s1c1 και αλληλουχία:CCGGCTGAAGATTCTCTACTTAGAACTCGAGTTCTAAGTAGAGAATCTTCAGTTTTTTG: shRNA (1)
- shRNA-TRCN0000158896 με κωδικό κλώνου NM\_145246.2-798s1c1 και αλληλουχία:CCGGGAAGAGAAATGCACTTGTTAACTCGAGTTAACAAGTGCATTTCTCTTTTTTTG: shRNA (2)
- shRNA-TRCN0000159265 με κωδικό κλώνου NM\_145246.2-585s1c1 και αλληλουχία:CCGGGAGACTTGCTAAGAAATACTACTCGAGTAGTATTTCTTAGCAAGTCTTTTTTTG: shRNA (3)
- shRNA-TRCN0000159795 με κωδικό κλώνου NM\_145246.2-830s1c1 και αλληλουχία:CCGGGCCAAGAATGTTCCATTAAATCTCGAGATTTAATGGAACATTTCTGGCTTTTTTTG: shRNA (4)

Όλοι οι παραπάνω κλώνοι shRNA στοχεύουν στην κωδική περιοχή του mRNA του γονιδίου *FRA10AC1* (Εικ. 28). Η αποτελεσματικότητα απενεργοποίησης της έκφρασης του γονιδίου ελέγχθηκε από την ίδια την εταιρία στην κυτταρική σειρά ινοβλαστών *MCH58*. Η δοκιμή αυτή έδειξε ότι τα τέσσερα διαφορετικά μόρια shRNA παρουσιάζουν κατ' αντιστοιχία ποσοστά αποτελεσματικότητας απενεργοποίησης 83%, 90%, 95%, και 72%.



**Εικ. 28:** Στόχοι κλώνων shRNA στο cDNA του γονιδίου *FRA10AC1*. Φαίνονται με μαύρα βέλη οι στόχοι των κλώνων TRCN0000158896 και TRCN0000160246 εντός των εξονίων 9 και 12, αντίστοιχα. Οι κλώνοι TRCN0000159265 και TRCN0000159795 στοχεύουν στα όρια εξονίου-ιντρονίου (κόκκινα βέλη).

### β) *FRA10AC1*-pEGFP-N1

Η πλασμιδιακή κατασκευή *FRA10AC1*-pEGFP-N1 είχε κατασκευαστεί παλιότερα (Sarafidou et al. 2004) και χρησιμοποιήθηκε για τον έλεγχο των πρωτεϊνικών αλληλεπιδράσεων της *FRA10AC1* με τα τρία υπονήφια αλληλεπιδρώντα μόρια (SAP30BP, THOC1, MFAP1) με τη μέθοδο μεταφοράς ενέργειας φθορισμού μέσω συντονισμού (FRET). Το τμήμα που αντιστοιχεί στα νουκλεοτίδια 189-1137 (διαθέτει κωδικόνιο έναρξης αλλά όχι λήξης) του cDNA του γονιδίου *FRA10AC1* (κωδικός νουκλεοτιδικής αλληλουχίας: NM\_145246.4), το οποίο προέκυψε με αλυσιδωτή αντίδραση πολυμεράσης (PCR) (Πιν. 2) σε cDNA από όρχεις ανθρώπου (της συλλογής Multiple Tissue Panels of cDNA της εταιρίας Clontech), είχε υποκλωνοποιηθεί αρχικά στο φορέα pGEM-T-Easy. Ο φορέας αυτός είναι γραμμικός και περιέχει προεξέχοντα νουκλεοτίδια θυμίνης στα άκρα του επιτρέποντας την εισαγωγή προϊόντων PCR τα

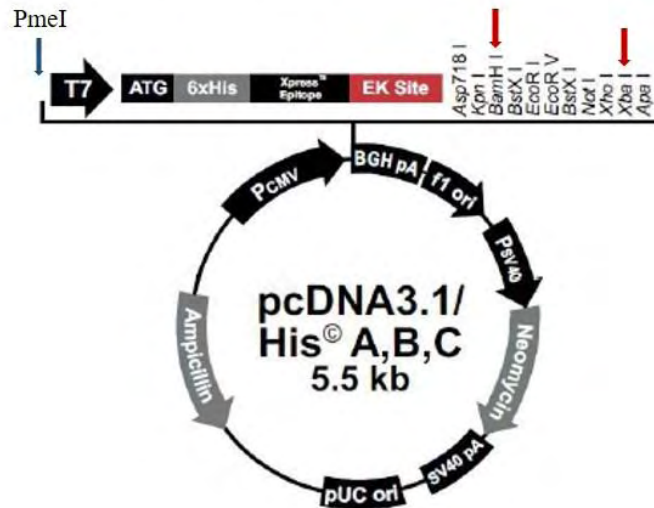
οποία έχουν ενισχυθεί με πολυμεράσες που παράγουν προεξέχοντα νουκλεοτίδια αδενίνης. Έπειτα, το συγκεκριμένο ένθεμα αποκόπηκε από το φορέα αυτό με τα ένζυμα περιορισμού *BamHI* και *EcoRI* και εισήχθη στο φορέα pEGFP-N1 στις θέσεις περιορισμού της κασέτας κλωνοποίησης (polylinker) *BglIII* και *EcoRI*, καθώς τα ένζυμα *BamHI* και *BglIII* έχουν συμβατά άκρα. Ο φορέας αυτός κατευθύνει την έκφραση μιας χημικτικής πρωτεΐνης που φέρει την EGFP στο καρβοξυτελικό της άκρο. Το μέγιστο φάσμα απορρόφησης και εκπομπής της EGFP είναι 488nm και 509nm, αντίστοιχα, και χρησιμοποιείται ευρέως σε σύντηξη με πολλές πρωτεΐνες για τεχνικές απεικόνισης (Day & Davidson 2012, Zhang et al. 1996).

#### **γ) pCMV-SPORT6/MFAP1, pCMV-SPORT6/SAP30BP, pDNR-LIB/THOC1**

Για την υποκλωνοποίηση των cDNA των γονιδίων *SAP30BP*, *THOC1*, *MFAP1* στο φορέα κλωνοποίησης pmCherry-C1 χρησιμοποιήθηκαν οι πλασμιδιακές κατασκευές από την εταιρία Dharmacon (<http://dharmacon.gelifesciences.com/>) που περιλαμβάνουν το πλήρους μήκους cDNA των γονιδίων *MFAP1* (Clone ID: 6200464, Cat. No MHS6278-202760123), *SAP30BP* (Clone ID: 6502502, Cat. No MHS6278-202760237) και *THOC1* (Clone ID: 4046908, Cat. No MHS6278-202839770) στους φορείς pCMV-SPORT6 (τα δύο πρώτα) και pDNR-LIB (το τελευταίο) οι οποίοι είναι απλοί φορείς κλωνοποίησης. Τα cDNA των γονιδίων αυτών προέρχονται από βιβλιοθήκες cDNA που είναι διαθέσιμες στο Mammalian Gene Collection (MGC) της εταιρίας Clontech (NIH\_MGC Library) και αντιστοιχούν στους παρακάτω κωδικούς νουκλεοτιδικής αλληλουχίας: για το γονίδιο *MFAP1* (BC050742.1, cDNA βιβλιοθήκη Lupski\_sympathetic\_trunk), για το γονίδιο *THOC1* (BC010381.1, cDNA βιβλιοθήκη NIH\_MGC\_54) και για το γονίδιο *SAP30BP* (BC063793.1, cDNA βιβλιοθήκη NIH\_MGC\_71).

#### **δ) pcDNA4/TO-DGCR14**

Η πλασμιδιακή κατασκευή pcDNA4/TO-DGCR14 δημιουργήθηκε στο πλαίσιο της μεταπτυχιακής διατριβής της κ. Ε. Καυκιά στο συνεργαζόμενο εργαστήριο Γενικής Βιολογίας, Ιατρική Σχολή του Πανεπιστήμιο Πατρών (επικεφαλής: Καθ. Ν. Μοσχονάς). Το cDNA του γονιδίου είχε απομονωθεί αρχικά με PCR από τον Image Clone #2823027 (Human ORFeome V7.1, <http://horfdb.dfci.harvard.edu>) με κατάλληλους εκκινητές (Πιν. 2) οι οποίοι περιείχαν θέσεις αναγνώρισης περιοριστικών ενζύμων (*BamHI* & *XbaI*) και υποκλωνοποιήθηκε στον φορέα pcDNA3.1\_His B (Εικ. 29).



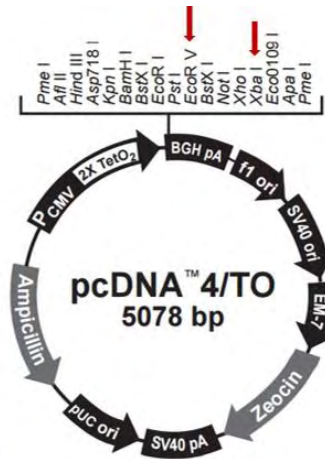
**Εικ. 29:** Χάρτης πλασμιδιακού φορέα pcDNA 3.1 His. Με κόκκινα βέλη φαίνονται οι θέσεις κλωνοποίησης του γονιδίου *DGCR14* στον φορέα pcDNA 3.1/His B. Με μπλε βέλος φαίνεται η αλληλουχία αναγνώρισης του ενζύμου *PmeI* που χρησιμοποιήθηκε μαζί με το *XbaI* για την εξαγωγή του cDNA του γονιδίου *DGCR14* και των επιτόπων (6xHis, Xpress) από το φορέα pcDNA 3.1 και την επακόλουθη ενσωμάτωση αυτού του τμήματος στο φορέα pcDNA4/TO.

**Πιν. 2:** Εκκινητές για την ενίσχυση του cDNA του γονιδίου *FRA10AC1* και *DGCR14*.

Εμπρόσθιοι εκκινητές (Forward) για τα cDNA των γονιδίων	
<i>FRA10AC1</i>	5'-CGG-GAT-CCG-AAA-GGA-TG-CAT-GGT-CAT-G-3'
<i>DGCR14</i>	5'-CGC-GGA-TCC-GAT-GGA-GAC-GCC-GGG-CGC-ATC-3'
Ανάστροφοι εκκινητές (Reverse) για τα cDNA των γονιδίων	
<i>FRA10AC1</i>	5'-CGG-AAT-TCG-AAA-CAA-ATC-CTG-AAA-ATA-CTC-3'
<i>DGCR14</i>	5'-TGC-TCT-AGA-CTA-AAA-GAA-GTC-CGA-AGC-TTT-GCG-3'

Με υπογραμμισμένη γραφή φαίνονται οι θέσεις αναγνώρισης των *BamHI* & *XbaI* και με έντονη γραφή το κωδικόνιο έναρξης της μετάφρασης.

Στη συνέχεια το cDNA του γονιδίου *DGCR14* αποκόπηκε από τον φορέα pcDNA3.1\_His B με τα ένζυμα περιορισμού *XbaI* και *PmeI* έτσι ώστε να ληφθούν μαζί οι επίτοποι Xpress και 6xHis από το φορέα pcDNA3.1\_His B (Εικ. 29) και ενσωματώθηκε προσανατολισμένα στο φορέα pcDNA4/TO στις θέσεις *XbaI* και *EcoRV* (δημιουργεί τυφλά άκρα όπως και η *PmeI*) της κασέτας κλωνοποίησής του (Εικ. 30). Το μήκος του cDNA του γονιδίου *DGCR14* μαζί με τους επίτοπους ήταν 1450bp.



**Εικ. 30: Χάρτης πλασμιδιακού φορέα pcDNA 4/TO.** Με κόκκινα βέλη φαίνονται οι θέσεις ενσωμάτωσης του cDNA του γονιδίου *DGCR14* και των επιτόπων του φορέα pcDNA3.1 (6xHis, Xpress). Ο φορέας έχει γονίδια ανθεκτικότητας στα αντιβιοτικά αμικιλίνη και ζεοσίνη σε κύτταρα βακτηρίων και θηλαστικών, αντίστοιχα.

### 2.5.3. Απομόνωση RNA

Η απομόνωση RNA από τα κύτταρα *HeLa* έγινε α) για την ανάλυση του μεταγραφικού προτύπου με τη μέθοδο των μικροσυστοιχιών και β) για την αντίστροφη μεταγραφή σε συνδυασμό με PCR ώστε να επιβεβαιωθεί η έκφραση του γονιδίου του καταστολέα της τετρακυκλίνης στους κλώνους των κυττάρων *HeLa*.

Στην πρώτη περίπτωση, η απομόνωση πραγματοποιήθηκε με κατάλληλο κιτ της εταιρίας QIAGEN (RNeasy micro kit, Cat. No 74004). Το πρωτόκολλο περιλαμβάνει τη λύση των κυττάρων και την επιλεκτική σύνδεση του RNA με μια μεμβράνη πυριτίου που βρίσκεται σε μια στήλη μέσω της προσθήκης ενός ειδικού διαλύματος (γουανιδίνης-θειοκυανικού οξέος). Η απομάκρυνση του DNA πραγματοποιείται με τη χρήση DNAσης και το RNA εκλούεται με νερό (RNase-free). Με αυτή τη διαδικασία απομονώνονται όλα τα μόρια RNA που έχουν μήκος μεγαλύτερο από 200 νουκλεοτίδια. Μικρότερου μήκους μόρια RNA, όπως το rRNA και tRNA συγκρατούνται στη στήλη. Η απόδοση του κιτ στην απομόνωση RNA από κύτταρα *HeLa* φτάνει στα 7,5μg ανά ~500.000 κύτταρα.

Στην δεύτερη περίπτωση η απομόνωση έγινε μέσω του αντιδραστηρίου TRI της εταιρίας Sigma (Cat. No T9424), σύμφωνα με τις οδηγίες του κατασκευαστή. Αρχικά προστέθηκε στους διαφορετικούς κλώνους κυττάρων (αριθμού ~10<sup>5</sup>) 1ml αντιδραστηρίου TRI για την επαναδιάλυσή τους και την επακόλουθη λύση τους. Το TRI περιέχει θειοκυανική γουανιδίνη (guanidinium thiocyanate) η οποία χρησιμεύει για την αποδιάταξη των πρωτεϊνών και για την απενεργοποίηση των ριβονουκλεασών και φαινόλη που συνεισφέρει στο διαχωρισμό του RNA από τα υπόλοιπα βιολογικά μακρομόρια (Rio et al. 2010). Έπειτα τα κύτταρα φυγοκεντρήθηκαν σε 12000g για 10min στους 4°C και το υπερκείμενο μεταφέρθηκε σε καθαρό σωληνάριο. Ακολούθησε επώαση σε θερμοκρασία δωματίου για 5min ώστε να διασπαστούν τα σύμπλοκα των νουκλεοπρωτεϊνών και προσθήκη 200μl χλωροφορμίου για τον διαχωρισμό των φάσεων. Στην υπερκείμενη φάση βρίσκεται το RNA, στην μεσόφαση το DNA και στην κατώτερη φάση οι πρωτεΐνες και τα λιπίδια. Το διάλυμα αναδεύτηκε



γρήγορα για 15sec, επώαστηκε σε θερμοκρασία δωματίου για 2-15min, και επαναλήφθηκε το βήμα της φυγοκέντρωσης. Στη συνέχεια προστέθηκε στο διάλυμα 0,5ml ισοπροπανόλης και ακολούθησε ανάδευση και επώαση σε θερμοκρασία δωματίου για 5-10min. Το διάλυμα φυγοκεντρήθηκε σε 12000g για 10min στους 4°C και το υπερκείμενο του διαλύματος αφαιρέθηκε. Έπειτα προστέθηκε αιθανόλη συγκέντρωσης 75% (1ml), ακολούθησε ανάδευση και φυγοκέντρωση σε 7500g για 5min στους 4°C. Τέλος, έγινε η απομάκρυνση του υπερκειμένου, η ξήρανση της πελλέτας του RNA, η επαναδιάλυσή της σε ~30μl νερό, και η αποθήκευση του δείγματος στους -80°C. Η απόδοση του πρωτοκόλλου φτάνει στα ~5μg ανά ~4-5x10<sup>5</sup> κύτταρα.

#### 2.5.3.1. Επεξεργασία του RNA με DNAση

Στα απομονωμένα δείγματα RNA από κάθε κλώνο κυττάρων που εκφράζουν τον καταστολέα της τετρακυκλίνης πραγματοποιήθηκε επεξεργασία με το ένζυμο DNAση μέσω της χρήσης κατάλληλου κιτ (DNA-free™ Kit, Cat. No AM1906, Ambion®/Thermo Fisher Scientific). Η επεξεργασία αυτή απαιτείται για την απαλλαγή του RNA από τυχόν γονιδιωματικό DNA που έχει απομείνει μετά τη διαδικασία απομόνωσης. Στα δείγματα RNA προστέθηκε 4μl ρυθμιστικό διάλυμα DNAσης (συγκέντρωσης 10x: 100 mM Tris-HCl pH 7.5, 25 mM MgCl<sub>2</sub>, 5 mM CaCl<sub>2</sub>), 1μl DNAσης (2 units/ποσότητα RNA έως 10μg) και ο τελικός όγκος του διαλύματος ρυθμίστηκε στα 40μl με νερό. Τα δείγματα επώαστηκαν στους 37°C για 30min και έπειτα προστέθηκαν σε αυτά 4μl από τον παράγοντα απενεργοποίησης της DNAσης (DNase Inactivation Reagent), ο οποίος, πέρα από την απενεργοποίηση της DNAσης χρησιμοποιείται και για την απομάκρυνση των δισθενών κατιόντων που υπάρχουν στο ρυθμιστικό διάλυμα DNAσης τα οποία μπορεί να επάγουν την αποικοδόμηση του RNA όταν θερμαίνεται. Στη συνέχεια ακολούθησε επώαση των δειγμάτων σε θερμοκρασία δωματίου με συνεχή ανάδευση για 3-4min και τα δείγματα φυγοκεντρήθηκαν στα 10.000g για 1,5min. Το υπερκείμενο των διαλυμάτων, που περιείχε το καθαρό πλέον RNA, συλλέχθηκε και αποθηκεύτηκε στους -80°C.

#### 2.5.3.2. Αντίστροφη μεταγραφή

Η αντίστροφη μεταγραφή πραγματοποιήθηκε για τη σύνθεση μονόκλωνου cDNA από τους κλώνους των κυττάρων που εκφράζουν τον καταστολέα της τετρακυκλίνης. Η διαδικασία που ακολουθήθηκε βασίστηκε στο πρωτόκολλο της εταιρίας Invitrogen. Σε 4μg ολικού RNA από κάθε κλώνο προστέθηκαν 250ng (1μl) εκκινητών (εξαμερή τυχαίας αλληλουχίας, 250ng/μl) σε τελικό όγκο 15μl. Τα δείγματα επώαστηκαν στους 70°C για 5min σε θερμοκυκλοποιητή PCR και έπειτα τοποθετήθηκαν σε πάγο για 5min. Σε κάθε δείγμα προστέθηκαν τα υπόλοιπα αντιδραστήρια: διάλυμα αντίστροφης μεταγραφής (συγκέντρωσης 5x: 250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM MgCl<sub>2</sub>) σε τελική συγκέντρωση 1x, μίγμα δεοξυριβονουκλεοτιδίων dNTPs (10mM) σε τελική συγκέντρωση 0,46mM, διθειοθρεϊτόλη (DTT, 0,1M) σε τελική συγκέντρωση 0,007M, 0,37 unit αναστολέα της RNAσης (RNase inhibitor, 40u/μl) και 5 units ενζύμου αντίστροφης μεταγραφής (200u/μl, SuperScript II Reverse transcriptase, Cat. No 18064-014, Invitrogen). Τέλος τα δείγματα επώαστηκαν στους 37°C για 1h σε

θερμοκυκλοποιητή PCR, και τα προϊόντα αντίστροφης μεταγραφής φυλάχθηκαν στους -80°C.

#### 2.5.4. Τεμαχισμός με περιοριστικές ενδονουκλεάσες

Η κλωνοποίηση DNA επιτυγχάνεται με την τεχνολογία του ανασυνδυασμένου DNA. Οι τεχνικές αυτές βασίζονται στη χρήση ειδικών ενζύμων που ονομάζονται περιοριστικές ενδονουκλεάσες. Τα ένζυμα αυτά απομονώθηκαν από διάφορους προκαρυωτικούς οργανισμούς, κυρίως βακτήρια, και η φυσιολογική τους λειτουργία είναι να προστατεύουν τους μικροοργανισμούς αυτούς από την εισβολή ξένου DNA μεθυλιώνοντας τις θέσεις στόχους των περιοριστικών ενδονουκλεασών. Έχουν την ιδιότητα να αναγνωρίζουν τετρα-, έξα- ή οκτανουκλεοτίδια του DNA και να το τεμαχίζουν αφήνοντας συμπληρωματικά άκρα (sticky ή cohesive ends) ή άκρα χωρίς αζευγάρωτες βάσεις (blunt ends) (Bloch & Grossmann 2001).

Στη συγκεκριμένη διατριβή χρησιμοποιήθηκαν: α) το ένζυμο *KpnI* (New England Biolabs, Cat No R0142S, 10units/μl) για τη γραμμοποίηση των διαγονιδίων pLKO.1-puro/shRNA που προορίζονταν για την καταστολή της έκφρασης του γονιδίου *FRA10AC1* και του αντίστοιχου διαγονιδίου αρνητικού ελέγχου β) το ένζυμο *AhdI* (New England Biolabs, Cat No R0584S, 20 unit/μl) για τη γραμμοποίηση του φορέα pcDNA6/TR, γ) το ένζυμο *PvuI* (New England Biolabs, Cat No R0150S, 20 unit/μl) για τη γραμμοποίηση του διαγονιδίου pcDNA4/TO-DGCR14 και δ) τα ένζυμα *XhoI* (New England Biolabs, Cat No R01465, 20 unit/μl) και *BamHI* (New England Biolabs, Cat No R01365, 20 unit/μl) για την υποκλωνοποίηση των υποψήφιων αλληλεπιδρώντων μορίων της πρωτεΐνης *FRA10AC1* στο φορέα κλωνοποίησης pmCherry-C1.

Οι πέψεις πραγματοποιήθηκαν με βάση τις άριστες συνθήκες στις οποίες λειτουργεί κάθε ένζυμο (θερμοκρασία, pH, μοριακότητα ρυθμιστικού διαλύματος σε άλατα). Ο τεμαχισμός του DNA πραγματοποιήθηκε με περίσσεια ενζύμου (3-5units/μg DNA) και με τη χρήση του ρυθμιστικού διαλύματος το οποίο ήταν συμβατό με το κάθε ένζυμο και είχε προμηθευτεί από την εταιρία μαζί με αυτό. Για το ένζυμο *KpnI* χρησιμοποιήθηκε το ρυθμιστικό διάλυμα NEBuffer 1.1 (B7201S), για το ένζυμο *BamHI* το CutSmart® Buffer (B7204S), για τα ένζυμα *PvuI*, *XhoI* και *BamHI* το NEBuffer 3.1 (B7203S) και για το *AhdI* το διάλυμα NEBuffer 4 (B7004S). Τα διαλύματα ήταν της εταιρίας New England Biolabs και χρησιμοποιήθηκαν σε τελική συγκέντρωση 1x. Στην περίπτωση του ενζύμου *AhdI* απαιτείται και προσθήκη BSA σε τελική συγκέντρωση 100μg/ml. Η επώαση με τα ένζυμα περιορισμού πραγματοποιήθηκε στους 37°C, ολονύκτια.

#### 2.5.5. Αντίδραση σύνδεσης

Για την αντίδραση σύνδεσης χρησιμοποιούνται τα ένζυμα λιγάσες. Τα ένζυμα αυτά διαφέρουν μεταξύ τους στην εξειδίκευση του υποστρώματος, στα είδη των συμπαραγόντων που απαιτούν και στη θερμική σταθερότητα (Lohman et al. 2011). Οι λιγάσες έχουν την ιδιότητα να επανασυνδέουν γραμμικά μόρια DNA που μπορεί να έχουν προέλθει από τον τεμαχισμό με ένζυμα περιορισμού ή με PCR δημιουργώντας ανασυνδυασμένα μόρια DNA που περιέχουν τμήματα από διαφορετικές πηγές

(Lohman et al. 2011). Για τη σύνδεση γραμμικών μορίων DNA χρησιμοποιείται η λιγάση DNA του βακτηριοφάγου T4, η οποία καταλύει το σχηματισμό φωσφοδιεστερικών δεσμών μεταξύ των νουκλεοτιδίων στο DNA (Western & Rose 1991).

Η αντίδραση σύνδεσης χρησιμοποιήθηκε για την υποκλωνοποίηση των cDNAs των υποψήφιων αλληλεπιδρώντων μορίων της πρωτεΐνης FRA10AC1 στο φορέα κλωνοποίησης pmCherry-C1. Συγκεκριμένα χρησιμοποιήθηκαν 1-3 units/μl της T4 DNA λιγάσης της εταιρίας Takara (Cat. No. 2011A, 350 units/μl) και το αντίστοιχο ρυθμιστικό διάλυμα 10x {10 mM Tris-HCl (pH 7.5), 50 mM KCl, 10 mM 2-mercaptoethanol, 0.1 mM EDTA, 50% glycerol} σε τελική συγκέντρωση 1x. Τα ενθέματα και ο φορέας κλωνοποίησης χρησιμοποιήθηκαν κατ' αντιστοιχία σε μοριακή αναλογία άκρων 1:1 ή 10:1 (προτεινόμενη αναλογία όταν χρησιμοποιούνται ένζυμα που δημιουργούν προεξέχοντα άκρα που μπορεί να είναι και συμβατά μεταξύ τους). Η συνολική ποσότητα DNA που χρησιμοποιήθηκε στην τελική αντίδραση κυμαινόταν από 1μg έως 10μg, σε τελικό όγκο 10μl. Η αντίδραση σύνδεσης επώαστηκε ολονύκτια στους 4°C. Οι αντιδράσεις σύνδεσης χρησιμοποιήθηκαν για το μετασχηματισμό βακτηρίων *E. coli*. Οι βακτηριακοί κλώνοι που αναπτύχθηκαν ελέγχθηκαν για την ενσωμάτωση των ανασυνδυασμένων ή μη πλασμιδίων. Ο έλεγχος έγινε με δοκιμαστικές πέψεις του πλασμιδιακού DNA, που απομονώθηκε από τους βακτηριακούς κλώνους, με τα ένζυμα περιορισμού τα οποία χρησιμοποιήθηκαν για την κλωνοποίηση ή με PCR μέσω της χρήσης εκκινητών (βλ. παρακάτω) για τα τρία γονίδια.

#### 2.5.6. Ηλεκτροφόρηση σε πηκτή αγαρόζης

Η ανίχνευση, ο έλεγχος ποιότητας και ο αδρός ποσοτικός προσδιορισμός μορίων DNA γίνεται με τη μέθοδο της ηλεκτροφόρησης (Lee et al. 2012). Η αρχή της μεθόδου βασίζεται στην ικανότητα μετακίνησης των μορίων DNA σε ένα ηλεκτρικό πεδίο, σύμφωνα με το φορτίο και το μέγεθός τους. Η ηλεκτροστατική δύναμη που αναπτύσσεται λόγω του ηλεκτρικού πεδίου κατευθύνει τα αρνητικά φορτισμένα μόρια προς τα ηλεκτρόδια της ανόδου της ηλεκτροφορητικής συσκευής, με ρυθμό αντιστρόφως ανάλογο του μοριακού τους βάρους. Ένας περιοριστικός παράγοντας της ταχύτητας μετακίνησης των μορίων, είναι το μέγεθος των πόρων της πηκτής. Όσο μεγαλύτερα είναι τα μόρια τόσο πιο αργά μετακινούνται διαμέσου μικρών πόρων. Η συγκέντρωση της αγαρόζης (0,8%-2,5%) (φυτικός πολυσακχαρίτης) που χρησιμοποιείται καθορίζει το μέγεθος των πόρων της πηκτής και συνεπώς ποικίλει ανάλογα με το μέγεθος και τις διαμορφώσεις (γραμμικό, κυκλικό, «nicked») του DNA που επιλέγεται να διαχωριστεί και να ανιχνευθεί κάθε φορά (Lee et al. 2012, Rill et al. 2002). Για την ανίχνευση του DNA χρησιμοποιείται βρωμιούχο αιθίδιο (EtBr), το οποίο έχει την ιδιότητα να παρεμβάλλεται μεταξύ των βάσεων και να καθιστά το DNA ορατό με έκθεση σε υπεριώδη ακτινοβολία. Η συγκέντρωση των πηκτών που χρησιμοποιήθηκαν κυμαινόταν από 1 έως 2% ανάλογα με το μέγεθος του DNA. Η διακριτική ικανότητα των πηκτών συγκέντρωσης 1-2% κυμαίνεται σε μεγέθη τμημάτων 50bp έως και 10kb. Για την παρασκευή της πηκτής αγαρόζης χρησιμοποιήθηκαν κατάλληλες ποσότητες αγαρόζης ανάλογα με τη σύσταση της

πηκτής, και ρυθμιστικό διάλυμα TAE 1x (40mM Tris, 20mM οξικό οξύ, 1mM EDTA pH=8). Το διάλυμα ζεσταίνεται σε φούρνο μικροκυμάτων μέχρι βρασμού ώστε να λιώσει η αγαρόζη και στη συνέχεια προστίθεται EtBr σε συγκέντρωση 0,1μg/ml, εφόσον έχει κρυώσει. Μετά τη στερεοποίηση, η πηκτή τοποθετείται σε συσκευή ηλεκτροφόρησης που περιέχει ρυθμιστικό διάλυμα 1xTAE. Για την τοποθέτηση των δειγμάτων στην πηκτή προστίθεται στο καθένα κατάλληλος όγκος διαλύματος φόρτωσης (6x loading buffer, 10 mM Tris-HCl (pH 7.6), 0.03% κυανό της βρωμοφαινόλης, 60% γλυκερόλη, 60 mM EDTA). Το κυανό της βρωμοφαινόλης χρησιμοποιείται για να παρακολουθείται η μετακίνηση των δειγμάτων στην πηκτή και η γλυκερόλη για να αποκτήσουν «βάρους» τα δείγματα και να μετακινηθούν στον πυθμένα των θέσεων της πηκτής. Η τάση που εφαρμόστηκε για τη μετακίνηση των τεμαχίων DNA ήταν 80-150V ανάλογα με το μέγεθός τους, τη συγκέντρωση της πηκτής και την απόσταση των ηλεκτροδίων.

Ο ποιοτικός έλεγχος της απομόνωσης RNA πραγματοποιήθηκε με ηλεκτροφόρηση σε πηκτή αγαρόζης 1%. Η διαδικασία είναι παρόμοια με την ηλεκτροφόρηση του DNA με μοναδική διαφορά τη χρήση ρυθμιστικού διαλύματος TBE 1x (89mM Tris, 89mM βορικό οξύ, 2mM EDTA pH=8). Πριν την τοποθέτηση των δειγμάτων στο πήκτωμα προηγήθηκε η θέρμανσή τους σε υδατόλουτρο στους 65°C για 15min για να παρεμποδιστεί ο σχηματισμός δευτεροταγών δομών του RNA, η μεταφορά τους σε πάγο και η στιγμιαία φυγοκέντρωσή τους.

#### 2.5.7. Απομόνωση τεμαχίων DNA και καθαρισμός προϊόντων PCR και πέψης

Η απομόνωση τεμαχίων DNA από πηκτή αγαρόζης προϋποθέτει τον επαρκή διαχωρισμό των εκάστοτε ζωνών στην πηκτή, έπειτα από την εφαρμογή χαμηλής ταχύτητας (30-50V) για τη μετακίνησή τους. Σε κάθε περίπτωση τα τεμάχια του DNA αποκόπηκαν από την πηκτή με τη χρήση καθαρής λεπίδας υπό υπεριώδες φως. Στη συνέχεια ακολούθησε ο καθαρισμός τους με κατάλληλο κιτ της εταιρίας Macherey-Nagel (Nucleospin Gel and PCR Clean-up, Cat. No 740609.250) που είναι κατάλληλο για καθαρισμό τμημάτων DNA μήκους 50 bp έως και ~ 20kb από πηκτή αγαρόζης. Στην περίπτωση καθαρισμού φορέων κλωνοποίησης, μετά την απομόνωση από την πηκτή ζυγίστηκαν τα τεμάχια που απομονώθηκαν, και ακολούθησε προσθήκη ενός διαλύματος του κιτ και επώαση στους 37°C μέχρι να λιώσει η αγαρόζη. Στη συνέχεια έγινε ο καθαρισμός σύμφωνα με το πρωτόκολλο της εταιρίας. Ο καθαρισμός των προϊόντων PCR έγινε απευθείας με το κιτ χωρίς να έχει προηγηθεί η απομόνωσή τους από πηκτή αγαρόζης. Το DNA συνδέεται σε μεμβράνη πυριτίου που είναι ειδική για τη σύνδεση μόνο των νουκλεϊκών οξέων μέσα σε μια στήλη (PCR Clean-up Column). Η ποσότητα του DNA που μπορεί να «συγκρατηθεί» από τη στήλη είναι μέχρι 25μg. Το DNA εκλύεται κάτω από συνθήκες χαμηλής αλατότητας με ένα ελαφρώς αλκαλικό διάλυμα. Ο όγκος του διαλύματος έκλυσης που χρησιμοποιήθηκε ήταν 10-60μl ανάλογα με το πόσο συγκεντρωμένο χρειαζόταν να είναι το DNA.

#### 2.5.8. Αλυσιδωτή αντίδραση πολυμεράσης (PCR)

Η μέθοδος της αλυσιδωτής αντίδρασης πολυμεράσης (PCR) επιτρέπει τον *in vitro* πολλαπλασιασμό αλληλουχιών DNA από ελάχιστες αρχικές ποσότητες δείγματος. Η

χρήση δύο ολιγονουκλεοτιδικών εκκινητών εκατέρωθεν της υπό εξέταση αλληλουχίας (5' και 3'), μετά από επαναλαμβανόμενους κύκλους αποδιάταξης του DNA, υβριδοποίησης των εκκινητών και επιμήκυνσης αυτών με τη βοήθεια μιας ειδικής θερμοανθεκτικής DNA πολυμεράσης, παρέχει τελικά την εκλεκτική ενίσχυση ενός τμήματος DNA (Garibyan & Avashia 2013, Valones et al. 2009). Η αποτελεσματικότητα της PCR καθορίζεται από την εξειδίκευση, την απόδοση και την ακρίβεια με την οποία πραγματοποιείται η ενίσχυση του DNA. Μία PCR υψηλής εξειδίκευσης ενισχύει μόνο ένα προϊόν, το επιθυμητό. Οι παράμετροι που καθορίζουν τα παραπάνω εξαρτώνται από διάφορους παράγοντες, όπως η θερμοκρασία υβριδοποίησης των εκκινητών, η συγκέντρωση ιόντων  $Mg^{+2}$ , το ρυθμιστικό διάλυμα του ενζύμου, η διάρκεια των κύκλων και η πολυμεράση (Garibyan & Avashia 2013, Valones et al. 2009).

Στην παρούσα διατριβή, η PCR χρησιμοποιήθηκε για: α) την ενίσχυση των cDNAs των τριών αλληλεπιδρώντων μορίων της FRA10AC1 ώστε να κλωνοποιηθούν σε κατάλληλο πλασμιδιακό φορέα, β) την ταυτοποίηση ανασυνδυασμένων αποικιών των μετασχηματισμένων βακτηριακών κυττάρων με τα παραπάνω διαγονίδια γ) την ενίσχυση τμήματος του cDNA του γονιδίου του καταστολέα TetR που είχε προκύψει από αντίστροφη μεταγραφή (RT-PCR) και δ) τον έλεγχο μυκοπλάσματος στο DNA των κυττάρων *HeLa* που χρησιμοποιήθηκαν.

Σε κάθε PCR από τις παραπάνω χρησιμοποιήθηκαν 100ng γενωμικού DNA κυττάρων *HeLa*, ~10-20% της αντίδρασης αντίστροφης μεταγραφής (cDNA από τους κλώνους κυττάρων που εκφράζουν τον TetR) ή 1ng πλασμιδιακού DNA. Η κάθε αντίδραση πραγματοποιήθηκε με το kit της εταιρίας Kapa Biosystems (KAPA Taq PCR Kit Cat. No. KK1015). Ειδικότερα χρησιμοποιήθηκαν: α) ρυθμιστικό διάλυμα αντίδρασης που περιέχει 15mM ιόντων  $Mg_2^+$  (10x) σε τελική συγκέντρωση 1,5mM, β) διάλυμα ιόντων  $Mg_2^+$  (25 mM) σε τελική συγκέντρωση 0.5mM, ώστε η συνολική τελική συγκέντρωση των ιόντων  $Mg_2^+$  να είναι 2mM και γ) Taq πολυμεράση (5 U/μl), που προέρχεται από το βακτήριο *Thermus aquaticus*, σε τελική συγκέντρωση 1unit/αντίδραση. Επιπλέον χρησιμοποιήθηκαν μίγμα δεοξυριβονουκλεοτιδίων dNTPs (50mM) σε τελική συγκέντρωση 0,2mM το καθένα και ζεύγος εκκινητών (εμπρόσθιος και ανάστροφος, 50 pmol/μl ο καθένας) για το εκάστοτε γονίδιο σε τελική συγκέντρωση 1pmol/μl ο καθένας (Eurofins Genomics). Ο τελικός όγκος της κάθε αντίδρασης ήταν 50μl.

Στην περίπτωση της ενίσχυσης των cDNAs των τριών αλληλεπιδρώντων μορίων της FRA10AC1 για την μετέπειτα κλωνοποίησή τους στον pmCherry-C1, τα ολιγονουκλεοτίδια-εκκινητές σχεδιάστηκαν έτσι ώστε να διαθέτουν αλληλουχίες αναγνώρισης για κατάλληλα περιοριστικά ένζυμα στο 5' άκρο τους, προκειμένου το τμήμα DNA που θα προκύψει να μπορεί στη συνέχεια να κλωνοποιηθεί στον φορέα (Πιν. 3). Αυτοί οι εκκινητές περιλάμβαναν επιπλέον έναν αριθμό νουκλεοτιδίων (5-6) στο 5' άκρο ώστε να αυξηθεί η αποτελεσματικότητα της πέψης. Στην περίπτωση ελέγχου μόλυνσης με μυκόπλασμα στο DNA των κυττάρων *HeLa* χρησιμοποιήθηκαν εκκινητές για: i) το DNA του μυκοπλάσματος {MGSO (εμπρόσθιος)- GPO-3 (ανάστροφος) σύμφωνα με τις ερευνητικές εργασίες (Young et al. 2010, Kong et al. 2001)} και ii) το γονίδιο της προθρομβίνης που αποτέλεσε το θετικό έλεγχο της PCR

λόγω της αποτελεσματικότητάς του και του παρόμοιου μεγέθους προϊόντος, που έδινε η ενίσχυσή του, με το προϊόν για το μυκόπλασμα {PRO\_F (εμπρόσθιος)- PRO\_R (ανάστροφος)} (Πιν. 3). Οι εκκινητές και η θερμοκρασία υβριδοποίησης που χρησιμοποιήθηκαν για το εκάστοτε γονίδιο φαίνονται στον Πίνακα 3.

**Πιν. 3: Αλληλουχίες εκκινητών.**

Όνομα	Αλληλουχία εκκινητών	Θερμοκρασία υβριδοποίησης (T <sub>m</sub> )
MFAP_F	5'-CAG <u>CGC TCG AGA TGT</u> CGG TCC CAA GCG CTC TC-3'	T <sub>m</sub> =64°C
MFAP_R	5'-CAG <u>CCG GAT CCT</u> AGT TTT CCG CTT CTT GGC AG 3'	T <sub>m</sub> =64°C
SAP30BP_F	5'-CAG <u>CGC TCG AGA TGG</u> CGG GGA AGA AGA ATG-3'	T <sub>m</sub> =62°C
SAP30BP_R:	5'-CAG <u>CCG GAT CCG</u> CTT GGC CTT CTT CAC AAT G-3'	T <sub>m</sub> =62°C
THOC1_F	5'-CAG <u>CGC TCG AGA TGT</u> CTC CGA CGC CGC CG-3'	T <sub>m</sub> =62°C
THOC1_R	5'-CAG <u>CCG GAT CCT</u> ATT TGT CTC ATT GTC ATT AGT TAG-3'	T <sub>m</sub> =55°C
MGSO	5'-CCR TGC ACC AYC TGT CWH HNB GWW AAC CTC-3'	T <sub>m</sub> =55°C
GPO-3	5'-GAA AGY GTG GGG AGC AAA YAG GAT TAG ATA CCC T-3'	T <sub>m</sub> =70°C
PRO_F	5'-TCT AGA AAC AGT TGC CTG GC-3'	T <sub>m</sub> =55°C
PRO_R	5'-ATA GCA CTG GGA GCA TTG AAG C-3'	T <sub>m</sub> =55°C
Tet_F1	5'-TGC CCT TTA GAA GGG GAA AG-3'	T <sub>m</sub> =62°C
Tet_R1	5'-CAC AGC GCT GAG TGC ATA TAA-3'	T <sub>m</sub> =62°C

Με υπογραμμισμένη και έντονη γραφή φαίνονται αντίστοιχα οι θέσεις αναγνώρισης των περιοριστικών ενζύμων και το κωδικόνιο έναρξης. Y:C ή T, W:A ή T, H:A ή C ή T

Οι συνθήκες της PCR που χρησιμοποιήθηκαν ήταν: α) αρχική αποδιάταξη του DNA με θέρμανση στους 95°C για 3min, β) αποδιάταξη του DNA στους 95°C για 15sec, με εξαίρεση την περίπτωση του μυκοπλάσματος που ήταν 45sec, γ) υβριδοποίηση των εκκινητών στο DNA για 15sec στη θερμοκρασία που φαίνεται στον Πίνακα 3, δ) επιμήκυνση του DNA από την Taq πολυμεράση στους 72 °C για 15sec στην περίπτωση του μυκοπλάσματος, της προθρομβίνης και του καταστολέα TetR, και για 1min όταν οι αλληλουχίες DNA που ενισχύονταν είχαν μεγαλύτερο μέγεθος, δηλαδή στην περίπτωση των cDNAs των γονιδίων *THOC1*, *MFAP1* και *SAP30BP*, ε) τελική επιμήκυνση του DNA για 10min. Οι κύκλοι ενίσχυσης σε κάθε αντίδραση ήταν 35 με εξαίρεση τις αντιδράσεις του μυκοπλάσματος και της προθρομβίνης στις οποίες ήταν 40.

Για να βρεθεί η κατάλληλη θερμοκρασία υβριδοποίησης για τους εκκινητές που χρησιμοποιήθηκαν για την κλωνοποίηση των cDNAs σε φορέα κλωνοποίησης, πραγματοποιήθηκε PCR διαβάθμισης θερμοκρασίας (gradient PCR) κατά την οποία δοκιμάστηκαν διαφορετικές θερμοκρασίες υβριδοποίησης των εκκινητών σε αντίγραφα της ίδιας αντίδρασης.

## 2.6. Απομόνωση πρωτεϊνών και ανάλυση western

Η ηλεκτροφορητική ανάλυση πρωτεϊνικών δειγμάτων μπορεί να παρέχει πληροφορίες σχετικά με τα μοριακά βάρη, τα ισοηλεκτρικά σημεία, και την ποσοτική εκπροσώπηση ορισμένων πρωτεϊνών. Η ευρέως χρησιμοποιούμενη μεθοδολογία για την ανίχνευση συγκεκριμένων πρωτεϊνών από σύνθετα δείγματα είναι η ανάλυση Western. Πρόκειται για μία αναλυτική μέθοδο που περιλαμβάνει τη μεταφορά των πρωτεϊνών-που έχουν διαχωριστεί ηλεκτροφορητικά σε μία αποδιατακτική πηκτή πολυακρυλαμίδης-σε ένα λεπτό και μεμβρανώδες υλικό και την επακόλουθη ανίχνευσή τους με κατάλληλα σημασμένα μονοκλωνικά ή πολυκλωνικά αντισώματα

(ανοσοανίχνευση) (Mahmood & Yang 2012). Η διαδικασία αυτή χρησιμοποιήθηκε για την επιλογή και τον έλεγχο των κλώνων υποέκφρασης του γονιδίου *FRA10AC1* που δημιουργήθηκαν, με τη χρήση αντισώματος έναντι της πρωτεΐνης *FRA10AC1* εξετάζοντας τα ποσοστά καταστολής της έκφρασης του γονιδίου. Επιπλέον, χρησιμοποιήθηκε για την επιλογή και τον έλεγχο των κλώνων υπερέκφρασης του γονιδίου *DGCR14*, μέσω της χρήσης αντισώματος έναντι της πρωτεΐνης *DGCR14*.

#### 2.6.1. Απομόνωση και ποσοτικοποίηση πρωτεϊνών

Η απομόνωση πρωτεϊνών (πυρηνικών, κυτταροπλασματικών, μεμβρανικών) από τα κύτταρα *HeLa* που χρησιμοποιήθηκαν έγινε με τη χρήση του διαλύματος 5x Laemmli (200 mM Tris-Cl pH 6,8, 8% SDS, 40% γλυκερόλη, 20% μερκαπτοαιθανόλη, 0,4% μπλε της βρωμοφαινόλης). Η μερκαπτοαιθανόλη που περιέχει το διάλυμα αυτό ανάγει τους δισουλφιδικούς δεσμούς που αναπτύσσονται ανάμεσα σε δύο ομάδες κυστεϊνών της ίδιας ή δύο διαφορετικών πολυπεπτιδικών αλυσίδων συνεισφέροντας στην αποδιάταξη των πρωτεϊνών. Η διαδικασία που ακολουθήθηκε περιλάμβανε την προσθήκη διαλύματος Laemmli ίσου όγκου (~50μl) με την πελλέτα των κυττάρων (~2x10<sup>6</sup> κύτταρα που προερχόταν από φλάσκα 25cm<sup>2</sup> κάλυψης 80%), το βρασμό των δειγμάτων στους 100°C για ~5-10min, την στιγμιαία φυγοκέντρωσή τους και την συλλογή του υπερκείμενου. Η επιλογή της χρήσης αυτού του διαλύματος έναντι άλλων διαλυμάτων (π.χ. διάλυμα RIPA) που χρησιμοποιούνται για την απομόνωση πρωτεϊνών έγινε διότι χρειάστηκε μικρότερος όγκος διαλύματος ώστε να μπορεί να τοποθετηθεί το κάθε δείγμα στις θέσεις υποδοχής της πηκτής πολυακρυλαμίδης και μεγαλύτερη συγκέντρωση πρωτεϊνών για να γίνει η ανίχνευση των πρωτεϊνών με ανάλυση western. Αντίθετα, το διάλυμα RIPA χρησιμοποιήθηκε για την ποσοτικοποίηση πρωτεϊνικών εκχυλισμάτων κυττάρων η οποία έγινε με τη μέθοδο Bradford. Ειδικότερα, απομονώθηκε αριθμός κυττάρων, 48h μετά την επίστρωση τρυβλίων με 10<sup>6</sup> κύτταρα που προορίζονταν για μεταβολομική ανάλυση, σε μορφή πελλέτας η οποία επαναδιαλύθηκε σε 200μl διάλυμα λύσης RIPA (RadioImmunoPrecipitation Assay buffer) {150 mM NaCl, 1.0% NP-40, 0.5% DOC (δεοξυχολικό νάτριο), 0.1% SDS (δωδεκυκλοθειικό νάτριο), 50 mM Tris, pH 8.0}. Τα δείγματα επώαστηκαν στους 4°C υπό ανάδευση, ολονύκτια, φυγοκεντρήθηκαν για 2-3min στις 13000rpm και συλλέχθηκε το υπερκείμενο.

Η μέθοδος Bradford βασίζεται στην πρόσδεση της μπλε χρωστικής Coomassie Brilliant G-250 με τις πρωτεΐνες και συγκεκριμένα με κατάλοιπα αργινίνης. Η χρωστική υπάρχει σε τρεις μορφές: κατιονική (κόκκινο), ουδέτερη (πράσινο) και ανιονική (μπλε). Υπό όξινες συνθήκες, η χρωστική είναι κατά κύριο λόγο στην πρωτονιωμένη ιονική μορφή (κόκκινο,  $A_{max} = 470 \text{ nm}$ ). Όταν δεσμεύεται με τις πρωτεΐνες, μετατρέπεται σε μια σταθερή μη πρωτονιωμένη μορφή (μπλε,  $A_{max} = 595 \text{ nm}$ ). Αυτή η μη πρωτονιωμένη μορφή της χρωστικής ανιχνεύεται στα 595nm και έτσι προσδιορίζεται η συγκέντρωση των πρωτεϊνών (Carlsson et al. 2011, Kruger 1994). Αρχικά παρασκευάστηκαν διαδοχικό εύρος συγκεντρώσεων (1-20 μg/ml) της πρωτεΐνης BSA (Bovine serum albumin) για την κατασκευή της πρότυπης καμπύλης. Στη συνέχεια, για το εκάστοτε δείγμα προετοιμάστηκαν διαφορετικές αραιώσεις (1:5, 1:10, 1:20) και σε αυτές προστέθηκε η χρωστική Coomassie Brilliant G-250 (Bio-Rad,

Cat. No 500-0006). Η αναλογία όγκου δείγματος και χρωστικής Bradford ήταν 1:5. Ακολούθησε ανάμειξη με ανάδευση των δειγμάτων και επώασή τους για 10min σε θερμοκρασία δωματίου. Επιλέχθηκε για φωτομέτρηση το/α δείγματα που είχαν χρώμα στο εύρος των χρωμάτων των διαφορετικών συγκεντρώσεων BSA. Ακολούθησε η φωτομέτρηση των διαφορετικών διαλυμάτων BSA και των υπό μελέτη δειγμάτων σε μήκος κύματος 595nm. Η πρότυπη καμπύλη σχεδιάστηκε σύμφωνα με τις τιμές οπτικής πυκνότητας των διαφορετικών συγκεντρώσεων BSA και προσδιορίστηκε η εξίσωση της μορφής  $y=ax-b$  (όπου  $x$  η συγκέντρωση της πρωτεΐνης και  $y$  οι τιμές της οπτικής πυκνότητας). Με βάση την εξίσωση προσδιορίστηκαν οι συγκεντρώσεις του πρωτεϊνικού εκχυλίσματος των υπό εξέταση δειγμάτων.

### 2.6.2. Ηλεκτροφόρηση πρωτεϊνών σε αποδιατακτική πηκτή πολυακρυλαμίδης

Η ηλεκτροφόρηση πρωτεϊνών σε πηκτή πολυακρυλαμίδης υπό αποδιατακτικές συνθήκες παρουσία SDS (Sodium Dodecyl Sulfate-PolyAcrylamide Gel Electrophoresis, SDS-PAGE) είναι η πιο διαδεδομένη μέθοδος που χρησιμοποιείται στη βιοχημεία, στην κυτταρική βιολογία και στην ανοσολογία για το διαχωρισμό ενός μίγματος πρωτεϊνών από ιστούς ή κυτταρικά εκχυλίσματα με βάση το μοριακό τους βάρος (Rath et al. 2013, Smith 1984). Η πολυακρυλαμίδη παρουσιάζει αρκετά πλεονεκτήματα, όπως είναι η χημική αδράνεια, η αντοχή σε υψηλή θερμοκρασία και η μεγάλη διακριτική ικανότητα των αποτελεσμάτων των ηλεκτροφορήσεων. Οι αποδιατακτικοί παράγοντες που χρησιμοποιούνται είναι το ανιονικό απορρυπαντικό SDS (δωδεκακυλοθειϊκό νάτριο) και η μερκαπτοαιθανόλη (Rath et al. 2013). Το SDS έχει διττό ρόλο: α) διασπά τους υδρογονικούς δεσμούς και τις υδρόφοβες αλληλεπιδράσεις αποδιατάσσοντας με αυτόν τον τρόπο τις πρωτεΐνες και προσδίδοντας σε αυτές αρνητικό φορτίο και β) δεσμεύεται στα αμινοξέα της πρωτεΐνης καταστρέφοντας τη φυσική δομή της, με αποτέλεσμα η πρωτεΐνη να παίρνει την μορφή ελάσματος και να δημιουργούνται επιμήκη μόρια, αρνητικά φορτισμένα (Rath et al. 2013, Gallagher 2006).

Ο σχηματισμός της πηκτής πολυακρυλαμίδης επιτυγχάνεται με τον πολυμερισμό των μονομερών της πολυακρυλαμίδης και της N,N-μεθυλενο-δισ-ακρυλαμίδης. Η N,N-μεθυλενο-δισ-ακρυλαμίδη συντελεί στο σχηματισμό γεφυρών μεταξύ των αλυσίδων των μονομερών της ακρυλαμίδης. Ο πολυμερισμός πραγματοποιείται με τη βοήθεια δύο πολυμεριστικών παραγόντων, του υπερθειϊκού αμμωνίου (ammonium persulfate, APS) που χρησιμοποιείται ως δότης των ελευθέρων ριζών, και του φωτοχημικού καταλύτη N,N,N,N-τετραμεθυλο-1,2-διαμινο-αιθάνιου (TEMED) ο οποίος καταλύει το σχηματισμό ελευθέρων ριζών από το APS. Κατ' αυτόν τον τρόπο, δημιουργείται ένα τρισδιάστατο πολυμερές πλέγμα μέσα στο οποίο κινούνται οι πρωτεΐνες, με μέγεθος πόρων που είναι αντιστρόφως ανάλογο της συγκέντρωσης των μονομερών της ακρυλαμίδης. Γενικά, πηκτές με μικρή συγκέντρωση πολυακρυλαμίδης έχουν μεγαλύτερους πόρους και το αντίστροφο. Οι μικρές σε μέγεθος πολυπεπτιδικές αλυσίδες μετακινούνται εύκολα διά μέσου των πόρων της πηκτής ενώ οι μεγαλύτερες μένουν στην κορυφή, κοντά στο σημείο εκκίνησης.

Το «καλούπι» στο οποίο σχηματίστηκε η πηκτή ήταν ένα «σάντουιτς» επίπεδων τζαμιών. Η πηκτή αποτελείται από δύο φάσεις: την ανώτερη (stacking gel) στην οποία



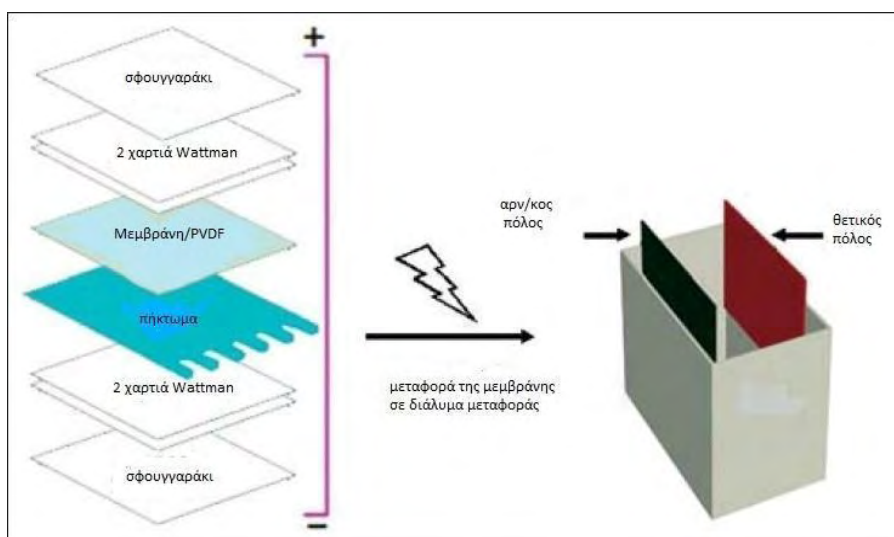
πραγματοποιείται η ευθυγράμμιση των πρωτεϊνών σε ενιαίο μέτωπο ώστε να μπορούν να φτάσουν ταυτόχρονα στην πηκτή διαχωρισμού που αποτελεί την υποκείμενη φάση (separating gel). Πρώτα παρασκευάζεται η πηκτή διαχωρισμού, η οποία στην παρούσα εργασία ήταν συγκέντρωσης 10%. Η σύσταση του διαλύματος, από το οποίο δημιουργείται η πηκτή είναι 10% (w/v) πολυακρυλαμίδη και N,N-μεθυλενο-δισ-ακρυλαμίδη (αναλογία 29:1 w/v), 0,56M Tris-Cl pH 8,8 (αρχική συγκέντρωση 1M), 1,5% (w/v) SDS (αρχική συγκέντρωση 20%), 1,5% (w/v) APS, 0,007% (v/v) TEMED. Το διάλυμα τοποθετείται στο καλούπι και κατευθείαν μετά την τοποθέτησή του προστίθεται ~1ml ισοπροπανόλης ώστε το πάνω μέρος του να ευθυγραμμιστεί. Στη συνέχεια παραμένει σε θερμοκρασία δωματίου για ~40min ώστε να πολυμεριστεί. Εφόσον έχει γίνει ο πολυμερισμός της πηκτής διαχωρισμού, αφαιρείται η ισοπροπανόλη, ακολουθεί έκπλυση με νερό και απομακρύνονται τυχόν υπολείμματα με διηθητικό χαρτί. Στη συνέχεια παρασκευάζεται η πηκτή ευθυγράμμισης, συγκέντρωσης 4,5%. Η σύσταση του διαλύματος είναι 4,5% (w/v) πολυακρυλαμίδη και N,N-μεθυλενο-δισ-ακρυλαμίδη (αναλογία 29:1), 0,125M Tris-Cl pH 8,8 (αρχική συγκέντρωση 1M), 0,1% (w/v) SDS (αρχική συγκέντρωση 20%), 0,1% (w/v) APS (αρχική συγκέντρωση 20%), 0,198% (v/v) TEMED. Η συσκευή της Biorad που χρησιμοποιήθηκε χρειάζεται 15ml για την πηκτή διαχωρισμού και 10ml για την πηκτή ευθυγράμμισης. Το διάλυμα τοποθετείται στο καλούπι και αμέσως μετά, η ειδική «χτένα», ώστε να σχηματιστούν οι θέσεις υποδοχής για τα δείγματα. Το διάλυμα παραμένει σε θερμοκρασία δωματίου για ~15min για να πολυμεριστεί. Η τοποθέτηση των δειγμάτων γίνεται στις θέσεις υποδοχής της πηκτής ευθυγράμμισης. Η ηλεκτροφόρηση πραγματοποιήθηκε με σταθερή ένταση πεδίου 12-25 mA σε ρυθμιστικό διάλυμα Tris-glycine 1x από αρχική συγκέντρωση 10x (250mM Tris, 1,92M γλυκίνη, 1% SDS) σε pH ρυθμισμένο στο 8,3 με πυκνό διάλυμα υδροχλωρίου (HCl). Παράλληλα με τα προς εξέταση δείγματα ηλεκτροφορήθηκε και ένας μάρτυρας γνωστών μοριακών μεγεθών (Nippon Genetics Cat. No MWD03).

### 2.6.3. Ηλεκτρομεταφορά πρωτεϊνών σε μεμβράνη

Έπειτα από την ηλεκτροφόρηση των πρωτεϊνών σε αποδιατακτική πηκτή ακολουθεί είτε χρώση με ειδικές χρωστικές ουσίες (π.χ. Coomassie) όταν χρειάζεται να γίνει παρατήρηση όλων των πρωτεϊνών είτε ηλεκτρομεταφορά των πρωτεϊνών σε περίπτωση που ακολουθεί ανάλυση western. Η μεταφορά των πρωτεϊνών πραγματοποιείται από την πηκτή πολυακρυλαμίδης σε μεμβράνη νιτροκυτταρίνης ή πολυβινυλιδενικού φθοριδίου (PVDF)-στη διατριβή χρησιμοποιήθηκε η δεύτερη (Macherey-Nagel Cat. No 741260)-με την εφαρμογή διαφοράς δυναμικού (Mahmood & Yang 2012). Οι πρωτεΐνες βρίσκονται με τη μορφή συμπλόκου με το SDS, το οποίο είναι αρνητικά φορτισμένο. Με την εφαρμογή ηλεκτρικού πεδίου το σύμπλοκο κινείται προς την άνοδο, εξέρχεται από την πηκτή και τελικά καθιλώνεται στο πλέγμα της μεμβράνης, εξαιτίας υδρόφοβων αλληλεπιδράσεων (Mahmood & Yang 2012).

Η διαδικασία μεταφοράς των πρωτεϊνών σε μεμβράνη περιλαμβάνει την αφαίρεση του πηκτώματος διαχωρισμού από το καλούπι στο οποίο ηλεκτροφορήθηκε, την μέτρηση των διαστάσεων του και την μεταφορά του σε διάλυμα ηλεκτρομεταφοράς (transfer buffer). Στη συνέχεια, στις διαστάσεις του πηκτώματος κόβονται τέσσερα

κομμάτια χαρτιού Wattman 3MM που τοποθετούνται σε διάλυμα ηλεκτρομεταφοράς και ένα κομμάτι μεμβράνης το οποίο, έπειτα από την εξισορρόπησή του σε μεθανόλη για 5sec, τοποθετείται και αυτό σε διάλυμα ηλεκτρομεταφοράς για 10min. Τέλος, στην κασέτα της ηλεκτρομεταφοράς (gel holder cassette) τοποθετούνται με τη σειρά: στον αρνητικό πόλο, ένα σφουγγάρι (sponge pad) το οποίο έχει εμποτιστεί σε νερό (dH<sub>2</sub>O), 2 χαρτιά Wattman 3MM, το πήκτωμα, η μεμβράνη PVDF, 2 χαρτιά Wattman 3MM και στο θετικό πόλο ένα σφουγγάρι (sponge pad) το οποίο έχει επίσης εμποτιστεί σε νερό, όπως φαίνεται και στην Εικόνα 31.



**Εικ. 31: Διαδικασία ηλεκτρομεταφοράς πρωτεϊνών σε μεμβράνη (προσαρμοσμένη από Mahmood & Yang, 2012).**

Η παραπάνω διάταξη τοποθετείται σε συσκευή ηλεκτροφόρησης και ηλεκτροφορείται σε διάλυμα ηλεκτρομεταφοράς 1x από πυκνό διάλυμα 10x το οποίο είναι ίδιο με το διάλυμα ηλεκτροφόρησης με τη διαφορά ότι περιέχει επιπλέον 20% μεθανόλη, στους 4°C σε σταθερή ένταση πεδίου 110mA ολονύκτια.

#### 2.6.4. Ανοσοαποτύπωση

Μετά το πέρας της ηλεκτρομεταφοράς των πρωτεϊνών ακολουθεί επώαση της μεμβράνης με διάλυμα παρεμπόδισης {blocking buffer: TBSTM (137mM NaCl, 20mM Tris, 0,1% Tween-20, 5% αποβουτυρωμένο γάλα-σκόνη) στην περίπτωση των αντισωμάτων α-FRA10AC1 και α-DGCR14 ή PBSTM (διάλυμα φωσφορικών αλάτων, KH<sub>2</sub>PO<sub>4</sub> 144mg/L, NaCl 9g/L, Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O 795mg/L, 0,1% Tween-20, 5% αποβουτυρωμένο γάλα-σκόνη) για τα αντισώματα της α-ακτίνης και της γ-τουμπουλίνης για την αποφυγή της μη ειδικής σύνδεσης των αντισωμάτων διότι οι πρωτεΐνες που μεταφέρονται από το πήκτωμα δεν είναι αρκετές για να καλύψουν όλες τις κενές θέσεις της μεμβράνης με αποτέλεσμα να καλύπτονται οι κενές θέσεις με πρωτεΐνες που προέρχονται από άλλη πηγή (όπως τον ορό του γάλακτος). Στη συνέχεια, αφού πραγματοποιηθεί έκπλυση με διάλυμα (TBST) που περιέχει μη ιονικό απορρυπαντικό (0,1% Tween-20) για 5min, ακολουθεί επώαση με ειδικό πρωτεΐον αντίσωμα για την εκάστοτε πρωτεΐνη ολονύκτια στους 4°C, τρεις εκπλύσεις με TBST

ή PBST για 10min η καθεμία. Στη συνέχεια ακολουθεί επώαση με το δευτερεύον αντίσωμα-το οποίο είναι συζευγμένο με το ένζυμο υπεροξειδάση-για μία ώρα σε θερμοκρασία δωματίου και πάλι τρεις εκπλύσεις με TBST ή PBST για 10min η καθεμία. Και στις δυο περιπτώσεις χρησιμοποιήθηκε κατάλληλη αραίωση αντισώματος σε TBSTM ή PBSTM. Τα αντισώματα που χρησιμοποιήθηκαν σε αυτήν την εργασία ήταν: α) αντίσωμα έναντι της πρωτεΐνης FRA10AC1 (αραίωση 1:500) το οποίο είχε απομονωθεί από πολυκλωνικό ορό κουνελιού (μεταπτυχιακή διατριβή: κ. Ε. Κατσαμόρα, Τμήμα Βιοχημείας & Βιοτεχνολογίας, Πανεπιστήμιο Θεσσαλίας, 2011) β) ορός που προέρχεται από κουνέλι ανοσοποιημένο με πρωτεΐνη DGCR14 σε τελική αραίωση 1:10000 και γ) αντισώματα α-ακτίνης (αραίωση 1:1000) (Cat. No MAB1501, Chemicon), και γ-τουμπουλίνης (αραίωση 1:2500, Cat. No T6557, Sigma) που χρησιμοποιήθηκαν για την κανονικοποίηση της ποσότητας του πρωτεϊνικού εκχυλίσματος που εξετάστηκε κάθε φορά. Το δευτερεύον αντίσωμα προερχόταν κάθε φορά από διαφορετικό είδος οργανισμού, από κουνέλι (αραίωση 1:15000, Cat. No 7074P2, Cell Signalling) στην περίπτωση των πρωτευόντων αντισωμάτων α-FRA10AC1 και α-DGCR14 ή από ποντικό (αραίωση 1:1000, Cat. No Cat. No 7076P2, Cell Signalling) στις περιπτώσεις της α-ακτίνης και γ-τουμπουλίνης.

#### 2.6.5. Ανοσοανίχνευση

Στη παρούσα εργασία, η ανοσοανίχνευση των πρωτεϊνών πραγματοποιήθηκε μέσω αντίδρασης χημειοφωταύγειας. Ως χημειοφωταύγεια ορίζεται η εκπομπή φωτός ως αποτέλεσμα μιας χημικής αντίδρασης (Chen et al. 2008, Mruk & Cheng 2011). Είναι μια αρκετά ευαίσθητη τεχνική κατά την οποία ένα δευτερεύον αντίσωμα που είναι ειδικό για το πρωτεύον αντίσωμα έναντι της πρωτεΐνης που εξετάζεται κάθε φορά, είναι συζευγμένο ομοιοπολικά με ένα ένζυμο-δείκτη (συγκεκριμένα σε αυτή την εργασία την υπεροξειδάση) το οποίο αντιδρώντας με το εξωγενώς προστιθέμενο υπόστρωμα, δημιουργεί προϊόν, εκπέμποντας φως. Η ανίχνευση των πρωτεϊνών πραγματοποιήθηκε με το Chemiluminescent HRP Substrate LumiSensor kit (GenScript, Cat. No. L00221V300) και η ανίχνευση σήματος έγινε μέσω του μηχανήματος FluorChem™ E System της εταιρίας Protein Simple. Η ποσοτικοποίηση των δειγμάτων και η κανονικοποίησή τους ως προς θετικό έλεγχο του πειράματος έγινε με το πρόγραμμα Image J (Schneider et al. 2012). Έπειτα από την κανονικοποίηση, συγκρίθηκαν οι τιμές των δειγμάτων μεταξύ τους για να ελεγχθεί η υποέκφραση της πρωτεΐνης FRA10AC1 στους κλώνους των κυττάρων που είχαν δημιουργηθεί ή η υπερέκφραση της πρωτεΐνης DGCR14. Κατά τη διαδικασία ποσοτικοποίησης των δειγμάτων στο πρόγραμμα αυτό, όταν γινόταν η μέτρηση κάθε δείγματος ξεχωριστά τα αποτελέσματα δεν ήταν επαναλήψιμα στις διαφορετικές μετρήσεις του ίδιου δείγματος, ενώ όταν μετρήθηκαν όλα τα δείγματα μαζί σε σειρά, οι τιμές ήταν επαναλήψιμες. Γι' αυτό το λόγο επιλέχθηκε ο δεύτερος τρόπος.

### **3. ΑΠΟΤΕΛΕΣΜΑΤΑ**

#### **3.1. *In silico* ανακατασκευή δικτύου άμεσων πρωτεϊνικών αλληλεπιδράσεων της πρωτεΐνης FRA10AC1 του ανθρώπου**

Ο προσδιορισμός των πρωτεϊνικών αλληλεπιδράσεων (Protein-Protein Interactions, PPIs) και των συμπλόκων που σχηματίζουν καθώς και η δυναμική τους όσον αφορά το χώρο και το χρόνο σε επίπεδο κυττάρου και οργανισμού είναι σημαντική για τη διερεύνηση της πρωτεϊνικής λειτουργίας καθώς και των βιοχημικών μονοπατιών που συμμετέχουν οι πρωτεΐνες. Επίσης, μπορεί να συμβάλλουν στην κατανόηση της συντονισμένης συνεργασίας των πρωτεϊνών οι οποίες συμμετέχουν σε διαφορετικές βιολογικές διαδικασίες και κατ' επέκταση της κυτταρικής λειτουργίας. Αυτές οι πληροφορίες είναι ακόμη πιο απαραίτητες στην περίπτωση που οι πρωτεΐνες δεν διαθέτουν κανένα καλά χαρακτηρισμένο μοτίβο ή δεν υπάρχει καμία λειτουργική πληροφορία γι' αυτές από τα αντίστοιχα ορθόλογα μόρια.

Η πρωτεΐνη FRA10AC1 του ανθρώπου δεν έχει χαρακτηρισμένο λειτουργικό ρόλο, και δεν διαθέτει κάποιο καλά συντηρημένο μοτίβο, εκτός από το σήμα πυρηνικού εντοπισμού, ωστόσο, το ορθόλογο μόριο της στο φύκος *Chlamydomonas reinhardtii* έχει δείχθει ότι συμμετέχει στην αναγνώριση των θέσεων συναρμογής. Η λειτουργία της FRA10AC1 στη διαδικασία της συναρμογής φαίνεται να είναι συντηρημένη καθώς η λειτουργία του ορθολόγου της εξετάστηκε σε μονοκύτταρο οργανισμό. Επιπλέον, η συμμετοχή της στη διαδικασία επεξεργασίας του mRNA υποδεικνύεται και από: α) την απομόνωσή της ως συστατικό των υποσυμπλόκων του σωματίου συναρμογής, β) τις αλληλεπιδράσεις της με δυο πρωτεΐνες που έχουν τεκμηριωμένο ρόλο στη συναρμογή, η ταυτοποίηση των οποίων πραγματοποιήθηκε με εναλλακτικές πειραματικές μεθόδους, γ) την πυρηνική υποκυτταρική τοπολογία της και δ) την ευρεία έκφρασή της (βλ. Ενότητα 1.3.1.).

Αξιοποιώντας τον μεγάλο όγκο πληροφοριών που είναι διαθέσιμος για τις πρωτεϊνικές αλληλεπιδράσεις είτε στις βάσεις δεδομένων ή σε μελέτες υψηλής απόδοσης, έγινε η αναζήτηση όλων των πρωτεϊνικών αλληλεπιδράσεων της FRA10AC1 ώστε να ανακατασκευαστεί *in silico* το πλήρες δίκτυο πρωτεϊνικών αλληλεπιδράσεών της. Η ανακατασκευή αυτού του δικτύου είχε ως στόχο την κατανόηση του λειτουργικού ρόλου της FRA10AC1, διερευνώντας τους πρώτους «γείτονές» της, και τις κυτταρικές διεργασίες στις οποίες αυτές συμμετέχουν δεδομένης της συνεργασίας των πρωτεϊνών μέσω αλληλεπιδράσεων για τη διεκπεραίωση μιας κυτταρικής λειτουργίας.

##### **3.1.1. Συλλογή πρωτεϊνικών αλληλεπιδράσεων από βάσεις δεδομένων**

Για την ανακατασκευή του δικτύου συλλέχθηκαν δεδομένα που αφορούσαν μόνο τις άμεσες πρωτεϊνικές αλληλεπιδράσεις της FRA10AC1 από πρωτογενείς βάσεις δεδομένων πρωτεϊνικών αλληλεπιδράσεων που στηρίζονται στη βιβλιογραφία τόσο στον άνθρωπο όσο και σε οργανισμούς-μοντέλα. Οι άμεσες αλληλεπιδράσεις λαμβάνουν χώρα μεταξύ ενός ζεύγους πρωτεϊνών χωρίς την μεσολάβηση μιας τρίτης πρωτεΐνης. Η επιλογή μόνο των άμεσων αλληλεπιδράσεων έγινε με στόχο τη διερεύνηση του λειτουργικού ρόλου των πρωτεϊνών που συνδέονται άμεσα με τη

FRA10AC1, ώστε να μπορούν να διεξαχθούν βιολογικά συμπεράσματα σχετικά με τις κυτταρικές διαδικασίες στις οποίες η FRA10AC1 μπορεί πιθανώς να συμμετέχει. Συγκεκριμένα, τα δεδομένα για τον άνθρωπο προήλθαν από τη μετα-βάση PICKLE (Klара et al. 2013, Gioutlakis et al. 2017) η οποία περιλαμβάνει πληροφορίες από πέντε βάσεις δεδομένων του ανθρώπου, HPRD, BioGRID, IntAct, MINT και DIP που στηρίζονται στη βιβλιογραφία (βλ. Υλικά & Μέθοδοι). Επιπλέον, πληροφορίες ανακτήθηκαν και από τις βάσεις δεδομένων DroID (<http://www.droidb.org/>) και Worm Interactome Database ([http://interactome.dfci.harvard.edu/C\\_elegans/](http://interactome.dfci.harvard.edu/C_elegans/)) για τους οργανισμούς-μοντέλα *D. melanogaster* και *C. elegans*, αντίστοιχα. Η επιλογή τους στηρίχτηκε στο σημαντικό αριθμό πρωτεϊνικών αλληλεπιδράσεων που είναι διαθέσιμος σε αυτές και στην ιδέα των ιντερόλογων (interologs) η οποία υποστηρίζει ότι όταν δυο πρωτεΐνες αλληλεπιδρούν σε ένα είδος, αυτή η αλληλεπίδραση πιθανώς είναι εξελικτικά συντηρημένη, με αποτέλεσμα την αλληλεπίδραση των αντίστοιχων ορθόλογων μορίων και σε άλλα είδη (Nguyen et al. 2013). Διαθέσιμα δεδομένα για τις πρωτεϊνικές αλληλεπιδράσεις της FRA10AC1 υπήρχαν μόνο στον οργανισμό *D. melanogaster* (Πιν. 4). Από όλες τις παραπάνω βάσεις δεδομένων ταυτοποιήθηκαν είκοσι επτά αλληλεπιδρώντα μόρια της FRA10AC1 (Πιν. 4). Οι αλληλεπιδράσεις αυτές προήλθαν από τρία πειράματα μεγάλης κλίμακας που αφορούν πρωτεΐνες του ανθρώπου (Hegele et al. 2012, Stelzl et al. 2005, Rolland et al. 2014) και ένα που αφορά τον οργανισμό μοντέλο *D. melanogaster* (Giot et al. 2003). Οι μελέτες αυτές περιλάμβαναν δοκιμές ελέγχου με το σύστημα δυο υβριδίων στο σακχαρομύκητα και πειράματα συν-ανοσοκατακρήμνισης των ανασυνδυασμένων υποψήφιων μορίων (Hegele et al. 2012). Δυο από τις 27 αλληλεπιδράσεις (FRA10AC1-DGCR14, FRA10AC1-SF3B2) ταυτοποιήθηκαν και από μέλη του εργαστηρίου μας με το σύστημα δυο υβριδίων στο σακχαρομύκητα και με συνκατακρήμνιση (GST pull-down assay) (Πιν. 4), και η αλληλεπίδραση του ζεύγους FRA10AC1-DGCR14 ταυτοποιήθηκε και με συν-ανοσοκατακρήμνιση των ενδογενών μορίων (Πιν. 4) (Sarafidou et al. υπό συγγραφή). Στις παραπάνω βάσεις δεδομένων ελέγχθηκαν και οι άμεσες αλληλεπιδράσεις μεταξύ των 27 αλληλεπιδρώντων μορίων της FRA10AC1. Από τους πρώτους «γείτονες» της FRA10AC1, οι 8 έχουν την ικανότητα σύνδεσης με το RNA (Πιν. 4). Ειδικότερα, η πλειοψηφία αυτών (6) συνδέεται με το mRNA, ενώ η NKAP και η DGCR14 με τα snRNAs. Στον Πίνακα 4 φαίνεται από ποια βάση δεδομένων ανακτήθηκε η κάθε αλληλεπίδραση και η αντίστοιχη βιβλιογραφία. Επίσης, παρουσιάζονται οι αλληλεπιδράσεις μεταξύ των πρώτων γειτόνων της πρωτεΐνης FRA10AC1, η βάση δεδομένων από την οποία προήλθαν, και η αντίστοιχη ερευνητική μελέτη, και οι πρωτεΐνες που συνδέονται με RNA.

Πιν. 4: Οι πρώτοι «γείτονες» της πρωτεΐνης FRA10AC1 και οι μεταξύ τους αλληλεπιδράσεις.

Πρώτοι Γείτονες				PICKLE						
Όνομα πρωτεΐνης 1	Κωδικός Πρωτεΐνης (UniProt ID)	Όνομα πρωτεΐνης 2	Κωδικός Πρωτεΐνης (UniProt ID)	HPRD	BIOGRID	INTACT	MINT	DIP	DroID	Βιβλιογραφία (PubMed ID)
FRA10AC1	Q70Z53	PRPF40A	Q75400		√		√			22365833
FRA10AC1	Q70Z53	SAP30BP	Q9UHR5		√		√			22365833
FRA10AC1	Q70Z53	**PRPF3	Q43395		√		√			22365833
FRA10AC1	Q70Z53	**U2AF1	Q01081		√		√			22365833
FRA10AC1	Q70Z53	IK	Q13123		√		√			22365833
FRA10AC1	Q70Z53	**DGCR14*	Q96DF8		√		√			22365833
FRA10AC1	Q70Z53	DGCR14	Q96DF8						√	14605208
FRA10AC1	Q70Z53	THOC1	Q96FV9		√		√			22365833
FRA10AC1	Q70Z53	ZCCHC10	Q8TBK6		√		√			22365833
FRA10AC1	Q70Z53	**NKAP	Q8N5F7		√		√			22365833
FRA10AC1	Q70Z53	**CHERP	Q8IWX8		√		√			22365833
FRA10AC1	Q70Z53	**MFAP1	P55081		√		√			22365833
FRA10AC1	Q70Z53	CWC27	Q6UX04		√		√			22365833
FRA10AC1	Q70Z53	ZNF830	Q96NB3		√		√			22365833
FRA10AC1	Q70Z53	TTC14	Q96N46		√		√			22365833
FRA10AC1	Q70Z53	**SF3B2*	Q13435		√		√			22365833
FRA10AC1	Q70Z53	ADPRHL2	Q9NX46	√	√					16169070
FRA10AC1	Q70Z53	CCDC155	Q8N6L0		√					25416956
FRA10AC1	Q70Z53	EEF1D	P29692	√	√					16169070
FRA10AC1	Q70Z53	TRIM41	Q8WV44		√					25416956
FRA10AC1	Q70Z53	HABP4	Q5JVS0	√						16169070
FRA10AC1	Q70Z53	C7orf25	Q9BPX7	√	√					16169070
FRA10AC1	Q70Z53	MAP3K12	Q12852	√	√					16169070
FRA10AC1	Q70Z53	VPS29	Q9UBQ0	√	√					16169070
FRA10AC1	Q70Z53	GDPD2	Q9HCC8	√						16169070
FRA10AC1	Q70Z53	PLPPR4	Q7Z2D5	√						16169070
FRA10AC1	Q70Z53	GATD3A	P30042	√						16169070
FRA10AC1	Q70Z53	**GLRX3	Q76003	√						16169070
Αλληλεπιδράσεις μεταξύ των πρώτων γειτόνων				PICKLE						
Όνομα πρωτεΐνης 1	Κωδικός Πρωτεΐνης (UniProt ID)	Όνομα πρωτεΐνης 2	Κωδικός Πρωτεΐνης (UniProt ID)	HPRD	BIOGRID	INTACT	MINT	DIP	DroID	Βιβλιογραφία (PubMed ID)
PRPF3	Q43395	SF3B2	Q13435		√		√			22365833
PRPF3	Q43395	MFAP1	P55081		√		√			22365833
PRPF3	Q43395	NKAP	Q8N5F7		√		√			22365833
PRPF3	Q43395	ZCCHC10	Q8TBK6		√		√			22365833
ZCCHC10	Q8TBK6	IK	Q13123		√		√			22365833
ZCCHC10	Q8TBK6	PRPF40A	Q75400		√		√			22365833
ZCCHC10	Q8TBK6	SF3B2	Q13435		√		√			22365833
ZCCHC10	Q8TBK6	U2AF1	Q01081		√		√			22365833
NKAP	Q8N5F7	PRPF40A	Q75400		√		√			22365833
NKAP	Q8N5F7	U2AF1	Q01081		√		√			22365833
NKAP	Q8N5F7	TTC14	Q96N46		√		√			22365833
CHERP	Q8IWX8	U2AF1	Q01081		√		√			22365833
CHERP	Q8IWX8	PRPF40A	Q75400		√		√			22365833
SF3B2	Q13435	PRPF40A	Q75400		√		√			22365833
SF3B2	Q13435	PRPF40A	Q75400		√		√			16055720
MFAP1	P55081	U2AF1	Q01081		√		√			22365833
MFAP1	P55081	IK	Q13123		√		√			22365833
ZNF830	Q96NB3	IK	Q13123		√		√			22365833

ZNF830	Q96NB3	SF3B2	Q13435		√		√			22365833
TTC14	Q96N46	SF3B2	Q13435		√		√			22365833
TTC14	Q96N46	CHERP	Q8IWX8		√		√			22365833
TTC14	Q96N46	IK	Q13123		√		√			22365833
TTC14	Q96N46	DGCR14	Q96DF8		√		√			22365833
U2AF1	Q01081	SF3B2	Q13435		√		√			22365833
U2AF1	Q01081	PRPF40A	O75400		√		√			22365833
DGCR14	Q96DF8	IK	Q13123						√	14605208
THOC1	Q96FV9	SAP30BP	Q9UHR5		√		√			22365833
THOC1	Q96FV9	IK	Q13123		√		√			22365833
THOC1	Q96FV9	SF3B2	Q13435		√		√			22365833
C7orf25	Q9BPX7	SF3B2	Q13435		√		√			21900206

\* Οι αλληλεπιδράσεις αυτές ταυτοποιήθηκαν και από μέλη του εργαστηρίου μας (Sarafidou et al. υπό συγγραφή). \*\* Πρωτεΐνες που συνδέονται με το RNA (RNA binding).

Επιπλέον, η βάση δεδομένων BioGRID περιλαμβάνει πληροφορίες και για τις αλληλεπιδράσεις της FRA10AC1 που προέρχονται από μία πρόσφατη ερευνητική μελέτη μεγάλης κλίμακας η οποία είχε ως στόχο την εύρεση του συνόλου των πρωτεϊνικών αλληλεπιδράσεων του ανθρώπου, δηλαδή αυτών που λαμβάνουν χώρα στο πλαίσιο πρωτεϊνικών συμπλόκων (Huttlin et al. 2017). Ειδικότερα, η μελέτη αυτή περιλαμβάνει μεγάλο αριθμό πειραμάτων χρωματογραφίας συγγένειας στα οποία χρησιμοποιήθηκαν πρωτεϊνικά εκχυλίσματα από εμβρυονικά νεφρικά κύτταρα (HEK293T) που είχαν σταθερά διαμολυνθεί ώστε να εκφράσουν 5892 διαφορετικές ανασυνδυασμένες πρωτεΐνες. Τα πειράματα χρωματογραφίας συγγένειας στόχευαν στις ανασυνδυασμένες πρωτεΐνες και οι πρωτεΐνες των συμπλόκων που απομονώθηκαν ταυτοποιήθηκαν με φασματομετρία μάζας. Η πρωτεΐνη FRA10AC1 δεν χρησιμοποιήθηκε ως πρωτεΐνη-«δολώμα» σε κάποιο από τα παραπάνω πειράματα αλλά ταυτοποιήθηκε σε 7 διαφορετικά πρωτεϊνικά σύμπλοκα υποδεικνύοντας ότι έχει αλληλεπιδράσεις με τις αντίστοιχες πρωτεΐνες-«δολώματα», οι οποίες φαίνονται στον Πίνακα 5.

**Πιν. 5: Αλληλεπιδράσεις της πρωτεΐνης FRA10AC1.**

Όνομα πρωτεΐνης-«δολώματος»	Κωδικός UniProt ID
ALOX5AP	P20292
DGCR14	Q96DF8
FGF9	P31371
FUOM	A2VDF0
LENG1	Q96BZ8
MTUS2	Q5JR59
WASF3	Q9UPY6

### 3.1.2. Κριτήρια αξιολόγησης των δεδομένων για τις άμεσες πρωτεϊνικές αλληλεπιδράσεις της FRA10AC1

Λαμβάνοντας υπόψη τον μικρό αριθμό πειραμάτων από τα οποία προήλθαν τα παραπάνω δεδομένα, το αυξημένο ποσοστό ψευδώς θετικών αποτελεσμάτων του συστήματος των δύο υβριδίων (Letteboer & Roerman 2008) και τον περιορισμένο αριθμό πειραμάτων μικρής κλίμακας τα οποία όμως είναι σαφώς πιο αξιόπιστα από τα αντίστοιχα μεγάλης κλίμακας, έγινε περαιτέρω αξιολόγηση των αλληλεπιδρώντων μορίων με βάση πέντε κριτήρια:

- α) τη διερεύνηση του λειτουργικού τους ρόλου,
- β) την ένταξή τους σε συγκεκριμένο πρωτεϊνικό σύμπλοκο,
- γ) τον υποκυτταρικό εντοπισμό τους,
- δ) τον έλεγχο της αξιοπιστίας των αλληλεπιδράσεων βάσει του αριθμού και της κλίμακας των πειραμάτων τα οποία υποστηρίζουν την εκάστοτε αλληλεπίδραση και
- ε) την εξελικτική συντήρηση των αλληλεπιδράσεων.

#### *α) Λειτουργικός ρόλος των πρώτων «γειτόνων» της FRA10AC1*

Όσον αφορά το πρώτο κριτήριο, για την αναζήτηση του λειτουργικού ρόλου καθενός από τα είκοσι επτά αλληλεπιδρώντα μόρια της πρωτεΐνης FRA10AC1 πραγματοποιήθηκε διερεύνηση της βιβλιογραφίας στις βάσεις δεδομένων PubMed (<https://www.ncbi.nlm.nih.gov/pubmed/>) και UniProt (<http://www.uniprot.org/>). Μελετώντας τα δεδομένα που προέκυψαν, παρατηρήθηκε ότι τα μόρια αυτά, αν εξαιρέσουμε αυτά που είναι άγνωστης λειτουργίας, συμμετέχουν σε κοινές λειτουργικές διαδικασίες και ομαδοποιούνται στις επτά παρακάτω κατηγορίες:

• Χαρακτηρισμένοι παράγοντες συναρμογής του pre-mRNA και πρωτεΐνες που σχετίζονται με τον έλεγχο αυτού του θεμελιώδους μηχανισμού:

- SF3B2: συμμετέχει στην αναγνώριση της θέσης διακλάδωσης στο ιντρόνιο και στη στρατολόγηση σε αυτήν πρωτεϊνών του U2 snRNP, οδηγώντας με αυτό τον τρόπο στο σχηματισμό του συμπλόκου A το οποίο αποτελεί ένα από τα αρχικά σύμπλοκα του σωματίου συναρμογής (prespliceosome) (Spadaccini et al. 2006, Golas et al. 2003, Lardelli et al. 2010).
- U2AF1: αναγνωρίζει τη συντηρημένη αλληλουχία AG στην 3' θέση συναρμογής η οποία είναι απαραίτητη τόσο για τη συναρμολόγηση των συστατικών του σωματίου συναρμογής όσο και για το δεύτερο καταλυτικό βήμα της συναρμογής (Voith von Voithenberg et al. 2016, Wu & Krainer 1999).
- PRPF3: αποτελεί συστατικό του U4/U6 snRNP το οποίο μαζί με το U5 snRNP, σχηματίζει το τριμερές σύμπλοκο U4/U6-U5 tri-snRNP που προσδένεται στο σύμπλοκο A (prespliceosome), με αποτέλεσμα το σχηματισμό του συμπλόκου B του σωματίου συναρμογής. Το τριμερές αυτό σύμπλοκο διευκολύνει τις αναδιατάξεις του U4 και U6 RNA και των U1 και U4 snRNPs έτσι ώστε να ξεκινήσει το πρώτο καταλυτικό βήμα της συναρμογής (Maita et al. 2005, Liu 2006).
- MFAP1: είναι απαραίτητη στη μετάβαση από το σύμπλοκο B στο σύμπλοκο B act του σωματίου συναρμογής για την πραγματοποίηση του πρώτου καταλυτικού βήματος της συναρμογής στη *D. melanogaster*. Συγκεκριμένα



- συμβάλλει στην απελευθέρωση του U4 snRNA από το U4 snRNP και στην ενσωμάτωση του U6 snRNA σε κατάλληλη θέση για να αλληλεπιδράσει με το U2 snRNA έτσι ώστε να πραγματοποιηθεί το πρώτο καταλυτικό βήμα (Andersen & Tapon 2008).
- IK: τροποποιεί την επιλογή της θέσης συναρμογής ρυθμίζοντας την εναλλακτική συναρμογή πολλών γονιδίων στον *C. elegans* (Spartz et al. 2004).
  - CHERP: ρυθμίζει την εναλλακτική συναρμογή γονιδίων που συμμετέχουν στη μεταφορά του ενδοκυττάριου ασβεστίου από το ενδοπλασματικό δίκτυο (Sasaki-Osugi et al. 2013).
  - DGCR14: συμβάλλει στην ακρίβεια της διαδικασίας της συναρμογής όταν υπάρχουν μη φυσιολογικές αλληλουχίες στις θέσεις της συναρμογής ενός γονιδίου στον *C. elegans* (Noma et al. 2014) και συνεισφέρει στην αναγνώριση των θέσεων συναρμογής στον μονοκύτταρο οργανισμό *Chlamydomonas reinhardtii* (Lin et al. 2018).
  - CWC27: πεπτιδυλο-προλυλο *cis-trans* ισομεράση δηλαδή ένζυμο που αλλάζει τη διαμόρφωση του πεπτιδικού δεσμού της προλίνης (η οποία παίζει ρόλο στις αλληλεπιδράσεις των πρωτεϊνών με πρωτεΐνες και μόρια RNA) κυρίως στις πρωτεΐνες που συμμετέχουν στη διαδικασία της συναρμογής στο σχηματισμό του συμπλόκου *B act*, πριν το πρώτο καταλυτικό βήμα της συναρμογής (Ulrich & Wahl 2014).
  - PRPF40A: το ορθόλογο μόριο στο σακχαρομύκητα αλληλεπιδρά με πρωτεΐνες του U1snRNP οι οποίες συνδέονται σε κρίσιμες θέσεις συναρμογής του πρόδρομου mRNA (θέση διακλάδωσης και περιοχή πολυπυριμιδίνης) (Becerra et al. 2015). Ωστόσο, στον άνθρωπο υπάρχουν δυο ομόλογα μόρια γι' αυτήν την πρωτεΐνη (PRPF40A και PRPF40B) και δεν έχει αποσαφηνιστεί ακόμη σε ποιο από τα δύο αντιστοιχεί.
  - HABP4: ρυθμίζει την εναλλακτική συναρμογή γονιδίων, όπως έχει δείχθει *in vitro* (Bressan et al. 2009).
- Πρωτεΐνες που σχετίζονται με τη ρύθμιση της μεταγραφής:
    - NKAP: ενεργοποιεί τον μεταγραφικό παράγοντα NF-kappaB ο οποίος ρυθμίζει την έκφραση πολλών γονιδίων τα οποία σχετίζονται με άμυνα, με απόκριση στη φλεγμονή και με αναστολή της απόπτωσης, ανταποκρινόμενος σε ποικίλου είδους ερεθίσματα (Chen et al. 2003).
    - SAP30BP: συμμετέχει στη μεταγραφική καταστολή ικών γονιδίων εμποδίζοντας τον πολλαπλασιασμό τους στο κύτταρο-ξενιστή (Chen et al. 2010).
    - EEF1D: παράγοντας επιμήκυνσης της μετάφρασης. Υπάρχουν δυο ισομορφές του παράγοντα αυτού. Υπό συγκεκριμένες συνθήκες στρες, η έκφραση της ισομορφής 2 (eEF1Bδ, 281 αμινοξέα, λειτουργεί ως παράγοντας επιμήκυνσης) μειώνεται και παράγεται η ισομορφή 1 (eEF1BδL, 647 αμινοξέα) η οποία μεταναστεύει στον πυρήνα και λειτουργεί ως μεταγραφικός παράγοντας γονιδίων τα οποία διαθέτουν στοιχεία θερμικού σοκ (heat-shock elements) (Kaitsuka et al. 2011, Kaitsuka & Matsushita 2015).

- ZNF830: συστατικό του συμπλόκου XAB2 που συμμετέχει στην επιμήκυνση της μεταγραφής (Kuraoka et al. 2008). Ειδικότερα, εμποδίζει το σχηματισμό υβριδικών μορίων DNA-RNA (R-loops) κατά τη μεταγραφή, τα οποία αναστέλλουν την εξέλιξη της επιμήκυνσης της μεταγραφής.
- Πρωτεΐνες που σχετίζονται με την έξοδο του mRNA από τον πυρήνα
  - THOC1: συστατικό του συμπλόκου THO, το οποίο στρατολογείται στο 5' άκρο του πρόδρομου mRNA κατά τη μεταγραφή και συμμετέχει στο σχηματισμό των mRNPs, δηλαδή συμπλόκων πρωτεϊνών (παραγόντων συναρμογής, προσθήκης καλύπτρας στο 5' άκρο και πολυαδενυλίωσης του 3' άκρου) και mRNA τα οποία εξέρχονται από τον πυρήνα του κυττάρου για να γίνει η μετάφραση (Cheng et al. 2006).
- Πρωτεΐνες που συμμετέχουν σε σηματοδοτικά μονοπάτια
  - MAP3K12: πρωτεϊνική κινάση σερίνης-θρεονίνης που φωσφορυλιώνει τις MAPK κινάσες (mitogen-activated-protein kinases) οι οποίες με τη σειρά τους ενεργοποιούν τα σηματοδοτικά μονοπατικά JNK (c-Jun N-terminal/kinase stress-activated-protein kinase) και p38 (Gallo & Johnson 2002) για την τελική ενεργοποίηση πολλαπλών γονιδίων-στόχων. Η σηματοδότηση μέσω της MAP3K12 συμβάλλει στο σχηματισμό του κύριου άξονα των νευρώνων, στη μετανάστευση τους, στην απόπτωση και στον εκφυλισμό τους.
  - GLRX3: Διαδραματίζει κύριο ρόλο στις διαδικασίες ομοιόστασης του σιδήρου συμμετέχοντας στη μεταφορά του συμπλόκου σιδήρου-θείου και λειτουργώντας ως «αισθητήρας» του σιδήρου στο κύτταρο (Haunhorst et al. 2010, 2013, Xia et al. 2015). Προστατεύει τα κύτταρα από το οξειδωτικό στρες και προωθεί την κυτταρική επιβίωση μέσω της ενεργοποίησης σηματοδοτικών μονοπατιών (NF-κB) στα καρκινικά κύτταρα (Pham et al. 2015, Qu et al. 2011). Κατά την εμβρυϊκή ανάπτυξη, ελέγχει την πρόοδο του κυτταρικού κύκλου σε απόκριση στο οξειδωτικό στρες (Cheng et al. 2011).
- Πρωτεΐνες που συμμετέχουν σε μετα-μεταφραστικές τροποποιήσεις
  - ADPRHL2: υδρολάση που καταργεί μία μετα-μεταφραστική τροποποίηση {αφαίρεση των μορίων ριβόζης-διφωσφορικής αδενοσίνης του νικοτιναμιδο-αδενινο-δινουκλεοτιδίου (NAD) από συγκεκριμένα αμινοξέα πρωτεΐνης} η οποία συμβάλλει στην αλλαγή της δομής και λειτουργίας πρωτεϊνών που συμμετέχουν σε διάφορες βιολογικές διεργασίες (π.χ. επιδιόρθωση του DNA, μεταγραφή) (Oka et al. 2006).
  - VPS29: πιθανή φωσφατάση (Swarbrick et al. 2011), και συστατικό ενός πρωτεϊνικού συμπλόκου (retromer complex) που συνδέεται με τα ενδοσώματα όπου ελέγχει τη μεταφορά του φορτίου τους στο σύστημα Golgi για την ανακύκλωσή του ή στα λυσοσώματα για την αποικοδόμησή του (Trousdale & Kim 2015).
  - PLPPR4: ανήκει στην οικογένεια των φωσφατάσεων των λιπιδίων (LPP) που καταλύουν την αποφωσφορυλίωση βιοδραστικών μεσολαβητών λιπιδίων (bioactive lipid mediators) που ρυθμίζουν μια ποικιλία κυτταρικών λειτουργιών. Εκφράζεται στους νευρώνες και συγκεκριμένα εντοπίζεται στις μεμβράνες των

- αναπτυσσόμενων αξόνων συνεισφέροντας στην ανάπτυξή τους (Bräuer et al. 2003).
- Πρωτεΐνες που συμμετέχουν σε μονοπάτια αποικοδόμησης
    - TRIM41: Ε3 λιγάση της ουβικουτίνης η οποία προσθέτει μόρια ουβικουτίνης στην πρωτεϊνική κινάση C για την αποικοδόμησή της από το πρωτεάσωμα (Chen et al. 2007).
  - Πρωτεΐνες που σχετίζονται με τον κυτταροσκελετό
    - CCDC155: εντοπίζεται στην πυρηνική μεμβράνη και αλληλεπιδρά με πρωτεΐνες του πυρήνα σχηματίζοντας σύμπλοκα, συμβάλλοντας στη δέσμευση των τελομερών στον πυρηνικό φάκελο κατά την αρχή της μείωσης (Viera et al. 2015) και στη σύνδεση του πυρηνοσκελετού με τον κυτταροσκελετό (Starr & Fridolfsson 2010).
    - GDPD2: υδρολάση που συμμετέχει στη διαφοροποίηση των οστεοβλαστών επάγοντας την αποδιοργάνωση της ακτίνης του κυτταροσκελετού (Corda et al. 2009).
  - Πρωτεΐνες άγνωστης λειτουργίας: TTC14, ZCCHC10, C7orf25 και GATD3A.

*β) Ενταξη των πρώτων «γειτόνων» της FRA10AC1 σε συγκεκριμένο πρωτεϊνικό σύμπλοκο και σχετική αφθονία*

Με βάση το δεύτερο κριτήριο αξιολόγησης, παρατηρήθηκε ότι δεκαπέντε από τις παραπάνω πρωτεΐνες εντάσσονται στο μεγαλύτερο ριβονουκλεοπρωτεϊνικό σύμπλοκο-το σωματίο συναρμογής (spliceosome)-καθώς έχουν ταυτοποιηθεί ως συστατικά του σε πρωτεομικές μελέτες οι οποίες χαρακτηρίζουν την πρωτεϊνική σύσταση ολόκληρου του σωματίου ή των επιμέρους υποσυμπλόκων του στην κυτταρική σειρά *HeLa* (Πιν. 4). Επιπλέον, παρατηρήθηκε ότι όλες αυτές οι πρωτεΐνες, με εξαίρεση τη SAP30BP, έχουν ταυτοποιηθεί σε κοινά υποσύμπλοκα με τη FRA10AC1, δηλαδή σε ένα από τα B act, C και P, ή στα δυο από αυτά ή ακόμα και σε όλα (Πιν. 6). Αυτό διαπιστώθηκε μέσω διερεύνησης τόσο της βιβλιογραφίας όσο και της βάσης δεδομένων του σωματίου συναρμογής (<http://spliceosomedb.ucsc.edu/>) η οποία περιλαμβάνει ερευνητικές μελέτες που αφορούν την απομόνωση των συστατικών των επιμέρους υποσυμπλόκων του. Το γεγονός αυτό δεν αποτελεί έκπληξη για δυο λόγους: i) διότι με βάση την παραπάνω λειτουργική κατάταξη δέκα από τα δεκαπέντε αυτά μόρια σχετίζονται με τη διαδικασία της συναρμογής και τα υπόλοιπα με διεργασίες που είναι λειτουργικά συνδεδεμένες με αυτή (π.χ. μεταγραφή), και ii) καθότι η πρωτεΐνη FRA10AC1 έχει απομονωθεί επανειλημμένα σε βιοχημικές μελέτες απομόνωσης των επιμέρους συμπλόκων του σωματίου συναρμογής και συγκεκριμένα του συμπλόκου B act (Bessonov et al. 2010), του C (Bessonov et al. 2008, Bessonov et al. 2010, Pagan et al. 2013, Schmidt et al. 2014) και του συμπλόκου P (Pagan et al. 2013). Στον Πίνακα 6 φαίνονται στην 1<sup>η</sup> στήλη τα αλληλεπιδρώντα μόρια της FRA10AC1, και στις επόμενες στήλες οι αντίστοιχες βιοχημικές μελέτες στις οποίες αυτά ταυτοποιήθηκαν ως συστατικά είτε των επιμέρους υποσυμπλόκων ή ολόκληρου του σωματίου συναρμογής.

Πιν. 6: Δεκαπέντε από τα αλληλεπιδρώντα μόρια της πρωτεΐνης FRA10AC1 αποτελούν συστατικά του σωματίου συναρμογής.

Όνομα πρωτεΐνης	Σύμπλοκο E	Σύμπλοκο A				Σύμπλοκο B				Σύμπλοκο B act			Σύμπλοκο C				Σύμπλοκο P	Ολόκληρο το σωματίο συναρμογής								
	Makarov et al. 2012	Hartmuth et al. 2002	Behzadnia et al. 2007	Agafonov et al. 2011	Makarova et al. 2004	Deckert et al. 2006	Bessonov et al. 2008	Herold et al. 2009	Agafonov et al. 2011	Schmidt et al. 2014	Makarova et al. 2004	Agafonov et al. 2011	Bessonov et al. 2010	Jurica et al. 2002	Bessonov et al. 2008	Ilgan et al. 2009	Agafonov et al. 2011	Bessonov et al. 2010	Ilgan et al. 2013	Schmidt et al. 2014	Ilgan et al. 2013	Makarov et al. 2002	Repsilber et al. 2002	Zhou et al. 2002	Chen et al. 2007	
PRPF40A	√		√	√		√	√			√														√	√	
SAP30BP	√		√																							
SF3B2	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
U2AF1	√	√	√	√		√	√	√	√	√		√											√	√	√	√
IK				√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
DGCR14													√	√	√	√	√	√	√	√	√		√	√		
THOC1						√			√			√		√		√		√		√			√	√	√	
ZCCHC10														√		√		√		√	√					
NKAP														√			√	√	√	√	√					
CHERP		√	√	√			√		√	√									√		√		√			
MFAP1				√	√	√	√	√	√	√	√	√		√		√	√	√	√	√	√	√	√	√	√	
CWC27						√	√	√			√	√	√	√	√	√	√	√	√	√	√	√	√	√		
ZNF830						√	√		√		√	√		√		√	√	√	√	√	√	√				
TTC14														√			√		√							
PRPF3	√		√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√

Η πρωτεΐνη FRA10AC1 έχει χαρακτηριστεί ως χαμηλής αφθονίας (non-core) στα υποσύμπλοκα B act, C και P, γεγονός που πιθανώς υποδεικνύει ότι έχει δευτερεύοντα ρόλο στη συναρμογή ή, εναλλακτικά, ότι λειτουργεί ως μόριο-προσαρμογέας που συνδέει τη συναρμογή με άλλες διαδικασίες της επεξεργασίας του mRNA, δεδομένου ότι, κατά κανόνα, οι κύριες πρωτεΐνες της συναρμογής ταυτοποιούνται σε υψηλή αφθονία. Με στόχο την κατάταξη των 15 αλληλεπιδρώντων πρωτεϊνών της FRA10AC1, εντός του σωματίου συναρμογής, ως συστατικά core ή non-core η οποία πιθανώς θα υποδείκνυε τον περιφερειακό ή κεντρικό τους ρόλο στη διαδικασία της συναρμογής, διερευνήθηκε η αφθονία τους στα διαφορετικά υποσύμπλοκα του σωματίου. Γι' αυτό το λόγο συλλέχθηκαν οι αντίστοιχες πληροφορίες από τις ερευνητικές εργασίες στις οποίες είχαν ταυτοποιηθεί ως συστατικά των υποσυμπλόκων του σωματίου συναρμογής οι 15 «γείτονες» της FRA10AC1. Ο προσδιορισμός της αφθονίας της εκάστοτε πρωτεΐνης υπολογίζεται με βάση τον αριθμό των πεπτιδίων που ταυτοποιήθηκαν στα αντίστοιχα βιοχημικά πειράματα. Οι 5 από τις 17 ερευνητικές εργασίες που αφορούσαν την ταυτοποίηση των συστατικών των υποσυμπλόκων A, B, B act και C περιείχαν πληροφορίες για την αφθονία τους. Στις 2 από τις 5 εργασίες (Hartmuth et al. 2002 και Agafonov et al. 2011) αποδίδονταν σε κάθε πρωτεΐνη μια τιμή αφθονίας, η οποία προσδιόριζε εάν αυτή είχε βρεθεί σε υψηλή, μέτρια ή χαμηλή αφθονία σε καθένα από τα υποσύμπλοκα στα οποία είχε ταυτοποιηθεί. Στις άλλες 3 στις οποίες δεν υπήρχε ο αντίστοιχος προσδιορισμός (Behzadnia et al. 2007, Bessonov et al. 2008, Bessonov et al. 2010), η αξιολόγηση του αριθμού των πεπτιδίων των

εκάστοτε πρωτεϊνών πραγματοποιήθηκε σύμφωνα με τη μέγιστη και ελάχιστη τιμή καθώς και με βάση το εύρος αριθμού πεπτιδίων όλων των πρωτεϊνών που είχαν ταυτοποιηθεί. Συγκεκριμένα, σε κάθε ερευνητική εργασία, βρέθηκε η πρωτεΐνη με το μεγαλύτερο και το μικρότερο αριθμό πεπτιδίων και στη συνέχεια προσδιορίστηκε η κατανομή των πεπτιδίων που αντιστοιχούσαν στις πρωτεΐνες του κάθε υποσύμπλοκου. Για παράδειγμα, σε μία ερευνητική εργασία για το υποσύμπλοκο *B act* (Bessonov et al. 2010), ο μέγιστος αριθμός πεπτιδίων ήταν 704, ο μικρότερος ήταν 1, ο συνολικός αριθμός των πρωτεϊνών ήταν 211, και το εύρος αριθμού πεπτιδίων για τις υπόλοιπες πρωτεΐνες κυμαινόταν από 100-200 (17 τιμές), 200-400 (14 τιμές), μεγαλύτερο από 400 (4 τιμές) και μικρότερο από 100 (176 τιμές). Με βάση αυτές τις τιμές, οι πρωτεΐνες που εμφάνιζαν αριθμό πεπτιδίων μικρότερο από 100 χαρακτηρίστηκαν ως χαμηλής αφθονίας, και οι υπόλοιπες ως μέτριας ή υψηλής αφθονίας ανάλογα με τον αριθμό πεπτιδίων τους. Η ίδια διαδικασία ακολουθήθηκε και στις υπόλοιπες ερευνητικές εργασίες. Με βάση αυτά τα δεδομένα προσδιορίστηκε η αφθονία της FRA10AC1 και των 15 πρώτων «γειτόνων» της εντός του σωματίου συναρμογής όπως φαίνεται και στον Πίνακα 7 (Πιν. 7, στήλες 2-5).

**Πιν. 7: Θερμικός πίνακας (heat-map) της αφθονίας της FRA10AC1 και των πρώτων «γειτόνων» της στα υποσύμπλοκα A, B, B act και C.**

Όνομα πρωτεΐνης	Σύμπλοκο A (Agafonov et al., 2011, Hartmuth et al., 2002, Behzadnia et al., 2007)	Σύμπλοκο B (Agafonov et al., 2011, Bessonov et al., 2008)	Σύμπλοκο B act (Agafonov et al., 2011, Bessonov et al., 2010)	Σύμπλοκο C (Agafonov et al., 2011, Bessonov et al., 2008, Bessonov et al., 2010)
FRA10AC1				
PRPF3				
MFAP1				
IK				
DGCR14				
THOC1				
CWC27				
ZNF830				
NKAP				
ZCCHC10				
PRPF40A				
TTC14				
SAP30BP				
U2AF1				
CHERP				
SF3B2				

Χρωματικός κώδικας: Πρωτεΐνες με υψηλή (κόκκινο γέμισμα), μέτρια (πορτοκαλί) και χαμηλή (πράσινο) αφθονία, πρωτεΐνες που δεν έχουν ταυτοποιηθεί σε αυτά τα υποσύμπλοκα (άσπρο γέμισμα).

Σύμφωνα με τα δεδομένα της αφθονίας των πρωτεϊνών παρατηρήθηκε ότι τόσο η FRA10AC1 όσο και οι 13 πρώτοι «γείτονές» της παρουσιάζουν χαμηλή αφθονία στα υποσύμπλοκα B act και C (Πιν. 7). Τρεις από αυτές έχουν ταυτοποιηθεί μόνο στα ίδια υποσύμπλοκα με τη FRA10AC1 (NKAP, ZCCHC10, PRPF40), δύο μόνο στο C (TTC14, CHERP) και μία μόνο στο B act (U2AF1). Οι υπόλοιπες ταυτοποιήθηκαν επιπλέον είτε στα υποσύμπλοκα A, B (PRPF3, IK, MFAP1) είτε στο υποσύμπλοκο B (DGCR14, THOC1, CWC27, ZNF830) σε μέτρια ή χαμηλή αφθονία, με εξαίρεση την

MFAP1 που βρέθηκε σε υψηλή αφθονία μόνο στο υποσύμπλοκο B. Η SAP30BP, η οποία ταυτοποιήθηκε στο υποσύμπλοκο A, εμφανίζει χαμηλή αφθονία. Οι πρωτεΐνες SAP30BP και FRA10AC1 ταυτοποιήθηκαν σε διαφορετικά υποσύμπλοκα διότι η αλληλεπίδρασή τους μπορεί να λαμβάνει χώρα εκτός του σωματίου συναρμογής ή επειδή ο χαρακτηρισμός της πρωτεϊνικής σύστασης των υποσυμπλόκων δεν είναι πλήρης. Μόνο μία πρωτεΐνη από όλους τους γείτονες (SF3B2) εμφανίζει υψηλή αφθονία στα υποσύμπλοκα A, B και B *act* και μέτρια στο C, το οποίο δεν συνάδει αναλογικά με την αφθονία της FRA10AC1, γεγονός που μπορεί να ερμηνευτεί από τη συμμετοχή της SF3B2 και σε άλλες λειτουργίες πέρα από την αλληλεπίδρασή της με την FRA10AC1. Συμπερασματικά, όλοι οι γείτονες της FRA10AC1, με εξαίρεση την SF3B2 που έχει βασικό και καλά μελετημένο ρόλο στη συναρμογή, εμφανίζουν χαμηλή αφθονία, δηλαδή, είναι non-core συστατικά του σωματίου συναρμογής τα οποία πιθανώς εμφανίζουν περιφερειακό ή δευτερεύοντα ρόλο στη διαδικασία της συναρμογής.

γ) *Έλεγχος υποκυτταρικού εντοπισμού των πρώτων «γειτόνων» της FRA10AC1*

Σχετικά με το τρίτο κριτήριο, ο έλεγχος του υποκυτταρικού εντοπισμού έγινε μέσω αναζήτησης στη βάση δεδομένων NCBI στην κατηγορία του γονιδίου (<https://www.ncbi.nlm.nih.gov/gene/>), στη βάση δεδομένων Protein Atlas (<http://www.proteinatlas.org/>), η οποία περιλαμβάνει πειράματα ανοσοεντοπισμού για τον προσδιορισμό της τοπολογίας των πρωτεϊνών σε διάφορες κυτταρικές σειρές (Huang et al. 2007, Pontén et al. 2011) και συμπληρωματικά στη βάση PubMed για την αναζήτηση σχετικής βιβλιογραφίας. Τα αποτελέσματα αυτής της διερεύνησης έδειξαν ότι η πλειοψηφία των μορίων εντοπίζονται, αποκλειστικά ή όχι, στον πυρήνα, με εξαίρεση τις πρωτεΐνες GDPD2, PLPR4, GATD3A, CCDC155 για τις οποίες δεν υπάρχει διαθέσιμη πληροφορία και την VPS29 η οποία δεν εμφανίζει πυρηνικό εντοπισμό.

δ) *Έλεγχος της αξιοπιστίας των αλληλεπιδράσεων της FRA10AC1 βάσει κλίμακας και αριθμού πειραμάτων*

Σύμφωνα με το τέταρτο κριτήριο, έγινε ο έλεγχος της αξιοπιστίας των πρωτεϊνικών αλληλεπιδράσεων σύμφωνα με τον αριθμό και την κλίμακα των διαφορετικών πειραματικών προσεγγίσεων που τις υποστηρίζουν. Όλες οι αλληλεπιδράσεις προέρχονται από πειραματικές προσεγγίσεις μεγάλης κλίμακας εκτός από δυο (FRA10AC1-DGCR14, FRA10AC1-SF3B2) που έχουν ταυτοποιηθεί και με πειράματα μικρής κλίμακας (Sarafidou et al. υπό συγγραφή). Όσον αφορά τον αριθμό των πειραμάτων, όλες οι πρωτεϊνικές αλληλεπιδράσεις της FRA10AC1 (27) έχουν ταυτοποιηθεί με το σύστημα δύο υβριδίων στο σακχαρομύκητα (Rolland et al. 2014, Hegele et al. 2012, Stelzl et al. 2005) και συμπληρωματικά, 4 από αυτές (FRA10AC1-SF3B2, FRA10AC1-DGCR14, FRA10AC1-IK, FRA10AC1-CHERP) με πειράματα συν-ανοσοκατακρήμνισης των ανασυνδυασμένων μορίων (Hegele et al. 2012). Δύο από αυτές, οι αλληλεπιδράσεις της FRA10AC1 με την DGCR14 και SF3B2, έχουν επιπλέον επιβεβαιωθεί με συνκατακρήμνιση (Sarafidou et al. υπό συγγραφή). Για το ζεύγος πρωτεϊνών FRA10AC1-DGCR14, η επιβεβαίωση έγινε και με πειράματα συν-

ανοσοκατακρήμιση των ενδογενών μορίων, και επιπλέον προσδιορίστηκαν οι περιοχές των πρωτεϊνών του ζεύγους που αλληλεπιδρούν (Sarafidou et al. υπό συγγραφή).

Σχετικά με τις αλληλεπιδράσεις μεταξύ των πρώτων γειτόνων της FRA10AC1, όλες προέρχονται από πειράματα μεγάλης κλίμακας. Όσον αφορά τον αριθμό των πειραμάτων, όλες οι πρωτεϊνικές αλληλεπιδράσεις μεταξύ των πρώτων γειτόνων (28) έχουν ταυτοποιηθεί με το σύστημα δύο υβριδίων στο σακχαρομύκητα (Hegele et al. 2012, Vinayagam et al. 2011) και συμπληρωματικά, 9 από αυτές (PRPF3-SF3B2, PRPF3-MFAP1, PRPF3-NKAP, NKAP-PRPF40A, NKAP-U2AF1, MFAP1-IK, ZNF830-IK, ZNF830-SF3B2, U2AF1-SF3B2) με πειράματα συνανοσοκατακρήμισης των ανασυνδυασμένων μορίων (Hegele et al. 2012). Επιπλέον, μία από αυτές (PRPF40A-SF3B2) έχει ταυτοποιηθεί και με πειράματα υγρής χρωματογραφίας σε συνδυασμό με φασματομετρία μάζας (LC-MS/MS), δηλαδή, έχει δειχθεί ότι οι πρωτεΐνες του ζεύγους εντοπίζονται στο ίδιο σύμπλοκο (Ingham et al. 2005).

#### ε) Έλεγχος εξελικτικής συντήρησης των αλληλεπιδράσεων της FRA10AC1

Σύμφωνα με το τελευταίο κριτήριο, βρέθηκε ότι οι αλληλεπιδράσεις των ζευγών FRA10AC1-DGCR14, και IK-DGCR14, πρώτοι «γείτονες» της FRA10AC1, είναι εξελικτικά συντηρημένες και στη *D. melanogaster*, όπως έχει δειχθεί με πειράματα μεγάλης κλίμακας (με το σύστημα δύο υβριδίων στο σακχαρομύκητα) (Giot et al. 2003).

Από τα τρία πρώτα κριτήρια αξιολόγησης, οι αλληλεπιδράσεις που θεωρήθηκαν πιο αξιόπιστες ήταν αυτές που αφορούν πρωτεΐνες οι οποίες: α) συμμετέχουν στις διαδικασίες επεξεργασίας του πρόδρομου mRNA, β) εμφανίζουν πυρηνικό εντοπισμό με δεδομένο ότι οι διεργασίες που αναφέρονται στο (α) λαμβάνουν χώρα στον πυρήνα ή/και σε αυτόν, και γ) αποτελούν συστατικά του σωματίου συναρμογής. Όσον αφορά τα δυο τελευταία κριτήρια, οι αλληλεπιδράσεις με αυξημένη αξιοπιστία θεωρήθηκαν αυτές που ταυτοποιήθηκαν με περισσότερες από μία πειραματικές μεθόδους και ειδικότερα σε μελέτες μικρής κλίμακας, καθώς και αυτές που ήταν εξελικτικά συντηρημένες.

#### 3.1.3. Κριτήρια αξιολόγησης των δεδομένων για τις έμμεσες πρωτεϊνικές αλληλεπιδράσεις της FRA10AC1

Τα κριτήρια που εφαρμόστηκαν για την αξιολόγηση των άμεσων αλληλεπιδράσεων της FRA10AC1 χρησιμοποιήθηκαν και για τις υπόλοιπες αλληλεπιδράσεις της (Huttlin et al. 2017). Όπως προαναφέρθηκε η FRA10AC1 απομονώθηκε ως συστατικό 7 πρωτεϊνικών συμπλόκων. Η DGCR14 ήταν η μοναδική από τις επτά πρωτεΐνες-«δολώματα» που χρησιμοποιήθηκαν για την απομόνωση των συμπλόκων αυτών που αποτελεί πρώτο «γείτονα» της FRA10AC1. Όσον αφορά την υποκυτταρική τοπολογία των επτά πρωτεϊνών που χρησιμοποιήθηκαν ως «δολώμα», οι δυο εντοπίζονται στον πυρήνα ή και σε αυτόν (LENG1, DGCR14), ενώ οι υπόλοιπες είτε δεν εντοπίζονται στον πυρήνα, γεγονός που υποδεικνύει ότι αυτές οι πρωτεΐνες σχετίζονται με

διεργασίες εκτός πυρήνα, είτε δεν υπάρχουν σχετικές πληροφορίες. Σχετικά με τον αριθμό και την κλίμακα του πειράματος, οι αλληλεπιδράσεις αυτές δεν μπορούν να θεωρηθούν αυξημένης αξιοπιστίας διότι ταυτοποιήθηκαν με μία μόνο πειραματική μέθοδο σε μία μελέτη μεγάλης κλίμακας, στο πλαίσιο ταυτοποίησης αλληλεπιδράσεων εντός πρωτεϊνικών συμπλόκων. Σχετικά με το λειτουργικό ρόλο των πρωτεϊνών-«δολωμάτων» που χρησιμοποιήθηκαν για την απομόνωση αυτών των συμπλόκων, ο ρόλος της DGCR14 έχει αναλυθεί παραπάνω (βλ. Ενότητα 1.3.1., 3.1.2.) και οι βιολογικές λειτουργίες των τεσσάρων πρωτεϊνών είναι οι ακόλουθες:

- ALOX5AP: συνεισφέρει στη βιοσύνθεση του λευκοτριένιου που αποτελεί βιολογικά ενεργό μεταβολίτη του αραχιδονικού οξέος, το οποίο συμμετέχει σε ποικιλία φλεγμονωδών αποκρίσεων. Ειδικότερα, μεταφέρει το αραχιδονικό οξύ στο ένζυμο 5-λιποξυγενάση για την περαιτέρω οξειδωσή του (Fang et al. 2014, Ferguson et al. 2007).
- FGF9: ανήκει στην οικογένεια των αυξητικών παραγόντων FGF. Οι παράγοντες αυτοί, κατά την εμβρυϊκή ανάπτυξη, ρυθμίζουν την οργανογένεση και, κατά την ενήλικη ζωή, διαδραματίζουν βασικό ρόλο στην ρύθμιση της ομοιόστασης, στην επούλωση των πληγών και στην επιδιόρθωση των ιστών (Frontini et al. 2011).
- MTUS2: αλληλεπιδρά με πρωτεΐνες που συνδέονται με τα άκρα των μικροσωληνίσκων και πιθανώς ρυθμίζει τη δυναμική τους χωρίς να είναι γνωστός ο μηχανισμός (Jiang et al. 2009).
- WASF3: συμμετέχει στο σηματοδοτικό μονοπάτι ρύθμισης του πολυμερισμού της ακτίνης και της αναδιαμόρφωσης του κυτταροσκελετού σε απόκριση σε εξωτερικά ερεθίσματα, προωθώντας το σχηματισμό ελασματοποδίων (lamellipodia) που λαμβάνουν μέρος στην κυτταρική μετανάστευση (Bai et al. 2011, Sossey-Alaoui et al. 2007).

Εξετάζοντας το λειτουργικό ρόλο των πρωτεϊνών που εντοπίστηκαν μαζί με την FRA10AC1 στο πρωτεϊνικό σύμπλοκο με την DGCR14 ως πρωτεΐνη-«δόλωμα», παρατηρήθηκε ότι οι περισσότερες από τις 22 πρωτεΐνες που συμμετέχουν στο σύμπλοκο σχετίζονται με τη διαδικασία της συναρμογής και ορισμένες από αυτές με τη μεταγραφή. Έτσι επιβεβαιώνεται με ανεξάρτητη προσέγγιση, η οποία εφαρμόστηκε σε διαφορετικό πλαίσιο και σε διαφορετική κυτταρική σειρά, η ένταξη της FRA10AC1 σε ένα πρωτεϊνικό σύμπλοκο το οποίο είναι εμπλουτισμένο με πρωτεΐνες που συμμετέχουν κυρίως στη συναρμογή και σε μικρότερο βαθμό στη μεταγραφή (συμπληρωματικός πίνακας 5 των Huttlin et al. 2017).

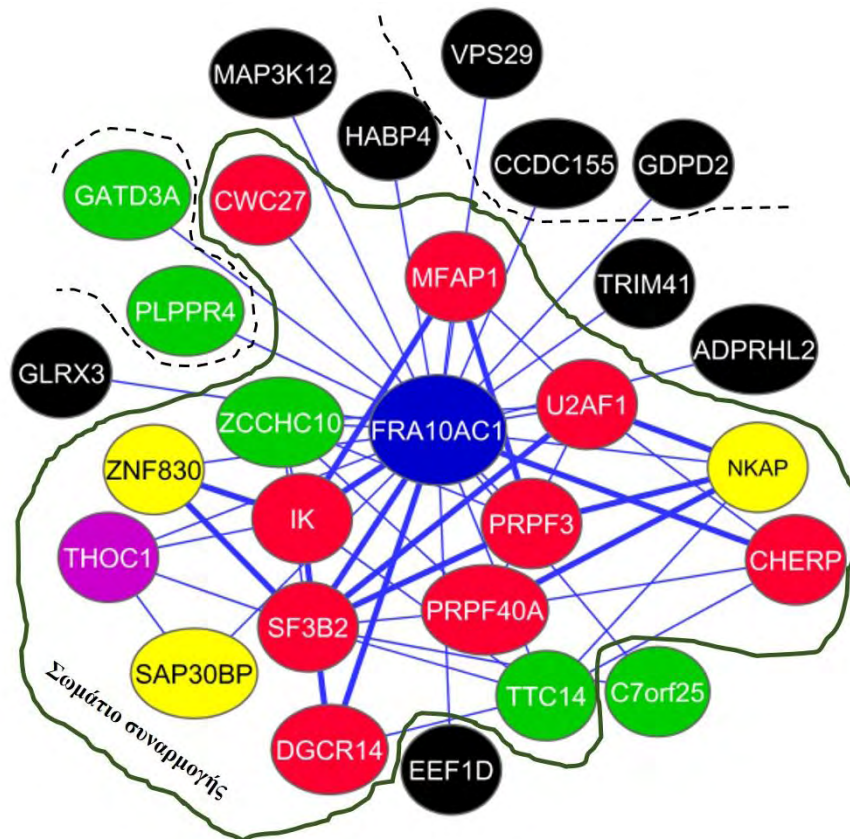
Επιπλέον, κατά τη διερεύνηση του λειτουργικού ρόλου των πρωτεϊνών που απομονώθηκαν στο σύμπλοκο με την πρωτεΐνη-«δόλωμα» FGF9, η οποία δεν συμμετέχει στη μεταγραφή, παρατηρήθηκε ότι υπάρχει εμπλουτισμός (4 από τις 9) στον οντολογικό όρο της διαδικασίας της μεταγραφής. Οι υπόλοιπες πρωτεΐνες-«δολώματα» με τις οποίες απομονώθηκε η FRA10AC1 συμμετέχουν σε διαφορετικές βιολογικές διαδικασίες, όπως περιγράφηκε και παραπάνω, αποτέλεσμα που πιθανώς μπορεί να εντάξει την FRA10AC1 σε επιπρόσθετο βιολογικό μονοπάτι εκτός από την επεξεργασία του RNA. Γι' αυτό το λόγο, και για κάθε σύμπλοκο όπου εντοπίστηκε η FRA10AC1, διερευνήθηκαν οι οντολογικοί όροι βιολογικής διαδικασίας των υπόλοιπων πρωτεϊνών που απομονώθηκαν (συμπληρωματικός πίνακας 5 των Huttlin et al. 2017). Ειδικότερα,



οι οντολογικοί όροι των πρωτεϊνών που απομονώθηκαν με την LENG1 (2 από τις 5), την WASF3 (11 από τις 19) και την MTUS2 (6 από τις 25) είναι εμπλουτισμένοι σε διάφορα σηματοδοτικά μονοπάτια (π.χ. σηματοδοτικό μονοπάτι κυτοκινών, μονοπάτι Wnt) που λαμβάνουν μέρος στη μεταφορά εξωκυττάρων σημάτων τα οποία καταλήγουν στον πυρήνα του κυττάρου όπου και ρυθμίζουν συγκεκριμένες πυρηνικές διεργασίες (π.χ. μεταγραφή). Επιπλέον, στην περίπτωση της MTUS2, ορισμένες πρωτεΐνες (7 από τις 25) συμμετέχουν στην μεταφορά ιόντων (π.χ. σιδήρου) και μορίων (π.χ. τρανσφερίνη) διαμέσου της μεμβράνης. Αν και μόνο η LENG1 από τις τρεις παραπάνω πρωτεΐνες-«δολώματα» εμφανίζει πυρηνικό εντοπισμό, η συμμετοχή της FRA10AC1 στα αντίστοιχα σύμπλοκα πιθανώς υποδεικνύει ότι συμμετέχει σε σημεία των σηματοδοτικών μονοπατιών που λαμβάνουν χώρα στον πυρήνα του κυττάρου. Για να εξαχθούν πιο αξιόπιστα συμπεράσματα για την πιθανή συμμετοχή της FRA10AC1 σε βιολογικά μονοπάτια πέραν της επεξεργασίας του RNA, τα παραπάνω αποτελέσματα θα αξιολογηθούν συνδυαστικά με τις αλλαγές που προκάλεσε η τροποποίηση της έκφρασης του γονιδίου *FRA10AC1* σε όλες τις παραπάνω πρωτεΐνες-«δολώματα» και στις πρωτεΐνες που κατακρημνίστηκαν με αυτές σε μεταγραφικό και πρωτεϊνικό επίπεδο.

#### 3.1.4. *In silico* ανακατασκευή του δικτύου πρωτεϊνικών αλληλεπιδράσεων της FRA10AC1

Το ανακατασκευασμένο δίκτυο δημιουργήθηκε με το πρόγραμμα Cytoscape (<http://www.cytoscape.org/>) και περιλαμβάνει τα είκοσι επτά αλληλεπιδρώντα μόρια της πρωτεΐνης FRA10AC1 καθώς και τις αλληλεπιδράσεις μεταξύ των πρώτων γειτόνων της (Εικ. 32 & Πιν. 4). Έτσι, περιλαμβάνει 28 πρωτεΐνες-κόμβους και 55 πρωτεϊνικές αλληλεπιδράσεις (edges) (Εικ. 32). Στην Εικόνα 32 οπτικοποιούνται και τα αποτελέσματα από την εφαρμογή των πέντε κριτηρίων αξιολόγησης των αλληλεπιδράσεων. Δηλαδή οι λειτουργικές διαδικασίες απεικονίζονται με διαφορετικό χρωματικό κώδικα, οι πρωτεΐνες που ανήκουν στο σωματίο συναρμογής διαχωρίζονται από τις υπόλοιπες με πράσινο πλαίσιο και οι πρωτεΐνες για τις οποίες δεν υπάρχει πληροφορία για πυρηνικό εντοπισμό οριοθετούνται από τις υπόλοιπες με διακεκομμένη γραμμή. Τέλος, οι αλληλεπιδράσεις του δικτύου με αυξημένη αξιοπιστία, σύμφωνα με την κλίμακα και τον αριθμό των πειραμάτων που υποστηρίζουν μία αλληλεπίδραση και την εξελικτική συντήρηση μίας αλληλεπίδρασης, φαίνονται με έντονες μπλε γραμμές.



**Εικ. 32:** Δίκτυο άμεσων πρωτεϊνικών αλληλεπιδράσεων της πρωτεΐνης **FRA10AC1**. Το δίκτυο περιλαμβάνει 28 πρωτεΐνες-κόμβους και 55 αλληλεπιδράσεις (edges). Οι έντονες μπλε γραμμές υποδεικνύουν i) την αυξημένη αξιοπιστία της συγκεκριμένης αλληλεπίδρασης, με βάση τον αριθμό των διαφορετικών πειραματικών προσεγγίσεων που τις υποστηρίζουν και ii) τις αλληλεπιδράσεις που είναι εξελικτικά συντηρημένες στη *D. melanogaster*. Οι αλληλεπιδράσεις με τις μπλε ανοιχτές και λεπτές γραμμές έχουν ταυτοποιηθεί μόνο με πειράματα δυο υβριδίων στον άνθρωπο (Stelzl et al. 2005, Hegele et al. 2012, Rolland et al. 2014) και στη *D. melanogaster* (Giot et al. 2003). Δεκαπέντε κόμβοι του δικτύου έχουν ταυτοποιηθεί ως συστατικά του σωματίου συναρμογής (Πιν. 6) (εντός του πράσινου πλαισίου). Το χρώμα των κόμβων υποδεικνύει τη λειτουργική κατηγορία των αντίστοιχων πρωτεϊνών (κόκκινο: παράγοντες συναρμογής, κίτρινο: μεταγραφικοί παράγοντες, μωβ: έξοδος του mRNA από τον πυρήνα, πράσινο: άγνωστης λειτουργίας, μαύρο: μη σχετιζόμενη λειτουργία με τη διαδικασία της επεξεργασίας του mRNA). Οι διακεκομμένες γραμμές διαχωρίζουν τα μόρια για τα οποία δεν υπάρχει διαθέσιμη πληροφορία για πυρηνικό εντοπισμό ή δεν εμφανίζουν πυρηνικό εντοπισμό (VPS29).

Συμπερασματικά, το δίκτυο πρωτεϊνικών αλληλεπιδράσεων της FRA10AC1 είναι εμπλουτισμένο με πρωτεΐνες που εντοπίζονται στον πυρήνα και συμμετέχουν στη συναρμογή και σε λειτουργικά συναφείς διαδικασίες. Η συναρμογή δεν λαμβάνει χώρα ανεξάρτητα από άλλες διαδικασίες στα κύτταρα, αλλά συνδέεται λειτουργικά με ποικίλες κυτταρικές διεργασίες (π.χ. μεταγραφή, έξοδος mRNA από τον πυρήνα κ.α.), όπως αναλύθηκε και στην Ενότητα 1.4.2. της Εισαγωγής.

Έτσι, με σκοπό να γίνει η ένταξη της FRA10AC1 σε μια συγκεκριμένη βιολογική διαδικασία της επεξεργασίας του mRNA πραγματοποιήθηκε η *in silico* ανακατασκευή του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής. Δεδομένης της ταυτοποίησης της FRA10AC1 στα υποσύμπλοκα B, C και P, ένας επιπρόσθετος στόχος ήταν ο προσδιορισμός της πιθανής λειτουργίας της σε αυτά τα υποσύμπλοκα, και γι' αυτό το λόγο έγινε η *in silico* ανασύσταση των δικτύων των υποσυμπλόκων του σωματίου συναρμογής.

Η *in silico* ανακατασκευή του δικτύου ολόκληρου του σωματίου συναρμογής θα συνεισέφερε: α) στον προσδιορισμό της τοπολογίας της πρωτεΐνης FRA10AC1 στο ευρύτερο σωματίο συναρμογής, εξετάζοντας τις τοπολογικές παραμέτρους του δικτύου και της FRA10AC1, β) στη διερεύνηση των βιολογικών διαδικασιών στις οποίες συμμετέχουν και οι δεύτεροι «γείτονες» της FRA10AC1 εντός του σωματίου συναρμογής ώστε να συλλεχθούν πληροφορίες για τον συγκεκριμένο ρόλο της και γ) στην πιθανή ένταξη της FRA10AC1 σε κάποιο σύμπλεγμα πρωτεϊνών (cluster) στο οποίο οι πρωτεΐνες είναι στενά συνδεδεμένες μεταξύ τους γεγονός που θα βοηθούσε στον προσδιορισμό του βιολογικού ρόλου της FRA10AC1 δεδομένου ότι οι πρωτεΐνες που αποτελούν ένα cluster συνήθως παρουσιάζουν κοινό λειτουργικό ρόλο.

Η *in silico* ανακατασκευή των δικτύων πρωτεϊνικών αλληλεπιδράσεων των υποσυμπλόκων του σωματίου θα εξυπηρετούσε στη διερεύνηση των βιολογικών διεργασιών στις οποίες συμμετέχουν τα μόρια της ευρύτερης «γειτονιάς» της FRA10AC1 στα δίκτυα των υποσυμπλόκων, δεδομένου ότι η πλειοψηφία των αλληλεπιδρώντων μορίων της εντός του σωματίου συναρμογής (80%) είναι παράγοντες συναρμογής με συγκεκριμένο ρόλο στα διαφορετικά στάδια συναρμολόγησης του σωματίου (π.χ. MFAP1, PRPF3) ή πρωτεΐνες οι οποίες συμμετέχουν στις διαδικασίες επεξεργασίας του mRNA (π.χ. THOC1) και στρατολογούνται στο mRNA σε συγκεκριμένο στάδιο της συναρμογής. Έτσι θα μπορούσε να γίνει η ένταξή της σε διεργασίες που λαμβάνουν χώρα σε κάποιο από τα ενδιάμεσα ή τελευταία στάδια της διαδικασίας (π.χ. συμμετοχή στη διεκπεραίωση των καταλυτικών βημάτων της συναρμογής).

### **3.2. *In silico* ανακατασκευή του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής (spliceosome) και των υποσυμπλόκων του**

Το σωματίο συναρμογής αποτελεί τη ριβονουκλεοπρωτεϊνική μηχανή με το μεγαλύτερο μοριακό βάρος (4–5 MDa) η οποία αποτελείται από εκατοντάδες πρωτεΐνες και μόρια RNA. Η συνεργασία των μορίων RNA και των πρωτεϊνών είναι απαραίτητη για την αναγνώριση με εξαιρετική ακρίβεια, μικρών αλληλουχιών (θέσεις συναρμογής) ανάμεσα σε πλήθος όμοιων αλληλουχιών στο πρόδρομο μετάγραφο και για τη σύζευξη αυτών των αλληλουχιών έτσι ώστε να πραγματοποιηθούν τα δυο καταλυτικά βήματα της συναρμογής (Wahl et al. 2009). Απαραίτητη προϋπόθεση για την ανακατασκευή του δικτύου πρωτεϊνικών αλληλεπιδράσεων ολόκληρου του σωματίου συναρμογής και των επιμέρους υποσυμπλόκων του ήταν η συλλογή όλων των πρωτεϊνών που έχουν ταυτοποιηθεί ως συστατικά του σωματίου συναρμογής από ερευνητικές μελέτες που αφορούσαν τη βιοχημική απομόνωση και ταυτοποίηση των συστατικών του.

#### 3.2.1. *In silico* ανακατασκευή του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής

##### 3.2.1.1. Δημιουργία του συνόλου δεδομένων των συστατικών του σωματίου συναρμογής

Τα συστατικά του σωματίου συναρμογής συλλέχθηκαν από δεκαοκτώ ερευνητικές εργασίες οι οποίες περιγράφουν πρωτεομικές μελέτες για την ταυτοποίηση των

συστατικών του, είτε για ολόκληρο το σωματίο συναρμογής ή για τα επιμέρους υποσύμπλοκα (Πιν. 8).

Η εύρεση αυτών των ερευνητικών εργασιών πραγματοποιήθηκε μέσω της βάσης δεδομένων του σωματίου συναρμογής (<http://spliceosomedb.ucsc.edu/>) η οποία περιλάμβανε δεκατρείς ερευνητικές μελέτες που αφορούσαν την απομόνωση των συστατικών των επιμέρους υποσυμπλόκων (E, A, B, B *act*, C και P) (Πιν. 8) χωρίς να συμπεριλαμβάνει μελέτες απομόνωσης ολόκληρου του σωματίου συναρμογής. Για να διασφαλιστεί η πλήρης κάλυψη του συνόλου των συστατικών του σωματίου συναρμογής έγινε περαιτέρω διερεύνηση της σχετικής βιβλιογραφίας η οποία οδήγησε στην εύρεση πέντε επιπλέον ερευνητικών εργασιών που αφορούσαν μελέτες απομόνωσης των συστατικών ολόκληρου του σωματίου συναρμογής αλλά και των επιμέρους υποσυμπλόκων B και B *act* (Πιν. 8). Στον Πίνακα 8 φαίνονται οι 17 ερευνητικές εργασίες (3<sup>η</sup> στήλη) και ο αριθμός των πρωτεϊνών που ταυτοποιήθηκαν σε καθεμία από αυτές (2<sup>η</sup> στήλη).

**Πιν. 8:** Ερευνητικές εργασίες απομόνωσης και ταυτοποίησης των συστατικών είτε ολόκληρου του σωματίου συναρμογής ή των επιμέρους υποσυμπλόκων του.

Υποσύμπλοκα	Αριθμός πρωτεϊνών	Ερευνητικές εργασίες
Υποσύμπλοκο E	267	Makarov <i>et al.</i> 2012
Υποσύμπλοκο A	113	Behzadnia <i>et al.</i> 2007
Υποσύμπλοκο A	70	Hartmuth <i>et al.</i> 2002
Υποσύμπλοκα A, B, B <i>act</i> , C	171	Agafonov <i>et al.</i> 2011
Υποσύμπλοκα B, B <i>act</i>	93	Makarova <i>et al.</i> 2004*
Υποσύμπλοκο B	135	Deckert <i>et al.</i> 2006
Υποσύμπλοκο B	231	Schmidt <i>et al.</i> 2014
Υποσύμπλοκο B <i>act</i>	140	Bessonov <i>et al.</i> 2010
Υποσύμπλοκα B, C	258	Bessonov <i>et al.</i> 2008
Υποσύμπλοκα B, C	166	Herold <i>et al.</i> 2008
Υποσύμπλοκο C	71	Jurica <i>et al.</i> 2002
Υποσύμπλοκο C	89	Ilagan <i>et al.</i> 2009
Υποσύμπλοκο C, P	115	Ilagan <i>et al.</i> 2013
Υποσύμπλοκο C	233	Schmidt <i>et al.</i> 2014
Σωματίο συναρμογής	160	Chen <i>et al.</i> 2007*
Σωματίο συναρμογής	272	Rappsilber <i>et al.</i> 2002*
Σωματίο συναρμογής	137	Zhou <i>et al.</i> 2002*
Σωματίο συναρμογής	83	Makarov <i>et al.</i> 2002*

\*Δεν συμπεριλαμβάνονται στη βάση δεδομένων του σωματίου συναρμογής

Σε πολλές περιπτώσεις, οι διαφορετικές μελέτες χρησιμοποιούσαν διαφορετικά ονόματα για την ίδια πρωτεΐνη ή τις ισομορφές μιας πρωτεΐνης, για το ίδιο γονίδιο ή για τα εναλλακτικά μετάγραφα του. Έτσι, αν γινόταν η άμεση καταγραφή όλων αυτών

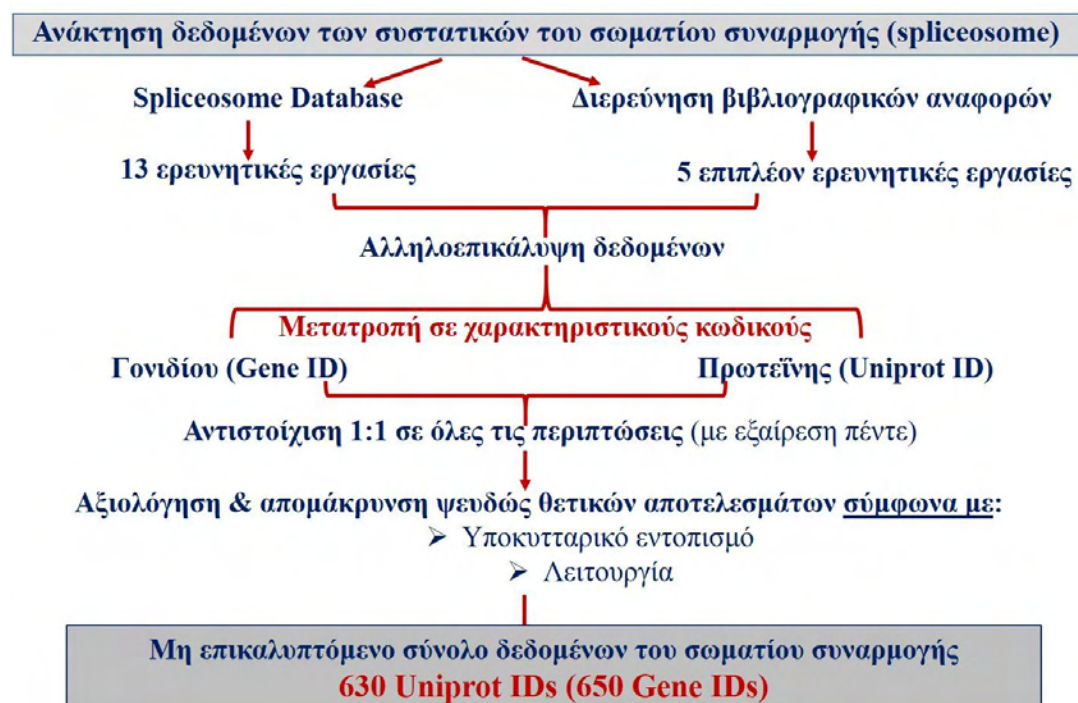
των ονομάτων χωρίς περαιτέρω επεξεργασία, θα επαναλαμβάνονταν οι ίδιες πρωτεΐνες ή γονίδια με διαφορετικά ονόματα και δεν θα ήταν εφικτό να διαφοροποιηθούν μεταξύ τους. Ο τρόπος λοιπόν αυτός θα οδηγούσε σε μεγάλη αλληλοεπικάλυψη μεταξύ των δεδομένων η οποία δεν θα ήταν διακριτή λόγω της διαφορετικής ονοματολογίας, και κατ' επέκταση στην υπερεκτίμηση του αριθμού των συστατικών του σωματίου συναρμογής.

Έτσι, για να επιτευχθεί η σύγκριση των δεδομένων που προέρχονταν από τις διαφορετικές μελέτες επιλέχθηκε μια κοινή ονοματολογία η οποία θα ήταν αντιπροσωπευτική για κάθε αναφερόμενο γονίδιο και την αντίστοιχη πρωτεΐνη. Συγκεκριμένα, όλοι οι αναφερόμενοι κωδικοί των πρωτεϊνών ή των μεταγράφων των γονιδίων μετατράπηκαν σε κωδικό Gene ID (κωδικός γονιδίου) (Genbank, <http://www.ncbi.nlm.nih.gov/gene>) ο οποίος είναι χαρακτηριστικός για κάθε γονίδιο, και σε κωδικό UniProt ID (κωδικός πρωτεΐνης) (UniProtKB <http://www.uniprot.org/>) ο οποίος είναι χαρακτηριστικός για κάθε πρωτεΐνη (Εικ. 33). Το Gene ID παρέχει τη δυνατότητα εύρεσης όλων των πληροφοριών για την νουκλεοτιδική αλληλουχία του εκάστοτε γονιδίου καθώς και των εναλλακτικών μεταγράφων αυτού, συμπεριλαμβάνοντας και τους διαφορετικούς κωδικούς που αντιστοιχούν σε καθένα από αυτά. Ο κωδικός αυτός επιλέχθηκε σε σύγκριση με τη χρήση του επίσημου συμβόλου-ονόματος (gene symbol), καθώς υπάρχουν περιπτώσεις απόδοσης ίδιου ονόματος σε διαφορετικά γονίδια με συνέπεια να μην ισχύει η αμφιμονοσήμαντη αντιστοίχιση συμβόλου-γονιδίου. Γι' αυτό η χρήση του επίσημου συμβόλου του γονιδίου όπως δίνεται από τη βάση δεδομένων NCBI (<http://www.ncbi.nlm.nih.gov/gene>) έγινε συμπληρωματικά. Το UniProt ID παρέχει όλες τις απαραίτητες πληροφορίες για την ακολουθία των πρωτεϊνών και των ισομορφών αυτών.

Σε όλες σχεδόν τις περιπτώσεις υπήρξε αντιστοιχία ένα προς ένα μεταξύ του Gene ID και του UniProt ID (Πιν 1. Παράρτημα). Ωστόσο, σε πέντε περιπτώσεις ένα συγκεκριμένο UniProt ID αντιστοιχούσε σε περισσότερα του ενός Gene IDs (Πιν. 1 Παράρτημα). Τα γονίδια αυτά ελέγχθηκαν ως προς την ακολουθία των πρωτεϊνών που κωδικοποιούν. Η διερεύνηση αυτή έδειξε ότι οι πρωτεΐνες που αντιστοιχούσαν στο ίδιο UniProt ID παρουσίαζαν 100% ομοιότητα. Αυτό συμβαίνει διότι με τη χρήση του κωδικού UniProt ID ομαδοποιούνται πρωτεΐνες με ταυτόσημη ακολουθία, όπως εξηγήθηκε και στην Ενότητα 2.1.2.2. των Μεθόδων. Τα πέντε αυτά UniProt IDs αφορούν κυρίως ιστόνες (3 από τα πέντε) (Πιν. 1 Παράρτημα). Για τις πέντε αυτές περιπτώσεις, οι αντίστοιχες ερευνητικές εργασίες χρησιμοποίησαν στην παράθεση των αποτελεσμάτων των συστατικών του σωματίου συναρμογής ένα συγκεκριμένο όνομα πρωτεΐνης ή γονιδίου ή έναν κωδικό νουκλεοτιδικής αλληλουχίας (nucleotide ID) ενώ δεν ήταν δυνατή η διάκριση καθώς η ταυτοποίηση γινόταν με φασματομετρία μάζας. Έτσι, στο σύνολο δεδομένων των συστατικών του σωματίου συναρμογής συμπεριελήφθησαν όλα τα Gene IDs που αντιστοιχούσαν σε ένα Uniprot ID (Πιν. 1 Παράρτημα). Στο Παράρτημα 1 φαίνονται οι 630 πρωτεΐνες-UniProt IDs (3<sup>η</sup> στήλη) του σωματίου συναρμογής, τα αντίστοιχα Gene IDs (2<sup>η</sup> στήλη) και σύμβολα γονιδίου (1<sup>η</sup> στήλη), οι ερευνητικές εργασίες (στήλες AF-BD) απομόνωσης και ταυτοποίησης των 630 πρωτεϊνών καθώς και το στάδιο της διαδικασίας στο οποίο αυτές έχουν

ταυτοποιηθεί (υποσύμπλοκα ή ολόκληρο το σωματίο συναρμογής, στήλες Y-AE) και οι κατηγορίες γονιδιακής οντολογίας (Gene Ontology, GO) στις οποίες ανήκουν τα συστατικά του σωματίου (βλ. Ενότητα 3.2.1.2.).

Σε επόμενο στάδιο, για να εξασφαλιστεί η πιστότητα του συνόλου δεδομένων, ακολούθησε περαιτέρω επεξεργασία του με στόχο τον καθαρισμό του από πρωτεΐνες που θεωρήθηκαν προσμίξεις της διαδικασίας απομόνωσης (ψευδώς θετικά αποτελέσματα) σύμφωνα με τις μελέτες της ερευνητικής ομάδας του Dr. R. Lührmann. Για την τελική απομάκρυνση αυτών των πρωτεϊνών προηγήθηκε η λεπτομερής διερεύνηση της σχετικής βιβλιογραφίας για το καθένα από αυτά όσον αφορά τον υποκυτταρικό τους εντοπισμό και το λειτουργικό τους ρόλο (Εικ. 33). Για παράδειγμα απομακρύνθηκαν οι ριβοσωμικές πρωτεΐνες οι οποίες θεωρούνται προσμίξεις. Επίσης, πρωτεΐνες που ήταν καλά χαρακτηρισμένες, με εδραιωμένο λειτουργικό ρόλο εκτός πυρήνα σε συνδυασμό με το μη πυρηνικό εντοπισμό τους, όπως η τουμπουλίνη και η δυνεΐνη, επιλέχθηκαν να απομακρυνθούν. Αντιθέτως, πρωτεΐνες παρόμοιας λειτουργίας οι οποίες εμφανίζουν πυρηνικό εντοπισμό (η ακτίνη ACTL6A ή έχουν συσχετιστεί και με τη συναρμογή ή με συναφείς λειτουργίες (π.χ. η ακτίνη ACTG1, Qi et al. 2011) συμπεριλήφθηκαν στο τελικό σύνολο δεδομένων. Στο ίδιο πλαίσιο εξέτασης των σχετικών βιβλιογραφικών αναφορών απομακρύνθηκε και το γονίδιο USP17L1P, το οποίο πλέον έχει χαρακτηριστεί ως ψευδογονίδιο. Έτσι, δημιουργήθηκε το τελικό μη επικαλυπτόμενο σύνολο δεδομένων που περιλάμβανε 630 UniProt IDs (Εικ. 33) (Πιν. 1 Παράρτημα). Στην Εικόνα 33 παρουσιάζονται διαγραμματικά τα στάδια επεξεργασίας των δεδομένων που συλλέχθηκαν μέχρι τη δημιουργία του τελικού συνόλου.



Εικ. 33: Δημιουργία συνόλου δεδομένων των συστατικών του σωματίου συναρμογής. Παρουσιάζονται τα βήματα που ακολουθήθηκαν για να δημιουργηθεί το μη επικαλυπτόμενο σύνολο δεδομένων των 630 πρωτεϊνών-συστατικών του σωματίου συναρμογής.

Την αξιοπιστία του τελικού συνόλου δεδομένων ενισχύει το γεγονός ότι το 62% του συνόλου των πρωτεϊνών (390 Uniprot IDs) είχαν ταυτοποιηθεί ως συστατικά του σωματίου συναρμογής σε περισσότερες από μία μελέτες. Ωστόσο, το 54% των υπολοίπων 240 πρωτεϊνών είχε ταυτοποιηθεί μόνο στο υποσύμπλοκο E (Πιν.1 παράρτημα), για το οποίο υπήρχε διαθέσιμη μια ερευνητική εργασία (Makarou et al. 2012). Πιθανώς, επιπλέον μελλοντικές εργασίες μπορεί να αυξήσουν την αξιοπιστία της συμμετοχής αυτών των πρωτεϊνών σε αυτό το σύμπλοκο. Επίσης, το 34% αυτών των 240 πρωτεϊνών είχαν ταυτοποιηθεί σε μελέτες απομόνωσης των συστατικών ολόκληρου του σωματίου συναρμογής (Πιν. 1 Παράρτημα). Οι παράγοντες αυτοί μπορεί να ταυτοποιήθηκαν στις μελέτες απομόνωσης ολόκληρου του σωματίου και όχι των επιμέρους υποσυμπλόκων πιθανώς λόγω της παροδικής τους συνεισφοράς ή της ασθενούς σύνδεσής τους σε ένα ή πολλά διαφορετικά στάδια της συναρμογής. Πρέπει να ληφθεί υπόψη ότι τα ενδιάμεσα στάδια έχουν καθοριστεί ως μεμονωμένα υποσύμπλοκα *in vitro* χωρίς να υπάρχει διαθέσιμη πληροφορία για τον τρόπο με τον οποίο η διαδικασία αυτή λαμβάνει χώρα *in vivo*. Βέβαια, η υπόθεση αυτή δεν αποτελεί την μόνη εξήγηση καθώς αυτές οι πρωτεΐνες μπορεί να αποτελούν και ψευδώς θετικά αποτελέσματα της βιοχημικής μεθόδου που χρησιμοποιήθηκε για τη απομόνωση τους. Ένας συμπληρωματικός τρόπος αξιολόγησης της αξιοπιστίας τους ως συστατικά του σωματίου συναρμογής είναι να γίνει η διερεύνηση των λειτουργικών διαδικασιών στις οποίες αυτές συμμετέχουν και να ελεγχθεί αν αυτές οι πρωτεΐνες ενσωματώνονται στο δίκτυο πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής τα οποία θα περιγράψουν παρακάτω (Ενότητα 3.2.1.3.).

#### 3.2.1.1.1. Λειτουργική κατηγοριοποίηση των συστατικών του σωματίου συναρμογής

Στη συνέχεια, πραγματοποιήθηκε ο καθορισμός της βιολογικής λειτουργίας όλων των συστατικών του τελικού συνόλου δεδομένων για τη διερεύνηση όλων των πιθανών βιολογικών διαδικασιών στις οποίες συμμετέχουν. Η λειτουργική αυτή κατηγοριοποίηση θα ήταν χρήσιμη για την ένταξη της πρωτεΐνης FRA10AC1 σε λειτουργικές διεργασίες σύμφωνα με το βιολογικό ρόλο των αλληλεπιδρώντων μορίων της «γειτονιάς» της.

Η λειτουργική αυτή κατάταξη επιτεύχθηκε, μέσω της βάσης δεδομένων DAVID η οποία προσφέρει πληροφορίες για την οντολογία των γονιδίων (βλ. Υλικά & Μέθοδοι). Η κατηγορία γονιδιακής οντολογίας που χρησιμοποιήθηκε ήταν αυτή της βιολογικής διαδικασίας (biological process). Η ανάλυση απέδωσε οντολογικούς όρους στις 488 πρωτεΐνες του συνόλου του σωματίου συναρμογής ώστε οι υπόλοιπες 142 να θεωρηθούν ότι δεν έχουν γνωστό βιολογικό ρόλο. Αναλύοντας τα αποτελέσματα παρατηρήθηκε ότι ορισμένες πρωτεΐνες δεν σχετίζονταν με το όρο γονιδιακής οντολογίας που τους είχε αποδοθεί αλλά συμμετείχαν σε διαφορετικές λειτουργικές διαδικασίες. Για παράδειγμα, μπορεί σε μία πρωτεΐνη να είχε αποδοθεί η οντολογία «Nuclear mRNA splicing via spliceosome» αλλά η μόνη πληροφορία για τη συμμετοχή της στη διαδικασία της συναρμογής να ήταν η απομόνωσή της ως συστατικό του σωματίου συναρμογής. Αυτό συμβαίνει διότι δυο από τους τρόπους προσδιορισμού των οντολογικών κατηγοριών των γονιδίων στη βάση δεδομένων DAVID είναι μερικώς ή τελείως αυτοματοποιημένοι, με την εποπτεία η όχι επιμελητή, αντίστοιχα

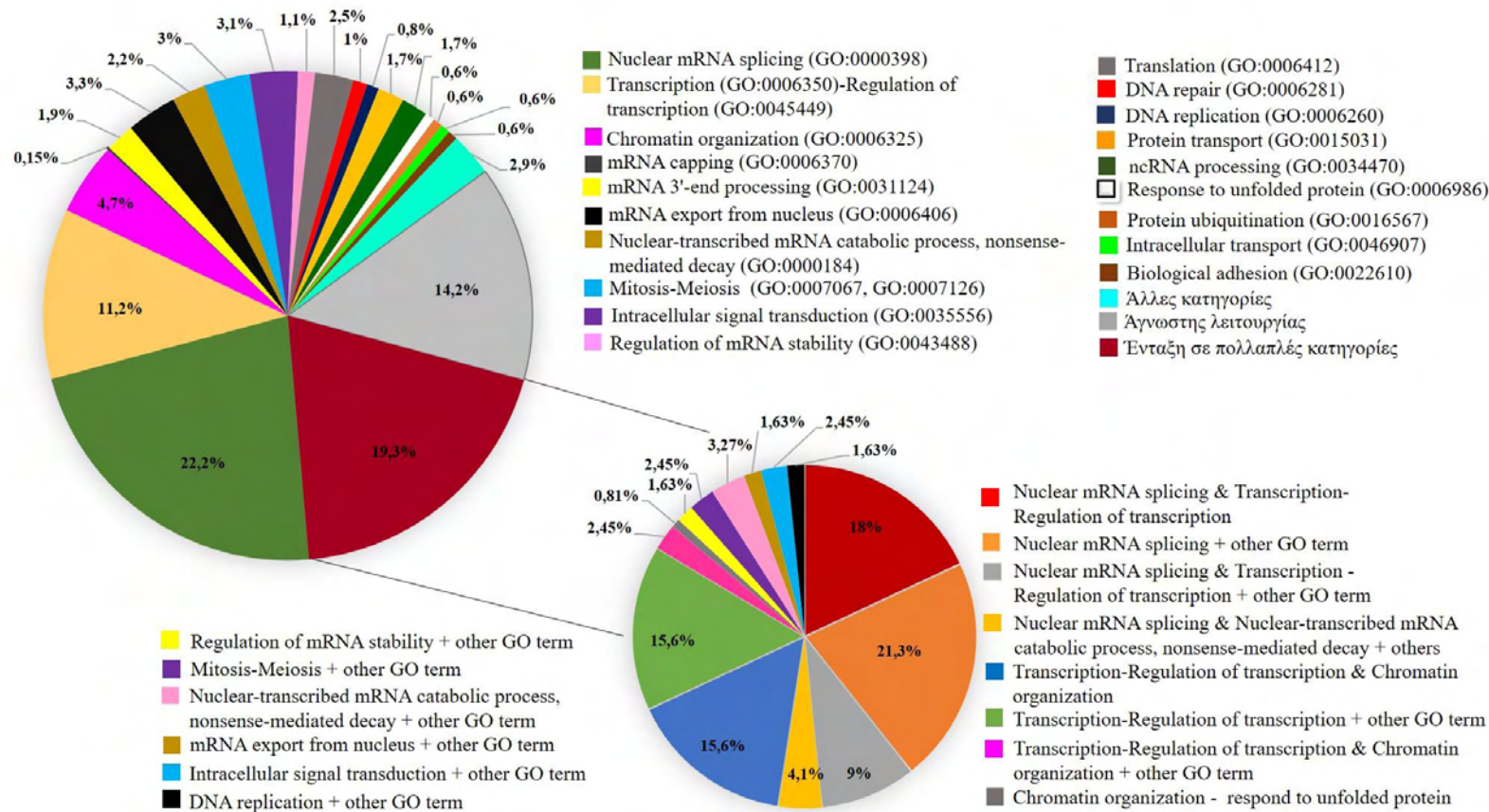
(βλ. Υλικά & Μέθοδοι). Η ανάκτηση των δεδομένων γίνεται μέσω αυτόματης επιμέλειας της βιβλιογραφίας και όχι μέσω διερεύνησης των βιβλιογραφικών αναφορών από τους επιμελητές της βάσης (manually). Γι' αυτό το λόγο τα δεδομένα που ανακτήθηκαν αξιολογήθηκαν περαιτέρω μέσω της λεπτομερούς διερεύνησης (manual curation) της σχετιζόμενης βιβλιογραφίας για κάθε μία πρωτεΐνη. Επιπλέον, επειδή ο αριθμός των πρωτεϊνών άγνωστης λειτουργίας ήταν σχετικά αυξημένος αναζητήθηκαν δεδομένα γονιδιακής οντολογίας γι' αυτές τις πρωτεΐνες στη βάση δεδομένων UniProt αλλά και στη βιβλιογραφία. Με βάση την παραπάνω αξιολόγηση, παρατηρήθηκε ότι 160 πρωτεΐνες δεν σχετίζονταν με τον όρο γονιδιακής οντολογίας που τους αποδίδονταν αλλά συμμετείχαν σε διαφορετικές λειτουργικές διαδικασίες. Ακόμη, με τη διερεύνηση της βιβλιογραφίας, 43 πρωτεΐνες συμπεριλήφθηκαν σε επιπρόσθετους όρους GO γεγονός που δείχνει ότι στις βάσεις που χρησιμοποιήθηκαν δεν ήταν καταχωρημένη η πλήρης πληροφορία σχετικά με τις βιολογικές διαδικασίες στις οποίες συμμετέχουν οι πρωτεΐνες. Οι πρωτεΐνες άγνωστης λειτουργίας μειώθηκαν στις 90. Η βιολογική λειτουργία 18 πρωτεϊνών προσδιορίστηκε μέσω της βιβλιογραφίας και όχι των βάσεων δεδομένων καθώς καμία από τις βάσεις δεν έδινε κάποια πληροφορία. Έτσι, 522 πρωτεΐνες κατηγοριοποιήθηκαν σε δεκαεννέα όρους γονιδιακής οντολογίας (GO terms) (Εικ. 34) (Πιν. 1 Παράρτημα). Αυτοί ήταν:

- Συναρμογή mRNA μέσω του σωματίου συναρμογής στον πυρήνα (Nuclear mRNA splicing via spliceosome, GO:0000398)
- Μεταγραφή και ρύθμιση της μεταγραφής (Transcription & Regulation of transcription, GO:0006350 & GO:0045449)
- Οργάνωση της χρωματίνης (Chromatin organization, GO:0006325)
- Προσθήκη καλύπτρας στο mRNA (mRNA capping, GO:0006370)
- Επεξεργασία του 3' άκρου του mRNA (mRNA 3'-end processing, GO:0031124)
- Έξοδος mRNA από τον πυρήνα (mRNA export from nucleus, GO:0006406)
- Καταβολική διαδικασία αποικοδόμησης του μεταγραφόμενου mRNA στον πυρήνα που διαθέτει πρόωρο κωδικόνιο λήξης (Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay, GO:0000184)
- Μίτωση και μείωση (Mitosis & Meiosis, GO:0007126 & GO:0007067)
- Ενδοκυτταρική μεταγωγή σήματος (Intracellular signal transduction, GO:0035556)
- Ρύθμιση σταθερότητας του mRNA (Regulation of mRNA stability GO:0043488)
- Μετάφραση (Translation, GO:0006412)
- Επιδιόρθωση DNA (DNA repair, GO:0006281)
- Αντιγραφή DNA (DNA replication, GO:0006260)
- Μεταφορά πρωτεΐνης (Protein transport, GO:0015031)
- Επεξεργασία μη κωδικών RNA (ncRNA processing, GO:0034470)
- Απόκριση στη μη αναδιπλωμένη πρωτεΐνη (Response to unfolded protein, GO:0006986)
- Ουβικουιτινίωση πρωτεϊνών (Protein ubiquitination, GO:0016567)
- Ενδοκυτταρική μεταφορά (Intracellular transport, GO:0046907)
- Βιολογική προσκόλληση (Biological adhesion, GO:0022610)



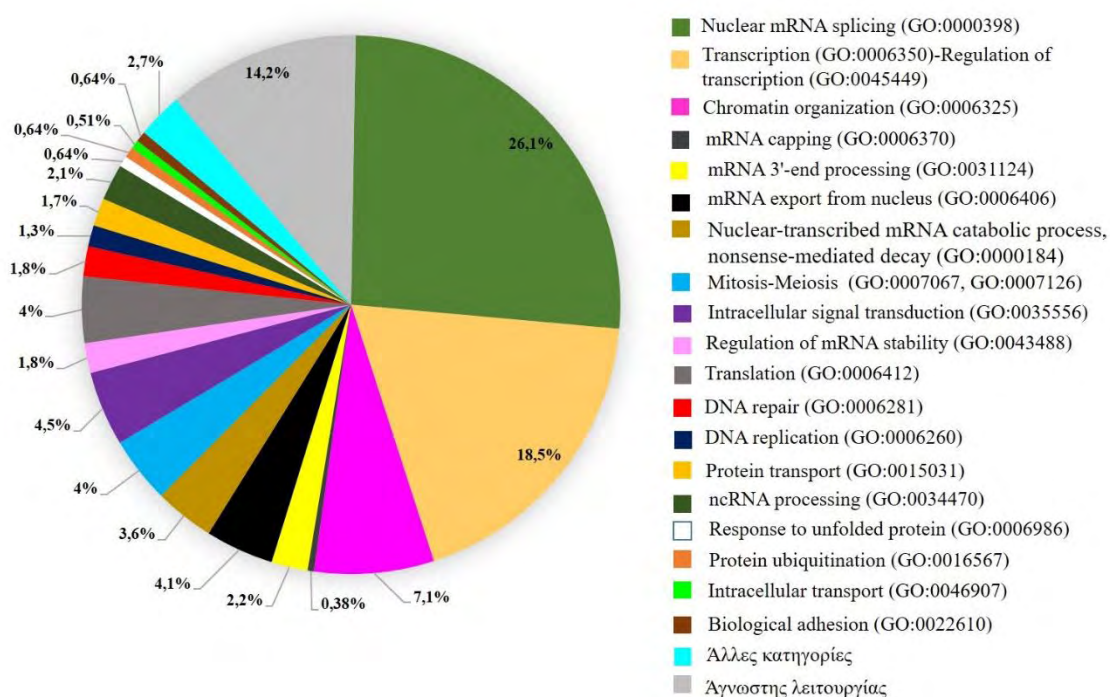
Σύμφωνα με την παραπάνω κατηγοριοποίηση των πρωτεϊνών, το 24% αυτών (122 πρωτεΐνες) ενσωματώθηκαν σε παραπάνω από μια λειτουργικές κατηγορίες (Εικ. 34). Περίπου οι μισές από τις πολυλειτουργικές πρωτεΐνες (39,3%) σχετίστηκαν με τη διαδικασία της συναρμογής και τη μεταγραφή ή με τη μεταγραφή και την οργάνωση της χρωματίνης και οι υπόλοιπες ενσωματώθηκαν μέχρι και σε τέσσερις διαφορετικές κατηγορίες, όπως φαίνεται και στην Εικόνα 34 και αναλυτικότερα στον Πίνακα 12 στην Ενότητα 3.2.1.3.

Έτσι, το σύνολο δεδομένων του σωματίου συναρμογής περιλάμβανε πρωτεΐνες που σχετίζονται με τη συναρμογή: α) παράγοντες συναρμογής, όπως πρωτεΐνες που είναι συστατικά των snRNPs ή σχετιζόμενες με αυτά, β) πρωτεΐνες hnRNPs οι οποίες βοηθούν στην συναρμογή και στην εναλλακτική συναρμογή, γ) πρωτεΐνες Sm και Lsm που βοηθούν στη συναρμολόγηση των snRNPs, και δ) ελικάσες που είναι απαραίτητες καθ' όλη τη διάρκεια συναρμολόγησης και αποσυναρμολόγησης του σωματίου συναρμογής στο πρόδρομο μετάγραφο, καθώς επιτρέπουν την αναδιαμόρφωση των μορίων RNA και των πρωτεϊνών για την επίτευξη των δυο καταλυτικών βημάτων της συναρμογής. Επίσης, περιείχε μεταγραφικούς παράγοντες, πρωτεΐνες που συμμετέχουν στην προσθήκη καλύπτρας στο 5' άκρο του μεταγράφου, στην επεξεργασία του 3' άκρου του μεταγράφου (προσθήκη ουράς πολυ-A), στην αναδιαμόρφωση της χρωματίνης, στην αποικοδόμηση των mRNA (στα οποία δεν έχει γίνει σωστή επεξεργασία για να μπορεί να συνεχιστεί η μετάφρασή τους) καθώς και συστατικά που συμμετέχουν στην έξοδο του mRNA από τον πυρήνα. Η ενσωμάτωση πρωτεϊνών σε αυτές τις κατηγορίες, ήταν αναμενόμενη καθώς οι διαδικασίες επεξεργασίας του mRNA είναι λειτουργικά σχετιζόμενες και αλληλορυθμιζόμενες μεταξύ τους, όπως αναφέρθηκε και στην Ενότητα 1.4.2. της Εισαγωγής. Τέλος, το σύνολο δεδομένων περιλάμβανε και 90 πρωτεΐνες που δεν έχουν γνωστό βιολογικό ρόλο.



**Εικ. 34: Λειτουργική ομαδοποίηση των 630 πρωτεϊνών-Uniprot IDs του σωματίου συναρμογής.** Στο πάνω αριστερό μέρος απεικονίζονται τα ποσοστά των πρωτεϊνών που ανήκουν σε κάθε όρο γονιδιακής οντολογίας, των πρωτεϊνών άγνωστης λειτουργίας και αυτών που ανήκουν σε άλλες βιολογικές κατηγορίες οι οποίες δεν εντάσσονται σε κάποιο από τους παραπάνω όρους, με αντίστοιχο χρωματικό κώδικα. Στο κάτω δεξί μέρος φαίνονται τα ποσοστά των πολυλειτουργικών πρωτεϊνών και οι διαφορετικοί όροι γονιδιακής οντολογίας στους οποίους ανήκουν. Other GO term: Επιπλέον όροι γονιδιακής οντολογίας.

Όπως αναφέρθηκε και προηγουμένως, το 24% των πρωτεϊνών ενσωματώθηκαν σε περισσότερους από έναν όρους γονιδιακής οντολογίας και η λειτουργική κατηγοριοποίησή αυτών έγινε ξεχωριστά από τις υπόλοιπες πρωτεΐνες οι οποίες ανήκουν σε μία κατηγορία GO (Εικ. 34). Η ξεχωριστή λειτουργική κατάταξη παρουσιάζει ένα μειονέκτημα καθώς δεν δίνει μία ολοκληρωμένη εικόνα της αντιπροσώπευσης κάθε οντολογικής ομάδας. Για παράδειγμα, μια πρωτεΐνη μπορεί να ανήκει στην οντολογική κατηγορία της συναρμογής και σε άλλες τρεις κατηγορίες GO και μια άλλη μόνο στην οντολογική κατηγορία της συναρμογής. Αν η πρώτη, επειδή είναι πολυλειτουργική, δεν προσμετρηθεί στο ποσοστό της συναρμογής, τότε το ποσοστό αυτής της κατηγορίας GO θα υποαντιπροσωπεύοταν. Σε αυτό το πλαίσιο, όλες οι πρωτεΐνες του τελικού συνόλου, ανεξάρτητα αν ήταν πολυλειτουργικές, προσμετρήθηκαν στην κάθε κατηγορία GO στην οποία είχαν ενσωματωθεί (Εικ. 35). Ο υπολογισμός των αντίστοιχων ποσοστών έγινε με βάση το άθροισμα των οντολογιών και όχι τον αριθμό του συνόλου των πρωτεϊνών του σωματίου συναρμογής (Εικ. 35). Ωστόσο, το ποσοστό των πρωτεϊνών άγνωστης λειτουργίας μετρήθηκε σύμφωνα με τον αριθμό του συνόλου των πρωτεϊνών διότι αυτές οι πρωτεΐνες δεν ενσωματώθηκαν σε κανέναν οντολογικό όρο. Έτσι ο υπολογισμός του αντίστοιχου ποσοστού ως προς τον αριθμό των οντολογιών δεν θα ήταν σωστός διότι το ποσοστό των πρωτεϊνών άγνωστης λειτουργίας θα φαινόταν μικρότερο δεδομένου του μεγαλύτερου αριθμού οντολογιών.



**Εικ. 35: Ομαδοποίηση των κατηγοριών γονιδιακής οντολογίας του συνόλου δεδομένων του σωματίου συναρμογής.** Απεικονίζονται τα ποσοστά των κατηγοριών γονιδιακής οντολογίας που αντιπροσωπεύονται από τις πρωτεΐνες του σωματίου συναρμογής, ακολουθώντας τον ίδιο χρωματικό κώδικα της ομαδοποίησης της Εικόνας 34.

Συγκρίνοντας τις Εικόνες 34 και 35 παρατηρείται, όπως ήταν αναμενόμενο, ότι το ποσοστό εκπροσώπησης κάθε οντολογικής κατηγορίας αυξάνεται.

Τέλος, ξεχωριστά μελετήθηκε η λειτουργική κατηγοριοποίηση των 240 πρωτεϊνών του σωματίου συναρμογής οι οποίες ταυτοποιήθηκαν σε μια μόνο ερευνητική εργασία. Με αυτόν τον τρόπο αξιολογήθηκε η συμμετοχή τους στο σύνολο δεδομένων. Παρατηρήθηκε ότι-με εξαίρεση τις 35 πρωτεΐνες που ήταν άγνωστης λειτουργίας-οι οντολογικοί όροι που κυριάρχησαν σε ποσοστό 60% ήταν αυτοί της συναρμογής και των σχετιζόμενων λειτουργιών, κυρίως της μεταγραφής ή της ρύθμισής της και της αναδιαμόρφωσης της χρωματίνης. Το γεγονός αυτό υποστήριξε την ταυτοποίησή τους ως συστατικά του σωματίου συναρμογής (Πιν. 1 Παράρτημα). Οι υπόλοιπες ενσωματώθηκαν σε ποικίλους οντολογικούς όρους, όπως φαίνεται και στο Παράρτημα 1.

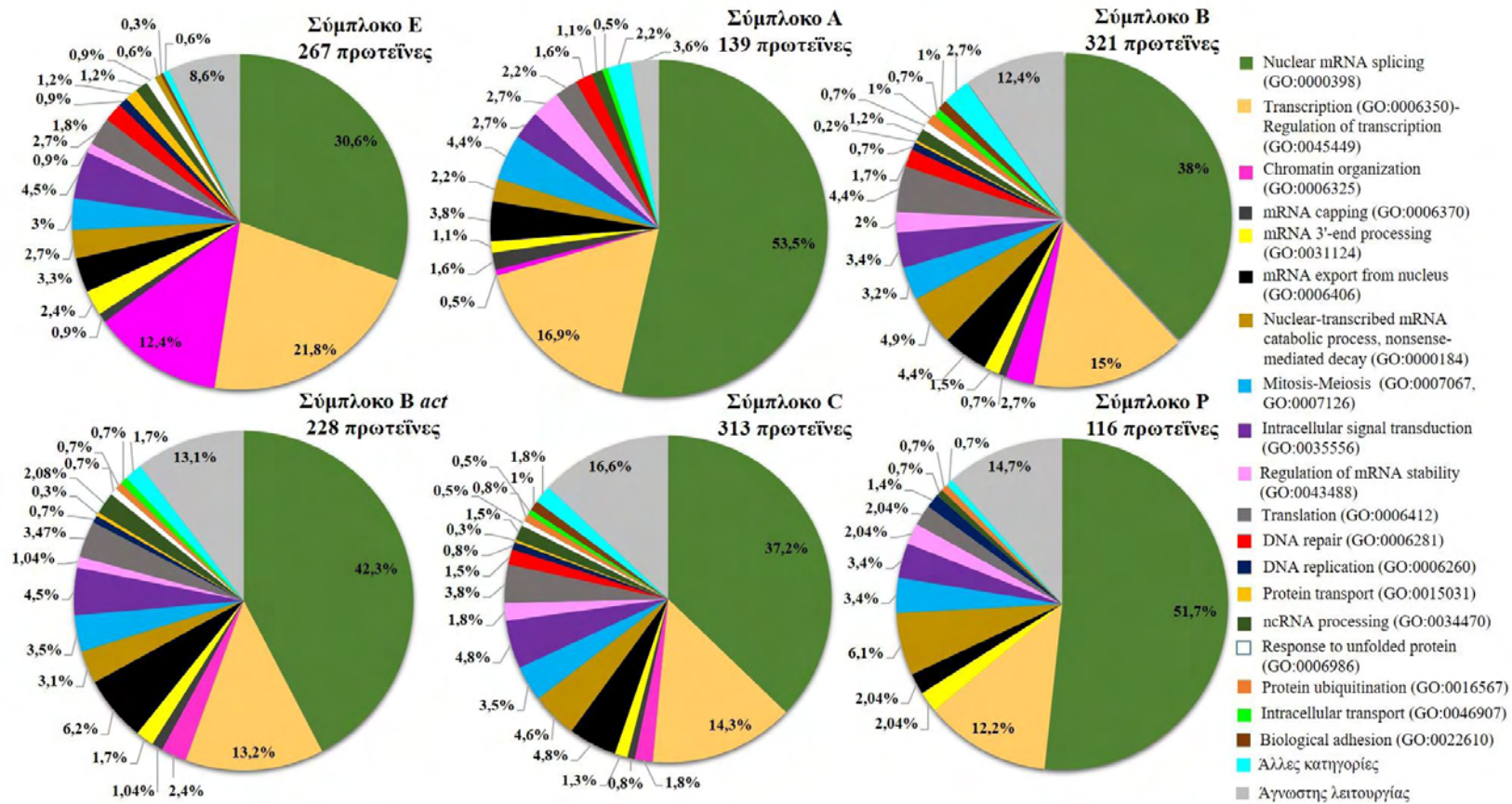
### *3.2.1.1.2. Η πρωτεϊνική σύσταση των υποσυμπλόκων του σωματίου συναρμογής*

Όπως αναφέρθηκε και στην εισαγωγή της ενότητας αυτής, η συναρμολόγηση των συστατικών του σωματίου συναρμογής αποτελεί μια δυναμική κατάσταση, δηλαδή είναι μια πολυβηματική διαδικασία κατά την οποία διαφορετικοί παράγοντες στρατολογούνται ή απελευθερώνονται από το mRNA σε διαφορετικό χρόνο με στόχο την εκτέλεση μιας κοινής ή διαφορετικής ενέργειας κάθε φορά έτσι ώστε να επιτευχθούν τα δυο καταλυτικά βήματα της συναρμογής. Τα διαφορετικά αυτά βήματα διεκπεραιώνονται από έξι επιμέρους υποσύμπλοκα (E, A, B, B *act*, C και P), όπως περιγράφηκε αναλυτικά και στην ενότητα 1.4. της Εισαγωγής. Οι ερευνητικές εργασίες που αφορούσαν την απομόνωση των συστατικών των επιμέρους αυτών υποσυμπλόκων ήταν δεκατρείς με το μεγαλύτερο αριθμό αυτών να αναφέρεται στα υποσύμπλοκα B και C (βλ. Πιν. 8 & Πιν. 1 παράρτημα). Για τα σύμπλοκα A και B *act* υπήρχαν τρεις ερευνητικές εργασίες για το καθένα ενώ για τα σύμπλοκα E και P μόνο μία για το καθένα (Πιν. 8 & Πιν. 1 παράρτημα). Έχοντας συγκεντρώσει και συγκρίνει όλα τα στοιχεία από όλες αυτές τις ερευνητικές εργασίες με τον τρόπο που αναφέρθηκε παραπάνω, η τελική σύσταση των υποσυμπλόκων E, A, B, B *act*, C και P σε πρωτεΐνες (UniProt IDs) ήταν κατ' αντιστοιχία 267, 139, 321, 228, 313 και 116 (Πιν. 8 & Πιν. 1 παράρτημα). Συνολικά, το 87% του συνόλου δεδομένων του σωματίου συναρμογής (549 πρωτεΐνες-UniProt IDs) κατηγοριοποιήθηκαν ως συστατικά ενός ή περισσότερων υποσυμπλόκων. Οι υπόλοιπες 81 πρωτεΐνες-UniProt IDs είχαν χαρακτηριστεί μόνο ως συστατικά ολόκληρου του σωματίου συναρμογής (Πιν. 1 παράρτημα), πιθανώς, λόγω της παροδικής τους σύνδεσης σε συγκεκριμένο βήμα της συναρμολόγησης του σωματίου ή της συμμετοχής τους σε μεταβατικό στάδιο από το ένα σύμπλοκο στο άλλο με αποτέλεσμα να μην είναι ανιχνεύσιμα κατά την απομόνωση των συστατικών του εκάστοτε συμπλόκου (Πιν. 1 παράρτημα). Όλες οι πρωτεΐνες αυτές έχουν ταυτοποιηθεί σε μία μόνο ερευνητική εργασία.

### *3.2.1.1.3. Λειτουργική κατηγοριοποίηση των συστατικών των υποσυμπλόκων του σωματίου συναρμογής*

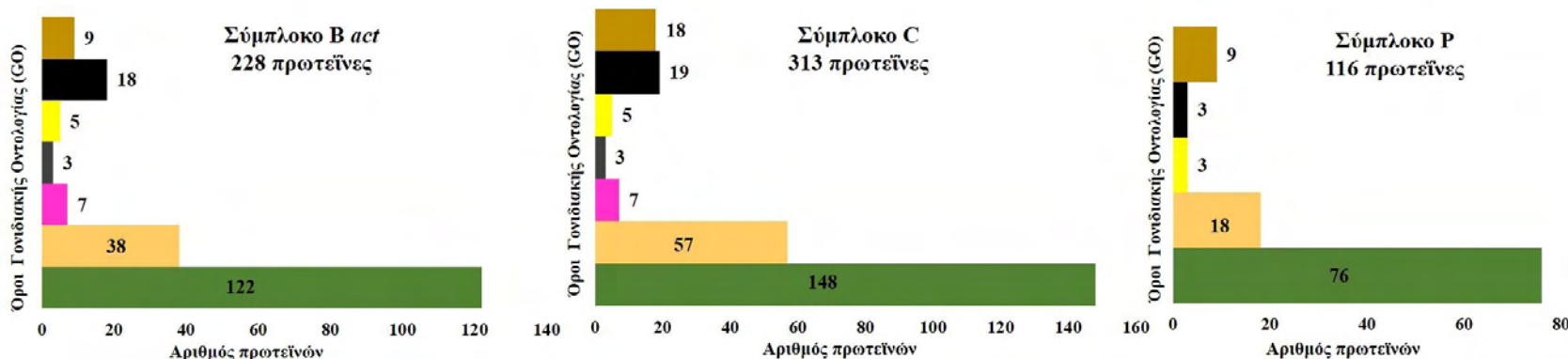
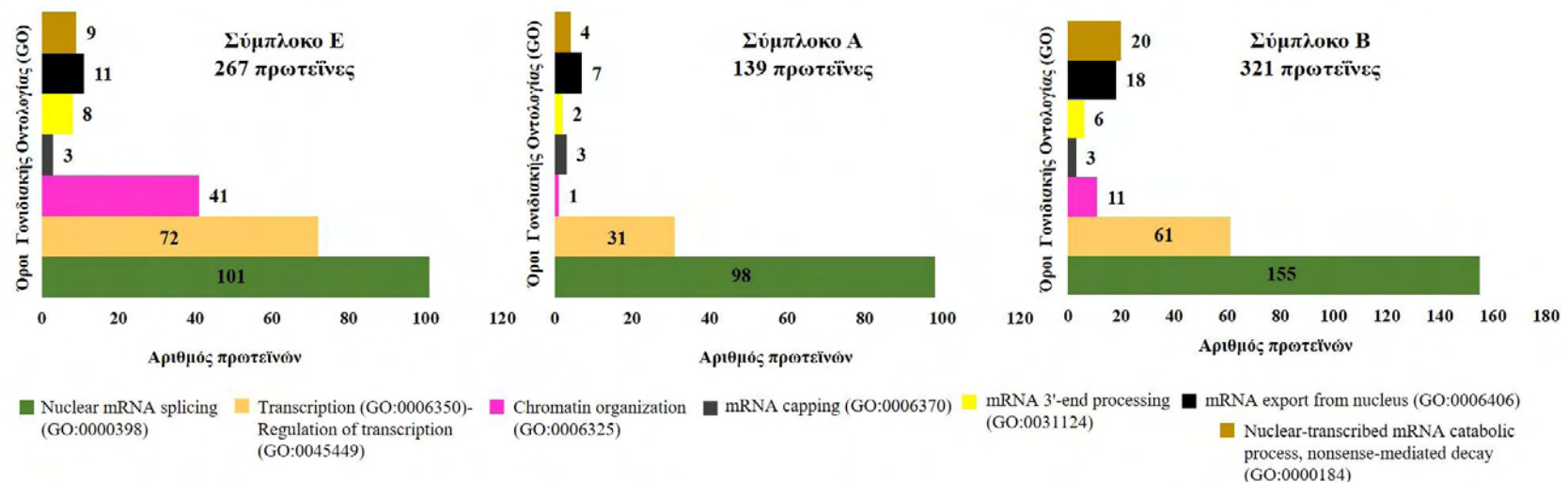
Όσον αφορά τις πρωτεΐνες που απομονώθηκαν ως συστατικά των επιμέρους υποσυμπλόκων, ακολούθησε η λειτουργική κατηγοριοποίησή τους σε κάθε

υποσύμπλοκο ξεχωριστά, έτσι ώστε να ελεγχθεί η εκπροσώπηση της εκάστοτε οντολογικής κατηγορίας στα διαφορετικά στάδια συναρμολόγησης του σωματίου συναρμογής (Εικ. 36). Στην Εικόνα 36 παρουσιάζονται τα αποτελέσματα αυτής της διερεύνησης, δηλαδή τα ποσοστά των οντολογικών όρων στα έξι διαφορετικά υποσύμπλοκα ως προς το συνολικό αριθμό αυτών διότι και στην περίπτωση των υποσυμπλόκων παρατηρήθηκε ότι υπάρχουν πολυλειτουργικές πρωτεΐνες με αποτέλεσμα αυτές να προσμετρώνται σε παραπάνω από μία κατηγορία.



**Εικ. 36: Ομαδοποίηση των κατηγοριών γονιδιακής οντολογίας των επιμέρους υποσυμπλόκων του σωματίου συναρμογής.** Απεικονίζονται τα ποσοστά των κατηγοριών γονιδιακής οντολογίας που αντιπροσωπεύονται σε κάθε επιμέρους υποσύμπλοκο του σωματίου συναρμογής, ακολουθώντας τον ίδιο χρωματικό κώδικα της ομαδοποίησης που χρησιμοποιήθηκε και σε ολόκληρο το σωματίο συναρμογής.

Κάθε υποσύμπλοκο περιέχει διαφορετικό αριθμό πρωτεϊνών, όπως φαίνεται και στην Εικόνα 36, γεγονός που τροποποιεί τα ποσοστά των εκάστοτε οντολογικών κατηγοριών. Δηλαδή η σύγκριση των ποσοστών δεν δίνει μια ξεκάθαρη εικόνα για τη λειτουργική κατηγοριοποίηση των συστατικών των υποσυμπλόκων διότι τα ποσοστά δεν αναλογούν σε απόλυτες τιμές. Συγκεκριμένο παράδειγμα αποτελεί η οντολογική κατηγορία της συναρμογής η οποία παρουσιάζεται ιδιαίτερα αυξημένη στα υποσύμπλοκα A, B *act* και P. Παρατηρώντας τα αντίστοιχα ποσοστά μπορεί να οδηγηθεί κανείς στο συμπέρασμα ότι τα υποσύμπλοκα αυτά είναι ιδιαίτερα εμπλουτισμένα σε παράγοντες συναρμογής. Ωστόσο, το γεγονός αυτό δεν ανταποκρίνεται στην πραγματικότητα διότι αυτά τα υποσύμπλοκα δεν διαθέτουν μεγαλύτερο αριθμό τέτοιων παραγόντων αλλά το ποσοστό τους αυξάνεται λόγω του συνολικού μικρότερου αριθμού πρωτεϊνών που περιέχουν σε σύγκριση με τα υπόλοιπα υποσύμπλοκα. Για να γίνει, λοιπόν, η ακριβέστερη σύγκριση μεταξύ των οντολογιών των επιμέρους υποσυμπλόκων, στην Εικόνα 37 παρουσιάζονται, σε μορφή ιστογράμματος, οι απόλυτοι αριθμοί των πρωτεϊνών στις επτά κύριες οντολογικές κατηγορίες.



**Εικ. 37:** Απόλυτες τιμές πρωτεϊνών κάθε οντολογικής κατηγορίας στα επιμέρους υποσύμπλοκα του σωματίου συναρμογής. Απεικονίζεται ο αριθμός των πρωτεϊνών των έξι υποσυμπλόκων του σωματίου συναρμογής που κατηγοριοποιούνται λειτουργικά στη συναρμογή και σε λειτουργικά σχετιζόμενες διαδικασίες (ίδιος χρωματικός κώδικας με την Εικόνα 36).



Μελετώντας συνδυαστικά τα αποτελέσματα που προέκυψαν από την παραπάνω λειτουργική κατηγοριοποίηση, όπως αυτά αποτυπώθηκαν τόσο ως ποσοστά (Εικ. 36) όσο και σε απόλυτους αριθμούς (Εικ. 37), παρατηρήθηκε, αρχικά, ότι το υποσύμπλοκο E ήταν εμπλουτισμένο σε παράγοντες οργάνωσης και αναδιαμόρφωσης της χρωματίνης (12,4%, 41 πρωτεΐνες) οι οποίοι μειώνονται σημαντικά στα υπόλοιπα υποσύμπλοκα. Οι αντίστοιχες τιμές είναι ιδιαίτερα χαμηλές στα υποσύμπλοκα A και P και αυξημένες στα B *act*, B και C. Η χαμηλή τιμή των παραγόντων αυτών στο υποσύμπλοκο P είναι αναμενόμενη διότι η ύπαρξή τους δεν συνάδει λειτουργικά με το τελικό στάδιο της συναρμολόγησης του σωματίου κατά το οποίο έχει απομακρυνθεί ο μεγαλύτερος αριθμός πρωτεϊνών από το mRNA. Ο μικρός αριθμός των παραγόντων οργάνωσης και αναδιαμόρφωσης της χρωματίνης στο υποσύμπλοκο A μπορεί να οδηγούσε στο συμπέρασμα ότι αυτές οι πρωτεΐνες απομακρύνονται και επαναστρατολογούνται σε συγκεκριμένο στάδιο της συναρμολόγησης του σωματίου συναρμογής. Ωστόσο, για το συγκεκριμένο υποσύμπλοκο έχουν πραγματοποιηθεί λιγότερες πειραματικές μελέτες απομόνωσης των συστατικών του γεγονός που μπορεί να σημαίνει ότι κάποιες πρωτεΐνες ανήκουν σε αυτό το υποσύμπλοκο αλλά δεν έχουν ταυτοποιηθεί έως τώρα ως συστατικά του. Τον σημαντικά αυξημένο αριθμό αυτών των πρωτεϊνών στο σύμπλοκο E (Εικ. 36 & Εικ. 37) μπορεί να εξηγήσει το γεγονός ότι παράγοντες αναδιαμόρφωσης της χρωματίνης έχουν συσχετιστεί με τη συναρμολόγηση των αρχικών υποσυμπλόκων (pre-spliceosome) (Allemand et al. 2016, Kfir et al. 2015, Haque & Oberdoerffer 2014) (Ενότητα 1.4.2.). Αυτή η συσχέτιση είναι έμμεση και γίνεται μέσω της διαδικασίας της μεταγραφής η οποία έχει μια καλά εδραιωμένη λειτουργική σχέση τόσο με την συναρμογή όσο και με τους παράγοντες οργάνωσης και αναδιαμόρφωσης της χρωματίνης (Merkhofer et al. 2014, Lee et al. 2016) (Ενότητα 1.4.2.).

Όσον αφορά τις πρωτεΐνες που συμμετέχουν στη διαδικασία της συναρμογής παρατηρήθηκε ότι ο αριθμός τους αυξάνεται στα υποσύμπλοκα B, B *act* και C σε σύγκριση με τα υπόλοιπα (Εικ. 39 & Εικ. 40). Ο αυξημένος αριθμός τους είναι αναμενόμενος διότι κατά το σχηματισμό αυτών, συγκεκριμένα στο B *act* και C, λαμβάνουν χώρα τα δυο καταλυτικά βήματα της συναρμογής. Στο υποσύμπλοκο B *act* ο αριθμός των πρωτεϊνών είναι μικρότερος σε σχέση με τα B και C, ίσως λόγω των λιγότερων ερευνητικών εργασιών ή του γεγονότος ότι αποτελεί μεταβατικό στάδιο κατά το οποίο απελευθερώνονται τα U1 και U4 snRNP για πραγματοποιηθεί η πρώτη αντίδραση μετεστεροποίησης και να σχηματιστεί το υποσύμπλοκο C. Στο υποσύμπλοκο P παρατηρείται ο μικρότερος αριθμός παραγόντων συναρμογής, το οποίο είναι αναμενόμενο διότι σε αυτό έχουν ήδη πραγματοποιηθεί τα δυο καταλυτικά βήματα, έχει γίνει η απομάκρυνση του ιντρονίου και η ένωση των εξονίων και έχει ξεκινήσει η αποσυναρμολόγηση του σωματίου συναρμογής. Αξίζει να σημειωθεί ότι συγκρίνοντας τις πρωτεΐνες αυτού του υποσυμπλόκου, παρατηρήθηκε ότι όλες οι πρωτεΐνες, με εξαίρεση τρεις που είναι παρούσες και στα άλλα υποσύμπλοκα, είναι κοινές με αυτές του υποσυμπλόκου C.

Τα υποσύμπλοκα B, B *act* και C παρατηρείται, επίσης, ότι είναι εμπλουτισμένα σε πρωτεΐνες που συμμετέχουν στην έξοδο του mRNA από τον πυρήνα και στην αποικοδόμησή του στον πυρήνα όταν δεν είναι κατάλληλο να μεταφραστεί (όπως

παρατηρείται και ως ποσοστά και σε απόλυτες τιμές) (Εικ. 36 & Εικ. 37). Το αποτέλεσμα αυτό μπορεί να εξηγηθεί από το γεγονός ότι οι πρωτεΐνες αυτών των δυο κατηγοριών στρατολογούνται κατά τη διάρκεια αυτών των σταδίων ώστε μετά το πέρας της συναρμογής να επιλέξουν και να κατευθύνουν το ώριμο πλέον mRNA προς τους πυρηνικούς πόρους για τη μετέπειτα μετάφρασή του στο κυτταρόπλασμα, και παράλληλα να ελέγξουν την καταλληλότητά του για μετάφραση (βλ. Ενότητα 1.4.2. & 4.2.6.).

Ολοκληρώνοντας τη σύγκριση των οντολογικών κατηγοριών των επιμέρους υποσυμπλόκων, παρατηρούμε ότι ο αριθμός των πρωτεϊνών που συμμετέχουν στη διαδικασία της μεταγραφής ή της ρύθμισης αυτής δεν αλλάζει σημαντικά σε κάποιο συγκεκριμένο υποσύμπλοκο (Εικ. 36 & Εικ. 37), το οποίο θεωρείται αναμενόμενο διότι η μεταγραφή συμβαίνει ταυτόχρονα με τη συναρμογή (βλ. Ενότητα 1.4.2.).

Ωστόσο, όλα τα παραπάνω θα πρέπει να εξεταστούν και να αξιολογηθούν και στα πλαίσια των δικτύων των πρωτεϊνικών αλληλεπιδράσεων των επιμέρους υποσυμπλόκων για να ανακτηθεί μια πλήρης εικόνα της αλλαγής του αριθμού των οντολογικών κατηγοριών στα διαφορετικά στάδια συναρμολόγησης του σωματίου συναρμογής.

#### *3.2.1.1.4. Προσδιορισμός και αφθονία των πρωτεϊνών με καθολική παρουσία κατά τη συναρμογή*

Όπως αναφέρθηκε και παραπάνω, η συναρμολόγηση του σωματίου συναρμογής δεν αποτελεί μια στατική κατάσταση, δηλαδή οι διαφορετικές πρωτεΐνες συμμετέχουν σε αυτή τη διαδικασία σε διαφορετικό στάδιο και χρόνο. Ωστόσο, κάποιοι παράγοντες είναι απαραίτητοι καθ' όλη τη διάρκειά της καθώς αποτελούν συστατικά των snRNPs, όπως οι πρωτεΐνες Sm και Lsm οι οποίες βοηθούν στη συναρμολόγηση αυτών. Επιπλέον, ομάδες πρωτεϊνών μπορεί να εμφανίζονται σε υψηλή αφθονία σε ένα συγκεκριμένο στάδιο διότι συμμετέχουν σε ένα πολύ συγκεκριμένο βήμα της συναρμογής (π.χ. 1<sup>ο</sup> καταλυτικό βήμα) και να μειώνεται η ποσότητα τους στα υπόλοιπα βήματα.

Για να πραγματοποιηθεί λοιπόν η διερεύνηση των παραγόντων που διαδραματίζουν σημαντικό ρόλο σε όλη τη διάρκεια της διαδικασίας έγινε, σε πρώτο στάδιο, ο έλεγχος των πρωτεϊνών που εμφανίζουν καθολική παρουσία καθ' όλη τη διάρκεια της συναρμογής. Ειδικότερα, πραγματοποιήθηκε η σύγκριση των πρωτεϊνών που είχαν ταυτοποιηθεί στα υποσύμπλοκα A, B, B *act* και C, αποκλείοντας τα υποσύμπλοκα E και P για τους εξής λόγους: στο υποσύμπλοκο E γίνεται η απαραίτητη «προετοιμασία» μέσω της αναδιαμόρφωσης της χρωματίνης για να δοθεί η κατάλληλη «προσβασιμότητα» για την έναρξη και εξέλιξη της μεταγραφής και την ταυτόχρονη στρατολόγηση των παραγόντων συναρμογής έτσι ώστε να ξεκινήσει η διαδικασία. Στην περίπτωση του υποσυμπλόκου P, ένας επιπλέον λόγος ήταν ότι αποτελούσε υποσύνολο του υποσυμπλόκου C, γεγονός το οποίο δεν μπορούσε να αποδώσει επιπλέον πληροφορία γι' αυτό το σκοπό. Επίσης, υπήρχε μόνο μια ερευνητική εργασία για την ταυτοποίηση των συστατικών του έτσι ώστε να μην είναι δυνατή η επαλήθευση της πληροφορίας που προσέφερε. Συνολικά 80 πρωτεΐνες-Uniprot IDs ήταν παρούσες και στα τέσσερα υποσύμπλοκα A, B, B *act* και C (Πιν. 9, 1<sup>η</sup> στήλη).

Επιπλέον πραγματοποιήθηκε η εκτίμηση της αφθονίας των ογδόντα αυτών πρωτεϊνών, σύμφωνα με τις πληροφορίες που παρείχαν οι αντίστοιχες ερευνητικές εργασίες. Η διαδικασία που ακολουθήθηκε ήταν ίδια με αυτή που περιγράφηκε για τους 15 πρώτους «γείτονες» της FRA10C1 εντός του σωματίου συναρμογής (βλ. Ενότητα 3.1.2.). Τα αποτελέσματα φαίνονται στον Πίνακα 9 (στήλες 2-5). Το 16% των 80 πρωτεϊνών εμφανίζουν υψηλή αφθονία σε όλα τα υποσύμπλοκα (13 πρωτεΐνες), το 21% αυτών υψηλή αφθονία στα τρία πρώτα υποσύμπλοκα (A, B, B *act*) ή μόνο στο υποσύμπλοκο A (17 πρωτεΐνες) και το 12% υψηλή αφθονία στα υποσύμπλοκα B, B *act* και C (10 πρωτεΐνες) (Πιν. 9).

Στη συνέχεια έγινε η διερεύνηση των όρων γονιδιακής οντολογίας στις οποίες ανήκουν αυτές οι 80 πρωτεΐνες. Η ανάλυση αυτή έδειξε ότι όλες οι πρωτεΐνες υψηλής αφθονίας συμμετέχουν στη διαδικασία της συναρμογής ενώ μικρό ποσοστό αυτών έχουν και επιπρόσθετες λειτουργίες (Πιν. 9, τελευταία στήλη). Το γεγονός αυτό υποδεικνύει ότι οι πρωτεΐνες αυτές αποτελούν βασικούς παράγοντες συναρμογής που είναι απαραίτητοι καθ' όλη τη διάρκεια της διαδικασίας. Όσον αφορά τις πρωτεΐνες με μέτρια ή χαμηλή αφθονία, παρατηρήθηκε ότι και αυτές συμμετέχουν κυρίως στη συναρμογή ή στη μεταγραφή/ρύθμιση της μεταγραφής (Πιν. 9), γεγονός που συνάδει με το ότι η συναρμογή πραγματοποιείται ταυτόχρονα με τη μεταγραφή. Επιπλέον, μόνο ένα πολύ μικρό ποσοστό των πρωτεϊνών μέτριας ή χαμηλής αφθονίας συμμετείχαν σε άλλες σχετιζόμενες λειτουργικές διαδικασίες με τη συναρμογή (Πιν. 9). Στον Πίνακα 9 φαίνονται όλες οι πρωτεΐνες που ήταν παρούσες σε όλα τα υποσύμπλοκα, η αφθονία τους σε κάθε υποσύμπλοκο με αναφορά στην αντίστοιχη ερευνητική εργασία και στην/στις κατηγορίες της βιολογικής διαδικασίας στις οποίες συμμετέχουν.

**Πιν. 9: Θερμικός πίνακας (heat-map) των πρωτεϊνών με καθολική παρουσία στα υποσύμπλοκα A, B, B act και C του σωματίου συναρμογής.**

Όνομα πρωτεΐνης	Σύμπλοκο A (Agafonov et al., 2011, Hartmuth et al., 2002, Behzadnia et al., 2007)	Σύμπλοκο B (Agafonov et al., 2011, Bessonov et al., 2008)	Σύμπλοκο B act (Agafonov et al., 2011, Bessonov et al., 2010)	Σύμπλοκο C (Agafonov et al., 2011, Bessonov et al., 2008, Bessonov et al., 2010)	GO (Βιολογική διαδικασία)	
SNRPA1	High	High	High	High	Συναρμογή του mRNA	
SNRPB2	High	High	High	High		
SNRPB	High	High	High	High		
SNRPD1	High	High	High	High		
SNRPD2	High	High	High	High		
SNRPD3	High	High	High	High		
SNRPE	High	High	High	High		
SNRPF	High	High	High	High		
SNRPG	High	High	High	High		
SNRNP40	High	High	High	High		
SRSF1	High	High	High	High	Συναρμογή του mRNA + άλλα GO	
NCBP2	High	High	High	High		
SF3A1	High	High	High	Low	Συναρμογή του mRNA	
SF3A2	High	High	High	Low		
SF3A3	High	High	High	Low		
SF3B1	High	High	High	Low		
SF3B2	High	High	High	Low		
SF3B3	High	High	High	Low		
SF3B4	High	High	High	Low		
SF3B6	High	High	High	Low		
PHF5A	High	High	High	Low		
DDX5	High	High	High	Low		
RBM17	High	High	High	Low	Συναρμογή του mRNA + Έξοδος του mRNA από τον πυρήνα	
SNRPA	High	High	High	Low		
SRSF6	High	High	High	Low		
HNRNPA2B1	High	High	High	Low		
HNRNPA1	High	High	High	Low		
SRSF3	High	High	High	Low		
SRSF7	High	High	High	Low		
PNN	High	High	High	Low		
PRPF8	High	High	High	Low		
SNRNP200	High	High	High	Low		Συναρμογή του mRNA
EFTUD2	High	High	High	Low		
PRPF19	High	High	High	Low		
CDC5L	High	High	High	Low		
BCAS2	High	High	High	Low		
CWC15	High	High	High	Low		
CTNBL1	High	High	High	Low		
DHX15	High	High	High	Low		
SNRNP70	High	High	High	Low		
SRSF9	High	High	High	Low	Συναρμογή του mRNA + Μεταγραφή-Ρύθμιση της μεταγραφής	
TRA2B	High	High	High	Low		
SAP18	High	High	High	Low		
HNRNPC	High	High	High	Low		
SRRM2	High	High	High	Low		
HNRNPK	High	High	High	Low		
NCBP1	High	High	High	Low		Συναρμογή του mRNA + Προσθήκη καλύπτρας στο mRNA
PRPF6	High	High	High	Low		
MFAP1	High	High	High	Low		Συναρμογή του mRNA
NHP2L1	High	High	High	Low		
TXNL4A	High	High	High	Low		
LSM3	High	High	High	Low		
PRPF3	High	High	High	Low		
USP39	High	High	High	Low		
LSM6	High	High	High	Low		
SART1	High	High	High	Low		
SRSF5	High	High	High	Low		
CD2BP2	High	High	High	Low		
WBP11	High	High	High	Low	Συναρμογή του mRNA + Αποικοδόμηση του mRNA	
U2SURP	High	High	High	Low		
RBMX	High	High	High	Low	Συναρμογή του mRNA + Επεξεργασία του 3' άκρου του mRNA	
RNPS1	High	High	High	Low		
SRRM1	High	High	High	Low	Συναρμογή του mRNA + Ρύθμιση της σταθερότητας του mRNA	
ELAVL1	High	High	High	Low		
HSPA8	High	High	High	Low	Συναρμογή του mRNA + Ενδοκυτταρική μεταγωγή σήματος	
SYNCRIP	High	High	High	Low		
DDX17	High	High	High	Low	Συναρμογή του mRNA + Μετάφραση	
RBM39	High	High	High	Low		
PCBP1	High	High	High	Low	Συναρμογή του mRNA + Μεταγραφή-Ρύθμιση της μεταγραφής	
DHX9	High	High	High	Low		
TCERG1	High	High	High	Low	Συναρμογή του mRNA + Μεταγραφή-Ρύθμιση της μεταγραφής	
YBX1	High	High	High	Low		
ACIN1	High	High	High	Low	Έξοδος του mRNA από τον πυρήνα	
ALYREF	High	High	High	Low		
HNRNPR	High	High	High	Low	Μεταγραφή-Ρύθμιση της μεταγραφής + other GO	
IK	High	High	High	Low		
SRRT	High	High	High	Low	Μίτωση-Μείωση	
SKIV2L2	High	High	High	Low		
ZCCHC8	High	High	High	Low	Προσθήκη 5' καλύπτρας στο mRNA	
PRCC	High	High	High	Low		
					Επεξεργασία ncRNA	
					Άγνωστης λειτουργίας	
					Άγνωστης λειτουργίας	

Χρωματικός κώδικας: Πρωτεΐνες με υψηλή (κόκκινο γέμισμα), μέτρια (πορτοκαλί) και χαμηλή (πράσινο) αφθονία.

### 3.2.1.1.5. Προσδιορισμός των πρωτεϊνών που εμφανίζονται μόνο σε συγκεκριμένα υποσύμπλοκα

Με στόχο να ανακτηθεί μια προκαταρκτική εικόνα της δυναμικής των αλληλεπιδράσεων κατά τη διάρκεια συναρμολόγησης του σωματίου συναρμογής χρησιμοποιήθηκαν τα δεδομένα που υπήρχαν για την πρωτεϊνική σύσταση των υποσυμπλόκων του. Συγκεκριμένα έγινε η σύγκριση των πρωτεϊνών των τεσσάρων υποσυμπλόκων (A, B, B *act* και C) και βρέθηκαν ποιες από αυτές ήταν μοναδικές σε καθένα υποσύμπλοκο, δεδομένα τα οποία χρησιμοποιήθηκαν στη συνέχεια για την εύρεση των αλληλεπιδράσεων αυτών των πρωτεϊνών στα συγκεκριμένα υποσύμπλοκα (βλ. Ενότητα 3.2.2). Ο αριθμός τους ήταν, κατ' αντιστοιχία, 10, 29, 17 και 30, γεγονός το οποίο υποδεικνύει ότι λίγες πρωτεΐνες στρατολογούνται σε συγκεκριμένα στάδια της διαδικασίας. Για να εκτιμηθεί, σε πρώτο στάδιο, εάν είναι αξιόπιστη η ταυτοποίησή τους σε αυτά τα υποσύμπλοκα, εξετάστηκε ο αριθμός των ερευνητικών εργασιών που τις ταυτοποιεί στα συγκεκριμένα υποσύμπλοκα. Σχεδόν όλες οι μοναδικές πρωτεΐνες των υποσυμπλόκων A (εκτός μίας), B (εκτός τριών), και B *act* ταυτοποιήθηκαν ως συστατικά αυτών σε μία μόνο ερευνητική εργασία. Ο μικρός αριθμός των ερευνητικών εργασιών που ταυτοποιούν τις μοναδικές πρωτεΐνες των υποσυμπλόκων A, B, B *act* μπορεί να υποδεικνύει ότι αυτές οι πρωτεΐνες αποτελούν ψευδώς θετικά αποτελέσματα των αντίστοιχων υποσυμπλόκων. Ωστόσο, η εύρεση των οντολογικών κατηγοριών βιολογικής διαδικασίας των πρωτεϊνών αυτών ενισχύει τη συμμετοχή τους στα συγκεκριμένα στάδια συναρμολόγησης του σωματίου συναρμογής. Ειδικότερα, 6 από τις 10 πρωτεΐνες του υποσυμπλόκου A, 13 από τις 29 πρωτεΐνες του υποσυμπλόκου B (45%) και 5 από τις 17 πρωτεΐνες του υποσυμπλόκου B *act* (30%) συμμετέχουν στη συναρμογή και στη μεταγραφή/ρύθμιση της μεταγραφής. Επιπλέον, 3 πρωτεΐνες του υποσυμπλόκου B είναι παράγοντες αναδιαμόρφωσης της χρωματίνης και 4 πρωτεΐνες των υποσυμπλόκων B και B *act* συμμετέχουν στην αποικοδόμηση του mRNA εάν αυτό δεν είναι κατάλληλο για να μεταφραστεί, δυο διαδικασίες οι οποίες σχετίζονται λειτουργικά με τη μεταγραφή και τη συναρμογή, αντίστοιχα. Όσον αφορά το υποσύμπλοκο C, το 63% των μοναδικών πρωτεϊνών του ταυτοποιήθηκαν σε περισσότερες από μία ερευνητικές εργασίες γεγονός που πιθανώς υποδεικνύει ότι αυτές οι πρωτεΐνες παρουσιάζουν ένα συγκεκριμένο ρόλο σε αυτό το στάδιο της συναρμολόγησης του σωματίου. Επιπλέον, η λειτουργική κατηγοριοποίηση των μοναδικών πρωτεϊνών του συνάδει με τον αυξημένο αριθμό των ερευνητικών εργασιών που τις ταυτοποιούν. Συγκεκριμένα, περισσότερες από τις μισές (57%) συμμετέχουν στη συναρμογή και στην αποικοδόμηση του mRNA εάν αυτό δεν είναι κατάλληλο για να μεταφραστεί, δηλαδή ο λειτουργικός τους ρόλος συνάδει με τις λειτουργίες που συμβαίνουν στο συγκεκριμένο υποσύμπλοκο.

### 3.2.1.2. Δημιουργία του συνόλου δεδομένων των πρωτεϊνικών αλληλεπιδράσεων των συστατικών του σωματίου συναρμογής

Μετά τη δημιουργία του συνόλου δεδομένων των 630 πρωτεϊνών του σωματίου συναρμογής, πραγματοποιήθηκε η διερεύνηση των πρωτεϊνικών αλληλεπιδράσεών τους (Εικ. 41) ώστε να ανακατασκευαστεί το δίκτυο του σωματίου συναρμογής. Όλα τα δεδομένα για τις αλληλεπιδράσεις των πρωτεϊνών (PPIs) του ανθρώπου

ανακτήθηκαν από την PICKLE (Klара et al. 2013, Gioutlakis et al. 2017, βλ. Υλικά & Μέθοδοι), με κριτήριο ότι και τα δυο μέλη της αλληλεπίδρασης αποτελούσαν συστατικά του σωματίου συναρμογής.

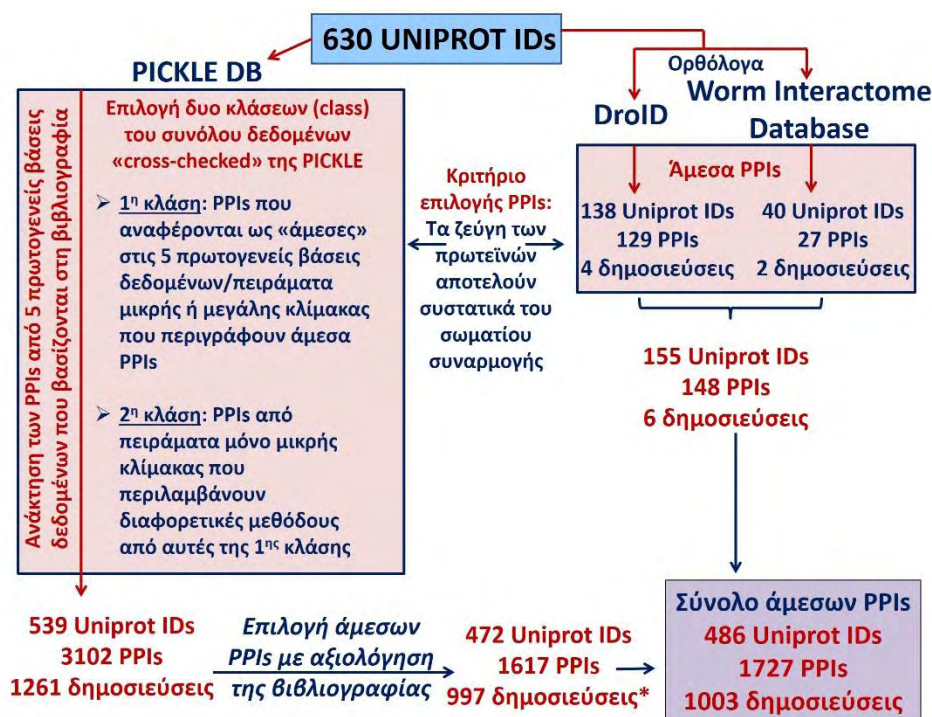
Σύμφωνα με τα κριτήρια κατάταξης των πρωτεϊνικών αλληλεπιδράσεων σε διαφορετικές κλάσεις (class), τα οποία εφάρμοσε η συγκεκριμένη μετα-βάση (βλ. Υλικά & Μέθοδοι), επιλέχθηκαν να χρησιμοποιηθούν οι 2 πρώτες κλάσεις («first» και «second» class) που περιέχουν PPIs με βαθμό αξιοπιστία (confidence score) 1 και 2, αντίστοιχα (βλ. Υλικά & Μέθοδοι). Η πρώτη κλάση περιλαμβάνει τις αλληλεπιδράσεις που φέρουν το χαρακτηρισμό «άμεσες» (direct) στις πρωτογενείς βάσεις δεδομένων ή έχουν ταυτοποιηθεί με πειραματικές μεθόδους, μεγάλης ή μικρής κλίμακας, που περιγράφουν άμεσες αλληλεπιδράσεις (βλ. Υλικά & Μέθοδοι). Η δεύτερη κλάση περιλαμβάνει τις αλληλεπιδράσεις που έχουν ταυτοποιηθεί με οποιαδήποτε μέθοδο (εκτός από αυτές που αναφέρονται στην πρώτη κλάση) σε τουλάχιστον μία μελέτη μικρής κλίμακας (βλ. Υλικά & Μέθοδοι). Οι δυο αυτές κλάσεις περιέχονταν στο «ελεγμένο» (cross-checked) σύνολο δεδομένων της μεταβάσης (βλ. Υλικά & Μέθοδοι), το οποίο και χρησιμοποιήθηκε για την αναζήτηση των αλληλεπιδράσεων των πρωτεϊνών του σωματίου συναρμογής. Τα δεδομένα που προέκυψαν από την αναζήτηση αυτή αφορούσαν 539 πρωτεΐνες-Uniprot IDs από το σύνολο των συστατικών του σωματίου συναρμογής (630 πρωτεΐνες-Uniprot IDs), 3102 PPIs και στηρίζονταν σε 1261 ερευνητικές εργασίες (Εικ. 38). Στο σύνολο αυτό των αλληλεπιδράσεων δεν ενσωματώθηκαν οι αλληλεπιδράσεις της ουβικουιτίνης διότι, στα πλαίσια του συνολικού δικτύου πρωτεϊνικών αλληλεπιδράσεων του ανθρώπου (human interactome, Klара et al. 2013), παρατηρήθηκε ότι «παραμορφώνει» το δίκτυο και προτάθηκε ότι η φύση των αλληλεπιδράσεων αυτής της πρωτεΐνης δεν την εντάσσει σε ένα δίκτυο πρωτεϊνικών αλληλεπιδράσεων αλλά σε ένα δίκτυο που περιλαμβάνει τις μετα-μεταφραστικές τροποποιήσεις των πρωτεϊνών (βλ. Ενότητα 4.2.9.).

Αν και οι βάσεις δεδομένων προσφέρουν μεγάλο όγκο πληροφοριών η αυτόματη ανάκτηση των δεδομένων τους μέσω της χρήσης λέξεων-κλειδιών και της αξιολόγησης των περιλήψεων των εργασιών μπορεί να αποδώσει πληροφορίες οι οποίες είναι λανθασμένες (βλ. Ενότητα 1.5.1.3.). Γι' αυτό το λόγο θεωρείται ότι η ανάκτηση δεδομένων από τις βάσεις θα πρέπει να συνοδεύεται από την αξιολόγηση των εκάστοτε βιβλιογραφικών αναφορών (βλ. Ενότητα 1.5.1.3.), όσο αυτό είναι εφικτό (με βάση τον όγκο των δεδομένων), ώστε οι πληροφορίες που θα προκύψουν να χαρακτηρίζονται από μεγαλύτερη εγκυρότητα και τα βιολογικά συμπεράσματα που θα εξαχθούν να είναι ορθά. Γι' αυτό πραγματοποιήθηκε αξιολόγηση συγκεκριμένων βιβλιογραφικών αναφορών ώστε να πιστοποιηθεί ότι όλα τα πειράματα που αναφέρονταν στις εργασίες αυτές αφορούσαν την ταυτοποίηση αποκλειστικά άμεσων PPIs. Ειδικότερα, επιλέχθηκε να γίνει πιο λεπτομερής αξιολόγηση των βιβλιογραφικών αναφορών: α) που αφορούσαν μεθόδους ανίχνευσης αλληλεπιδράσεων της 1<sup>ης</sup> κλάσης για τις οποίες δεν ήταν ξεκάθαρο αν αναφέρονταν σε άμεση ή έμμεση αλληλεπίδραση και β) που περιλάμβαναν πειράματα συν-ανοσοκατακρήμνισης (co-IP) για την ταυτοποίηση των αλληλεπιδράσεων της 2<sup>ης</sup> κλάσης. Σχετικά με την 1<sup>η</sup> κλάση, η αξιολόγηση έγινε για αλληλεπιδράσεις που ταυτοποιήθηκαν με τρεις μεθόδους, οι οποίες χαρακτηρίζονται από την πρωτογενή βάση δεδομένων BioGRID ως: «reconstituted complex», «gel shift

assay» και «surface plasmon resonance». Ο όρος «reconstituted complex» αναφέρεται σε μεθόδους ταυτοποίησης αλληλεπιδράσεων μεταξύ μιας πρωτεΐνης-δολώματος που μπορεί να είναι ανασυνδυασμένη ή/και καθαρισμένη με πρωτεΐνες που προέρχονται από κυτταρικό εκχύλισμα ή είναι ανασυνδυασμένες. Οι άλλοι δυο όροι αναφέρονται σε πειράματα που εξετάζουν αλληλεπιδράσεις νουκλεϊκών οξέων και πρωτεϊνών (gel shift assay) καθώς και πειράματα φασματοσκοπίας επιφανειακού συντονισμού πλασμονίου (surface plasmon resonance) τα οποία περιγράφουν τις αλληλεπιδράσεις μορίων (πρωτεϊνών, νουκλεϊκών οξέων) εκ των οποίων το ένα είναι ακινητοποιημένο σε μια επιφάνεια. Έτσι, όταν οποιαδήποτε από τις δυο παραπάνω μεθόδους αναφερόταν στην ταυτοποίηση αλληλεπιδράσεων είτε μεταξύ νουκλεϊκών οξέων και πρωτεϊνών είτε μεταξύ πολλών πρωτεϊνών και δεν αφορούσε ένα ζεύγος πρωτεϊνών που ήταν ανασυνδυασμένες ή καθαρισμένες, τότε αυτές οι αλληλεπιδράσεις δεν περιλαμβάνονταν στο τελικό σύνολο δεδομένων. Όσον αφορά τη δεύτερη περίπτωση η 2<sup>η</sup> κλάση της μεταβάσης επιλέχθηκε για να ελεγχθούν οι ερευνητικές εργασίες που περιλάμβαναν πειράματα συν-ανοσοκατακρήμνισης, ανασυνδυασμένων και καθαρισμένων πρωτεϊνών καθώς και οι περιπτώσεις στις οποίες ελέγχονταν, επιπλέον, οι αλληλεπιδράσεις επιμέρους περιοχών των εξεταζόμενων πρωτεϊνών με την ίδια μέθοδο. Αξίζει να σημειωθεί ότι κατά τη διερεύνηση των ερευνητικών εργασιών, παρατηρήθηκε ότι αρκετές πρωτεϊνικές αλληλεπιδράσεις οι οποίες ήταν καταχωρημένες στις πρωτογενείς βάσεις δεδομένων, δεν αναφέρονταν στην αντίστοιχη ερευνητική εργασία. Το ποσοστό αυτών κυμαινόταν στο 5% (~150 PPIs) του συνόλου των αλληλεπιδράσεων. Αυτό πιθανόν συμβαίνει διότι η αξιολόγηση των ερευνητικών εργασιών από τους επιμελητές κάθε πρωτογενούς βάσης γίνεται μέσω λέξεων-κλειδιών και μέσω μελέτης των περιλήψεων αλλά όχι των ερευνητικών εργασιών στην πλειοψηφία των περιπτώσεων, με αποτέλεσμα να προκύπτουν ψευδώς θετικά αποτελέσματα.

Έπειτα από την εξονυχιστική αξιολόγηση των PPIs, παρατηρήθηκε ότι ο αριθμός των πρωτεϊνικών αλληλεπιδράσεων αφορούσε πλέον 472 πρωτεΐνες-Uniprot IDs και μειώθηκε σχεδόν στο μισό (1617 PPIs). Επιπλέον και ο αριθμός των ερευνητικών μελετών μειώθηκε σημαντικά (997) (Εικ. 38) (Πιν. 2 Παράρτημα). Όλες αυτές οι αλληλεπιδράσεις υποστηρίζονται από τουλάχιστον μία ερευνητική εργασία που περιγράφει την ταυτοποίηση αποκλειστικά άμεσων PPIs (Πιν. 2 Παράρτημα). Ωστόσο, ορισμένες από αυτές υποστηρίζονται και από πειραματικές μελέτες που τις ταυτοποιούν και σε επίπεδο συμπλόκου (έμμεσα PPIs) (Πιν. 2 Παράρτημα). Έτσι, οι 997 ερευνητικές εργασίες περιλαμβάνουν 573 ερευνητικές εργασίες που περιγράφουν αποκλειστικά άμεσες αλληλεπιδράσεις και 424 που περιλαμβάνουν δεδομένα απομόνωσης πρωτεϊνικών συμπλόκων (Εικ. 38) (Πιν. 2 Παράρτημα). Στο Παράρτημα 2 φαίνονται οι άμεσες αλληλεπιδράσεις των πρωτεϊνών του σωματίου συναρμογής, η προέλευση της καθεμίας από συγκεκριμένη βάση δεδομένων (στήλες E-J), και η βιβλιογραφική αναφορά (PubMed ID, στήλη M) στην οποία ταυτοποιήθηκε. Επίσης επισημαίνονται (γκρι απόχρωση) τα PPIs τα οποία έχουν ταυτοποιηθεί και σε πειράματα προσδιορισμού έμμεσων PPIs. Τέλος, υπάρχουν και ορισμένα PPIs που

ταυτοποιήθηκαν με προσωπική αναζήτηση της σχετικής βιβλιογραφίας (manually-curated PPIs, στήλη L).



Εικ. 38: Διαδικασία ανάκτησης και αξιολόγησης των δεδομένων των PPIs που προέκυψαν από τις βάσεις δεδομένων του ανθρώπου και των οργανισμών μοντέλων.

\*:Περιλαμβάνονται και οι ερευνητικές εργασίες που αφορούν έμμεσες αλληλεπιδράσεις.

Επιπλέον, το σύνολο δεδομένων των PPIs του ανθρώπου εμπλουτίστηκε με τις πρωτεϊνικές αλληλεπιδράσεις των ορθόλογων πρωτεϊνών από τους οργανισμούς-μοντέλα *D. melanogaster* και *C. elegans*, οι οποίες ταυτοποιήθηκαν μέσω του συστήματος των δύο υβριδίων (Εικ. 38). Αυτή η προσέγγιση βασίστηκε στη ιδέα των ιντερόλογων (interologs), όπως αναφέρθηκε και στην ανακατασκευή του δικτύου της πρωτεΐνης FRA10AC1 (βλ. Ενότητα 3.1.) (Nguyen et al. 2013, Huang et al. 2007, Walhout et al. 1999). Τα δεδομένα αυτά ανακτήθηκαν από τις βάσεις δεδομένων Droid (<http://www.droidb.org/>, Yu et al. 2008) και Worm Interactome Database ([http://interactome.dfci.harvard.edu/C\\_elegans/](http://interactome.dfci.harvard.edu/C_elegans/), Simonis et al. 2009) αντίστοιχα (Εικ. 38). Η διερεύνηση της Droid έγινε χρησιμοποιώντας το επίσημο όνομα των γονιδίων του ανθρώπου. Για την ανάκτηση των δεδομένων από την Worm Interactome Database πραγματοποιήθηκε μετατροπή των κωδικών Gene ID του ανθρώπου σε κωδικούς γονιδίων του *C. elegans* χρησιμοποιώντας τη βάση δεδομένων Homologene (<http://www.ncbi.nlm.nih.gov/homologene>), εφόσον υπήρχε πιθανό ορθόλογο γονίδιο. Οι κωδικοί των αλληλεπιδρώντων πρωτεϊνών μετατράπηκαν σε κωδικούς Gene IDs και Uniprot IDs του ανθρώπου, εφόσον υπήρχαν και πάλι τα ορθόλογα, και το αντίστοιχο ζεύγος αλληλεπιδρώντων ενσωματώθηκε στο σύνολο δεδομένων με την προϋπόθεση ότι και το δεύτερο μέλος του ζεύγους των αλληλεπιδρώντων ανήκε στη λίστα των συστατικών του σωματίου συναρμογής. Από την παραπάνω αναζήτηση ανακτήθηκαν 148 πρωτεϊνικές αλληλεπιδράσεις, 27 από αυτές προέρχονταν από τον *C.*



*elegans* και 129 από την *D. melanogaster*, με τις 8 να είναι κοινές και στα δυο είδη οργανισμών. Οι αλληλεπιδράσεις αφορούσαν 155 πρωτεΐνες και βασίστηκαν σε 6 ερευνητικές εργασίες, δυο και τέσσερις αντίστοιχα για το κάθε είδος (Εικ. 38). Πρέπει να σημειωθεί ότι αυτές οι 155 πρωτεΐνες δεν αθροίζονται σε αυτές που βρέθηκαν στον άνθρωπο (472) διότι υπάρχουν κοινές μεταξύ τους (Εικ. 38)

Συγκεντρώνοντας τα παραπάνω δεδομένα, το τελικό σύνολο δεδομένων των PPIs περιλάμβανε 1727 άμεσα PPIs μεταξύ 486 πρωτεϊνών (UniProt IDs) (Εικ. 38). Από αυτές, οι 207 εμφάνιζαν αλληλεπίδραση και με τον εαυτό τους (self-loops), ενώ οι 30 από αυτές αλληλεπιδρούσαν μόνο με τον εαυτό τους (Πιν. 2 Παράρτημα).

Για να γίνει ο έλεγχος της αξιοπιστίας του παραπάνω συνόλου των άμεσων αλληλεπιδράσεων εφαρμόστηκαν δυο κριτήρια: α) το πλήθος των ερευνητικών εργασιών που υποστηρίζουν μια άμεση αλληλεπίδραση λαμβάνοντας υπόψη και την ύπαρξη των επιπρόσθετων ερευνητικών εργασιών οι οποίες περιγράφουν, επιπλέον, και έμμεσες αλληλεπιδράσεις μεταξύ των πρωτεϊνών και β) η εξελικτική συντήρηση των αλληλεπιδράσεων. Σχετικά με το πρώτο κριτήριο, διερευνήθηκε ο αριθμός των δημοσιεύσεων που υποστήριζαν κάθε αλληλεπίδραση. Ο έλεγχος αυτός έδειξε ότι 524 από τις 1727 PPI (30,4% του συνόλου) ταυτοποιήθηκαν σε περισσότερες από μία ερευνητικές εργασίες, σε κλιμακούμενο αριθμό, από 2-44 (Πιν. 2 παράρτημα). Επίσης, για 101 από τις 524 PPI υπήρχε επιπλέον πληροφορία από πειράματα που τις εξέταζαν σε επίπεδο συμπλόκου. Ειδικότερα, τα δεδομένα πρωτεϊνικών αλληλεπιδράσεων του ανθρώπου και των οργανισμών-μοντέλων στηρίζονταν σε 1003 ερευνητικές μελέτες (Εικ. 38) (Πιν. 2 παράρτημα), οι 579 από τις οποίες αναφέρονταν σε άμεσες PPI και οι υπόλοιπες (424) σε PPI που προέρχονταν από δεδομένα απομόνωσης πρωτεϊνικών συμπλόκων. Όσον αφορά το δεύτερο κριτήριο, ελέγχθηκαν ποιες αλληλεπιδράσεις είχαν ταυτοποιηθεί τόσο μεταξύ των ζευγών πρωτεϊνών στον άνθρωπο όσο και στους οργανισμούς μοντέλα. Έτσι 46 πρωτεϊνικές αλληλεπιδράσεις βρέθηκαν να είναι εξελικτικά συντηρημένες (Πιν. 2 παράρτημα).

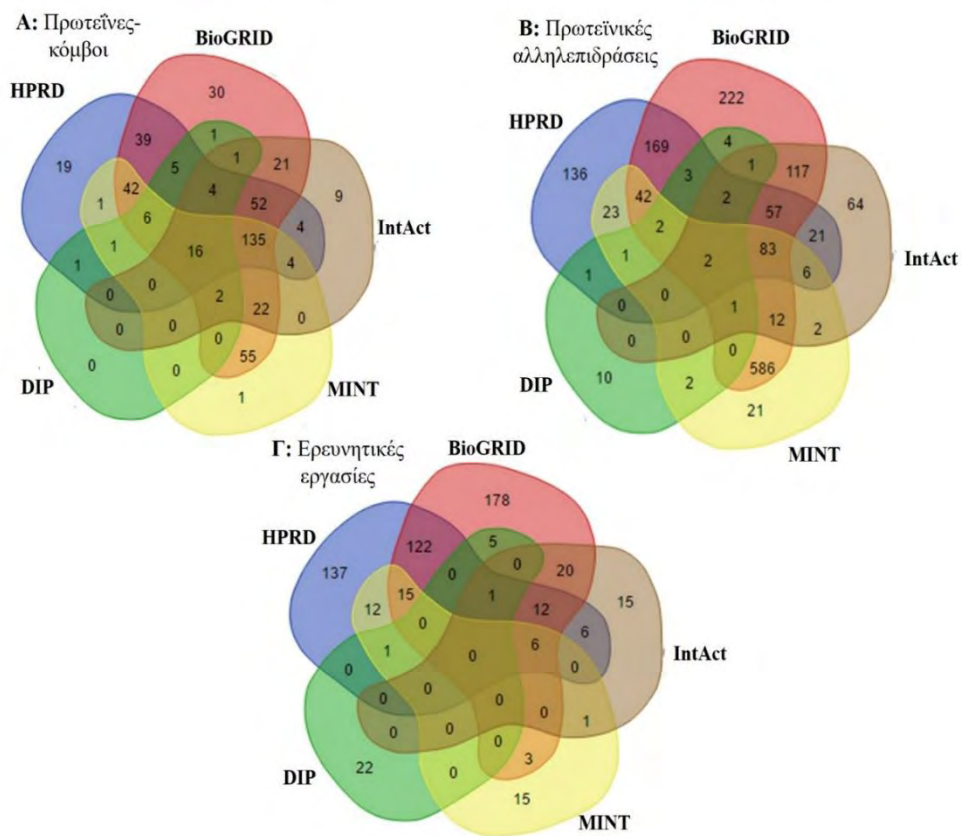
Συμπερασματικά, οι 524 από τις 1727 άμεσες αλληλεπιδράσεις του τελικού συνόλου δεδομένων πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής παρουσιάζουν αυξημένη αξιοπιστία διότι υποστηρίζονται από παραπάνω από μία ερευνητική εργασία και περιλαμβάνουν αλληλεπιδράσεις που είναι εξελικτικά συντηρημένες. Ειδικότερα, οι 424 PPIs από αυτές υποστηρίζονται από μία ερευνητική εργασία που περιγράφει άμεση αλληλεπίδραση και ταυτόχρονα, από άλλες εργασίες που αφορούν έμμεσα PPIs. Επίσης, οι υπόλοιπες 100 υποστηρίζονται από περισσότερες από μία ερευνητικές εργασίες που περιγράφουν μόνο άμεσες αλληλεπιδράσεις. Τέλος, 46 από τις 524 αλληλεπιδράσεις έχουν ταυτοποιηθεί στον άνθρωπο και στους οργανισμούς-μοντέλα *C. elegans* και *D. melanogaster*.

### 3.2.1.2.1. Αλληλοεπικάλυψη βάσεων δεδομένων σε πρωτεΐνες, αλληλεπιδράσεις και ερευνητικές εργασίες

Η PICKLE, όπως περιγράφηκε και παραπάνω, είναι μια μετα-βάση που περιέχει πληροφορίες από πέντε διαφορετικές βάσεις δεδομένων. Ωστόσο, με σκοπό να διαπιστωθεί το ποσοστό αλληλοεπικάλυψης των δεδομένων τα οποία προέρχονταν από κάθε βάση και κατ' επέκταση να ελεγχθεί αν η ανάκτηση δεδομένων από διαφορετικές

πηγές πληροφόρησης είναι αναγκαία πραγματοποιήθηκε η σύγκριση των δεδομένων που αφορούν τις αλληλεπιδράσεις, τις πρωτεΐνες και τις ερευνητικές εργασίες της καθεμιάς βάσης.

Η συνεισφορά της κάθε βάσης δεδομένων στον άνθρωπο σε πρωτεϊνικές αλληλεπιδράσεις, πρωτεΐνες και ερευνητικές εργασίες φαίνεται στον Πίνακα 10. Φαίνεται η αξιοσημείωτα μικρή αλληλοεπικάλυψη μεταξύ τους και υποδεικνύεται η σημασία του συνδυασμού των δεδομένων που προέρχονται από αυτές (Πιν. 10 & Εικ. 39). Συγκεκριμένα, η αλληλοεπικάλυψη ανάμεσα σε όλες τις βάσεις δεδομένων, όσον αφορά τον αριθμό των πρωτεϊνών και των PPIs ήταν κατ' αντιστοιχία δεκαεπτά και δύο ενώ δεν υπήρχε καμία κοινή ερευνητική εργασία (Πιν. 10 & Εικ. 39). Επίσης, λαμβάνοντας υπόψη τις τέσσερις μεγαλύτερες βάσεις δεδομένων (HPRD, BioGRID, IntAct και MINT), παρατηρήθηκε και πάλι ότι η αλληλοεπικάλυψη ήταν μικρή τόσο στον αριθμό των PPIs (86) όσο και των ερευνητικών εργασιών (6) (Πιν. 10 & Εικ. 39). Συγκρίνοντας τα δεδομένα που ανακτήθηκαν από τις βάσεις δεδομένων HPRD και BioGRID με τα συνολικά δεδομένα από όλες τις βάσεις, παρατηρήθηκε ότι αυτές συγκεντρώνουν τον μεγαλύτερο αριθμό πρωτεϊνών, πρωτεϊνικών αλληλεπιδράσεων και ερευνητικών εργασιών (Πιν. 10 & Εικ. 39). Με τη χρήση μόνο της BioGRID, που θεωρείται η μεγαλύτερη βάση δεδομένων, παρατηρήθηκε ότι είχε ανακτηθεί ένα μεγάλο ποσοστό των πρωτεϊνών (91%) και των PPIs (80%) αλλά ένα μειωμένο ποσοστό των ερευνητικών εργασιών (36%) (Πιν. 10 & Εικ. 39). Επιπλέον, χρησιμοποιώντας τις βάσεις HPRD και BioGRID, αν και το 95% των πρωτεϊνών είχε ανακτηθεί από αυτές, το 12% των PPIs και το 20% των ερευνητικών εργασιών δεν περιλαμβάνονταν στο τελικό σύνολο δεδομένων (Πιν. 10 & Εικ. 39). Στην Εικόνα 39 και στον Πίνακα 10 φαίνεται η συνεισφορά της κάθε βάσης δεδομένων σε πρωτεΐνες, PPIs και ερευνητικές εργασίες και οι αντίστοιχες τιμές που προέκυψαν από τη σύγκριση των βάσεων δεδομένων ανά δύο, ανά τρεις, ανά τέσσερις και όλων μεταξύ τους.



**Εικ. 39: Αλληλοεπικάλυψη πρωτεϊνών-αλληλεπιδράσεων-ερευνητικών εργασιών των 5 βάσεων δεδομένων PPIs μέσω διαγράμματος Venn.**

**Πιν. 10: Συγκριτική συνεισφορά των βάσεων δεδομένων PPIs στον άνθρωπο.**

Συνεισφορά των βάσεων δεδομένων	Βάσεις δεδομένων PPI	PPI					Προτεΐνες					Ερευνητικές εργασίες					
		HPRD	BIOGRID	MINT	INTACT	DIP	HPRD	BIOGRID	MINT	INTACT	DIP	HPRD	BIOGRID	MINT	INTACT	DIP	
Συνεισφορά κάθε μίας βάσης δεδομένων	HPRD	548					329					312					
	BIOGRID		1303					431					362				
	MINT			783					285					53			
	INTACT				368						270				61		
	DIP					29						37				29	
Αλληλεπικάλυψη δεδομένων (από δυο βάσεις)	HPRD-BIOGRID	360	360				260	260				156	156				
	HPRD-MINT	159		159			205		205			34		34			
	HPRD-INTACT	171			171		215			215		25			25		
	HPRD-DIP	11				11	33				33	2				2	
	BIOGRID-INTACT		275		275			231		231			39			39	
	BIOGRID-MINT		728	728				278	278				24	24			
	BIOGRID-DIP		15			15		35			35		6			6	
	MINT-INTACT			106	106				179	179				7	7		
	MINT-DIP			6		6			25		25			1		1	
INTACT-DIP				6	6					23	23			1	1		
Αλληλεπικάλυψη δεδομένων (από τρεις βάσεις)	HPRD-BIOGRID-MINT	129	129	129			199	199	199			21	21	21			
	HPRD-BIOGRID-INTACT	144	144		144		207	207		207		19	19		19		
	HPRD-BIOGRID-DIP	9	9			9	33	33			33	1	1			1	
	HPRD-MINT-INTACT	91		91	91		155		155	155		6		6	6		
	HPRD-INTACT-DIP	4			4	4	20			20	20	1			1	1	
	HPRD-MINT-DIP	5		5		5	23			23	23	1		1		1	
	BIOGRID-MINT-INTACT		98	98	98			175	175	175			6	6	6		
	BIOGRID-MINT-DIP		5	5		5		26	26		26		0	0		0	
	BIOGRID-INTACT-DIP		6		6	6		24			24	24		1		1	1
	MINT-INTACT-DIP			3	3	3				18	18	18			0	0	0
Αλληλεπικάλυψη δεδομένων (από τέσσερις βάσεις)	HPRD-BIOGRID-MINT-INTACT	85	85	85	85		151	151	151	151		6	6	6	6		
	HPRD-BIOGRID-MINT-DIP	4	4	4		4	22	22	22		22	0	0	0		0	
	HPRD-BIOGRID-INTACT-DIP	4	4		4	4	20	20			20	1	1		1	1	
	HPRD-MINT-INTACT-DIP	2		2	2	2	16		16	16	16	0		0	0	0	
	BIOGRID-MINT-INTACT-DIP		3	3	3	3		18	18	18	18		0	0	0	0	
Αλληλεπικάλυψη δεδομένων (από όλες τις βάσεις)	HPRD-BIOGRID-MINT-INTACT-DIP	2	2	2	2	2	16	16	16	16	16	0	0	0	0	0	

Έντονη γραφή: αποτελέσματα της σύγκρισης όλων των βάσεων.

### 3.2.1.3. Οπτικοποίηση του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής

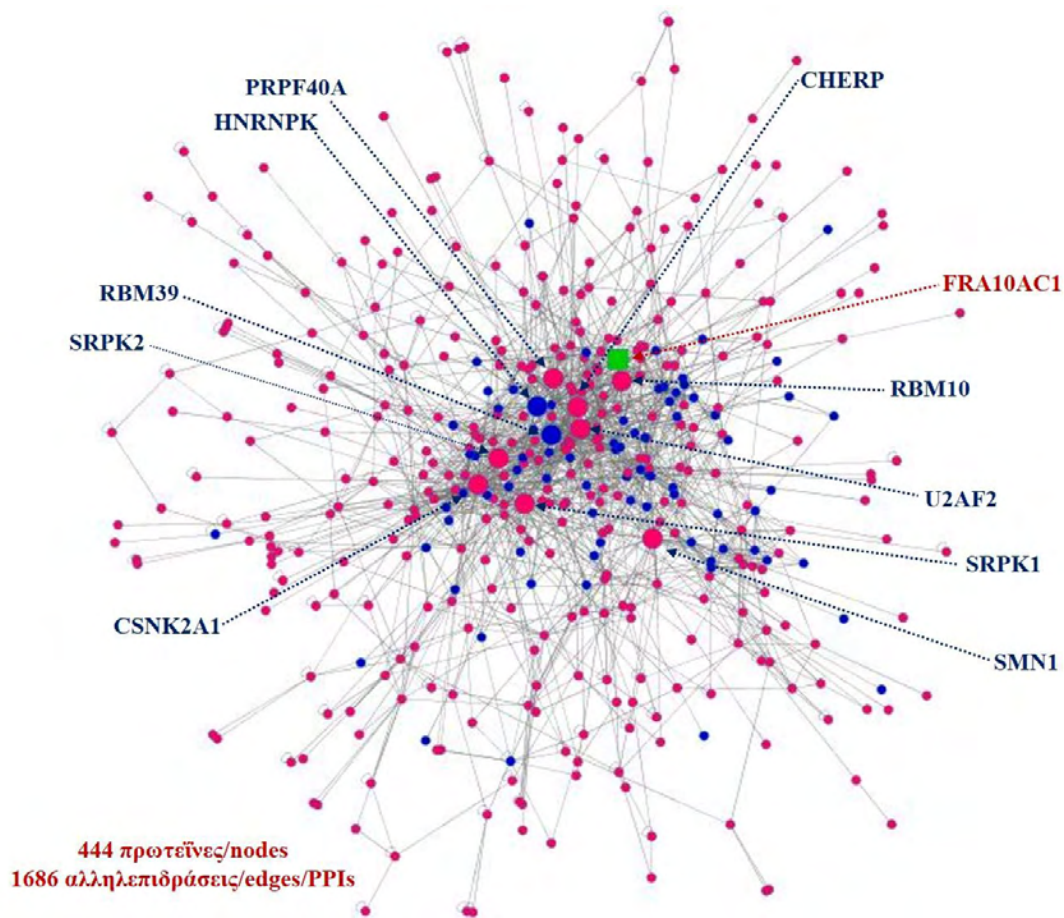
Με βάση τα παραπάνω δεδομένα για τα PPIs, 444 πρωτεΐνες-κόμβοι σχημάτισαν ένα ενιαίο δίκτυο 1686 πρωτεϊνικών αλληλεπιδράσεων, 173 από τις οποίες αλληλεπιδρούσαν και με τον εαυτό τους (self-loops) (Εικ. 40). Οι υπόλοιπες 44 πρωτεΐνες σχημάτισαν ομοδιμερή ή ετεροδιμερή εκτός του ενιαίου δικτύου. Οι 1581 από τις συνολικές 1686 πρωτεϊνικές αλληλεπιδράσεις προέρχονταν από δεδομένα για τον άνθρωπο, 48 εκ των οποίων αποτελούσαν και «αντερόλογα», ενισχύοντας την αξιοπιστία τους (Matthews et al. 2001). Οι υπόλοιπες 105 αλληλεπιδράσεις προέρχονταν αποκλειστικά από δεδομένα PPI στους οργανισμούς-μοντέλα.

Η οπτικοποίηση του ανακατασκευασμένου δικτύου έγινε με τη χρήση του προγράμματος Cytoscape ([www.cytoscape.org](http://www.cytoscape.org), version 3.2.1) (Shannon et al. 2003) σε διάταξη «Edge-weighted Spring-Embedded» που είναι παρόμοια με τη διάταξη «force-directed» η οποία δεν χρησιμοποιήθηκε διότι δεν εμφανιζόταν ως προεπιλογή σε αυτήν την έκδοση του Cytoscape. Οι διατάξεις αυτές ήταν πιο πληροφοριακές σε σύγκριση με άλλες όσον αφορά την απεικόνιση των αλληλεπιδράσεων μεταξύ των πρωτεϊνών-κόμβων και την τοπολογία αυτών στο συνολικό δίκτυο. Η διάταξη που χρησιμοποιήθηκε λειτουργεί με τον εξής τρόπο: Οι πρωτεΐνες-κόμβοι του δικτύου ορίζονται ως φυσικά αντικείμενα που απωθούνται μεταξύ τους και οι συνδέσεις ανάμεσα στις πρωτεΐνες-κόμβους ορίζονται σαν μεταλλικά «ελατήρια» που συνδέονται με το ζεύγος των κόμβων. Αυτά τα «ελατήρια» απωθούν ή έλκουν τα καταληκτικά τους σημεία (end points) βάσει των δυνάμεων των συνδέσεων ανάμεσα στους κόμβους με στόχο να οριστούν οι θέσεις των πρωτεϊνών-κόμβων με τέτοιο τρόπο ώστε να ελαχιστοποιείται το σύνολο των δυνάμεων του δικτύου (Kamada & Kawai 1989).

Στο τελικό ανακατασκευασμένο δίκτυο ενσωματώθηκε το 70% των πρωτεϊνών του συνόλου δεδομένων του σωματίου συναρμογής (Πιν. 1 Παράρτημα). Όσον αφορά τις υπόλοιπες πρωτεΐνες, το 30% (57) ήταν πρωτεΐνες άγνωστης λειτουργίας ενώ το ~60% (115) αυτών είχαν ταυτοποιηθεί σε μία μόνο ερευνητική εργασία. Όλες οι πρωτεΐνες άγνωστης λειτουργίας που δεν συμμετείχαν στο δίκτυο και οι περισσότερες πρωτεΐνες που είχαν ταυτοποιηθεί σε μία ερευνητική εργασία είχαν απομονωθεί είτε ως συστατικά του υπο-συμπλόκου E (~51%) ή σε μελέτες απομόνωσης ολόκληρου του σωματίου συναρμογής (~30%) (Πιν. 1 Παράρτημα).

Σχετικά με τις ογδόντα πρωτεΐνες που ήταν παρούσες καθ' όλη τη διάρκεια της συναρμολόγησης του σωματίου συναρμογής παρατηρήθηκε ότι όλες ενσωματώνονται στο τελικό δίκτυο πρωτεϊνικών αλληλεπιδράσεων, ενισχύοντας την άποψη ότι αποτελούν σημαντικούς παράγοντες για τη διαδικασία της συναρμογής (Εικ. 40).

Επίσης, αξίζει να σημειωθεί ότι η πρωτεΐνη FRA10AC1 τοποθετείται κοντά στο κέντρο του δικτύου και όχι στην περιφέρεια αυτού (Εικ. 40)-που σημαίνει, πιθανώς, ότι διαδραματίζει βασικό ρόλο στη διαδικασία της συναρμογής-γεγονός που επιβεβαιώθηκε και από την ανάλυση των τοπολογικών παραμέτρων του δικτύου, όπως αναλύεται και παρακάτω. Στην Εικόνα 40 παρουσιάζεται το ανακατασκευασμένο δίκτυο PPI του σωματίου συναρμογής, οι πρωτεΐνες που εμφάνισαν τις περισσότερες αλληλεπιδράσεις (hubs), η πρωτεΐνη FRA10AC1 καθώς και οι πρωτεΐνες που ήταν παρούσες σε όλα τα στάδια συναρμολόγησης του σωματίου συναρμογής.



**Εικ. 40:** Δίκτυο πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής. Φαίνονται οι 444 πρωτεΐνες του σωματίου συναρμογής (nodes) με κυκλικό σχήμα και κόκκινο χρωματισμό και οι αλληλεπιδράσεις μεταξύ τους (edges) με γκρι γραμμές. Τα βέλη υποδεικνύουν τις 10 πρωτεΐνες με τις περισσότερες αλληλεπιδράσεις (hubs) (κόκκινοι κύκλοι σε μεγαλύτερο μέγεθος). Με πράσινο τετράγωνο φαίνεται η πρωτεΐνη FRA10AC1. Οι πρωτεΐνες με καθολική παρουσία στα υποσύμπλοκα A, B, B *act* και C φαίνονται με μπλε χρωματισμό. Διάταξη δικτύου: «Edge-weighted Spring-Embedded».

Επιπλέον, πραγματοποιήθηκε η κατηγοριοποίηση των πρωτεϊνών του δικτύου στις δεκαεννέα οντολογικές κατηγορίες οι οποίες χρησιμοποιήθηκαν και κατά την κατηγοριοποίηση όλων των πρωτεϊνών του σωματίου. Αυτή έγινε για να ελεγχθεί, σε πρώτο στάδιο, ο αριθμός των πρωτεϊνών της καθεμίας από τις δεκαεννέα κατηγορίες GO που ενσωματώθηκαν στο δίκτυο σε σύγκριση με τα αντίστοιχα στοιχεία του συνόλου δεδομένων του σωματίου συναρμογής και, σε επόμενο στάδιο, να εξεταστεί ο εμπλουτισμός της κάθε οντολογικής κατηγορίας σε πρωτεΐνες-κόμβους στα πλαίσια του δικτύου.

Όσον αφορά την πρώτη περίπτωση, στον Πίνακα 11 φαίνεται ο αριθμός των πρωτεϊνών καθεμίας από τις δεκαεννέα κατηγορίες GO βιολογικής διαδικασίας που ενσωματώθηκαν στο δίκτυο και ο αντίστοιχος αριθμός στο σύνολο δεδομένων του σωματίου συναρμογής (Πιν. 11). Ακόμη, και στα πλαίσια του δικτύου εμφανίζεται μεγάλος αριθμός πολυλειτουργικών πρωτεϊνών (22%), για τις οποίες γίνεται αναφορά στην επόμενη ενότητα. Μια σημαντική διαφορά που παρατηρείται είναι ότι ο αριθμός των πρωτεϊνών άγνωστης λειτουργίας μειώνεται κατά τουλάχιστον 50% (Πιν. 11), γεγονός που πιθανώς σημαίνει ότι αυτές οι πρωτεΐνες μπορεί να αποτελούν ψευδώς θετικά αποτελέσματα της διαδικασίας απομόνωσης των συστατικών του σωματίου συναρμογής. Το 92% των πρωτεϊνών-κόμβων του δικτύου έχουν χαρακτηριστεί λειτουργικά.

**Πιν. 11: Ομαδοποίηση των πρωτεϊνών-κόμβων του δικτύου σε οντολογικές κατηγορίες σε σύγκριση με το σύνολο δεδομένων του σωματίου συναρμογής.**

Κατηγορίες γονιδιακής οντολογίας (GO terms)	Πρωτεΐνες του συνόλου δεδομένων του σωματίου συναρμογής	Πρωτεΐνες του δικτύου PPI του σωματίου συναρμογής
Nuclear mRNA splicing (GO:0000398)	140	132
Transcription (GO:0006350)-Regulation of transcription (GO:0045449)	71	42
Chromatin organization (GO:0006325)	30	25
mRNA capping (GO:0006370)	1	1
mRNA 3'-end processing (GO:0031124)	12	6
mRNA export from nucleus (GO:0006406)	21	16
Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (GO:0000184)	14	12
Mitosis-Meiosis (GO:0007067, GO:0007126)	19	11
Intracellular signal transduction (GO:0035556)	20	13
Regulation of mRNA stability (GO:0043488)	7	7
Translation (GO:0006412)	16	10
DNA repair (GO:0006281)	6	5
DNA replication (GO:0006260)	5	3
Protein transport (GO:0015031)	11	9
ncRNA processing (GO:0034470)	11	3
Response to unfolded protein (GO:0006986)	4	2
Protein ubiquitination (GO:0016567)	4	3
Intracellular transport (GO:0046907)	4	0
Biological adhesion (GO:0022610)	4	4
Άλλες κατηγορίες	18	9
Άγνωστης λειτουργίας	90	33
Πολυλειτουργικές πρωτεΐνες	122	98
<b>Σύνολο πρωτεϊνών</b>	<b>630</b>	<b>444</b>

Παρατηρώντας τα αποτελέσματα του Πίνακα 11, φαίνεται ότι περίπου οι μισές πρωτεΐνες που σχετίστηκαν λειτουργικά με τη μεταγραφή-ρύθμιση της μεταγραφής, δεν ενσωματώθηκαν στο τελικό δίκτυο. Αυτό όμως δεν ανταποκρίνεται στην πραγματικότητα, διότι το 33% των πολυλειτουργικών πρωτεϊνών συμμετείχαν στη συναρμογή και τη μεταγραφή-ρύθμιση της μεταγραφής ή στην οργάνωση και αναδιαμόρφωση της χρωματίνης και τη μεταγραφή-ρύθμιση της μεταγραφής. Το γεγονός αυτό εξηγεί το χαμηλό αριθμό των πρωτεϊνών που συμμετείχαν αποκλειστικά στη μεταγραφή (Πιν. 12) αλλά και τη μείωση του αριθμού των πρωτεϊνών στη συναρμογή και στην αναδιαμόρφωση της χρωματίνης.

Στον Πίνακα 12 παρουσιάζονται αναλυτικά όλες οι οντολογικές κατηγορίες στις οποίες ενσωματώθηκαν οι πολυλειτουργικές πρωτεΐνες σε σύγκριση με τα αντίστοιχα στοιχεία του συνόλου δεδομένων του σωματίου συναρμογής.

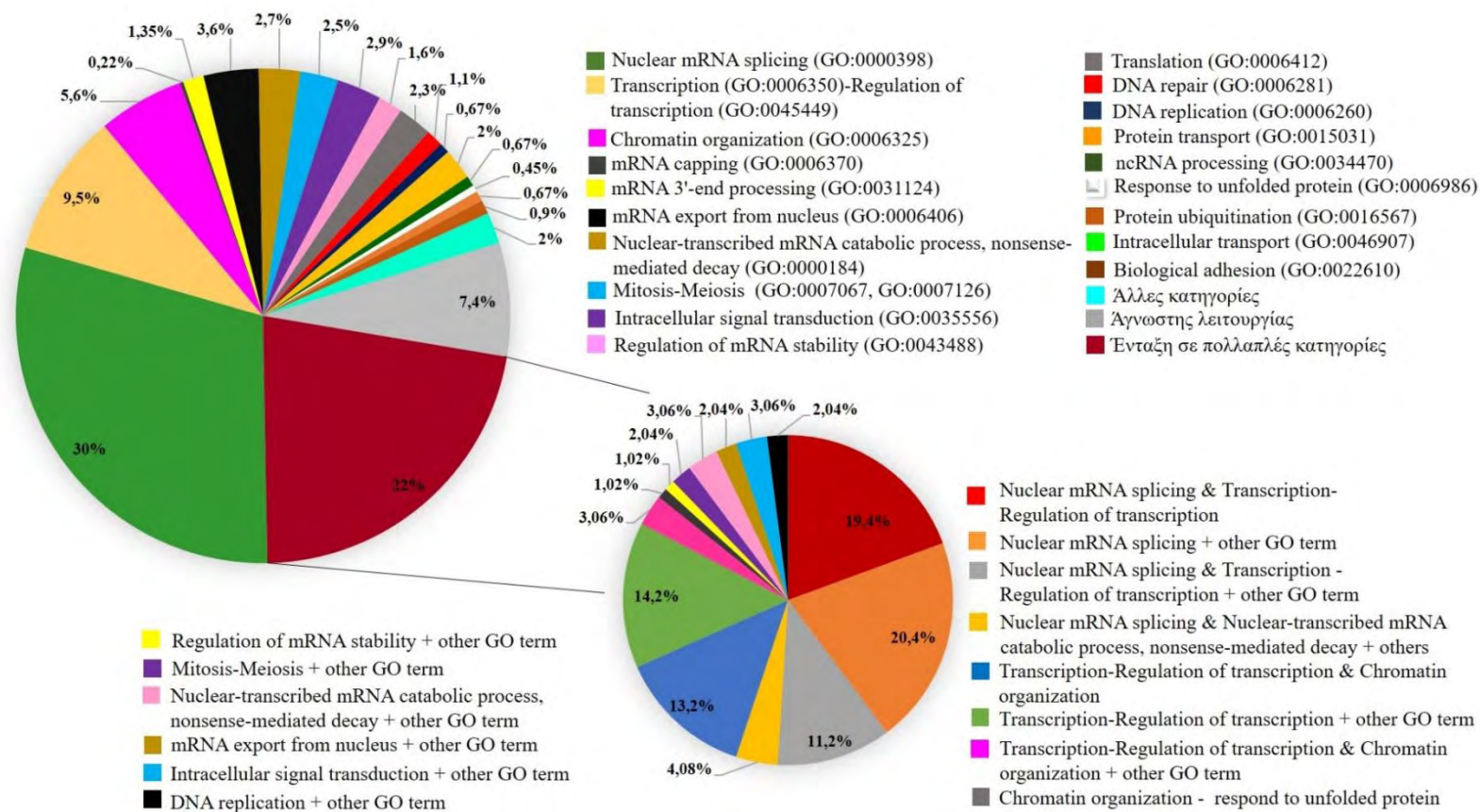


**Πιν. 12: Κατηγορίες οντολογίας των πολυλειτουργικών πρωτεϊνών-κόμβων σε σύγκριση με αυτές του συνόλου δεδομένων του σωματίου συναρμογής.**

Λειτουργικές κατηγορίες των πολυλειτουργικών πρωτεϊνών	# πρωτεϊνών στο σωματίο συναρμογής	# πρωτεϊνών στο δίκτυο PPI του σωματίου συναρμογής
Splicing*/ Transc., reg. of*	22	19
Splicing*/ Chromatin*	2	1
Splicing*/ mRNA capping	2	2
Splicing*/ mRNA 3'-end processing	3	2
Splicing*/ mRNA export from nucleus	5	5
Splicing*/ mRNA catabolic process, NMD*	2	2
Splicing*/ Mitosis-Meiosis	2	2
Splicing*/ Intracellular signal transduction	3	1
Splicing*/ Reg. of mRNA stability*	3	2
Splicing*/ Translation	3	2
Splicing*/ Άλλη βιολογική κατηγορία	1	1
Transc., reg. of*/ Chromatin*	19	13
Transc., reg. of*/ mRNA 3'-end processing	1	1
Transc., reg. of*/ Mitosis-Meiosis	3	2
Transc., reg. of*/ Intracellular signal transduction	3	2
Transc., reg. of*/ Translation	2	1
Transc., reg. of*/ DNA repair	2	1
Transc., reg. of*/ ncRNA processing	4	3
Transc., reg. of*/ Άλλη βιολογική κατηγορία	1	1
Transc., reg. of*/ Intracellular signal transduction/ DNA repair	2	2
Transc., reg. of*/ mRNA catabolic process, NMD*/ Reg. of mRNA stability*	1	1
mRNA catabolic process, NMD*/ mRNA 3'-end processing	1	1
mRNA catabolic process, NMD*/ Translation	1	1
mRNA catabolic process, NMD*/ DNA repair	1	1
mRNA catabolic process, NMD*/ ncRNA processing	1	0
mRNA export from nucleus/ Intracellular signal transduction	1	1
mRNA export from nucleus/ Translation	1	1
Mitosis-Meiosis/ Intracellular signal transduction	1	0
Mitosis-Meiosis/ Protein transport	1	1
Mitosis-Meiosis/ Chromatin*	1	1
Intracellular signal transduction/ Translation/ Response to unfolded protein	1	1
Intracellular signal transduction/ Translation	1	1
Intracellular signal transduction/ Άλλη βιολογική κατηγορία	1	1
Reg. of mRNA stability*/ Translation	1	1
Reg. of mRNA stability*/ DNA replication	1	0
DNA replication/ Translation	1	1
DNA replication/ DNA repair	1	1
Splicing*- Transc., reg. of*/ mRNA export from nucleus	2	2
Splicing*- Transc., reg. of*/ ncRNA processing	1	1
Splicing*- Transc., reg. of*/ Intracellular signal transduction	2	2
Splicing*- Transc., reg. of*/ Translation	1	1
Splicing*- Transc., reg. of*/ mRNA catabolic process, NMD*	2	2
Splicing*- Transc., reg. of*/ DNA repair	1	1
Splicing*- Transc., reg. of*/ Mitosis-Meiosis - DNA repair	1	1
Splicing*/ Transc., reg. of*/ mRNA 3'-end processing/ Reg. of mRNA stability*	1	1
Splicing*/ mRNA export from nucleus/ mRNA catabolic process, NMD*	2	2
Splicing*/ mRNA catabolic process, NMD*/ Intracellular signal transduction	1	1
Splicing*/ mRNA catabolic process, NMD*/ Translation	2	1
Transc., reg. of*/ Chromatin*/ Protein transport	1	1
Transc., reg. of*/ Chromatin*/ mRNA catabolic process, NMD*	1	1
Transc., reg. of*/ Chromatin*/ Mitosis-Meiosis/ DNA replication	1	1
Chromatin*/ Response to unfolded protein	1	1
<b>Σύνολο πρωτεϊνών</b>	<b>122</b>	<b>98</b>

Splicing\*: Nuclear mRNA splicing, Transc., reg. of\*: Transcription-Regulation of transcription, Chromatin\*: Chromatin organization, mRNA catabolic process, NMD\*: Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay, Reg. of mRNA stability\*: Regulation of mRNA stability

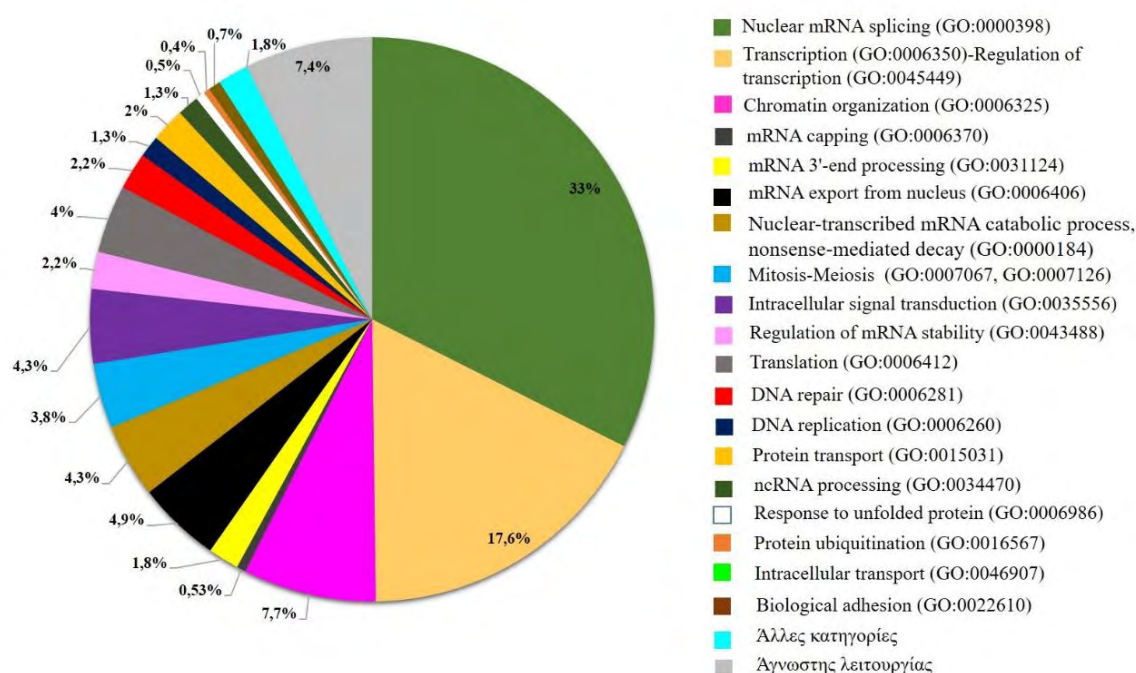
Για την καλύτερη οπτικοποίηση των παραπάνω αποτελεσμάτων, παρουσιάζονται στην Εικόνα 41, σε μορφή ποσοστών, τόσο τα αποτελέσματα της λειτουργικής κατηγοριοποίησης όλων των πρωτεϊνών-κόμβων του δικτύου (πάνω αριστερό μέρος) όσο και αυτά των πολυλειτουργικών πρωτεϊνών (κάτω δεξί μέρος).



**Εικ. 41: Λειτουργική ομαδοποίηση των 444 πρωτεϊνών-κόμβων του δικτύου του σωματίου συναρμογής.** Στο πάνω αριστερό μέρος απεικονίζονται τα ποσοστά των πρωτεϊνών που ανήκουν στον εκάστοτε όρο γονιδιακής οντολογίας, των πρωτεϊνών άγνωστης λειτουργίας και αυτών που ανήκουν σε άλλες βιολογικές κατηγορίες που δεν εντάσσονται σε κάποιο από τους παραπάνω όρους, με αντίστοιχο χρωματικό κώδικα. Στο κάτω δεξί μέρος φαίνονται τα ποσοστά των πολυλειτουργικών πρωτεϊνών και οι διαφορετικοί όροι γονιδιακής οντολογίας στους οποίους ανήκουν. Other GO term: Επιπλέον όροι γονιδιακής οντολογίας.

Στη συνέχεια, για να ελεγχθεί το ποσοστό εμπλουτισμού των δεκαεννέα οντολογικών κατηγοριών σε πρωτεΐνες του δικτύου, έγινε η συλλογή των πρωτεϊνών που ενσωματώθηκαν σε καθεμία από αυτές χωρίς να ληφθεί υπόψη το γεγονός ότι πολλές από αυτές ήταν πολυλειτουργικές (Εικ. 42), στο ίδιο πλαίσιο με την αντίστοιχη διερεύνηση των ποσοστών των οντολογικών κατηγοριών των πρωτεϊνών του σωματίου συναρμογής (Ενότητα 3.2.1.1.1.).

Στην Εικόνα 42 φαίνονται τα ποσοστά των οντολογικών κατηγοριών του δικτύου πρωτεϊνικών αλληλεπιδράσεων που υπολογίστηκαν με τον τρόπο που εξηγήθηκε εκτενώς και στην περίπτωση του συνόλου δεδομένων του σωματίου συναρμογής (Ενότητα 3.2.1.1.1.).



**Εικ. 42: Ομαδοποίηση των πρωτεϊνών σε κατηγορίες GO στο δίκτυο πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής.** Απεικονίζονται τα ποσοστά των κατηγοριών γονιδιακής οντολογίας που περιλαμβάνουν τις πρωτεΐνες-κόμβους του δικτύου, ακολουθώντας τον ίδιο χρωματικό κώδικα της Εικόνας 41.

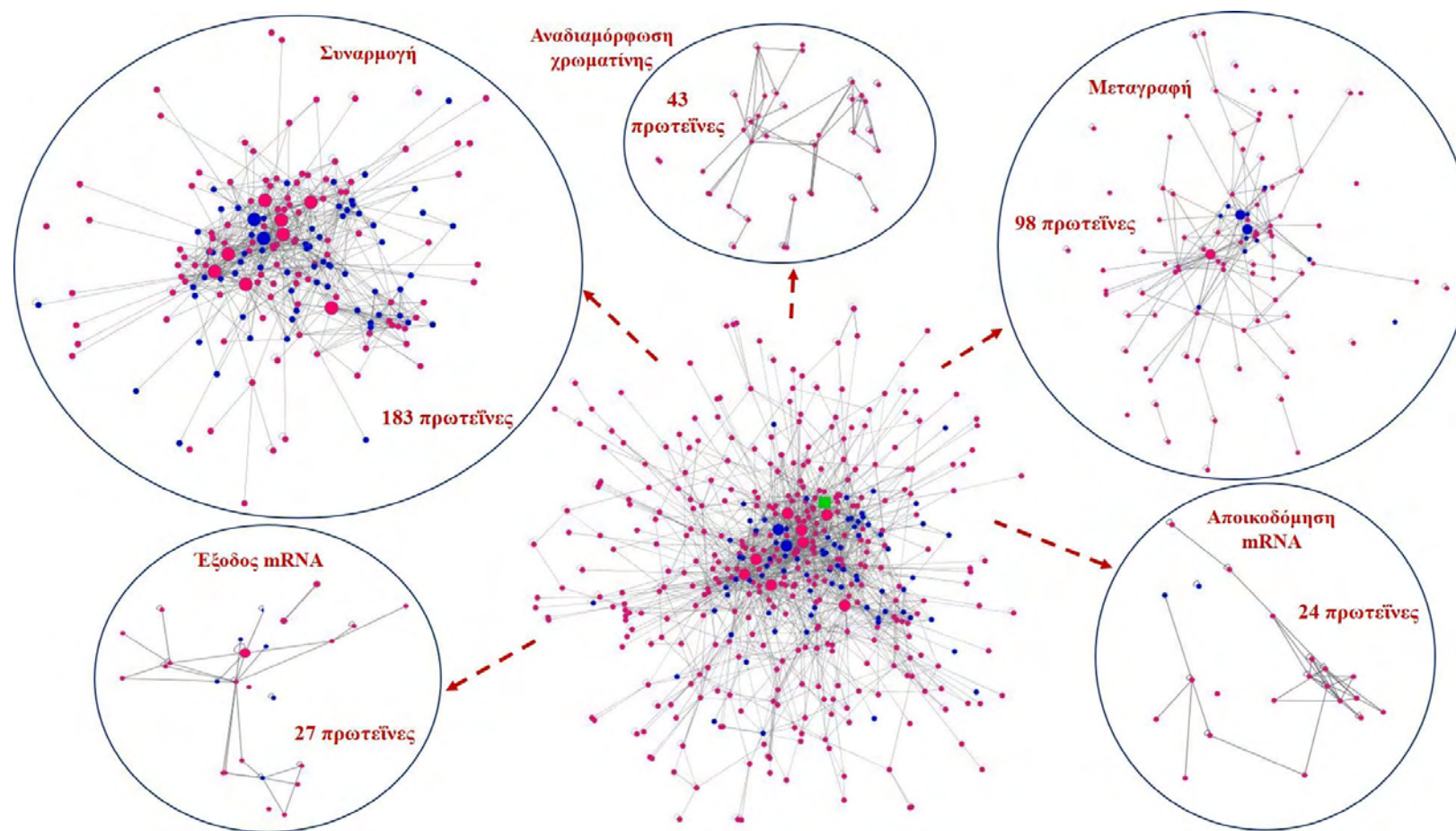
Για να γίνει η σύγκριση των παραπάνω δεδομένων με τα αντίστοιχα του συνόλου δεδομένων του σωματίου συναρμογής, παρουσιάζονται στον Πίνακα 13 τα σχετικά στοιχεία σε μορφή απόλυτων αριθμών. Η σύγκριση δεν ήταν δυνατή σε μορφή ποσοστών διότι το άθροισμα των οντολογικών κατηγοριών διέφερε σε καθεμία από τις δύο περιπτώσεις (Πιν. 13). Αυτό που παρατηρήθηκε ήταν ότι στο δίκτυο πρωτεϊνικών αλληλεπιδράσεων ενσωματώθηκαν ο μεγαλύτερος αριθμός των οντολογικών κατηγοριών (Πιν. 13). Ενδεικτικά ενσωματώθηκε το 89% της οντολογικής κατηγορίας της συναρμογής, το 67% της μεταγραφής-ρύθμισης της μεταγραφής, το 77% της οργάνωσης και αναδιαμόρφωσης της χρωματίνης, και κατ' αντιστοιχία το 84% και 86% των παραγόντων που συμμετείχαν στην έξοδο του mRNA από τον πυρήνα και στην αποικοδόμηση μη κατάλληλων μεταγράφων για τη μετάφραση. Αξίζει να σημειωθεί ότι από τις 186 πρωτεΐνες που δεν ενσωματώθηκαν στο δίκτυο πρωτεϊνικών

αλληλεπιδράσεων, η οντολογική κατηγορία της συναρμογής ήταν πολύ μειωμένη, μόλις στο 10% ενώ ήταν σχετικά αυξημένες, σε ποσοστό 30%, οι κατηγορίες της μεταγραφής-ρύθμισης της μεταγραφής και της οργάνωσης και αναδιαμόρφωσης της χρωματίνης. Το αποτέλεσμα αυτό μπορεί να θεωρηθεί αναμενόμενο διότι το 31% των πρωτεϊνών που δεν περιλαμβάνονταν στο δίκτυο είχαν απομονωθεί μόνο ως συστατικά του συμπλόκου E το οποίο είναι εμπλουτισμένο σε τέτοιου είδους παράγοντες για τους λόγους που έχουν ήδη αναλυθεί στην Ενότητα 3.2.1.1.1. των αποτελεσμάτων.

**Πιν. 13: Ομαδοποίηση των πρωτεϊνών σε GO του δικτύου PPI σε σύγκριση με το σύνολο δεδομένων του σωματίου συναρμογής.**

Κατηγορίες γονιδιακής οντολογίας (GO terms)	Αριθμός οντολογιών του συνόλου δεδομένων του σωματίου συναρμογής	Αριθμός οντολογιών του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής
Nuclear mRNA splicing (GO:0000398)	204	183
Transcription (GO:0006350)-Regulation of transcription (GO:0045449)	145	98
Chromatin organization (GO:0006325)	56	43
mRNA capping (GO:0006370)	3	3
mRNA 3'-end processing (GO:0031124)	17	10
mRNA export from nucleus (GO:0006406)	32	27
Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (GO:0000184)	28	24
Mitosis-Meiosis (GO:0007067, GO:0007126)	31	21
Intracellular signal transduction (GO:0035556)	35	24
Regulation of mRNA stability (GO:0043488)	14	12
Translation (GO:0006412)	31	22
DNA repair (GO:0006281)	14	12
DNA replication (GO:0006260)	10	7
Protein transport (GO:0015031)	13	11
ncRNA processing (GO:0034470)	17	7
Response to unfolded protein (GO:0006986)	5	3
Protein ubiquitination GO:0016567	5	2
Intracellular transport GO:0046907	4	0
Biological adhesion GO:0022610	5	4
Άλλες κατηγορίες	21	10
Άγνωστης λειτουργίας	90	33
<b>Σύνολο πρωτεϊνών</b>	<b>780</b>	<b>556</b>

Τα παραπάνω αποτελέσματα που αφορούσαν τις πέντε κύριες οντολογικές κατηγορίες (συναρμογή & σχετιζόμενες λειτουργικά διαδικασίες) απεικονίζονται στο δίκτυο πρωτεϊνικών αλληλεπιδράσεων (Εικ. 43) διότι αυτές οι κατηγορίες περιλαμβάνουν τον μεγαλύτερο αριθμό πρωτεϊνών. Στην Εικόνα 43 απομονώθηκαν τα συστατικά των πέντε αυτών κατηγοριών από το συνολικό δίκτυο διατηρώντας όμως την αρχική διάταξη του δικτύου. Με αυτόν τον τρόπο φαίνεται και η τοπολογία των πρωτεϊνών της καθεμίας κατηγορίας στο συνολικό δίκτυο.

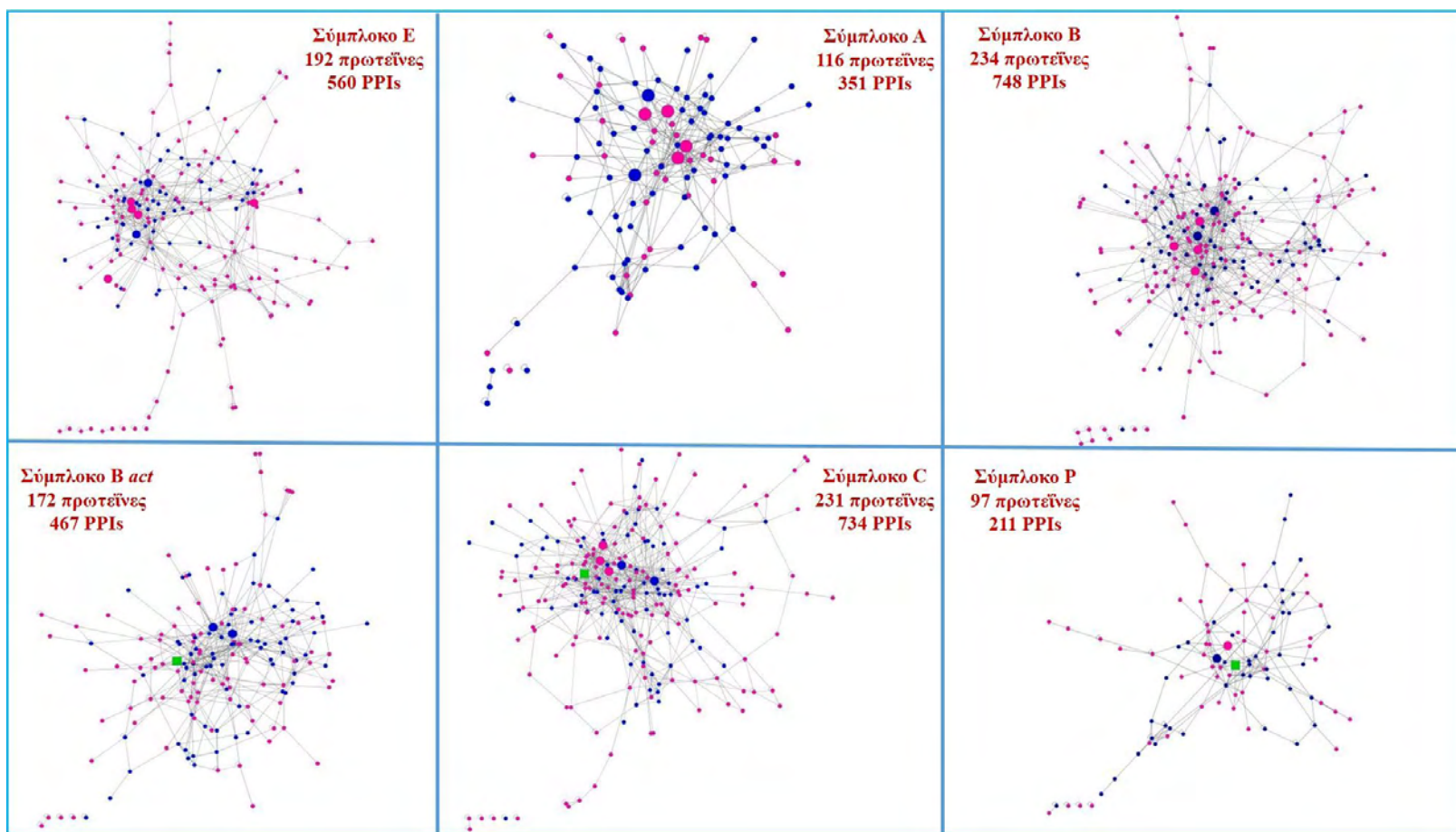


**Εικ. 43:** Δίκτυο πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής και επιμέρους δίκτυα των κύριων οντολογικών κατηγοριών. Στο κέντρο της εικόνας φαίνεται το δίκτυο PPI του σωματίου συναρμογής, και στη περιφέρεια τα δίκτυα που σχηματίζονται από τις πέντε κύριες οντολογικές κατηγορίες, διατηρώντας την αρχική διάταξη του δικτύου. Οι πρωτεΐνες-κόμβοι αναπαρίστανται με κυκλικό σχήμα και κόκκινο χρωματισμό και οι συνδέσεις μεταξύ τους με γκρι γραμμές. Οι πρωτεΐνες-hubs φαίνονται με μεγαλύτερο κόκκινο κυκλικό σχήμα και οι πρωτεΐνες με καθολική παρουσία στα υποσύμπλοκα A, B, B *act* και C με μπλε χρωματισμό. Η πρωτεΐνη FRA10AC1 φαίνεται με πράσινο τετράγωνο. Διάταξη: «Edge-weighted Spring-Embedded».

### 3.2.2. In silico ανακατασκευή των δικτύων πρωτεϊνικών αλληλεπιδράσεων των υποσυμπλόκων του σωματίου συναρμογής

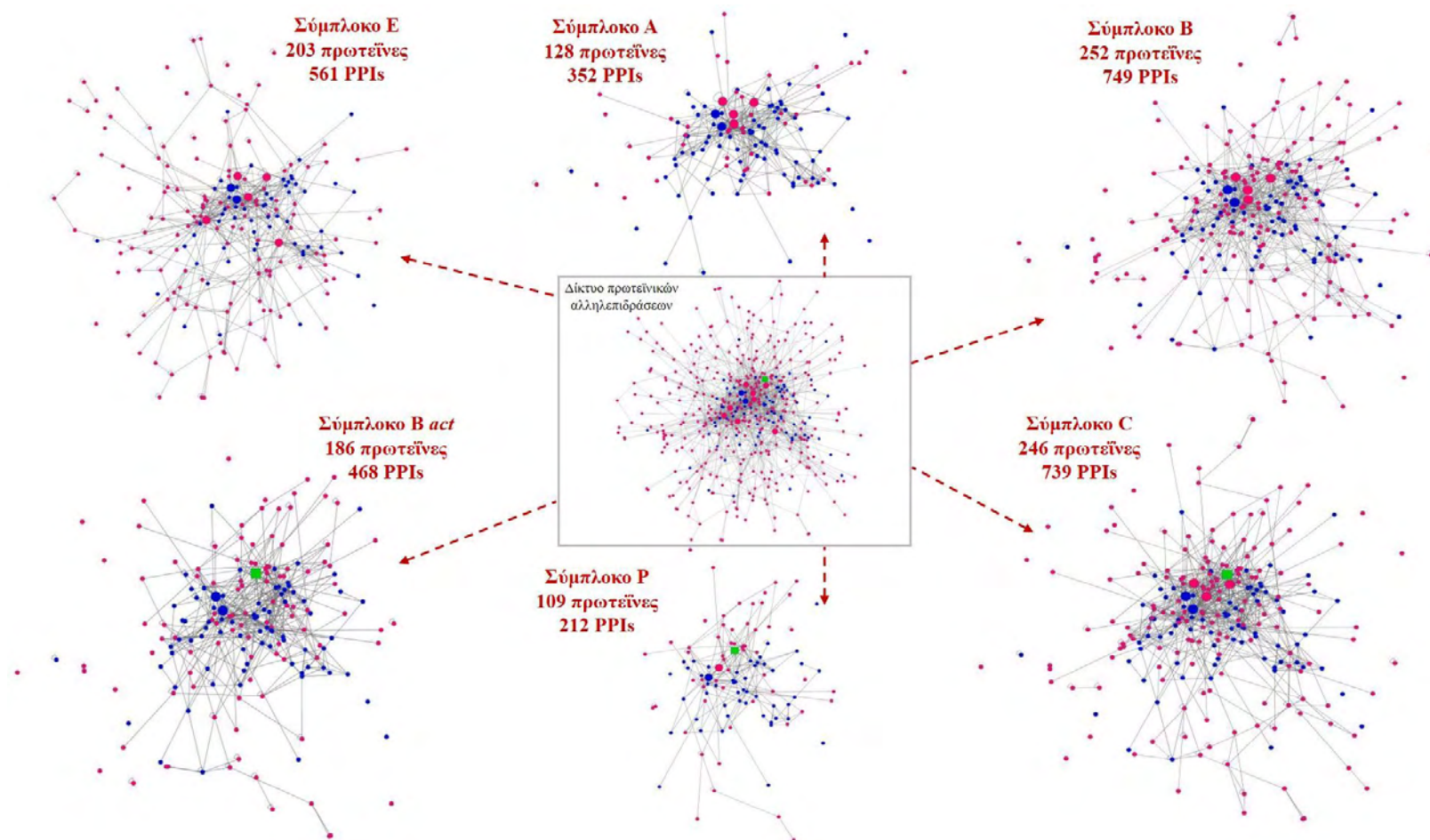
Τα επιμέρους σύμπλοκα του σωματίου συναρμογής αντιπροσωπεύουν μία κατάσταση που λαμβάνει χώρα *in vitro*. Ωστόσο, επειδή η συναρμολόγηση των επιμέρους συμπλόκων στο πρόδρομο μετάγραφο δεν συμβαίνει ταυτόχρονα και η σύστασή τους διαφέρει, η κατασκευή των δικτύων PPI των επιμέρους υποσυμπλόκων είχε ως στόχο μία αρχική και αδρή εκτίμηση της δυναμικής του σωματίου συναρμογής μέσω της αξιολόγησης της αλλαγής των αλληλεπιδράσεων των πρωτεϊνών στα υποσύμπλοκα και της αλλαγής των οντολογικών κατηγοριών των αντίστοιχων πρωτεϊνών πριν και μετά την ενσωμάτωσή τους στα δίκτυα.

Με βάση τη σύσταση του καθενός υποσυμπλόκου του σωματίου συναρμογής (E, A, B, B *act*, C και P), πραγματοποιήθηκε η ανακατασκευή των δικτύων PPI αυτών και η οπτικοποίησή τους με το πρόγραμμα cytoscape ([www.cytoscape.org](http://www.cytoscape.org)) σε διάταξη «Edge-weighted Spring-Embedded», χωρίς την ενσωμάτωση των πρωτεϊνών που είχαν ταυτοποιηθεί μόνο ως συστατικά ολόκληρου του σωματίου συναρμογής (81 πρωτεΐνες) (Πιν. 1 παράρτημα) (Εικ. 44). Η σύσταση των δικτύων σε πρωτεΐνες και PPI φαίνονται στην Εικόνα 44. Σε όλα τα δίκτυα των υποσυμπλόκων υπάρχουν πρωτεΐνες που βρίσκονται εκτός του ενιαίου δικτύου (Εικ. 44 & 45). Σε όλες τις περιπτώσεις ενσωματώθηκαν στα τελικά δίκτυα οι πρωτεΐνες που είχαν ταυτοποιηθεί ως συστατικά των επιμέρους συμπλόκων του σωματίου συναρμογής σε ποσοστά μεγαλύτερα από 70%. Η ανασύσταση των δικτύων έγινε ανεξάρτητα από το συνολικό δίκτυο, λαμβάνοντας υπόψη μόνο τα συστατικά του εκάστοτε συμπλόκου και τις αλληλεπιδράσεις που λαμβάνουν χώρα μεταξύ τους, εντός του συμπλόκου. Έτσι, στο εκάστοτε δίκτυο εφαρμόζονται διαφορετικές «δυνάμεις» μεταξύ των πρωτεϊνών-κόμβων με αποτέλεσμα αυτές να εμφανίζουν διαφορετική τοπολογία σε σύγκριση με αυτήν που είχαν στο συνολικό δίκτυο (Εικ. 44). Συμπληρωματικά, για να είναι δυνατή η σύγκριση μεταξύ των δικτύων των υποσυμπλόκων, έγινε η ανακατασκευή τους διατηρώντας τη διάταξη και την τοπολογία των πρωτεϊνών του συνολικού δικτύου, απομονώνοντας τις πρωτεΐνες του κάθε υποσυμπλόκου (Εικ. 45). Έτσι, σχηματίστηκαν έξι διαφορετικά δίκτυα, στα οποία οι πρωτεΐνες-κόμβοι διατηρούσαν τη θέση του αρχικού συνολικού δικτύου με τη διαφορά όμως ότι κάποιες από αυτές παρέμεναν ασύνδετες διότι χανόταν οι αλληλεπιδράσεις τους (Εικ. 45). Ο αριθμός αυτών των πρωτεϊνών βέβαια ήταν ελάχιστος. Στις Εικόνες 44 και 45 φαίνονται τα δίκτυα των πρωτεϊνικών αλληλεπιδράσεων των υποσυμπλόκων και στις δυο περιπτώσεις ανακατασκευής των δικτύων που αναφέρθηκαν, ο αριθμός των πρωτεϊνών-κόμβων (nodes) και των συνδέσεων (edges) σε κάθε ένα από αυτά.



**Εικ. 44:** Δίκτυα PPI των υποσυμπλόκων του σωματίου συναρμογής ανεξάρτητα από το συνολικό δίκτυο PPI. Φαίνονται τα έξι δίκτυα των υποσυμπλόκων, στα οποία οι πρωτεΐνες εμφανίζονται σε διαφορετική τοπολογία τόσο ως προς το συνολικό δίκτυο όσο και μεταξύ τους διότι επιδέχονται διαφορετικές δυνάμεις στη κάθε περίπτωση. Η χρωματική και σχηματική απεικόνιση των πρωτεϊνών και των αλληλεπιδράσεων είναι ίδιες με τις Εικόνες 40 και 43.

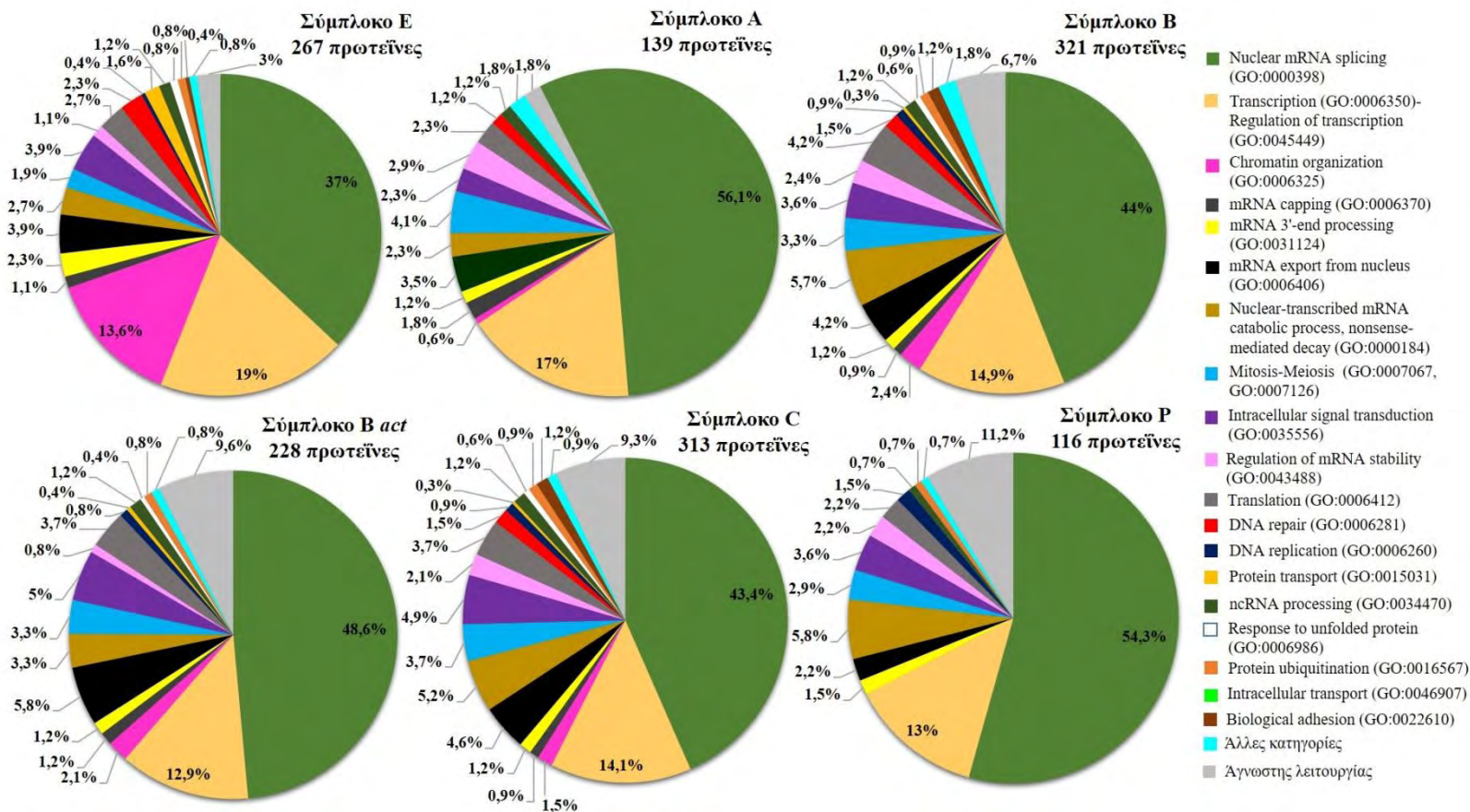




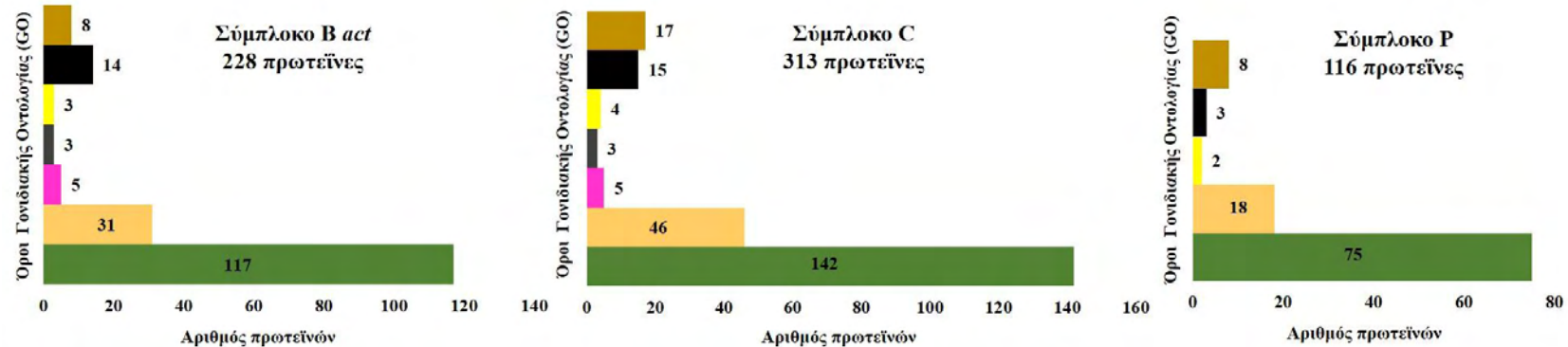
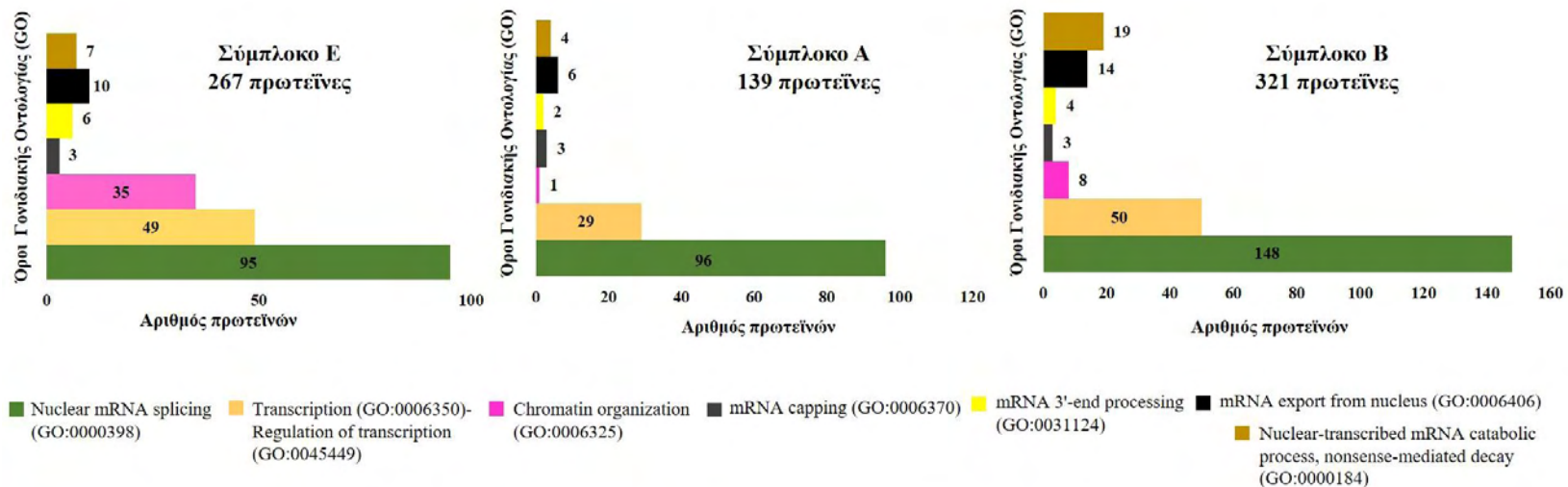
**Εικ. 45:** Δίκτυα PPI των υποσυμπλόκων του σωματίου συναρμογής διατηρώντας την τοπολογία των πρωτεϊνών-κόμβων του συνολικού δικτύου PPI. Φαίνονται τα έξι δίκτυα των υποσυμπλόκων, στα οποία οι πρωτεΐνες εμφανίζονται στην ίδια θέση που είχαν στο συνολικό δίκτυο καθώς και στο δίκτυο των άλλων υποσυμπλόκων. Η χρωματική και σχηματική απεικόνιση των πρωτεϊνών και των αλληλεπιδράσεων είναι ίδιες με τις 40 και 43.

Συγκρίνοντας τις αλληλεπιδράσεις των πρωτεϊνών στα δίκτυα των υποσυμπλόκων βρέθηκε ότι 156 αλληλεπιδράσεις ήταν ίδιες στα υποσύμπλοκα A, B, B *act* και C. Δεν επιλέχθηκαν τα υποσύμπλοκα E και P για τους λόγους που αναφέρθηκαν στην Ενότητα 3.2.1.1.4.. Αυτές οι αλληλεπιδράσεις λάμβαναν χώρα μόνο μεταξύ των 80 πρωτεϊνών που ήταν παρούσες και στα τέσσερα υποσύμπλοκα. Επιπρόσθετα, στην προσπάθεια εξέτασης της δυναμικής του σωματίου συναρμογής, βρέθηκαν 25, 70, 29 και 104 μοναδικές αλληλεπιδράσεις στα υποσύμπλοκα A, B, B *act* και C, αντίστοιχα. Μια αλληλεπίδραση ήταν μοναδική σε ένα υποσύμπλοκο όταν το ζεύγος των πρωτεϊνών είχε ταυτοποιηθεί μόνο σε αυτό το υποσύμπλοκο ή, εναλλακτικά, οι πρωτεΐνες του ζεύγους μπορεί να είχαν ταυτοποιηθεί σε περισσότερα από ένα υποσύμπλοκα, αλλά μόνο το υποσύμπλοκο που βρέθηκαν να αλληλεπιδρούν μοναδικά ήταν κοινό μεταξύ τους. Για παράδειγμα, η αλληλεπίδραση U2AF2-SRSF11 ήταν μοναδική στο υποσύμπλοκο B, αν και οι πρωτεΐνες του ζεύγους δεν ήταν μοναδικές στο υποσύμπλοκο αυτό, δηλαδή, η πρώτη είχε ταυτοποιηθεί στα υποσύμπλοκα A και B ενώ η δεύτερη στα B και C. Επιπλέον πρέπει να σημειωθεί ότι αρκετές από τις μοναδικές πρωτεΐνες του κάθε υποσυμπλόκου (Ενότητα 3.2.1.1.5.) δεν είχαν αλληλεπιδράσεις ώστε να ενσωματωθούν στα αντίστοιχα δίκτυα. Ο αριθμός αυτών των πρωτεϊνών ήταν 1, 13, 9 και 11 για τα υποσύμπλοκα A, B, B *act* και C, αντίστοιχα. Από τα παραπάνω γίνεται κατανοητό ότι ο προσδιορισμός των αλληλεπιδράσεων ως καθολικές ή μοναδικές εξαρτάται άμεσα από τη σύσταση των υποσυμπλόκων σε πρωτεΐνες (βλ. Ενότητα 4.2.4.).

Στη συνέχεια, ακολούθησε η διερεύνηση του εμπλουτισμού των επιμέρους συμπλόκων σε οντολογικές κατηγορίες, για να ελεγχθεί αν αυτά τα αποτελέσματα έρχονται σε συμφωνία με τα αντίστοιχα αποτελέσματα της σύστασης των επιμέρους συμπλόκων του σωματίου συναρμογής (Ενότητα 3.2.1.1.3.). Και σε αυτήν την περίπτωση, λόγω της ύπαρξης πολυλειτουργικών πρωτεϊνών, τα τελικά ποσοστά υπολογίστηκαν με βάση το σύνολο των οντολογικών κατηγοριών και όχι των πρωτεϊνών του εκάστοτε υποσυμπλόκου. Όπως αναλύθηκε και στην Ενότητα 3.2.1.1.3., για να επιτευχθεί η περαιτέρω σύγκρισή τους τα αντίστοιχα δεδομένα βρέθηκαν σε απόλυτους αριθμούς (Εικ. 47). Στις Εικόνες 46 και 47 φαίνονται τα αποτελέσματα της λειτουργικής ομαδοποίησης σε οντολογικές κατηγορίες των πρωτεϊνών του κάθε υποσυμπλόκου σε μορφή ποσοστών (Εικ. 46) και, επιπλέον, για τις επτά κυριότερες οντολογικές κατηγορίες, τα σχετικά αποτελέσματα σε απόλυτους αριθμούς (Εικ. 47).



**Εικ. 46: Ομαδοποίηση των κατηγοριών GO των δικτύων PPI των επιμέρους συμπλόκων του σωματίου συναρμογής.** Απεικονίζονται τα ποσοστά των κατηγοριών GO που αντιπροσωπεύονται στο δίκτυο του κάθε επιμέρους συμπλόκου, ακολουθώντας τον ίδιο χρωματικό κώδικα που χρησιμοποιήθηκε και στο συνολικό δίκτυο του σωματίου συναρμογής. Σε κάθε σύμπλοκο φαίνεται ο αριθμός των πρωτεϊνών που περιέχει.



**Εικ. 47: Ομαδοποίηση των πρωτεϊνών των δικτύων PPI των επιμέρους συμπλόκων του σωματίου συναρμογής σε επτά οντολογικές κατηγορίες.** Απεικονίζονται, σε μορφή ιστογράμματος, ο αριθμός των πρωτεϊνών των δικτύων των έξι υποσυμπλόκων που ανήκουν στις κατηγορίες GO της συναρμογής και σε λειτουργικά σχετιζόμενες διαδικασίες. Και σε αυτήν την εικόνα ακολουθείται ο χρωματικός κώδικας της Εικόνας 37.

Με βάση τα αποτελέσματα των Εικόνων 46 & 47 παρατηρήθηκε (και στο πλαίσιο των δικτύων των PPIs των υποσυμπλόκων του σωματίου συναρμογής) το ίδιο πρότυπο αλλαγής των οντολογικών κατηγοριών με αυτό που είχε παρατηρηθεί στο σύνολο των πρωτεϊνών τους πριν την ενσωμάτωσή τους σε δίκτυα (βλ. Ενότητα 3.2.1.1.3.). Όπως εξηγήθηκε αναλυτικά και στην Ενότητα. Ο εμπλουτισμός του υποσυμπλόκου E σε παράγοντες οργάνωσης και αναδιαμόρφωσης της χρωματίνης ήταν αναμενόμενος καθώς και ο εμπλουτισμός των υποσυμπλόκων B, B *act* και C σε παράγοντες συναρμογής και πρωτεΐνες που συμμετείχαν στην αποικοδόμηση του mRNA και στην έξοδο αυτού από τον πυρήνα. Επιπλέον, και στα πλαίσια των δικτύων, η κατηγορία της μεταγραφής-ρύθμισης της μεταγραφής δεν παρουσίασε διαφορές ανάμεσα στα δίκτυα των υποσυμπλόκων όπως παρατηρήθηκε και στο σύνολο των συστατικών των υποσυμπλόκων πριν ενσωματωθούν στα δίκτυα (Ενότητα 3.2.1.1.3.).

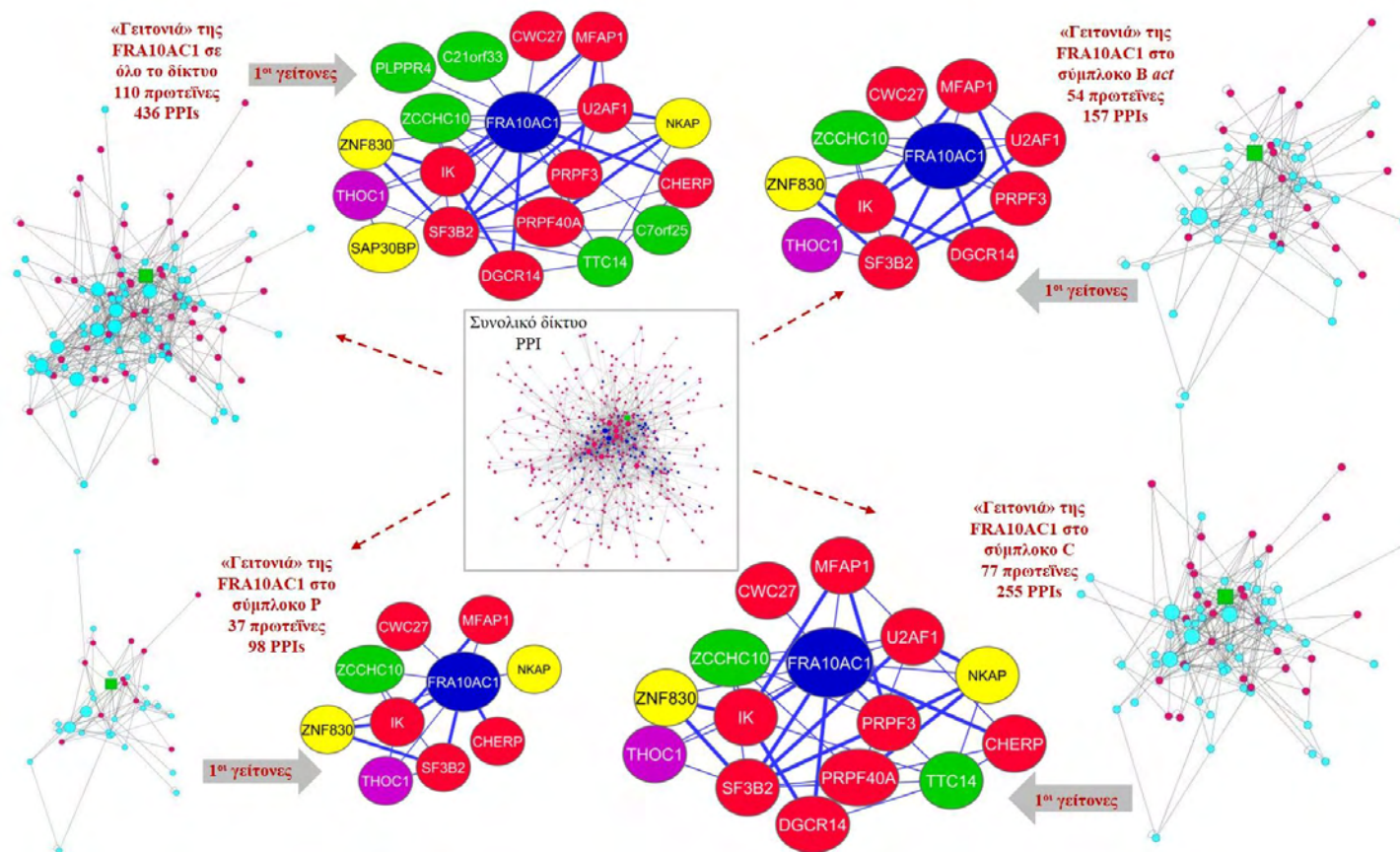
### 3.2.3. Διερεύνηση της τοπολογίας της πρωτεΐνης FRA10AC1 στο συνολικό δίκτυο του σωματίου συναρμογής και στα δίκτυα των υποσυμπλόκων

Όπως έχει αναφερθεί και παραπάνω, η πρωτεΐνη FRA10AC1 έχει απομονωθεί ως συστατικό των υποσυμπλόκων B *act*, C και P του σωματίου συναρμογής. Με στόχο την πιο λεπτομερή εξέταση των δεύτερων «γειτόνων» της FRA10AC1 ως προς τον αριθμό των PPI τους και ως προς τις λειτουργικές διαδικασίες που συμμετέχουν στο κάθε υποσύμπλοκο ξεχωριστά και την εξαγωγή συμπερασμάτων για την πιθανή ένταξή της σε κάποια συγκεκριμένη βιολογική διεργασία (π.χ. συμμετοχή στο πρώτο καταλυτικό βήμα της συναρμογής), πραγματοποιήθηκε η ανακατασκευή των δικτύων PPI των έξι υποσυμπλόκων και η διερεύνηση της «γειτονιάς» της FRA10AC1 σε αυτά τα υποσύμπλοκα, δηλαδή οι πρώτοι και δεύτεροι γείτονές της.

Μελετώντας, αρχικά τη «γειτονιά» της στο συνολικό δίκτυο, παρατηρήθηκε ότι στους πρώτους και δεύτερους γείτονες της υπήρχαν οι 8 από τις 10 πρωτεΐνες-«hubs» του συνολικού δικτύου, οι δύο από αυτές (PRPF40A, CHERP) ήταν πρώτοι «γείτονες» της. Και οι 8 αυτές πρωτεΐνες συμμετείχαν στη διαδικασία της συναρμογής (Εικ. 48). Συμπληρωματικά, αξίζει να αναφερθεί ότι περίπου το 60% της «γειτονιάς» της (πρώτοι και δεύτεροι γείτονες), συμμετείχαν στη διαδικασία της συναρμογής (Εικ. 48) και, επιπλέον, το 25% αυτών σε σχετιζόμενες λειτουργικά διαδικασίες με τη συναρμογή, όπως η μεταγραφή, η επεξεργασία του 3' άκρου του μεταγράφου, η έξοδος των ώριμων mRNA από τον πυρήνα και η αποικοδόμηση των μη κατάλληλων για μετάφραση μεταγράφων.

Παρόμοια διερεύνηση πραγματοποιήθηκε και στα υποσύμπλοκα B *act*, C και P, όπου και πάλι η «γειτονιά» της FRA10AC1 ήταν εμπλουτισμένη σε πρωτεΐνες που συμμετείχαν στη συναρμογή, σε ποσοστό μεγαλύτερο από 50% και στα τρία υποσύμπλοκα (Εικ. 48). Επιπλέον, ένα μη αμελητέο ποσοστό (~20%) από τις συνολικές πρωτεΐνες των οντολογικών κατηγοριών της εξόδου του ώριμου mRNA από τον πυρήνα και της αποικοδόμησης των μη κατάλληλων μεταγράφων για μετάφραση τοποθετήθηκαν στη «γειτονιά» της FRA10AC1 και στα τρία υποσύμπλοκα. Ακόμη, και στο πλαίσιο των δικτύων των υποσυμπλόκων, υπήρχαν πρωτεΐνες που αποτελούσαν «hubs», με δυο από αυτές να είναι πρώτοι γείτονες της FRA10AC1 (PRPF40A και CHERP), τις οποίες διατηρούσε από το συνολικό δίκτυο. Μια τελευταία

διαπίστωση ήταν ότι η FRA10AC1 διατηρούσε τους πρώτους γείτονες της από το συνολικό δίκτυο και στα τρία υποσύμπλοκα, με τη μεγαλύτερη πλειοψηφία αυτών να παρουσιάζεται στο σύμπλοκο C (13 από τους 15 που βρέθηκαν) (Εικ. 48). Στην Εικόνα 48 φαίνονται οι πρώτοι και δεύτεροι γείτονες της FRA10AC1 σε όλο το δίκτυο και στα δίκτυα των υποσυμπλόκων *B act*, C και P, διατηρώντας την αρχική τους θέση στο δίκτυο, οι πρωτεΐνες-«hubs» και αυτές που συμμετείχαν στη διαδικασία της συναρμογής. Ταυτόχρονα, στην ίδια εικόνα, φαίνονται οι πρώτοι γείτονες της FRA10AC1 που διατηρήθηκαν στα υποσύμπλοκα *B act*, C και P, χρησιμοποιώντας αυτή τη φορά τη διάταξη του δικτύου PPI της FRA10AC1 (Ενότητα 3.1.4.).



**Εικ. 48:** Δίκτυα PPI της «γειτονιάς» της FRA10AC1 στο συνολικό δίκτυο PPI του σωματίου συναρμογής και στα δίκτυα των υποσυμπλόκων B *act*, C και P. Στο κέντρο της εικόνας φαίνεται το συνολικό δίκτυο, και στην περιφέρεια τα δίκτυα των πρώτων και δεύτερων «γειτόνων» της FRA10AC1 τόσο στο συνολικό δίκτυο όσο και στα επιμέρους σύμπλοκα B *act*, C και P, διατηρώντας την αρχική θέση που κατείχαν οι πρώτοι και δεύτεροι γείτονες στο συνολικό δίκτυο. Στα δίκτυα των πρώτων γειτόνων με κόκκινο χρωματισμό και με σχήμα κύκλου απεικονίζονται οι πρωτεΐνες-κόμβοι των δικτύων, με γκρι γραμμές οι συνδέσεις μεταξύ τους, με μεγαλύτερου μεγέθους κύκλο οι πρωτεΐνες-«hubs» του δικτύου, με γαλάζια απόχρωση και με κύκλο οι πρωτεΐνες που συμμετέχουν στη διαδικασία της συναρμογής και με πράσινο τετράγωνο η FRA10AC1. Στα δίκτυα των πρώτων γειτόνων φαίνονται οι πρωτεΐνες που συμμετείχαν στη συναρμογή (κόκκινο), στη μεταγραφή (κίτρινο) και στην έξοδο του mRNA από τον πυρήνα (μωβ), και οι πρωτεΐνες άγνωστης λειτουργίας (πράσινο) όπως και στην Εικόνα 32. Με μπλε έντονες γραμμές απεικονίζονται οι αλληλεπιδράσεις με αυξημένη αξιοπιστία.

### 3.2.4. Ανάλυση τοπολογικών παραμέτρων του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής

Η αξιολόγηση και ο χαρακτηρισμός των βιολογικών δικτύων μέσω των τοπολογικών ιδιοτήτων του γράφου μπορεί να οδηγήσει στην ανάκτηση πληροφοριών σχετικά με τη συνολική δομή του (Dong & Horvath 2007, Almaas 2007, Zhu et al. 2007). Ο προσδιορισμός των απλών (simple) και των σύνθετων (complex) τοπολογικών παραμέτρων του δικτύου έγινε με την εφαρμογή Network Analyzer plugin (Assenov et al. 2008) του προγράμματος Cytoscape. Οι τιμές των απλών παραμέτρων φαίνονται στον Πίνακα 14.

Πιν. 14: Τιμές απλών τοπολογικών παραμέτρων του δικτύου PPI του σωματίου συναρμογής.

Απλές παράμετροι	Τιμή
Ακτίνα δικτύου (Network radius)	5
Διάμετρος δικτύου (Network diameter)	8
Χαρακτηριστικό μήκος μονοπατιού (Characteristic path length)	3,527
Μέσος όρος αριθμού γειτόνων (Average number of neighbor)	6,815
Μικρότερα μονοπάτια (Shortest paths)	196692 (100%)
Συντελεστής ομαδοποίησης (Clustering coefficient)	0,14
Πυκνότητα δικτύου (Network density)	0,015
Κεντροκοποίηση δικτύου (Network centralization)	0,159
Ετερογένεια δικτύου (Network heterogeneity)	1,021

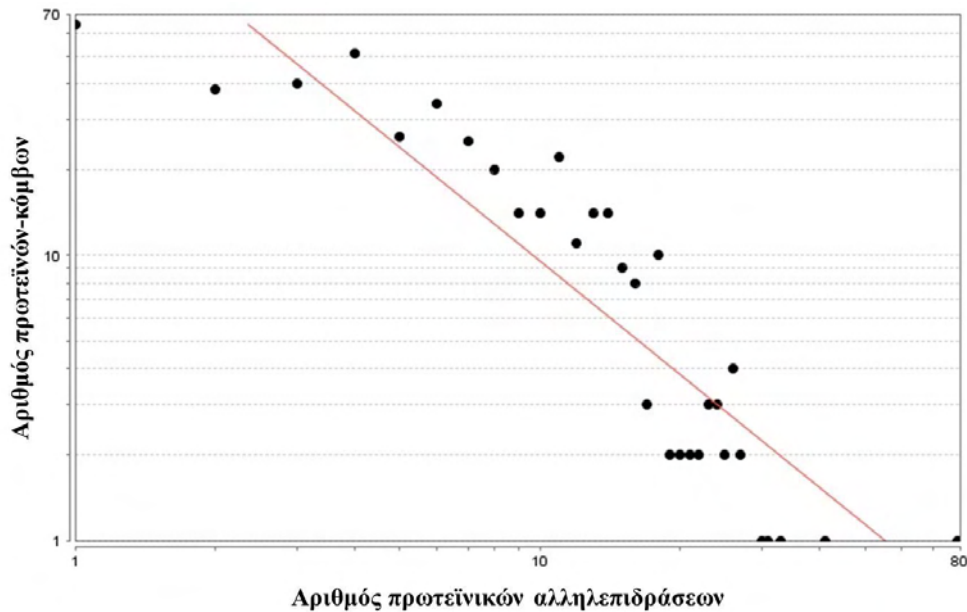
Η συνδεσιμότητα και η πυκνότητα του δικτύου εξαρτάται από τρεις παραμέτρους: α) τον μέσο αριθμό των πρωτεϊνών-«γειτόνων» (average number of neighbors), που δείχνει την κατά μέσο όρο συνδεσιμότητα των πρωτεϊνών-κόμβων (αριθμός άμεσα συνδεδεμένων «γειτόνων») (Dong & Horvath 2007), β) την πυκνότητα του δικτύου (network density=0,015)-μια κανονικοποιημένη εκδοχή της παραπάνω παραμέτρου- η οποία μετράει πόσο πυκνά το δίκτυο συμπληρώνεται με συνδέσεις (edges). Οι τιμές της πυκνότητας κοντά στο 1 δείχνουν ότι δεν υπάρχει μεγάλος αριθμός πρωτεϊνών-κόμβων «απομονωμένων» από τους υπόλοιπους του δικτύου (Doncheva et al. 2012) και γ) την παράμετρο «clustering coefficient» (=0,14) η οποία υπολογίζει το λόγο του αριθμού των αλληλεπιδράσεων που υπάρχουν μεταξύ των πρώτων «γειτόνων» μιας πρωτεΐνης προς τις πιθανές αλληλεπιδράσεις που μπορεί να υπάρξουν μεταξύ αυτών. Όταν η τιμή της πλησιάζει το 1, υποδηλώνει ότι οι πρωτεΐνες-κόμβοι βρίσκονται στο κέντρο ομάδων πρωτεϊνών που είναι στενά συνδεδεμένες μεταξύ τους (Barabasi & Oltvai 2004) (Πιν. 14).

Η ομοιομορφία της συνδεσιμότητας του δικτύου υποδηλώνεται από τις υπόλοιπες τέσσερις παραμέτρους. Το μικρό χαρακτηριστικό μήκος του μονοπατιού από μια πρωτεΐνη-κόμβο σε μια άλλη (characteristic ή shortest path length), δείχνει το μικρότερο αριθμό των συνδέσεων που ενώνουν μια πρωτεΐνη-κόμβο με οποιαδήποτε άλλη στο δίκτυο. Έτσι, η αύξηση της τιμής του υποδεικνύει ένα δίκτυο το οποίο δεν είναι στενά συνδεδεμένο καθώς οι πρωτεΐνες-κόμβοι βρίσκονται «απομακρυσμένες»



μεταξύ τους. Στην ανάλυση του συγκεκριμένου δικτύου η τιμή αυτή ήταν 3,527 κατά μέσο όρο συνδέσεις (edges) μεταξύ δυο πρωτεϊνών-κόμβων του δικτύου, το οποίο υποδηλώνει στενή συνδεσιμότητα μεταξύ τους (Πιν. 14) (Doncheva et al. 2012). Συμπληρωματικά, η παράμετρος «shortest paths metric», που υπολογίζει τον αριθμό των μονοπατιών με το μικρότερο μήκος, ήταν-στην περίπτωση αυτού του δικτύου-100% (196692 shortest paths) (Πιν. 14). Ο αριθμός αυτός υποδηλώνει τον αριθμό των μονοπατιών τα οποία ενώνουν όλες τις πρωτεΐνες του δικτύου με όλες τις υπόλοιπες και το αντίστοιχο ποσοστό υποδεικνύει ότι σχηματίζεται ένα ενιαίο δίκτυο. Ωστόσο, το ποσοστό αυτό θα μειώνονταν στο 96% αν είχαμε συμπεριλάβει στην ανάλυση και τις πρωτεΐνες-κόμβους που σχημάτιζαν ομοδιμερή και ετεροδιμερή εκτός του ενιαίου δικτύου. Η συνδεσιμότητα του δικτύου υποστηρίζεται από τη σχετικά μικρή διάμετρο του δικτύου (network diameter=8) (Πιν. 14), η οποία δείχνει τη μεγαλύτερη τιμή της μεγαλύτερης απόστασης μεταξύ δυο οποιονδήποτε πρωτεϊνών-κόμβων. Επιπρόσθετα, υποστηρίζεται και από την ακτίνα του δικτύου (network radius=5) (μικρότερη τιμή της μεγαλύτερης απόστασης μεταξύ δυο οποιονδήποτε πρωτεϊνών-κόμβων), η οποία αν είχε μικρότερη τιμή θα συνέβαλλε περισσότερο στην αύξηση της συνδεσιμότητας μεταξύ των κόμβων του δικτύου (Doncheva et al. 2012). Τέλος, η τιμή της παραμέτρου «network centralization», το εύρος τιμών της οποίας κυμαίνεται από 0-1, είναι κοντά στο μηδέν (0,159). Οι τιμές κοντά στο μηδέν υποδεικνύουν ένα πιο ομοιόμορφα συνδεδεμένο δίκτυο, το οποίο εμφανίζει αλληλεπιδράσεις τόσο στο κέντρο όσο και στην περιφέρειά του. Αντίθετα, αν οι τιμές πλησίαζαν στο 1, τότε η δομή του δικτύου θα έμοιαζε σαν «αστέρι», δηλαδή ο μεγαλύτερος αριθμός των αλληλεπιδράσεων θα συγκεντρώνονταν στο κέντρο του δικτύου (Πιν. 14).

Η παράμετρος «network heterogeneity» δείχνει την ετερογένεια στον αριθμό των PPI των κόμβων του δικτύου και συνδέεται άμεσα με την σύνθετη παράμετρο της κατανομής του αριθμού των PPIs των πρωτεϊνών-κόμβων (node degree distribution). Η μέτρηση της ετερογένειας του δικτύου βασίζεται στη διακύμανση της συνδεσιμότητας των πρωτεϊνών-κόμβων (node degree) και δείχνει την τάση που έχει ένα δίκτυο να περιέχει πρωτεΐνες-κόμβους με υψηλή συνδεσιμότητα (hubs) (Dong & Horvath 2007). Τα βιολογικά δίκτυα έχουν την τάση να είναι ετερογενή, δηλαδή λίγες πρωτεΐνες-κόμβοι να είναι υψηλά συνδεδεμένες (hubs) και η πλειοψηφία των πρωτεϊνών-κόμβων να τείνουν να έχουν λίγες συνδέσεις (low node degree) (Barabasi & Oltvai 2004, Barabási 2009, Barabási & Albert 1999). Σύμφωνα με την παράμετρο της κατανομής του αριθμού των PPIs των πρωτεϊνών-κόμβων (node degree distribution), το δίκτυο PPI του σωματίου συναρμογής περιλαμβάνει πρωτεΐνες-κόμβους με ευρεία κλίμακα αριθμού συνδέσεων, ξεκινώντας από μια ή δυο συνδέσεις μέχρι και 79 συνδέσεις. Αυτό το πρότυπο συνάδει με την κατανομή του νόμου των δυνάμεων (power-law), δηλαδή, με το γεγονός ότι η πλειοψηφία των πρωτεϊνών-κόμβων παρουσιάζουν μικρό αριθμό συνδέσεων ενώ λίγες πρωτεΐνες-κόμβοι (hubs) έχουν μεγάλο αριθμό συνδέσεων (Εικ. 49).



**Εικ. 49: Κατανομή του αριθμού των αλληλεπιδράσεων των πρωτεϊνών-κόμβων (node degree distribution) του δικτύου του σωματίου συναρμογής.** Το δίκτυο ακολουθεί την κατανομή του νόμου των δυνάμεων (power-law), καθώς η πλειοψηφία των πρωτεϊνών-κόμβων παρουσιάζουν μικρό αριθμό συνδέσεων ενώ λίγες πρωτεΐνες-κόμβοι (hubs) έχουν μεγάλο αριθμό συνδέσεων.

Στον Πίνακα 15 παρουσιάζεται ο κατάλογος των δέκα πρωτεϊνών με τον υψηλότερο αριθμό συνδέσεων (Εικ. 40) καθώς και η αντίστοιχη οντολογική κατηγορία στην οποία ανήκουν. Όλες αυτές οι πρωτεΐνες συμμετέχουν κυρίως στη διαδικασία της συναρμογής του mRNA, και σε μικρότερο ποσοστό, σε άλλες σχετιζόμενες λειτουργικά διαδικασίες. Ειδικότερα, η U2AF2 συνδέεται με την περιοχή που είναι πλούσια σε πυριμιδίνες και βρίσκεται ανοδικά της 3' θέσης συναρμογής και βοηθάει στη σύνδεση του U2snRNP στο πρόδρομο μετάγραφο (Agrawal et al. 2016). Η SMN1 αποτελεί συστατικό του συμπλόκου SMN που είναι υπεύθυνο για τη συναρμολόγηση των snRNPs (Paushkin et al. 2002). Η SRPK2 συμμετέχει στο σχηματισμό του U4/U6.U5 snRNP πριν το 1<sup>ο</sup> καταλυτικό βήμα της συναρμογής. Η CSNK2A1 αποτελεί μια κινάση σερίνης/θρεονίνης η οποία συμμετέχει σε κάθε βήμα συναρμολόγησης και αποσυναρμολόγησης του σωματίου συναρμογής, φωσφορυλιώνοντας πρωτεΐνες που στρατολογούνται στα διαφορετικά στάδια, ορισμένες από τις οποίες αποτελούν συστατικά των snRNPs (Bian et al. 2013). Επιπλέον, συμμετέχει στη μεταγραφή-ρύθμιση της μεταγραφής φωσφορυλιώνοντας παράγοντες έναρξης της μεταγραφής και πρωτεΐνες που αποτελούν ρυθμιστές της RNA πολυμεράσης (Moir et al. 2012). Τέλος, παίζει ρόλο και στην έξοδο του mRNA από τον πυρήνα, φωσφορυλιώνοντας συστατικά του συμπλόκου ένωσης εξονίου-εξονίου τα οποία συνεισφέρουν στην καθοδήγηση του ώριμου μεταγράφου στους πυρηνικούς πόρους ώστε το mRNA να εξέλθει στο κυτταρόπλασμα (Trembley et al. 2005). Η hnRNPK συμμετέχει στην εναλλακτική συναρμογή (Expert-Bezancon et al. 2002), τη μεταγραφή (Moumen et al. 2005) και τη μετάφραση (Bomsztyk et al. 2004).

**Πιν. 15: Οι δέκα πρωτεΐνες του δικτύου του σωματίου συναρμογής με τον υψηλότερο αριθμό συνδέσεων.**

Όνομα πρωτεΐνης	Uniprot ID	Αριθμός PPI των πρωτεϊνών-κόμβων	BC	CC	Βιολογική περιγραφή (GO terms)
CSNK2A1	P68400	79	0.25072236	0.43948413	Συναρμογή του mRNA στον πυρήνα μέσω σωματίου συναρμογής GO:0000398, Μεταγραφή GO:0006350 Ρύθμιση της μεταγραφής GO:0045449, Έξοδος του mRNA από τον πυρήνα GO:0006406
HNRNPK	P61978	41	0.04501201	0.37638063	Εναλλακτική συναρμογή του mRNA στον πυρήνα (Expert-Bezanon A. et al., 2002), Μεταγραφή GO:0006350 Ρύθμιση της μεταγραφής GO:0045449, Μετάφραση GO:0006412
SRPK2	P78362	33	0.03473966	0.35047468	Συναρμογή του mRNA στον πυρήνα μέσω σωματίου συναρμογής GO:0000398
SMN1	Q16637	31	0.05823477	0.36104319	Συναρμογή του mRNA στον πυρήνα μέσω σωματίου συναρμογής GO:0000398
U2AF2	P26368	30	0.03531282	0.37447168	Συναρμογή του mRNA στον πυρήνα μέσω σωματίου συναρμογής GO:0000398
RBM10	P98175	27	0.02599537	0.38189655	Εναλλακτική συναρμογή του mRNA στον πυρήνα (Wang Y. et al., 2013)
SRPK1	Q96SB4	27	0.01787086	0.33433962	Συναρμογή του mRNA στον πυρήνα μέσω σωματίου συναρμογής GO:0000398
PRPF40A	O75400	26	0.02937814	0.37638063	Συναρμογή του mRNA στον πυρήνα μέσω σωματίου συναρμογής GO:0000398
RBM39	Q14498	26	0.02663831	0.35298805	Συναρμογή του mRNA στον πυρήνα μέσω σωματίου συναρμογής GO:0000398, Μεταγραφή GO:0006350 Ρύθμιση της μεταγραφής GO:0045449
CHERP	Q8IWX8	26	0.02525497	0.36702568	Εναλλακτική συναρμογή του mRNA στον πυρήνα (Sasaki-Osugi K. et al., 2013)
FRA10AC1	Q70Z53	15	0.00795795	0.32863501	Άγνωστης λειτουργίας

BC=betweenness centrality, CC=closeness centrality

### 3.2.4.1. Σύνθετες (complex) παράμετροι

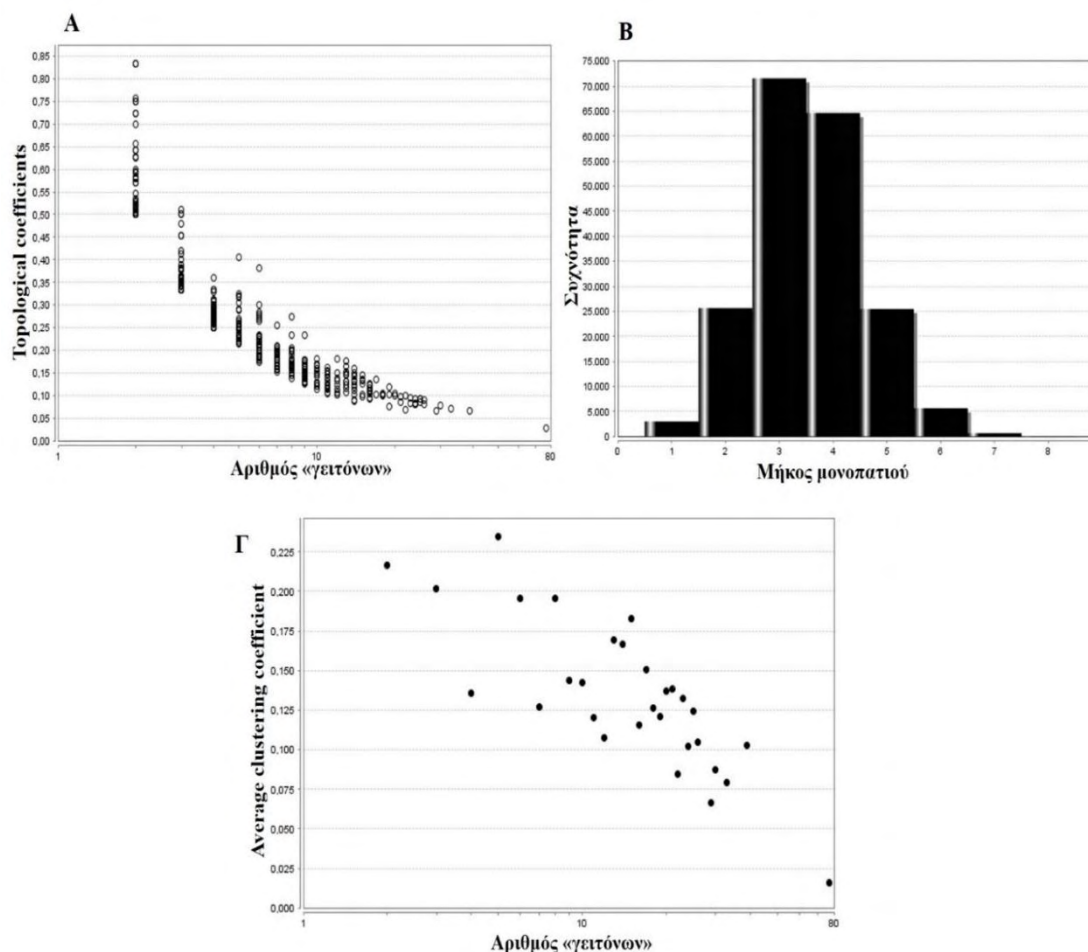
Οι πρωτεΐνες-«hubs» εμφάνισαν επίσης υψηλές τιμές της παραμέτρου «betweenness centrality» (υπολογίζει τον αριθμό των μικρότερων μονοπατιών που περνούν από κάθε πρωτεΐνη-κόμβο του δικτύου) και της «closeness centrality» (δείχνει την ικανότητα κάθε πρωτεΐνης-κόμβου να επικοινωνεί με οποιαδήποτε άλλη μέσα στο δίκτυο, με το μικρότερο πιθανό μονοπάτι) (Πιν. 15). Η υψηλή τιμή τους υποδεικνύει ότι αυτές οι πρωτεΐνες-κόμβοι ασκούν έλεγχο στις άλλες πρωτεΐνες του δικτύου και μπορούν να διαδώσουν γρήγορα τις πληροφορίες μεταξύ των κόμβων του δικτύου (Yoon et al. 2006, Freeman et al. 1979). Αξίζει να σημειωθεί ότι δυο από τις πρωτεΐνες-«hubs», η PRPF40A και η CHERP, είναι πρώτοι «γείτονες» της πρωτεΐνης FRA10AC1, και φαίνεται να έχουν κεντρικό ρόλο στο πλαίσιο του δικτύου PPI εξαιτίας των υψηλών τιμών των παραπάνω παραμέτρων. Οι αντίστοιχες τιμές αυτών των παραμέτρων για την πρωτεΐνη FRA10AC1, την κατατάσσουν στην 85<sup>η</sup> σειρά για την παράμετρο «closeness centrality» και στην 100<sup>η</sup> σειρά για την «betweenness centrality», γεγονός

που υποδεικνύει ότι βρίσκεται στο 25% των πρωτεϊνών που εμφανίζουν τις υψηλότερες τιμές αυτών των παραμέτρων. Αυτό σημαίνει ότι έχει σχετικά αυξημένο ποσοστό ικανότητας να επικοινωνεί με άλλες πρωτεΐνες με το μικρότερο πιθανό μονοπάτι (Πιν. 15). Στον Πίνακα 15 φαίνονται το επίσημο όνομα της πρωτεΐνης και το Uniprot ID των δέκα πρωτεϊνών-«hubs» και της πρωτεΐνης FRA10AC1. Επίσης, φαίνονται οι τιμές των παραμέτρων «betweenness centrality» και «closeness centrality», ο αριθμός των PPIs των πρωτεϊνών-«hubs» (node degree) και οι βιολογικές οντολογίες αυτών.

Σε αντίθεση με τα παραπάνω, η παράμετρος «average clustering coefficient distribution» (κατανομή μέσου όρου των συντελεστών ομαδοποίησης) μειώνεται όσο αυξάνεται ο αριθμός των PPIs των πρωτεϊνών-κόμβων του δικτύου (Εικ. 50). Η παράμετρος αυτή υπολογίζει, για τους κόμβους του δικτύου που έχουν τον ίδιο αριθμό PPIs, το μέσο όρο των τιμών «clustering coefficient» (συνδεσιμότητα των πρώτων «γειτόνων» κάθε πρωτεΐνης του δικτύου) και κατανέμει αυτούς τους μέσους όρους σε συνάρτηση με τον αριθμό των PPIs των κόμβων του δικτύου (Εικ. 50). Μια πρωτεΐνη λοιπόν που έχει μικρό αριθμό πρώτων «γειτόνων» (π.χ. 3 PPIs), για να έχει τη μέγιστη τιμή «clustering coefficient», πρέπει αυτοί οι 3 «γείτονες» να συνδέονται όλοι μεταξύ τους, δηλαδή να υπάρχουν 3 PPIs μεταξύ αυτών. Μια πρωτεΐνη που έχει μεγάλο αριθμό πρώτων «γειτόνων» (π.χ. 30 PPIs), για να έχει τη μέγιστη τιμή «clustering coefficient», πρέπει αυτοί οι 30 «γείτονες» να συνδέονται όλοι μεταξύ τους, δηλαδή να υπάρχουν 435 PPIs μεταξύ αυτών. Έτσι γίνεται κατανοητό ότι όσο αυξάνεται ο αριθμός των PPIs μιας πρωτεΐνης, τόσο μειώνεται η πιθανότητα αλληλεπίδρασης μεταξύ όλων των πρώτων «γειτόνων» λόγω της ύπαρξης μεγάλου αριθμού πιθανών αλληλεπιδράσεων που πρέπει να λάβουν χώρα. Στο ίδιο πλαίσιο, η μείωση της τιμής του συντελεστή τοπολογίας «topological coefficient» για μια πρωτεΐνη (ποσοστό κοινών «γειτόνων» μεταξύ των πιθανών ζευγών αυτής της πρωτεΐνης με όλες του δικτύου) καθώς αυξάνεται ο αριθμός των PPIs της αποτελεί ένα χαρακτηριστικό που παρατηρείται γενικότερα στα δίκτυα (Ravasz et al. 2002) και το οποίο ταυτοποιήθηκε και σε αυτό το δίκτυο (Εικ. 50). Ο συντελεστής τοπολογίας κάθε πρωτεΐνης του δικτύου υπολογίζεται ως ο μέσος όρος των τιμών των συντελεστών τοπολογίας που εμφανίζει με καθεμία από τις υπόλοιπες πρωτεΐνες του δικτύου (Εικ. 50). Οι πρωτεΐνες με μεγάλο αριθμό αλληλεπιδράσεων έχουν μικρή τιμή συντελεστή τοπολογίας διότι οι «γείτονες» τους έχουν λιγότερες αλληλεπιδράσεις σε σχέση με αυτές, με αποτέλεσμα να μειώνεται η πιθανότητα να εμφανίζουν πολλούς κοινούς «γείτονες» μεταξύ τους.

Τέλος, με βάση την σύνθετη παράμετρο κατανομής του μήκους του μικρότερου μονοπατιού (shortest path length distribution)-η οποία δίνει πληροφορίες για την κατανομή του μήκους του μικρότερου μονοπατιού (ο μικρότερος αριθμός συνδέσεων που ενώνουν δυο πρωτεΐνες) που συνδέει όλα τα πιθανά ζεύγη πρωτεϊνών-κόμβων-παρατηρείται ότι το πιο συχνό μικρότερο μήκος μονοπατιού του δικτύου είναι οι 3 συνδέσεις (shortest path length 3.0) (Εικ. 50). Η διαγραμματική απεικόνιση αυτής της παραμέτρου αναπαριστά πόσα μονοπάτια από τα συνολικά του δικτύου (196692) εμφανίζουν ένα συγκεκριμένο μήκος μικρότερου μονοπατιού (Εικ. 50). Η μικρή τιμή της παραμέτρου «shortest path length distribution» (κατανομή του μήκους του μικρότερου μονοπατιού) υποδεικνύει την «ανθεκτικότητα» του δικτύου (βλ. Ενότητα 4.2.9.). Στην Εικόνα 50 φαίνονται τα διαγράμματα κατανομής των παραμέτρων

«average clustering coefficient distribution» (κατανομή μέσου όρου των συντελεστών ομαδοποίησης), του συντελεστή τοπολογίας (topological coefficient) και του μήκους του μικρότερου μονοπατιού (shortest path length distribution), όπως αυτά προέκυψαν από την ανάλυση του δικτύου με το Network Analyzer plugin του προγράμματος Cytoscape.



**Εικ. 50:** Διαγράμματα σύνθετων παραμέτρων του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής. A: Η παράμετρος «topological coefficient» σε σχέση με τον αριθμό των γειτόνων των πρωτεϊνών-κόμβων του δικτύου, B: Η συχνότητα του μήκους των συνδέσεων (μήκος μονοπατιού) μεταξύ των πρωτεϊνών-κόμβων του δικτύου (shortest path length distribution) και Γ: Η παράμετρος «average clustering coefficient distribution» σε σχέση με τον αριθμό των «γειτόνων» των πρωτεϊνών-κόμβων του δικτύου.

3.2.4.2. Οι επιδράσεις της απομάκρυνσης των πρωτεϊνών-«hubs» και της FRA10AC1 στις τοπολογικές παραμέτρους το δικτύου του σωματίου συναρμογής και των υποσυμπλόκων του

Μια προκαταρκτική προσπάθεια επιβεβαίωσης της «ανθεκτικότητας» του δικτύου στην απομάκρυνση κόμβων και αλληλεπιδράσεων πραγματοποιήθηκε, σε πρώτο στάδιο, μέσω της απομάκρυνσης πρωτεϊνών-«hubs» οι οποίες είχαν τις περισσότερες αλληλεπιδράσεις (CSNK2A1 και HNRNPK) και, σε δεύτερο στάδιο, μέσω της απομάκρυνσης της πρωτεΐνης FRA10AC1. Σε όλες τις περιπτώσεις παρατηρήθηκε ότι ο συνολικός αριθμός των κόμβων και των αλληλεπιδράσεων του ενιαίου δικτύου

μειώθηκε. Ειδικότερα, και στις 3 περιπτώσεις ο αριθμός των κόμβων έγινε 443 από 444 και ο αριθμός των αλληλεπιδράσεων έγινε 1608, 1646 και 1671 από 1686 για την CSNK2A1, HNRNPK και FRA10AC1, αντίστοιχα (Πιν. 16). Μια αλλαγή που βρέθηκε ήταν ότι, και στις τρεις περιπτώσεις, υπήρχαν πρωτεΐνες που δεν συνδεόταν με το ενιαίο δίκτυο, όπως υποδεικνύεται και από την τιμή των μικρότερων μονοπατιών (shortest paths) που ανήκουν στο ενιαίο δίκτυο (Πιν. 16). Δηλαδή, οι πρωτεΐνες του ενιαίου δικτύου ήταν κατ' αντιστοιχία 436 και 442 για την πρωτεΐνη CSNK2A1 και τις HNRNPK και FRA10AC1. Ο αριθμός των απομονωμένων πρωτεϊνών ήταν μεγαλύτερος (7) στην περίπτωση της CSNK2A1, όπως φαίνεται και από τον αριθμό των shortest paths (95%), και μικρότερος στις άλλες δυο περιπτώσεις (1) (Πιν. 16). Γενικότερα με την αφαίρεση της CSNK2A1 παρατηρήθηκαν οι μεγαλύτερες αλλαγές στις απλές τοπολογικές παραμέτρους (Πιν. 16). Η μεγαλύτερη διαφορά ήταν στην ακτίνα του δικτύου η οποία μειώθηκε στο 1, και στις τιμές των παραμέτρων «network centralization» και «network heterogeneity» οι οποίες μειώθηκαν αρκετά (Πιν. 16). Συνδυαστικά, η μείωση αυτών των τριών τιμών υποδεικνύει ότι η αφαίρεση ενός «hub» με πολύ μεγάλο αριθμό συνδέσεων συμβάλλει στην αύξηση της συνδεσιμότητας μεταξύ των κόμβων του δικτύου και στην αύξηση της ομοιομορφίας της συνδεσιμότητας των κόμβων, αν ληφθεί υπόψη η ερμηνεία αυτών των παραμέτρων στα πλαίσια του δικτύου (βλ. προηγούμενη ενότητα). Αντίθετα, η αφαίρεση ενός «hub» με λιγότερες συνδέσεις, όπως η HNRNPK δεν επηρεάζει την «αρχιτεκτονική» του δικτύου, όπως αποτυπώνεται στις τοπολογικές παραμέτρους του δικτύου (Πιν. 16). Τέλος, εξετάζοντας τις απλές παραμέτρους του δικτύου έπειτα από την απομάκρυνση της FRA10AC1, δεν παρατηρήθηκε κάποια σημαντική αλλαγή (Πιν. 16). Τα παραπάνω υποδεικνύουν ότι το δίκτυο είναι «ανθεκτικό» σε τυχόν διαταραχές δεδομένου του γεγονότος ότι ακόμα και αν αφαιρεθούν πρωτεΐνες με πολύ μεγάλο αριθμό συνδέσεων (CSNK2A1), οι υπόλοιπες του δικτύου παραμένουν καλά συνδεδεμένες μεταξύ τους, και υπάρχει αμελητέα απώλεια κόμβων και αλληλεπιδράσεων.

**Πιν. 16: Τιμές απλών τοπολογικών παραμέτρων του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής έπειτα από την απομάκρυνση των πρωτεϊνών-«hubs» CSNK2A1 και HNRNPK με τις περισσότερες συνδέσεις και της πρωτεΐνης FRA10AC1.**

Απλές παράμετροι	Τιμή απουσία της CSNK2A1 (79 PPIs)	Τιμή απουσία της HNRNPK (41 PPIs)	Τιμή απουσία της FRA10AC1 (15 PPIs)	Τιμή παρουσία των «hubs» και της FRA10AC1
Αριθμός πρωτεϊνών (nodes)	443	443	443	444
Αριθμός αλληλεπιδράσεων (edges)	1608	1646	1671	1686
Ακτίνα δικτύου (Network radius)	1	5	5	5
Διάμετρος δικτύου (Network diameter)	8	8	8	8

Χαρακτηριστικό μήκος μονοπατιού (Characteristic path length)	3,686	3,544	3,528	3,527
Μέσος όρος αριθμού γειτόνων (Average number of neighbor)	6,483	6,655	6,763	6,815
Μικρότερα μονοπάτια (Shortest paths)	187924 (95%)	194922 (99%)	194922 (99%)	196692 (100%)
Συντελεστής ομαδοποίησης (Clustering coefficient)	0,134	0,136	0,139	0,14
Πυκνότητα δικτύου (Network density)	0,015	0,015	0,015	0,015
Κεντρικοποίηση δικτύου (Network centralization)	0,074	0,160	0,160	0,159
Ετερογένεια δικτύου (Network heterogeneity)	0,942	1,013	1,022	1,021

Η απομάκρυνση της FRA10AC1 πραγματοποιήθηκε και στα πλαίσια των δικτύων των υποσυμπλόκων *B act* και *C*. Το υποσύμπλοκο *P* δεν επιλέχθηκε διότι οι αλληλεπιδράσεις του συμπεριλαμβάνονται σε αυτές του *C*. Και στο πλαίσιο αυτό, ο συνολικός αριθμός των κόμβων μειώθηκε κατά μία πρωτεΐνη (την FRA10AC1) και στα δυο υποσύμπλοκα και ο αριθμός των αλληλεπιδράσεων του δικτύου των υποσυμπλόκων *B act* και *C* έγινε 459 και 722 από 468 και 739, αντίστοιχα (Πιν. 17). Ο αριθμός των «απομονωμένων» κόμβων αυξήθηκε κατά 1 και στα δυο υποσύμπλοκα. Πριν την απομάκρυνση της FRA10AC1 οι «απομονωμένοι» κόμβοι ήταν 15 και 17 για τα υποσύμπλοκα *B act* και *C*, αντίστοιχα. Ελέγχοντας τις υπόλοιπες απλές παραμέτρους των δικτύων αυτών, δεν παρατηρήθηκε και σε αυτήν την περίπτωση καμία ιδιαίτερη αλλαγή (Πιν. 17). Στον Πίνακα 17 φαίνονται οι απλές παράμετροι των δικτύων των υποσυμπλόκων *B act* και *C* πριν και μετά την απομάκρυνση της FRA10AC1.

**Πιν. 17: Τιμές απλών τοπολογικών παραμέτρων των δικτύων πρωτεϊνικών αλληλεπιδράσεων των υποσυμπλόκων *B act* και *C* πριν και μετά την απομάκρυνση της FRA10AC1.**

Απλές παράμετροι	Τιμές			
	<i>B act</i>	<i>B act</i> χωρίς FRA10AC1	<i>C</i>	<i>C</i> χωρίς FRA10AC1
Αριθμός πρωτεϊνών (nodes)	186	185	246	245
Αριθμός αλληλεπιδράσεων (edges)	468	459	739	722
Ακτίνα δικτύου (Network radius)	1	1	1	1
Διάμετρος δικτύου (Network diameter)	9	9	10	10
Χαρακτηριστικό μήκος μονοπατιού (Characteristic path length)	3,572	3,577	3,574	3,578
Μέσος όρος αριθμού γειτόνων (Average number of neighbor)	4,398	4,324	5,333	5,249
Μικρότερα μονοπάτια (Shortest paths)	27724 (80%)	27062 (79%)	50402 (83%)	49508 (82%)

Συντελεστής ομαδοποίησης (Clustering coefficient)	0,103	0,100	0,158	0,155
Πυκνότητα δικτύου (Network density)	0,024	0,024	0,022	0,022
Κεντρικοποίηση δικτύου (Network centralization)	0,074	0,075	0,089	0,090
Ετερογένεια δικτύου (Network heterogeneity)	0,841	0,836	0,898	0,894

### 3.2.4.3. Ανάλυση σχηματισμού συμπλεγμάτων πρωτεϊνών εντός του δικτύου του σωματίου συναρμογής

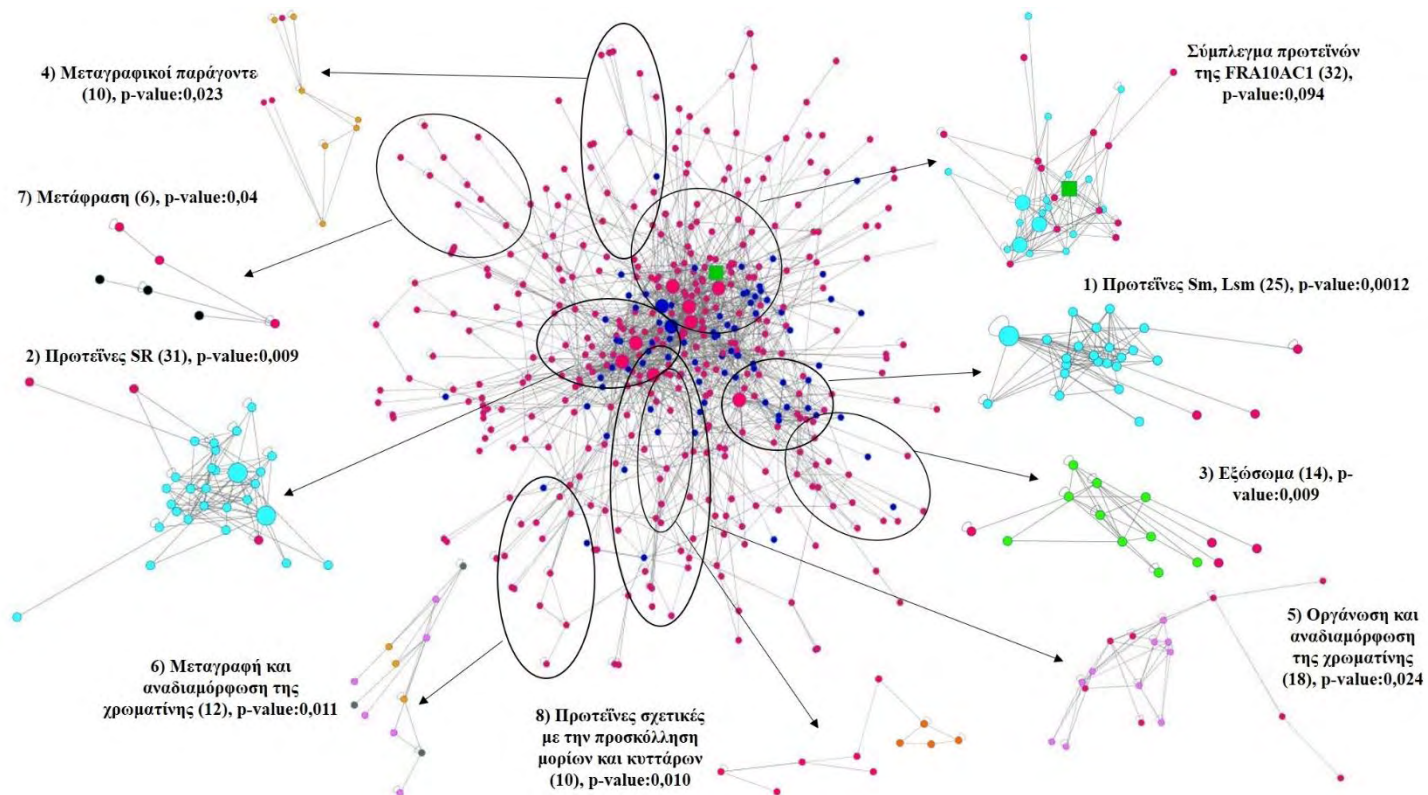
Το τελευταίο βήμα στην ανάλυση του δικτύου πρωτεϊνικών αλληλεπιδράσεων ήταν ο έλεγχος της ικανότητάς του να περιλαμβάνει συμπλέγματα πρωτεϊνών (clusters), δηλαδή, ομάδες πρωτεϊνών που είναι στενά συνδεδεμένες μεταξύ τους και εμφανίζουν αραή συνδεσιμότητα με τις υπόλοιπες (Spirin & Mirny 2003, Mitra et al. 2013). Η ανάλυση αυτή έγινε με την εφαρμογή Cluster one, η οποία περιγράφεται αναλυτικά στην Ενότητα 2.2.2. των Μεθόδων, στο Cytoscape (Nepusz et al. 2012). Εάν οι πρωτεΐνες του «cluster» συμμετέχουν σε κοινές λειτουργικές διαδικασίες αλλά αλληλεπιδρούν σε διαφορετικό χρόνο και τόπο τότε αυτά χαρακτηρίζονται ως λειτουργικά (functional modules). Ενώ, αν οι πρωτεΐνες που το απαρτίζουν αλληλεπιδρούν μεταξύ τους την ίδια χρονική στιγμή και στον ίδιο τόπο, τότε τα «clusters» χαρακτηρίζονται ως πρωτεϊνικά σύμπλοκα (protein complexes) (Spirin & Mirny 2003, Chen et al. 2014). Τα λειτουργικά συμπλέγματα συχνά περιέχουν ένα ή περισσότερα πρωτεϊνικά σύμπλοκα σε συγκεκριμένο χρόνο και τόπο (Chen et al. 2014).

Η εφαρμογή αυτής της ανάλυσης είχε ως αποτέλεσμα τη δημιουργία οκτώ στατιστικά σημαντικών συμπλεγμάτων (clusters) (τιμή p-value<0,05) τα οποία δεν ήταν αλληλοεπικαλυπτόμενα μεταξύ τους (Εικ. 51). Αυτά περιείχαν με σειρά μειωμένης στατιστικής σημαντικότητας: 1) τις πρωτεΐνες Sm και Lsm που βοηθούν στη συναρμολόγηση των snRNPs (p-value=0.0012, 25 πρωτεΐνες), 2) τις πρωτεΐνες SR που συνεισφέρουν στη συναρμογή και στην εναλλακτική συναρμογή (p-value=0,009, 31 πρωτεΐνες), 3) τις πρωτεΐνες του εξωσώματος που συμμετέχουν στην αποικοδόμηση του mRNA στον πυρήνα (p-value=0,009, 14 πρωτεΐνες), 4) μεταγραφικούς παράγοντες (p-value=0,023, 10 πρωτεΐνες), 5) παράγοντες που συμμετέχουν στην οργάνωση και στην αναδιαμόρφωση της χρωματίνης (p-value=0,019, 18 πρωτεΐνες), 6) πρωτεΐνες που συμβάλλουν στη μεταγραφή και στην αναδιαμόρφωση της χρωματίνης (p-value=0,011, 12 πρωτεΐνες), 7) πρωτεΐνες που σχετίζονται με την μετάφραση (αμινοακυλίωση των tRNA, π.χ. RARS) και την μεταφορά των πρωτεϊνών (p-value=0,040, 6 πρωτεΐνες), 8) πρωτεΐνες που συμμετέχουν στην αντιγραφή του DNA (KIN), στη μεταγραφή και στην προσκόλληση κυττάρων μέσω του σχηματισμού διακυτταρικών κόμβων (intercellular junctions) (JUP), στη σταθερότητα του mRNA (PABPC4), την πρωτεΐνη EEF1A1, η οποία αποτελεί μία ισομορφή του παράγοντα επιμήκυνσης της μετάφρασης EEF1A και συμμετέχει στη μεταγραφή γονιδίων σε απόκριση στο θερμικό σοκ, και 4 πρωτεΐνες που αποτελούν συστατικά των δεσμοσωμάτων και σχετίζονται με την προσκόλληση μορίων και κυττάρων (DSG1, PKP1, PKP2, DSP) (Patel & Green 2014) (p-value=0,010, 9 πρωτεΐνες) (Εικ. 51). Η ύπαρξη των συμπλεγμάτων (7) και (8) εντός του δικτύου του σωματίου συναρμογής



φαίνεται μη αναμενόμενη λόγω του λειτουργικού ρόλου των πρωτεϊνών που περιλαμβάνουν, ωστόσο, η περαιτέρω διερεύνησή αυτών των πρωτεϊνών για επιπρόσθετες λειτουργίες των ισομορφών τους δικαιολογούν την παρουσία τους (βλ. συζήτηση). Η πρωτεΐνη FRA10AC1 ταυτοποιήθηκε στα όρια στατιστικής σημαντικότητας ( $p$ -value=0.094) σε σύμπλεγμα 32 πρωτεϊνών. Οι μισές από αυτές συμμετέχουν στη διαδικασία της συναρμογής, γεγονός το οποίο ενισχύει την υπόθεση για τη συμμετοχή της FRA10AC1 σε αυτή τη διαδικασία. Στο σύμπλεγμα αυτό συμμετείχαν οι 13 (εκτός από την SAP30BP και την ZNF830) από τους 15 πρώτους γείτονες της FRA10AC1 (Εικ. 51).

Τα αποτελέσματα αυτής της ανάλυσης έδειξαν ότι οι πρωτεΐνες του κάθε συμπλέγματος (cluster) που δημιουργήθηκε, συμμετείχαν κατά το πλείστον σε κοινή λειτουργική διαδικασία (π.χ. συναρμογή/εναλλακτική συναρμογή, μεταγραφή, οργάνωση και αναδιαμόρφωση της χρωματίνης), δηλαδή είναι λειτουργικά συμπλέγματα (functional modules). Ορισμένα «clusters» αποτελούνταν από πρωτεΐνες οι οποίες είναι γνωστό ότι είναι συστατικά πρωτεϊνικών συμπλόκων (π.χ. εξώσωμα, κύρια συστατικά των snRNPs). Δεδομένου ότι αυτές οι πρωτεΐνες έχουν απομονωθεί πειραματικά από τον ίδιο κυτταρικό τύπο, τα συγκεκριμένα «clusters» μπορούν να χαρακτηριστούν ως «protein complexes». Στην Εικόνα 51 φαίνονται τα συμπλέγματα πρωτεϊνών (clusters) που προέκυψαν από την ανάλυση του δικτύου, ο αριθμός πρωτεϊνών του καθενός από αυτά, η οντολογική κατηγορία που εμφανίζεται εμπλουτισμένη και η τιμή στατιστικής σημαντικότητας.

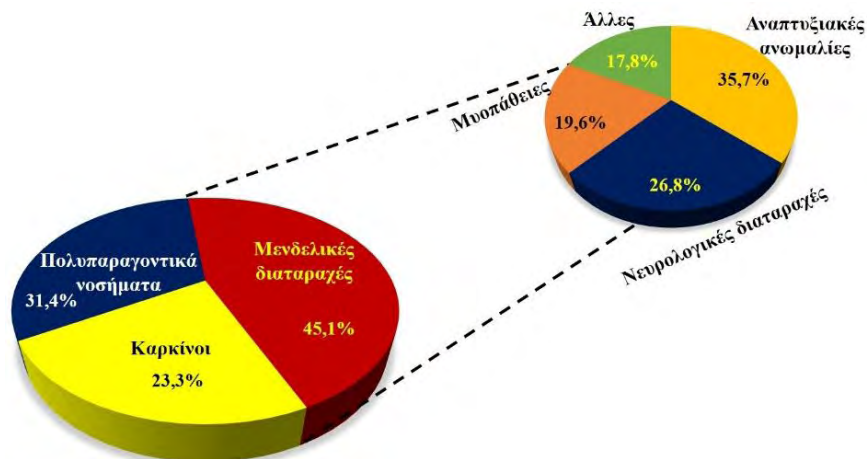


**Εικ. 51: Συμπλέγματα πρωτεϊνών του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής.** Στο κέντρο της εικόνας φαίνεται το δίκτυο PPI και στην περιφέρεια τα οκτώ συμπλέγματα πρωτεϊνών που προέκυψαν από την ανάλυση με το Cluster one plugin του Cytoscape, διατηρώντας τη θέση τους στο αρχικό δίκτυο. Ο αριθμός των πρωτεϊνών κάθε συμπλέγματος φαίνεται στις παρενθέσεις. Στα συμπλέγματα οι διαφορετικοί όροι γονιδιακής οντολογίας φαίνονται με διαφορετικό χρώμα (συναρμογή/εναλλακτική συναρμογή=γαλάζιο, μεταγραφή-ρύθμιση της μεταγραφής=κίτρινο, αναδιαμόρφωση και οργάνωση της χρωματίνης=ροζ, αποικοδόμηση των μη κατάλληλων mRNA για μετάφραση=πράσινο, μετάφραση=μαύρο, προσκόλληση μορίων και κυττάρων=πορτοκαλί, πολυλειτουργικές πρωτεΐνες=γκρι). Με μεγαλύτερου μεγέθους κυκλικό σχήμα απεικονίζονται οι πρωτεΐνες που αποτελούν «hubs» του δικτύου και με πράσινο τετράγωνο η FRA10AC1. Τα διαφορετικά συμπλέγματα είναι εμπλουτισμένα με πρωτεΐνες των λειτουργικών κατηγοριών που αναγράφονται στην εικόνα.

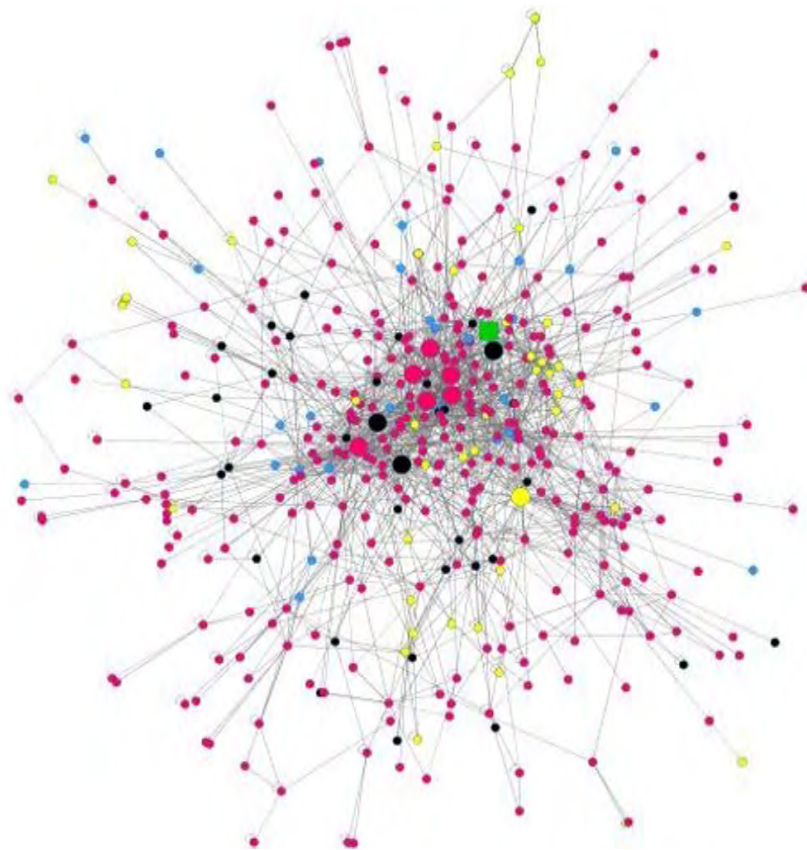
### 3.2.5. Συσχετισμός των συστατικών του σωματίου συναρμογής με γενετικά νοσήματα

Ένας από τους στόχους της βιολογίας δικτύων είναι η «απλοποίηση» των πολύπλοκων βιολογικών συστημάτων μέσω της κατασκευής δικτύων αλληλεπιδράσεων πρωτεϊνών, γονιδίων, μεταβολιτών και άλλων μορίων. Τα δίκτυα αυτά έχουν αποτελέσει χρήσιμα εργαλεία για την ανάλυση των λειτουργικών ιδιοτήτων των συστημάτων αυτών (Vidal 2009). Μια άλλη προσέγγιση που ακολουθείται και αναπτύσσεται αρκετά τα τελευταία χρόνια είναι η χρήση των δικτύων ως εργαλεία για τη διερεύνηση ομάδων γονιδίων ή πρωτεϊνών που αλληλεπιδρούν μεταξύ τους και πιθανόν σχετίζονται με την εμφάνιση παρόμοιων φαινοτύπων (Vidal et al. 2011). Η επέκταση αυτής της προσέγγισης είναι η εύρεση συμπλεγμάτων (clusters), εντός των πρωτεϊνικών δικτύων, τα οποία περιέχουν πρωτεΐνες που συμμετέχουν σε κοινή λειτουργική διαδικασία και παράλληλα προκαλούν ή σχετίζονται με την ίδια γενετική ασθένεια ή την ίδια κατηγορία γενετικών ασθενειών, με στόχο τη συσχέτιση των υπολοίπων πρωτεϊνών του συμπλέγματος με τους ίδιους ή παρόμοιους φαινοτύπους, όπως εφαρμόζεται και σε μία σχετικά πρόσφατη ερευνητική εργασία (Liu et al. 2015) (βλ. Ενότητα 1.5.1.5.).

Στο πλαίσιο αυτό, σε πρώτο στάδιο, πραγματοποιήθηκε αναζήτηση για τη συσχέτιση των πρωτεϊνών του σωματίου συναρμογής με γενετικά νοσήματα. Τα δεδομένα αυτά ανακτήθηκαν από τη βάση δεδομένων OMIM (<http://www.ncbi.nlm.nih.gov/omim/>) και συμπληρωματικά για τα έτη μετά την τελευταία ενημέρωσή της, από τη βάση δεδομένων PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>). Η διερεύνηση αυτή έδειξε ότι μεταλλάξεις σε 124 συστατικά του σωματίου συναρμογής σχετίζονται με 80 γενετικά νοσήματα. Το 45% αυτών των νοσημάτων είναι μενδελικά, το 24% πολυπαραγοντικά και το 31% σχετίζονται με διάφορα είδη καρκίνων, δηλαδή, έχουν βρεθεί σωματικές μεταλλάξεις διαφόρων γονιδίων σε καρκινικά κύτταρα (Εικ. 55 & 56). Όσον αφορά τα μενδελικά νοσήματα, το 27% αυτών αφορά νευρολογικές διαταραχές, το 36% αναπτυξιακές ανωμαλίες, το 20% μυοπάθειες και το 17% άλλου τύπου διαταραχές (Εικ. 52). Εκατό δώδεκα από αυτές τις πρωτεΐνες ενσωματώθηκαν στο τελικό ανακατασκευασμένο δίκτυο (Εικ. 53). Οι τέσσερις από τις δέκα πρωτεΐνες-«hubs» του δικτύου έχουν συσχετιστεί με νοσήματα (Εικ. 53), η SMN1 με νωτιαία μυϊκή ατροφία (Spinal Muscular Atrophy) και οι SRPK1, SRPK2 και RBM10 με διάφορα είδη καρκίνου. Στις Εικόνες 52 και 53 φαίνονται η κατάταξη των συστατικών ολόκληρου του συνόλου δεδομένων του σωματίου συναρμογής στις διάφορες κατηγορίες νοσημάτων και οι πρωτεΐνες που σχετίζονταν με τις τρεις διαφορετικές κατηγορίες νοσημάτων στο πλαίσιο του δικτύου, με διαφορετικό χρωματικό κώδικα.



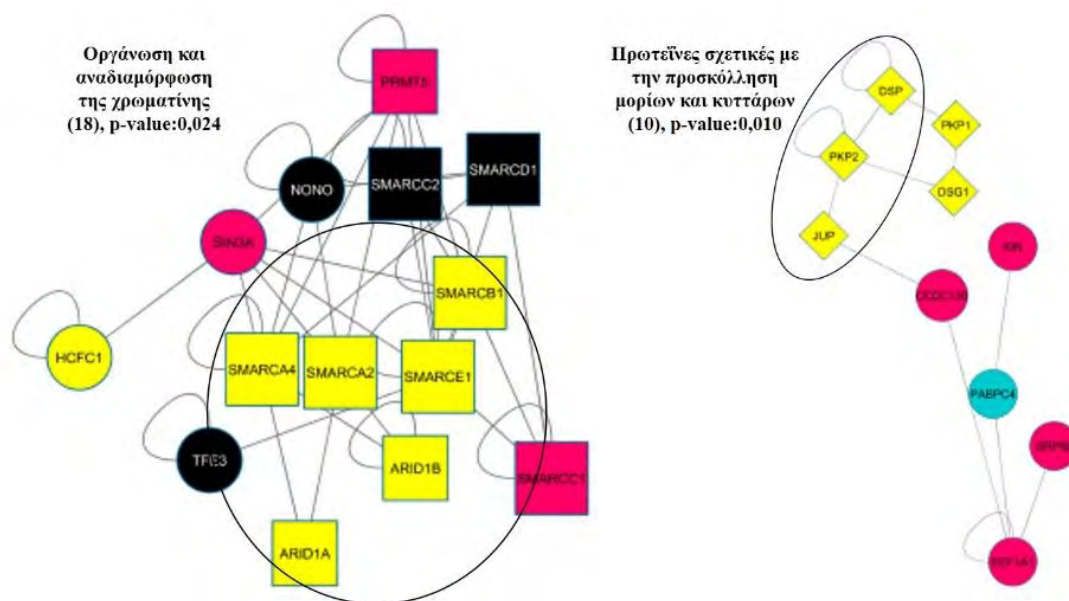
Εικ. 52: Κατηγορίες γενετικών νοσημάτων που σχετίζονται με μεταλλάξεις συστατικών του σωματίου συναρμογής.



Εικ. 53: Συσχετισμός των πρωτεϊνών-κόμβων με γενετικά νοσήματα στο πλαίσιο του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής. Οι πρωτεΐνες-κόμβοι που συνδέονται με μενδελικά νοσήματα απεικονίζονται με κίτρινο χρώμα (46), αυτές που συσχετίζονται με πολυπαραγοντικά νοσήματα με μπλε (30) και αυτές που σχετίζονται με καρκίνο με μαύρο χρώμα (36). Με μεγαλύτερο μέγεθος κύκλους φαίνονται οι πρωτεΐνες που αποτελούν «hubs» του δικτύου (10), και με πράσινο τετράγωνο η FRA10AC1. Οι πρωτεΐνες με κόκκινο χρώμα δεν έχουν συσχετιστεί έως τώρα με κάποιο νόσημα.

Σε δεύτερο στάδιο, με στόχο την εύρεση πρωτεϊνών, εντός των συμπλεγμάτων (clusters) του δικτύου συναρμογής, που έχουν κοινό λειτουργικό ρόλο και παράλληλα προκαλούν ή σχετίζονται με την ίδια γενετική ασθένεια ή ίδια κατηγορία γενετικών

ασθενειών, εξετάστηκαν τα clusters του δικτύου ως προς τη βιολογική λειτουργία των πρωτεϊνών που περιλαμβάνουν και τη συσχέτισή τους με το ίδιο γενετικό νόσημα ή την ίδια κατηγορία νοσημάτων. Η διερεύνηση αυτή οδήγησε στην εύρεση πρωτεϊνών που είχαν αυτά τα χαρακτηριστικά εντός δυο συμπλεγμάτων (clusters), στο 5<sup>ο</sup> (εμπλουτισμένο σε πρωτεΐνες οργάνωσης και αναδιαμόρφωσης της χρωματίνης) και στο 8<sup>ο</sup> (εμπλουτισμένο σε πρωτεΐνες που συμμετέχουν στην προσκόλληση μορίων και κυττάρων) (Εικ. 54). Ειδικότερα, στο 5<sup>ο</sup> cluster βρέθηκε ότι μεταλλάξεις των γονιδίων που κωδικοποιούν 6 πρωτεΐνες του συμπλόκου αναδιαμόρφωσης της χρωματίνης (σύμπλοκο SWI/SNF) (Εικ. 54) προκαλούν το ίδιο μενδελικό νόσημα, το σύνδρομο Coffin-Siris, το οποίο χαρακτηρίζεται από αναπτυξιακή υστέρηση, σοβαρή δυσχέρεια λόγου και ήπια νοητική υστέρηση (Kosho et al. 2013). Ακόμη, στο 8<sup>ο</sup> cluster, μεταλλάξεις σε γονίδια που κωδικοποιούν 3 πρωτεΐνες (Εικ. 54) οι οποίες συμμετέχουν στην προσκόλληση μορίων, προκαλούν την ίδια μενδελική μορφή καρδιοπάθειας (Arrhythmogenic Right Ventricular Cardiomyopathy, ARVC) (Boyce et al. 2012, Roberts et al. 2013). Συνεπώς, οι υπόλοιπες πρωτεΐνες του 5<sup>ου</sup> και του 8<sup>ου</sup> συμπλέγματος που αλληλεπιδρούν και έχουν την ίδια βιολογική λειτουργία με τις πρωτεΐνες του συμπλόκου SWI/SNF και με τις πρωτεΐνες που συμμετέχουν στην προσκόλληση μορίων και κυττάρων, αντίστοιχα, θα μπορούσαν να αποτελέσουν υποψήφια γονίδια για τα μενδελικά νοσήματα που προκαλούν τα γονίδια που κωδικοποιούνται από αυτές τις δυο ομάδες πρωτεϊνών.



**Εικ. 54: Συσχετισμός των πρωτεϊνών-κόμβων των δύο συμπλεγμάτων του δικτύου συναρμογής με γενετικά νοσήματα.** Φαίνονται με κίτρινο χρώμα οι πρωτεΐνες-κόμβοι που συνδέονται με μενδελικά νοσήματα, με μπλε αυτές που συσχετίζονται με πολυπαραγοντικά νοσήματα και με μαύρο αυτές που σχετίζονται με καρκίνο. Οι πρωτεΐνες με κόκκινο χρώμα δεν έχουν συσχετιστεί έως τώρα με κάποιο νόσημα. Σε κύκλο απεικονίζονται οι πρωτεΐνες που προκαλούν το ίδιο μενδελικό νόσημα. Οι όροι γονιδιακής οντολογίας με τους οποίους είναι εμπλουτισμένα τα δυο συμπλέγματα φαίνονται με τετράγωνο (οργάνωση και αναδιαμόρφωση της χρωματίνης) και με ρομβοειδές σχήμα (προσκόλληση μορίων και κυττάρων). Οι αριθμοί στις παρενθέσεις αντιπροσωπεύουν τον αριθμό των πρωτεϊνών των συμπλεγμάτων.

### 3.2.6. Πρωτεΐνες στην «γειτονιά» της πρωτεΐνης FRA10AC1 που σχετίζονται με γενετικά νοσήματα

Όπως έχει αναφερθεί και στην Ενότητα 3.1. της Εισαγωγής, το γονίδιο *FRA10AC1* χαρτογραφείται στην εύθραυστη θέση *FRA10A*, η έκφραση της οποίας έχει αναφερθεί ότι σχετίζεται με νοητική υστέρηση, νευρολογικές διαταραχές και αναπτυξιακές ανωμαλίες (Mavrou et al. 1991, Petit et al. 1986, Sutherland 1982 & 1985, Kahkonen et al. 1989, Sarafidou et al. 2004). Επίσης, σε μια πρόσφατη μελέτη έχει δειχθεί συσχέτιση πολυμορφισμού του γονιδίου *FRA10AC1* με την νόσο Alzheimer (Li et al. 2015).

Λαμβάνοντας υπόψη τα παραπάνω στοιχεία, έγινε η διερεύνηση του συσχετισμού των πρωτεϊνών-κόμβων της «γειτονιάς» της *FRA10AC1* με γενετικά νοσήματα μελετώντας συγκεκριμένα τους πρώτους και δεύτερους γείτονες αυτής, με στόχο να γίνει η πιθανή συσχέτισή της με κάποια κατηγορία νοσημάτων από τα παραπάνω που προαναφέρθηκαν. Το υποδίκτυο που δημιουργήθηκε περιλάμβανε 108 πρωτεΐνες-κόμβους, εκ των οποίων οι τριάντα δυο σχετίζονται με γενετικά νοσήματα (Εικ. 55). Οι 16 από αυτές συνδέονται με μενδελικά νοσήματα, κυρίως με αναπτυξιακές ανωμαλίες και με την μελαγχρωστική αμφιβληστροειδοπάθεια. Από τις υπόλοιπες πρωτεΐνες, οι δέκα έχουν συσχετιστεί με πολυπαραγοντικά νοσήματα και οι οκτώ με μη κληρονομικούς καρκίνους (Εικ. 55). Από τους πρώτους γείτονες της *FRA10AC1*, τρεις έχουν συσχετιστεί με γενετικά νοσήματα, η *U2AF1* με καρκίνο (μυελοδυσπλαστικό σύνδρομο), η *PRPF3* με μελαγχρωστική αμφιβληστροειδοπάθεια (μενδελικό νόσημα) και η *DGCR14* με σχιζοφρένεια (πολυπαραγοντικό νόσημα) (Wang et al. 2006). Αξίζει να σημειωθεί ότι τρεις δεύτεροι «γείτονες» της *FRA10AC1* έχουν επίσης συνδεθεί με μελαγχρωστική αμφιβληστροειδοπάθεια (*PRPF6*, *PRPF8*, *PRPF31*). Η ύπαρξη μεγάλου ποσοστού αναπτυξιακών ανωμαλιών στη «γειτονιά» της *FRA10AC1*, μπορεί να υποδεικνύει έναν πιθανό συσχετισμό της με τέτοιου είδους γενετικά νοσήματα. Επίσης, μελετώντας τους πρώτους γείτονες της πρωτεΐνης *FRA10AC1* που δεν ανήκουν στο σωματίο συναρμογής, παρατηρήθηκε ότι τρεις από αυτούς, η *MAP3K12*, η *VPS29* και η *EEF1D* έχουν συσχετιστεί με πολυπαραγοντικά νοσήματα (Parkinson και Alzheimer) (Schulte et al. 2014, Tedeschi & Bradke 2013, Berman et al. 2015) γεγονός που συνάδει με τη συσχέτιση που έχει αναφερθεί και για την *FRA10AC1* με Alzheimer (Li et al. 2015). Στην Εικόνα 55 φαίνονται οι πρώτοι και δεύτεροι γείτονες της *FRA10AC1* και η κατηγοριοποίησή τους στις τρεις κατηγορίες γενετικών νοσημάτων (μενδελικά, πολυπαραγοντικά, καρκίνοι).



**Εικ. 55:** Συσχετισμός πρώτων και δεύτερων γειτόνων της πρωτεΐνης **FRA10AC1** με γενετικά νοσήματα εντός του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής. Στη «γειτονιά» της πρωτεΐνης **FRA10AC1** (πράσινο τετράγωνο) ταυτοποιήθηκαν 32 πρωτεΐνες που σχετίζονται με μενδελικά νοσήματα (16) (κίτρινο), πολυπαραγοντικά νοσήματα (10) (μπλε) και με καρκίνο (8) (μαύρο). Με μεγαλύτερου μέγεθος κύκλους φαίνονται οι πρωτεΐνες-«hubs» (8). Οι πρωτεΐνες με κόκκινο χρώμα δεν έχουν συσχετιστεί έως τώρα με κάποιο νόσημα.

### 3.3. Έλεγχος επιλεγμένων πρωτεϊνικών αλληλεπιδράσεων του ανακατασκευασμένου δικτύου της πρωτεΐνης FRA10AC1

Με βάση το δίκτυο πρωτεϊνικών αλληλεπιδράσεων της πρωτεΐνης FRA10AC1 παρατηρήθηκε ότι η FRA10AC1 αλληλεπιδρά με πρωτεΐνες που συμμετέχουν σε διαφορετικές λειτουργικές διαδικασίες (Εικ. 32). Αξιολογώντας την αξιοπιστία των αλληλεπιδράσεων με βάση το λειτουργικό ρόλο των αλληλεπιδρώντων μορίων, την ένταξή τους σε συγκεκριμένο πρωτεϊνικό σύμπλοκο, τον υποκυτταρικό εντοπισμό τους, τον αριθμό και την κλίμακα των πειραμάτων τα οποία υποστηρίζουν την εκάστοτε αλληλεπίδραση, και την εξελικτική συντήρηση των αλληλεπιδράσεων (βλ. Ενότητα 3.1.2.), επιλέχθηκαν συγκεκριμένες πρωτεΐνες από τις λειτουργικές κατηγορίες που αφορούσαν την επεξεργασία του mRNA για τον περαιτέρω έλεγχό τους. Η εξέταση αυτών των αλληλεπιδράσεων αποτέλεσε έναν τρόπο για να ενισχυθούν οι ενδείξεις για την συμμετοχή της πρωτεΐνης στη διαδικασία της επεξεργασίας του mRNA, αλλά και να ελεγχθεί για πρώτη φορά ο συσχετισμός της με λειτουργικές διαδικασίες όπως η μεταγραφή και η έξοδος του mRNA από τον πυρήνα. Γι' αυτό το λόγο, οι πρωτεΐνες που επιλέχθηκαν ήταν οι παρακάτω:

- Η πρωτεΐνη MFAP1, η οποία αποτελεί έναν παράγοντα συναρμογής (Andersen et al. 2008) του πρώτου καταλυτικού βήματος της διαδικασίας. Η πρωτεΐνη αυτή έχει ταυτοποιηθεί ως συστατικό του σωματίου συναρμογής σε κοινά υποσύμπλοκα με την FRA10AC1 (*B act* και *C*) και η αφθονία της σε αυτά τα υποσύμπλοκα συνάδει με αυτή της FRA10AC1. Η επιλογή της έναντι της PRPF3 που συμμετέχει και αυτή στο ίδιο καταλυτικό βήμα έγινε εξαιτίας του μικρότερου μεγέθους του cDNA της MFAP1 το οποίο θα διευκόλυνε την επακόλουθη υποκλωνοποίησή του σε φορέα έκφρασης. Η U2AF1, ένας ακόμη παράγοντας συναρμογής, που αναγνωρίζει την 3' θέση συναρμογής δεν επιλέχθηκε διότι έχει ταυτοποιηθεί σε διαφορετικά υποσύμπλοκα του σωματίου συναρμογής από την FRA10AC1. Οι πρωτεΐνες IK, CHERP, SF3B2 και DGCR14 δεν επιλέχθηκαν διότι οι αλληλεπιδράσεις τους με την FRA10AC1 έχουν επιβεβαιωθεί με περισσότερες από μία πειραματικές μεθόδους. Τέλος, οι πρωτεΐνες CWC27 και PRPF40A δεν επιλέχθηκαν διότι δεν έχουν καλά χαρακτηρισμένο ρόλο στη συναρμογή και η HABP4 δεν αποτελεί συστατικό του σωματίου συναρμογής.
- Η πρωτεΐνη SAP30BP, η οποία συμμετέχει στη ρύθμιση της μεταγραφής και ειδικότερα στη μεταγραφική καταστολή γονιδίων (Chen et al. 2010). Η SAP30BP είναι η μοναδική από τις 3 πρωτεΐνες που χρησιμοποιήθηκαν η οποία δεν έχει ταυτοποιηθεί στα ίδια υποσύμπλοκα του σωματίου συναρμογής με την FRA10AC1, αλλά έχει απομονωθεί ως συστατικό του υποσυμπλόκου A σε χαμηλή αφθονία. Η επιλογή της SAP30BP έναντι της NKAP η οποία συμμετέχει κι αυτή στη ρύθμιση της μεταγραφής (Chen et al. 2003) στηρίχτηκε αποκλειστικά στο σημαντικά μικρότερο μέγεθος της πρώτης. Η επιλογή της SAP30BP αντί της ZNF830 έγινε διότι παρόλο που η δεύτερη έχει ταυτοποιηθεί ως συστατικό ενός συμπλόκου που συνεισφέρει στην επιμηκύνση της μεταγραφής δεν είναι γνωστός ο βιολογικός της ρόλος στη διαδικασία της μεταγραφής. Τέλος, η πρωτεΐνη EEF1D δεν επιλέχθηκε διότι δεν αποτελεί συστατικό του σωματίου συναρμογής.



- Η πρωτεΐνη THOC1, η οποία ήταν η μοναδική που συμμετέχει στην έξοδο των ώριμων mRNAs από τον πυρήνα. Η THOC1 έχει ταυτοποιηθεί ως συστατικό του σωματίου συναρμογής σε κοινά υποσύμπλοκα με την FRA101AC1 (B *act* και C) και στην ίδια αναλογία αφθονίας (χαμηλή). Η επιβεβαίωση αυτής της αλληλεπίδρασης θα μπορούσε να συσχετίσει για πρώτη φορά την FRA10AC1 με αυτήν την διαδικασία της επεξεργασίας του mRNA.

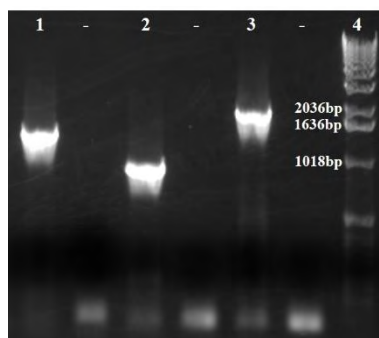
Ο έλεγχος των αλληλεπιδράσεων αυτών πραγματοποιήθηκε με τη μέθοδο Μεταφοράς Ενέργειας Φθορισμού μέσω Συντονισμού (Fluorescence Resonance Energy Transfer - FRET). Στην παρούσα εργασία επιλέχθηκαν οι φθορίζουσες χρωστικές GFP (Green Fluorescent protein, Tsien et al. 1998) και m-Cherry (Red Fluorescent protein, Shaner et al. 2004) οι οποίες παράγονται ως επίτοποι από τους αντίστοιχους πλασμιδιακούς φορείς έκφρασης pEGFP-N1 και pmCherry-C1.

Η διαδικασία που ακολουθήθηκε ήταν, σε πρώτο στάδιο, η υποκλωνοποίηση του πλήρους μήκους cDNA των γονιδίων *SAP30BP*, *THOC1* και *MFAP1* στο φορέα pmCherry-C1 (μόρια-δέκτες). Η πλασμιδιακή κατασκευή pEGFP-N1/*FRA10AC1* που χρησιμοποιήθηκε ως μόριο-δότης, είχε κατασκευαστεί παλιότερα (Sarafidou et al. 2004) (βλ. Υλικά & Μέθοδοι). Τα ανασυνδυασμένα πλασμίδια που χρησιμοποιήθηκαν ως μόριο-δότης και μόρια-δέκτες επιλέχθηκαν βάσει των φασματικών χαρακτηριστικών απορρόφησης και εκπομπής τους. Για τη βέλτιστη μεταφορά ενέργειας συντονισμού το φάσμα εκπομπής του δότη πρέπει να αλληλεπικαλύπτεται σημαντικά με το φάσμα απορρόφησης του δέκτη. Το εύρος του φάσματος εκπομπής της EGFP που είναι από τα 480nm έως τα 600nm (μέγιστη τιμή: 509nm) αλληλοεπικαλύπτεται με το εύρος του φάσματος απορρόφησης της mCherry το οποίο ξεκινά από τα 500nm και φτάνει έως τα 720nm (μέγιστη τιμή: 610nm) (Albertazzi et al. 2009). Σε δεύτερο στάδιο, έγινε η συν-διαμόλυνση των τριών διαγονιδίων pmCherry/*SAP30BP*, pmCherry/*THOC1*, pmCherry/*MFAP1* με το διαγονίδιο pEGFP-N1/*FRA10AC1* σε κύτταρα *HeLa*. Τέλος, έγινε ο έλεγχος των αλληλεπιδράσεων των τριών ζευγών πρωτεϊνών με ανάλυση FRET σε συνεστιακό μικροσκόπιο σάρωσης με ακτίνες Laser (Confocal Laser Scanning Microscope).

### 3.3.1. Υποκλωνοποίηση των cDNAs των γονιδίων *SAP30BP*, *THOC1* και *MFAP1* του ανθρώπου στον πλασμιδιακό φορέα έκφρασης σε κύτταρα θηλαστικών pmCherry-C1

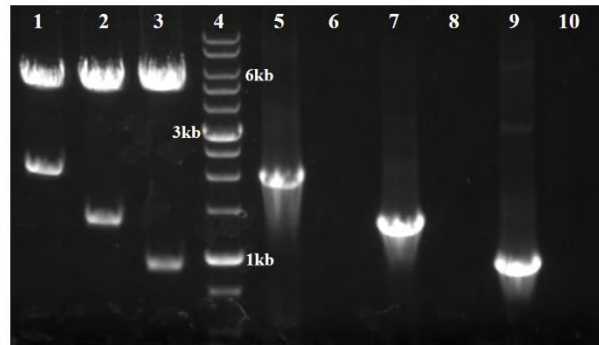
Η υποκλωνοποίηση των μορίων cDNA των γονιδίων *SAP30BP*, *THOC1* και *MFAP1* στο φορέα pmCherry-C1 σχεδιάστηκε με τέτοιο τρόπο ώστε να ακολουθούν το ίδιο πλαίσιο ανάγνωσης της αλληλουχίας mCherry του φορέα κλωνοποίησης (διατηρώντας παράλληλα και το δικό τους πλαίσιο ανάγνωσης συμπεριλαμβανομένου του κωδικονίου τερματισμού). Έτσι, παράγονται σε σύντηξη με την πρωτεΐνη mCherry στο αμινοτελικό τους άκρο. Οι κωδικές περιοχές των cDNAs των τριών γονιδίων ενισχύθηκαν με PCR χρησιμοποιώντας ως υπόστρωμα DNA από cDNA κλώνους πλήρους μήκους (βλ. Υλικά & Μέθοδοι). Τα προϊόντα της PCR μεγέθους 921bp, 1314bp και 1968bp για τα γονίδια *SAP30BP*, *MFAP* και *THOC1*, αντίστοιχα φαίνονται στην Εικ. 56. Πραγματοποιήθηκαν πολλαπλές PCR και η τελική ποσότητα που εκτιμήθηκε με φωτομέτρηση ήταν αντιστοίχως για τα cDNAs των γονιδίων *MFAP*, *SAP30BP* και *THOC1* 64μg, 62μg και 53μg. Το επόμενο βήμα ήταν ο τεμαχισμός των

προϊόντων των τριών γονιδίων με τα ένζυμα περιορισμού *XhoI* και *BamHI* που είχαν οι εκκινητές στα 5' άκρα τους και οι θέσεις κλωνοποίησης στον φορέα (βλ. Υλικά & Μέθοδοι). Σε κάθε αντίδραση χρησιμοποιήθηκε περίσσεια ενζύμου (5-7units/συνολικά μl αντίδρασης). Ακολούθησε ο καθαρισμός των αντίστοιχων πέψων με κατάλληλο κιτ.



**Εικ. 56: Προϊόντα PCR για τα cDNAs των γονιδίων *MFAP*, *SAP30BP* και *THOC1*.** Στις διαδρομές 1, 2 και 3 φαίνονται αντίστοιχα τα προϊόντα των γονιδίων *MFAP* (1314bp), *SAP30BP* (921bp) και *THOC1* (1968bp). (-): αρνητικοί έλεγχοι των PCR, 4: δείκτης μοριακών μεγεθών (Cat. No. 15615-016, Invitrogen).

Παράλληλα, πραγματοποιήθηκε απομόνωση μεγάλης ποσότητας πλασμιδιακού DNA του φορέα pmCherry-C1 με κατάλληλο κιτ απομόνωσης και διπλή πέψη του πλασμιδιακού DNA του pmCherry-C1 με τα ίδια ένζυμα περιορισμού χρησιμοποιώντας συνολική ποσότητα 300μg και περίσσεια ενζύμου (10 units/συνολικά μl αντίδρασης για κάθε ένζυμο). Μετά τον καθαρισμό ακολούθησε αντίδραση σύνδεσης του γραμμικού πλασμιδιακού φορέα με τα cDNAs των τριών γονιδίων με την T4 λιγάση (1-3 units). Στη συνέχεια έγινε μετασηματισμός δεκτικών βακτηριακών κυττάρων *DH5a* με τις τρεις αντιδράσεις σύνδεσης με αποτέλεσμα την ανάπτυξη ~30-50 αποικιών για το κάθε ανασυνδυασμένο μόριο. Ακολούθησε απομόνωση πλασμιδιακού DNA σε μικρή κλίμακα από ~20-30 αποικίες. Αυτές ελέγχθηκαν με PCR με τους εκκινητές των τριών cDNAs για την ύπαρξη ή όχι του ενθέματος γεγονός που οδήγησε στην ταυτοποίηση τριών θετικών αποικιών, μία για κάθε ανασυνδυασμένο πλασμίδιο (Εικ. 57). Συμπληρωματικά, πραγματοποιήθηκαν δοκιμαστικές πέψεις στο πλασμιδιακό DNA των αποικιών αυτών με τα ένζυμα περιορισμού των θέσεων κλωνοποίησης (*XhoI* και *BamHI*). Τα αναμενόμενα μεγέθη των προϊόντων πέψης για τον pmCherry-C1, το *SAP30BP*, το *MFAP* και το *THOC1* ήταν 4722bp, 921bp, 1314bp, 1968bp, αντίστοιχα (Εικ. 57). Επιπλέον, πραγματοποιήθηκε απομόνωση μεγάλης ποσότητας πλασμιδιακού DNA του διαγονιδίου pEGFP-N1/*FRA10AC1* με κατάλληλο κιτ απομόνωσης για τη μετέπειτα χρήση του στα πειράματα που ακολούθησαν.



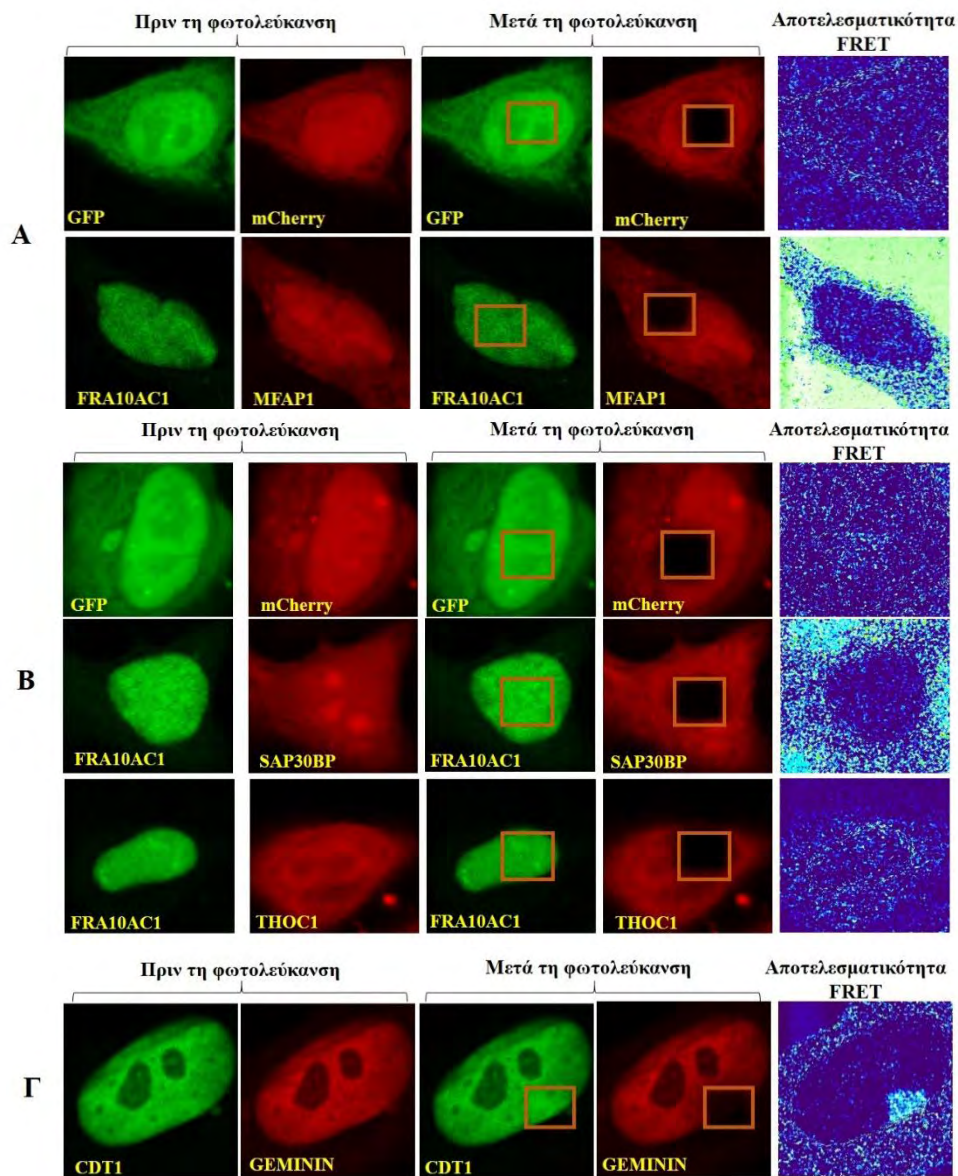
**Εικ. 57:** Έλεγχος ανασυνδυασμένων πλασμιδίων με τεμαχισμό με ένζυμα περιορισμού και PCR. 1 & 5: pmCherry/THOC1 (πέψη και προϊόν PCR, 1968bp), 2 & 7: pmCherry/MFAP1 (πέψη και προϊόν PCR, 1314bp), 3 & 9: pmCherry/SAP30BP (πέψη και προϊόν PCR, 921bp), 6, 8, 10: αρνητικοί έλεγχοι PCR για τα pmCherry/THOC1, pmCherry/MFAP1 και pmCherry/SAP30BP, αντίστοιχα, 4: δείκτης μοριακών μεγεθών (Cat. No. 305-105, GeneON).

### 3.3.2. Έλεγχος πρωτεϊνικών αλληλεπιδράσεων με τη μέθοδο Μεταφοράς Ενέργειας Φθορισμού μέσω Συντονισμού (FRET)

Ο έλεγχος των αλληλεπιδράσεων των ζευγών πρωτεϊνών FRA10AC1-THOC1, FRA10AC1-MFAP και FRA10AC1-SAP30BP έγινε από το συνεργαζόμενο εργαστήριο Γενικής Βιολογίας, Ιατρική Σχολή, Πανεπιστήμιο Πατρών (Επικεφαλής: Καθ. Ν. Μοσχονάς). Το πρώτο βήμα ήταν η συν-διαμόλυνση των ζευγών διαγονιδίων (pEGFP-N1-FRA10AC1 με τα pmCherry/THOC1, pmCherry/MFAP και pmCherry/SAP30BP) χρησιμοποιώντας ποσότητα DNA των δυο διαγονιδίων που εξετάστηκαν κάθε φορά σε στοιχειομετρική αναλογία 1:1 καθώς έχει δειχθεί ότι η τήρηση αυτής της αναλογίας αυξάνει την αποτελεσματικότητα της μεθόδου FRET (Chen et al. 2006). Στη συνέχεια, έγινε η μονιμοποίηση των κυττάρων στις καλυπτρίδες που είχαν αναπτυχθεί, η τοποθέτηση αυτών σε αντικειμενοφόρους πλάκες, η παρατήρηση σε συνεστιακό μικροσκόπιο σάρωσης με ακτίνες Laser (Confocal Laser Scanning Microscope) και ο έλεγχος των πρωτεϊνικών αλληλεπιδράσεων μέσω της μεθόδου FRET στη Μονάδα Λειτουργικής Μικροσκοπίας & Πληροφορικής Επιστημών Ζωής (Πανεπιστήμιο Πατρών).

Η μέθοδος που χρησιμοποιήθηκε για τη μέτρηση της μεταφοράς ενέργειας φθορισμού είναι αυτή της φωτολεύκανσης του μορίου-δέκτη (acceptor photobleaching) (Wouters et al. 1998, Kenworthy & Edidin 1998). Φωτολεύκανση ορίζεται ως η μη αναστρέψιμη αποσύνθεση των φθορίζοντων μορίων στη διεγερμένη κατάσταση εξαιτίας της αλληλεπίδρασής τους με μοριακό οξυγόνο πριν από την εκπομπή φθορισμού. Η φωτολεύκανση προκαλείται όταν εφαρμόζεται υψηλή ένταση του λέιζερ σε μια μικρή περιοχή του φθορίζοντος ζωντανού κυττάρου. Επειδή τα φθορίζοντα μόρια καταστρέφονται μέσω της φωτολεύκανσης, η διαδικασία χρησιμοποιείται μόνο μία φορά στα κύτταρα που εξετάζονται και δεν είναι κατάλληλη για δυναμικές αλληλεπιδράσεις (Piston & Kremers 2007). Κατά την ανάλυση FRET, πραγματοποιείται απόσβεση του φθορισμού του πρώτου φθοριοχρώματος (μόριο-δότης) διότι δεν μεταπίπτει από τη διεγερμένη κατάσταση στη θεμελιώδη εκπέμποντας ενέργεια, αλλά η ενέργεια αυτή μεταφέρεται στο μόριο-δέκτη όταν τα δυο

φθοριοχρώματα βρίσκονται σε κοντινή απόσταση μεταξύ τους. Έτσι, το δεύτερο φθοριόχρωμα (μόριο-δέκτης), το οποίο λαμβάνει την ενέργεια, εκπέμπει φθορισμό. Όταν πραγματοποιείται η φωτολεύκανση του μορίου-δέκτη (των ανασυνδυασμένων πλασμιδίων με Cherry-C1) δεν μπορεί να γίνει μεταφορά της ενέργειας του μορίου-δότη (pEGFP-N1/*FRA10AC1*) στο μόριο-δέκτη. Έτσι, η απόσβεση του φθορισμού του μορίου-δότη απελευθερώνεται με αποτέλεσμα να αυξάνεται η ενέργεια φθορισμού που εκπέμπει. Η αποτελεσματικότητα της ανάλυσης FRET υπολογίζεται ως η διαφορά της έντασης φθορισμού του μορίου-δότη παρουσία του μορίου δέκτη πριν τη φωτολεύκανση από την ένταση φθορισμού του μορίου-δότη μετά τη φωτολεύκανση του μορίου-δέκτη, προς την ένταση φθορισμού του μορίου-δότη μετά τη φωτολεύκανση (Albertazzi et al. 2009, Karpova et al. 2003). Πραγματοποιήθηκαν δυο αναλύσεις FRET μέσω της εφαρμογής της μεθόδου της φωτολεύκανσης στα κύτταρα που είχε γίνει συν-διαμόλυνση των ζευγών διαγονιδίων. Στο ένα πείραμα εξετάστηκε το ζεύγος pEGFP-N1/*FRA10AC1* και pmCherry/*MFAP* (Εικ. 58 A & Πιν. 18) και στο δεύτερο τα ζεύγη pEGFP-N1/*FRA10AC1*, pmCherry/*THOC1* και pEGFP-N1/*FRA10AC1*, pmCherry/*SAP30BP* (Εικ. 58 B & Πιν. 18). Και στα δυο πειράματα χρησιμοποιήθηκε ως αρνητικός έλεγχος το ζεύγος των πλασμιδιακών φορέων pEGFP-N1/pmCherry-C1 (Εικ. 58). Ως θετικός έλεγχος χρησιμοποιήθηκε το ζεύγος πρωτεϊνών GFP/CDT1-dsRed/GEMININ (Εικ. 58 Γ & Πιν. 18). Ο αριθμός των κυττάρων που εξετάστηκαν κάθε φορά σε κάθε ζεύγος πρωτεϊνών ήταν μεγαλύτερος του δέκα (Πιν. 18), ο οποίος είναι προτεινόμενος για την εξαγωγή των συμπερασμάτων σχετικά με την αλληλεπίδραση των πρωτεϊνών (Karpova et al. 2003). Στη συνέχεια μετρήθηκε ο μέσος όρος της αποτελεσματικότητας της ανάλυσης FRET από όλα τα κύτταρα που εξετάστηκαν με τον τρόπο που αναφέρθηκε παραπάνω (Πιν. 18). Το λογισμικό που χρησιμοποιήθηκε ήταν το Leica TCS SP5 LAS AF Version 1.7.0. Για την ύπαρξη αλληλεπίδρασης θα πρέπει η αποτελεσματικότητα της ανάλυσης FRET να είναι μεγαλύτερη από 10% διότι μικρότερες τιμές μπορεί να προκύπτουν λόγω ψευδώς θετικού σήματος που ανακτάται από την ένταση του φθορισμού εκτός του κυττάρου (background) (Karpova et al. 2003).



**Εικ. 61: Μέτρηση αποτελεσματικότητας της ανάλυσης FRET σε συνεστιακό μικροσκόπιο σάρωσης με ακτίνες Laser με τη μέθοδο της φωτολεύκανσης. Φαίνονται τα ζεύγη πρωτεϊνών που εξετάστηκαν, FRA10AC1-MFAP1 (A), FRA10AC1-SAP30BP και FRA10AC1-THOC1 (B) πριν και μετά τη φωτολεύκανση καθώς και η αποτελεσματικότητα της ανάλυσης FRET σε ένα κύτταρο που αποτέλεσε ενδεικτική εικόνα του αποτελέσματος των πειραμάτων 1 (A) και 2 (B) Σε κάθε πείραμα χρησιμοποιήθηκε ως αρνητικός έλεγχος το ζεύγος των πλασμιδιακών φορέων pEGFP-N1/pmCherry-C1 (A,B) και ως θετικός έλεγχος το ζεύγος CTD1-GEMININ (Γ).**

**Πιν. 18: Αποτελεσματικότητα της ανάλυσης FRET.**

Ζεύγη πρωτεϊνών	Αριθμός κυττάρων που εφαρμόστηκε η φωτολεύκανση	Μέσος όρος (%) αποτελεσματικότητας της ανάλυσης FRET
GFP-mCherry (αρνητικός έλεγχος 1)	10	0,1%
GFP-mCherry (αρνητικός έλεγχος 2)	15	0,37%
FRA10AC1-MFAP1	10	2,8%
FRA10AC1-SAP30BP	15	3,32%
FRA10AC1-THOC1	15	1,5%
GFP/CDT1-dsRed/GEMININ	10	13,17%

Συμπερασματικά, η χαμηλή αποτελεσματικότητα της ανάλυσης αυτής υποδεικνύει ότι τα ζεύγη FRA10AC1-MFAP1, FRA10AC1-SAP30BP και FRA10AC1-THOC1 δεν αλληλεπιδρούν τη δεδομένη χρονική στιγμή που εξετάστηκαν. Ωστόσο, το αποτέλεσμα αυτό δε συνεπάγεται αποκλειστικά ότι τα δυο μόρια-πρωτεΐνες που εξετάστηκαν δεν αλληλεπιδρούν μέσα στο κύτταρο. Αυτό μπορεί να συμβαίνει διότι το μεγάλο μέγεθος των επιτόπων μπορεί να αλλάζει την δομή των χιμαιρικών πρωτεϊνών με τέτοιο τρόπο ώστε τα δίπολα του μορίου-δότη να μην βρίσκονται στον σωστό προσανατολισμό, δηλαδή να μην είναι παράλληλα, με τα δίπολα του μορίου-δέκτη (βλ. Ενότητα 4.1.1.).

Αν και ο έλεγχος στοχευμένων άμεσων αλληλεπιδράσεων του ανακατασκευασμένου δικτύου της FRA10AC1, μέσω της ανάλυσης FRET, δεν είχε θετικά αποτελέσματα, η συνεισφορά της στις διαδικασίες επεξεργασίας του mRNA, υποδεικνύεται τόσο στα πλαίσια του δικτύου των άμεσων «γειτόνων» της όσο και στο ευρύτερο δίκτυο του σωματίου συναρμογής και των υποσυμπλόκων του. Η τοπολογία της και η διερεύνηση του λειτουργικού ρόλου των πρωτεϊνών της ευρύτερης «γειτονιάς» της στο δίκτυο του σωματίου συναρμογής και των υποσυμπλόκων του υποδεικνύουν ότι η FRA10AC1 μπορεί να συμμετέχει στα τελευταία στάδια συναρμολόγησης του σωματίου συναρμογής ή να αποτελεί έναν συνδεδετικό κρίκο της συναρμογής με άλλες διαδικασίες επεξεργασίας του mRNA. Ένας από τους πολλούς εναλλακτικούς τρόπους διερεύνησης της βιολογικής λειτουργίας μίας πρωτεΐνης είναι η εξέταση των αλλαγών που μπορεί να επιφέρει η τροποποίηση της έκφρασής της στην κυτταρική φυσιολογία, σε πολλαπλά μοριακά επίπεδα, από τα μετάγραφα στις πρωτεΐνες και στους μεταβολίτες, μέσω της χρήσης ομικών τεχνολογιών (βλ. Ενότητα 1.6.). Λαμβάνοντας υπ' όψιν τις ενδείξεις συμμετοχής της FRA10AC1 στην επεξεργασία του mRNA, πραγματοποιήθηκε η δημιουργία ενός κυτταρικού μοντέλου με τροποποιημένη την έκφρασή της ώστε να ελεγχθούν οι επιδράσεις που φέρει αυτή στα μόρια του δικτύου αλληλεπιδράσεων του σωματίου συναρμογής. Οι αλλαγές αυτές ελέγχθηκαν σε τρία επίπεδα, σε μεταγραφικό, πρωτεϊνικό και μεταβολικό επίπεδο μέσω της εφαρμογής ομικών τεχνολογιών.

### **3.4. Συνδυαστικές ομικές αναλύσεις της τροποποιημένης έκφρασης του FRA10AC1 σε κυτταρικό μοντέλο**

Για τη διερεύνηση των μοριακών μηχανισμών στους οποίους συμμετέχει η πρωτεΐνη FRA10AC1 και κατ' επέκταση για την κατανόηση της λειτουργίας της, επιλέχθηκε να

εφαρμοστούν ομικές αναλύσεις μεγάλης κλίμακας (high-throughput) οι οποίες ερευνούν τη συμπεριφορά σύνθετων βιολογικών συστημάτων σε πολλαπλά επίπεδα. Τα διαφορετικά επίπεδα αφορούν τα μετάγραφα (transcriptomics), τις πρωτεΐνες (proteomics) και τις μεταβολικές διεργασίες και τα μεταβολικά μονοπάτια (metabolomics). Συνδυάζοντας τα αποτελέσματα των αναλύσεων αυτών είναι εφικτό να γίνει η καθολική εξέταση των χαρακτηριστικών του υπο-μελέτη συστήματος. Στην παρούσα εργασία το σύστημα που δημιουργήθηκε ήταν ένα κυτταρικό μοντέλο *HeLa* με τροποποιημένη την έκφραση του γονιδίου *FRA10AC1*. Η ανθρώπινη κυτταρική σειρά *HeLa* έχει χρησιμοποιηθεί σε όλες τις μελέτες απομόνωσης των συστατικών του σωματίου συναρμογής και γι' αυτό επιλέχθηκε για τις ομικές αναλύσεις. Επιπλέον, τα κύτταρα *HeLa* αποτελούν ένα απλό και εύκολο στο χειρισμό πρότυπο σύστημα για τη μελέτη της κυτταρικής φυσιολογίας (Landry et al. 2013) και έχουν χρησιμοποιηθεί, τα τελευταία χρόνια, ως σύστημα-μοντέλο για ομικές προσεγγίσεις (Chaudhry et al. 2002, Whitfield 2002, Nagaraj et al. 2011). Γι' αυτούς τους λόγους επιλέχθηκε να δημιουργηθεί το κυτταρικό μοντέλο της εργασίας αυτής στη συγκεκριμένη κυτταρική σειρά. Πριν την δημιουργία του υπο-μελέτη κυτταρικού μοντέλου προηγήθηκε η επιλογή του κατάλληλου πρωτοκόλλου για την επεξεργασία και συλλογή των κυττάρων που προορίζονται για ομικές αναλύσεις το οποίο έχει λιγότερες επιδράσεις στην κυτταρική φυσιολογία. Ακολούθησε η δημιουργία του κυτταρικού μοντέλου με τροποποιημένη την έκφραση του γονιδίου *FRA10AC1* και στη συνέχεια εξετάστηκαν οι επιδράσεις της αλλαγής αυτής στα βιολογικά χαρακτηριστικά του κυτταρικού μοντέλου σε μεταγραφικό, πρωτεϊνικό και μεταβολικό επίπεδο μέσω της χρήσης ομικών τεχνολογιών.

#### 3.4.1. Επεξεργασία και συλλογή κυττάρων *HeLa* για ομικές αναλύσεις

Οι υψηλής απόδοσης ομικές αναλύσεις απαιτούν υψηλής ακρίβειας ποσοτική διερεύνηση της φυσιολογίας των κυττάρων που χρησιμοποιούνται ώστε να μπορούν να εξαχθούν σωστά βιολογικά συμπεράσματα. Μια σημαντική παράμετρος στην ανάλυση της κυτταρικής λειτουργίας είναι η γρήγορη αδρανοποίηση των ενζυμικών διεργασιών έτσι ώστε οι μετρήσεις που θα προκύψουν να αντικατοπτρίζουν την πραγματική κυτταρική φυσιολογία του υπό μελέτη συστήματος. Αυτό μπορεί να επιτευχθεί μέσω κατάλληλου πειραματικού σχεδιασμού ώστε οι απαιτούμενες, πριν την ανάλυση, διαδικασίες χειρισμού και συλλογής των κυττάρων να έχουν ελάχιστη επίδραση στη φυσιολογία τους.

Σε αυτό το πλαίσιο, πραγματοποιήθηκε η δοκιμή διαφορετικών πρωτοκόλλων τα οποία αφορούν την επεξεργασία και συλλογή των κυττάρων που προορίζονται για υψηλής απόδοσης βιομοριακές αναλύσεις. Το κριτήριο επιλογής ενός συγκεκριμένου πρωτοκόλλου ήταν ο έλεγχος των επιδράσεων του στην κυτταρική φυσιολογία με την ανάλυση του μεταβολικού προτύπου των εκάστοτε κυττάρων από το συνεργαζόμενο εργαστήριο Μεταβολικής Μηχανικής και Βιολογίας Συστημάτων (Επικεφαλής: Δρ. Μ. Κλάπα), στο Ινστιτούτο Χημικών Μηχανικών Επιστημών (Ίδρυμα Έρευνας και Τεχνολογίας, FORTH/ICE-HT, Πάτρα). Η επιλογή αυτή έγινε διότι θεωρείται ότι η μεταβολομική ανάλυση αποτυπώνει τις βιοχημικές επιδράσεις, στο υπο-μελέτη σύστημα, των «σημάτων» που προκύπτουν από τα άλλα λειτουργικά επίπεδα

(μεταγράφομα και πρωτεϊνώμα) (Nielsen & Oliver 2005). Ένα μεταβολικό δίκτυο περιλαμβάνει πολλές διαφορετικές βιοχημικές αντιδράσεις οι οποίες συνδέονται μεταξύ τους. Το 15% των μεταβολιτών συμμετέχουν σε περισσότερες από 10 βιοχημικές αντιδράσεις και το ~5% σε περισσότερες από 30 (Förster et al. 2003). Επιπλέον, ο αριθμός των διαφορετικών ενζύμων που καταλύουν αυτές τις βιοχημικές αντιδράσεις είναι εξίσου μεγάλος. Όταν λοιπόν συμβαίνουν ακόμα και μικρές διαταραχές σε μεταγραφικό και πρωτεϊνικό επίπεδο, αυτές μπορεί να προκαλέσουν αλλαγές στις συγκεντρώσεις κάποιων ενζύμων. Οι αλλαγές αυτές, αν και μπορεί να μην έχουν μεγάλη επίδραση στις μεταβολικές αντιδράσεις, μπορεί να προκαλέσουν σημαντικές αλλαγές στις συγκεντρώσεις πολλών μεταβολιτών (Goodacre et al. 2004). Επίσης, η ανάλυση του μεταβολικού προτύπου επιλέχθηκε διότι αποτελούσε βελτιστοποιημένη μέθοδο στο συνεργαζόμενο εργαστήριο και παρόλο που απαιτεί αρκετή ανάλυση μετέπειτα, είναι πιο οικονομική σε αναλώσιμα.

Το πρωτόκολλο που τελικά επιλέχθηκε προσαρμόστηκε με τέτοιο τρόπο έτσι ώστε το τελευταίο βήμα της συλλογής των κυττάρων να είναι κατάλληλο για την απομόνωση επαρκούς ποσότητας RNA και πρωτεϊνών που χρησιμοποιήθηκαν μετέπειτα για τη ανάλυση του μεταγραφώματος και του πρωτεϊνώματος, αντίστοιχα.

#### 3.4.1.1. Μεταβολομική ανάλυση κυττάρων HeLa που συλλέχθηκαν με διαφορετικά πρωτόκολλα

Τα πρωτόκολλα που ελέγχθηκαν ήταν τέσσερα και στηρίχτηκαν σε δύο μεθόδους συλλογής. Η πρώτη αποτελεί την πιο συνήθη διαδικασία συλλογής των κυττάρων και περιλαμβάνει μηχανική (με τη χρήση ειδικής ξύστρας) αποκόλληση των κυττάρων από το τρυβλίο καλλιέργειας και πολλαπλά στάδια φυγοκεντρήσεων και επαναιωρήσεων σε ρυθμιστικό διάλυμα μέχρι τη συλλογή μίας καθαρής από το θρεπτικό μέσο πελλέτας κυττάρων (βλ. Υλικά & Μέθοδοι). Η δεύτερη στηρίζεται σε μία διαφορετική μέθοδος συλλογής κυττάρων για μεταβολομική ανάλυση (Teng et. al. 2009) κατά την οποία τα πολλαπλά βήματα των φυγοκεντρήσεων/επαναιωρήσεων αποφεύγονται και η συλλογή των κυττάρων γίνεται με χρήση μεθανόλης (βλ. Υλικά & Μέθοδοι). Οι παραπάνω μέθοδοι τροποποιήθηκαν σε συγκεκριμένα στάδια της διαδικασίας έτσι ώστε να ελαχιστοποιηθούν οι χειρισμοί και οι χρόνοι που ενδέχεται να επηρεάζουν τη φυσιολογία των κυττάρων (βλ. Υλικά & Μέθοδοι). Ένας επιπρόσθετος στόχος ήταν η επιλογή του βέλτιστου πρωτοκόλλου κατά το οποίο θα απομονώνονταν μεγαλύτερος αριθμός ενδογενών μεταβολιτών που θα εμφάνιζαν και υψηλότερη συγκέντρωση. Οι ομοιότητες και οι διαφορές των τεσσάρων πρωτοκόλλων παρουσιάζονται στον Πίνακα 19.



**Πιν. 19: Ομοιότητες και διαφορές των τεσσάρων πρωτοκόλλων συλλογής κυττάρων.**

		Πρωτόκολλο 1	Πρωτόκολλο 2	Πρωτόκολλο 3A	Πρωτόκολλο 3B
<b>Κύτταρα HeLa</b>	<b>Είδος αποκόλλησης</b>	Μηχανική	Μηχανική	Μηχανική	Μηχανική
	<b>Έκπλυση</b>	Στο τρυβλίο καλλιέργειας	Ενδιάμεσα των φυγοκεντρήσεων (στο σωληνάριο των 15ml)	Στο τρυβλίο καλλιέργειας	Στο τρυβλίο καλλιέργειας
	<b>Διάλυμα έκπλυσης</b>	1xPBS	1xPBS ή 0,9% w/v NaCl	1xPBS ή 0,9% NaCl ή DMEM	1xPBS ή 0,9% NaCl
	<b>Χρονική στιγμή αποκόλλησης</b>	Μετά την 3 <sup>η</sup> έκπλυση	Πριν τις εκπλύσεις	Μετά την 3 <sup>η</sup> έκπλυση	Μετά την 3 <sup>η</sup> έκπλυση
	<b>Διάλυμα αποκόλλησης</b>	1xPBS ή 0,9% w/v NaCl	DMEM + FBS + P/S	MeOH	MeOH
	<b>Τρόπος Συλλογής</b>	Φυγοκέντρωση & συλλογή πελλέτας	Φυγοκέντρωση & συλλογή πελλέτας	Χρήση μεθανόλης (MeOH)	Χρήση μεθανόλης (MeOH)
	<b>Διάλυμα συλλογής</b>	1xPBS ή 0,9% w/v NaCl	0,9% w/v NaCl	MeOH	MeOH
<b>Διάρκεια πρωτοκόλλου</b>	30min	45min	10min	1min	

Για κάθε πρωτόκολλο ελέγχθηκαν διαφορετικός αριθμός κυττάρων (αριθμός τρυβλίων), διαφορετικό στάδιο ανακαλλιέργειας και διαφορετικά διαλύματα έκπλυσης και συλλογής (Πιν. 20, 21). Το στάδιο ανακαλλιέργειας των κυττάρων αφορά τον αριθμό των θρυψινοποιήσεων που έγιναν από τη στιγμή της απόψυξής τους μέχρι τη στιγμή της επίστρωσής τους στα τρυβλία. Τα δείγματα δηλαδή, προήλθαν από την ίδια αρχική καλλιέργεια κυττάρων τα οποία ανακαλλιεργούνταν ανά 2-3 ημέρες (Πιν. 20, 21). Οι εκπλύσεις ήταν τρεις και πραγματοποιήθηκαν είτε με το ίδιο διάλυμα είτε με συνδυασμό διαλυμάτων (Πιν. 20, 21). Η συλλογή των κυττάρων έγινε 48 ώρες μετά την επίστρωσή τους στα τρυβλία. Τα κύτταρα συλλέχθηκαν είτε σε μορφή πελλέτας είτε με πιπέτα μετά την προσθήκη μεθανόλης (Πιν. 20, 21) και αποθηκεύθηκαν σε ξηρό πάγο στους -80°C μέχρι να γίνει η αποστολή τους στο συνεργαζόμενο εργαστήριο για μεταβολομική ανάλυση. Στο 1<sup>ο</sup> πρωτόκολλο η συλλογή των κυττάρων είχε πραγματοποιηθεί και στις 24 και στις 72h. Ωστόσο, οι ώρες αυτές δεν επιλέχθηκαν διότι ο αριθμός των μεταβολιτών ήταν μικρός στις 24h, και στις 72h, εφόσον δεν εμφανίστηκε σημαντική διαφορά στον αριθμό των μεταβολιτών με τις 48h, η παράταση του χρόνου θεωρήθηκε ότι θα «ταλαιπωρούσε» τα κύτταρα.

**Πιν. 20: Πειραματικές συνθήκες των πρωτοκόλλων 1 και 2.**

Πρωτόκολλο 1		
Αριθμός ανακαλλιεργειών	Διάλυμα πλύσης: Διάλυμα συλλογής	
	<i>PBS:NaCl</i>	<i>PBS:PBS</i>
Δείγματα κυττάρων		
2	pool 2 τρυβλίων	-
4	pool 2 τρυβλίων	pool 2 τρυβλίων
6	pool 2 τρυβλίων	pool 2 τρυβλίων
Πρωτόκολλο 2		
Αριθμός ανακαλλιεργειών	Διάλυμα πλύσης: Διάλυμα συλλογής	
	<i>PBS+NaCl:NaCl</i>	<i>NaCl:NaCl</i>
Δείγματα κυττάρων		
2	pool 4 τρυβλίων	pool 4 τρυβλίων
4	pool 2 & 5 τρυβλίων	pool 2 & 5 τρυβλίων
6	pool 2 & 5 τρυβλίων	pool 2 & 5 τρυβλίων

Τα βέλη υποδεικνύουν την πορεία της αρχικής καλλιέργειας και σε ποιο στάδιο ανακαλλιέργειας έγινε η εκάστοτε επανάληψη του πειράματος. Pool: κύτταρα τα οποία προέρχονται από την ένωση διαφορετικού αριθμού τρυβλίων. Σημείωση: όταν ο αριθμός ανακαλλιέργειας ορίζεται με τον αριθμό 2 σημαίνει ότι έχουν μεσολαβήσει δυο θρυψινοποιήσεις, μία αφού προσκολληθούν ώστε να μοιραστούν σε περισσότερες φιάσκες και μία κατά την επιστροφή τους στα τρυβλία. Αντίστοιχα ορίζονται και τα άλλα στάδια ανακαλλιέργειας.

**Πιν. 21: Πειραματικές συνθήκες των πρωτοκόλλων 3A και 3B**

Πρωτόκολλο 3					
	Αριθμός ανακαλλιεργειών	Διάλυμα πλύσης: Διάλυμα συλλογής			
		<i>PBS+NaCl: MeOH</i>	<i>PBS: MeOH</i>	<i>NaCl: MeOH</i>	<i>DMEM*: MeOH</i>
Δείγματα κυττάρων					
<b>Πρωτόκολλο 3A (10min)</b>	5	2 τεχνικές επαναλήψεις	2 τεχνικές επαναλήψεις	1 τεχνική επανάληψη	2 τεχνικές επαναλήψεις
	8	2 τεχνικές επαναλήψεις	-	-	-
<b>Πρωτόκολλο 3B (1min)</b>	8	2 τεχνικές επαναλήψεις	2 τεχνικές επαναλήψεις	-	-
	2	1 τεχνική επανάληψη	2 τεχνικές επαναλήψεις	-	-

Το κόκκινο βέλος υποδεικνύει από ποιο στάδιο ανακαλλιέργειας κυττάρων ξεκίνησαν οι επαναλήψεις του πειράματος και με ποιο συνεχίστηκε. Το πορτοκαλί βέλος δείχνει ότι τα κύτταρα της 8<sup>ης</sup> ανακαλλιέργειας πάγωσαν και έπειτα από 4 ημέρες ξεπάγωσαν, ανακαλλιεργήθηκαν 2 φορές και χρησιμοποιήθηκαν γι' αυτή την επανάληψη. DMEM\* = χωρίς εμβρυικό ορό βοοειδούς και αντιβιοτικά. MeOH: μεθανόλη

Στις υποενότητες που ακολουθούν περιγράφονται ποιοι μεταβολίτες συμπεριλαμβάνονται στο θρεπτικό μέσο ανάπτυξης της καλλιέργειας των κυττάρων, στα διαφορετικά διαλύματα έκπλυσης και στα διαφορετικά πρωτόκολλα συλλογής των κυττάρων. Επίσης, αναλύεται η μεταβολή του πλήθους και της συγκέντρωσης των μεταβολιτών κατά τη σύγκριση των διαφορετικών τρόπων συλλογής για την επιλογή εκείνου του πρωτοκόλλου επεξεργασίας και συλλογής κυττάρων το οποίο θα έχει λιγότερες επιδράσεις στην κυτταρική φυσιολογία και θα προσομοιάζει με τον καλύτερο δυνατό τρόπο την φύση των κυττάρων *HeLa*.

➤ Μεταβολομική ανάλυση των διαλυμάτων έκπλυσης και των θρεπτικών μέσων καλλιέργειας

Σε πρώτο στάδιο, το τελικό διάλυμα έκπλυσης από όλα τα δείγματα των κυττάρων συλλέχθηκε για να ελεγχθεί η αποτελεσματικότητα της έκπλυσης των κυττάρων από το θρεπτικό μέσο καλλιέργειας. Στη συνέχεια, ακολούθησε η μεταβολομική ανάλυση των δειγμάτων του κάθε πρωτοκόλλου με GC-MS από το συνεργαζόμενο εργαστήριο Μεταβολικής Μηχανικής και Βιολογίας Συστημάτων (Επικεφαλής: Δρ. Μ. Κλάπα). Τα αποτελέσματα αυτής της ανάλυσης έδειξαν ότι στα τελικά διαλύματα έκπλυσης δεν ανιχνεύονται μεταβολίτες ούτε και αυτοί που υπάρχουν σε υψηλή συγκέντρωση στο θρεπτικό μέσο ανάπτυξης όπως για παράδειγμα αμινοξέα, γλυκόζη και πυροσταφυλικό οξύ. Έτσι, αποδείχθηκε ότι τα κύτταρα δεν έφεραν εξωγενώς συστατικά του θρεπτικού μέσου ανάπτυξης γεγονός που σημαίνει ότι η έκπλυσή τους με όλα τα πρωτόκολλα γινόταν αποτελεσματικά και ότι η μέτρηση των ενδογενών μεταβολιτών δεν θα επηρεαζόταν.

Επίσης, πραγματοποιήθηκε η ανάλυση του θρεπτικού μέσου DMEM καθώς και του πλήρους θρεπτικού μέσου ανάπτυξης το οποίο επιπλέον περιέχει εμβρυϊκό ορό βοοειδούς και τα αντιβιοτικά πενικιλίνη και στρεπτομυκίνη. Η ανάλυση αυτή έδειξε ότι 42 μεταβολίτες είναι κοινί μεταξύ του DMEM και του πλήρους θρεπτικού υλικού, ενώ 26 μεταβολίτες εντοπίστηκαν αποκλειστικά στο πλήρες θρεπτικό μέσο ανάπτυξης γεγονός που σημαίνει ότι αυτοί αποτελούν συστατικά του εμβρυϊκού ορού βοοειδούς. Όσον αφορά τους πρώτους, οι 17 μόνο από αυτούς αναγράφονται στη σύσταση του DMEM και οι υπόλοιποι είναι το βενζοϊκό οξύ, αμινοξέα (4), σάκχαρα (4), λιπαρά οξέα (4), μεταβολίτες άγνωστης ταυτότητας που ανήκουν στην κατηγορία των σακχάρων και των πυρανοζών καθώς και μεταβολίτες άγνωστης ταυτότητας και χημικής κατηγορίας. Όσον αφορά τους μεταβολίτες του εμβρυϊκού ορού βοοειδούς, μεγαλύτερη συγκέντρωση παρουσίασαν το γαλακτικό οξύ, η ουρία, η σορβιτόλη, η φρουκτόζη, η ερυθριτόλη και δυο άγνωστοι μεταβολίτες που ανήκουν στη χημική κατηγορία των πυρανοζών. Συγκρίνοντας τις αναλογίες των μεταβολιτών στο DMEM και στο πλήρες θρεπτικό μέσο παρατηρήθηκε ότι είναι ίδιες για τους περισσότερους μεταβολίτες εκτός από τη γλουταμίνη, το φωσφορικό οξύ και τη μυοϊνοσιτόλη. Τα δυο τελευταία βρέθηκαν σε αυξημένη συγκέντρωση στο πλήρες θρεπτικό μέσο που σημαίνει ότι τροφοδοτούνται σε ένα ποσοστό από τον ορό. Η συγκέντρωση της γλουταμίνης ήταν μικρότερη στο πλήρες θρεπτικό μέσο το οποίο εξηγείται από το γεγονός ότι στα περισσότερα θρεπτικά μέσα η γλουταμίνη είναι πολύ ασταθής σε ουδέτερο pH και διασπάται σε αμμώνιο και πυρογλουταμικό οξύ (πιθανώς μέσω της επίδρασης κάποιων ενζύμων του ορού) με αποτέλεσμα να αυξάνεται η συγκέντρωση του γλουταμικού οξέος. Παρ' όλα αυτά, το άθροισμα γλουταμίνης και γλουταμινικού οξέος παρέμενε σταθερό μεταξύ του DMEM και του πλήρους θρεπτικού μέσου ανάπτυξης.

➤ Σύγκριση διαλυμάτων έκπλυσης

Στη συνέχεια συγκρίθηκαν τα αποτελέσματα της επεξεργασίας των κυττάρων με το διάλυμα φωσφορικών ιόντων (PBS), το διάλυμα φυσιολογικού ορού (NaCl) και το θρεπτικό μέσο DMEM. Το ρυθμιστικό διάλυμα φωσφορικών ιόντων παρουσιάζει υψηλή συγκέντρωση φωσφορικού οξέος γεγονός που δικαιολογεί την αύξηση στις

μετρήσεις του ενδογενούς φωσφορικού οξέος επειδή πιθανώς εισάγεται στα κύτταρα με διάχυση. Γι' αυτό το λόγο χρησιμοποιήθηκε το διάλυμα του φυσιολογικού ορού ως εναλλακτική επιλογή. Επίσης, το DMEM χρησιμοποιήθηκε ώστε το περιβάλλον που εκτίθενται τα κύτταρα κατά τη διάρκεια της έκπλυσης να προσομοιάζει κατά πολύ το περιβάλλον ανάπτυξής τους και να μην διαταράσσει την κυτταρική φυσιολογία.

Στα δείγματα όλων των επαναλήψεων που είχαν εκπλυθεί με PBS ή NaCl ταυτοποιήθηκε μεγαλύτερο πλήθος μεταβολιτών σε σχέση με τα κύτταρα που είχαν εκπλυθεί με DMEM, γεγονός που δεν είναι αναμενόμενο καθώς το DMEM χρησιμοποιήθηκε διότι θεωρητικά είναι πιο πλούσιο και προσομοιάζει το περιβάλλον ανάπτυξης των κυττάρων. Στα κύτταρα των οποίων η έκπλυση έγινε με το διάλυμα PBS ή NaCl παρατηρήθηκε αυξημένη συγκέντρωση των μεταβολιτών, γαλακτικό οξύ, αλανίνη, μηλικό οξύ, φουμαρικό οξύ, και μειωμένη συγκέντρωση της φρουκτόζης. Τα αποτελέσματα αυτά υποδεικνύουν αύξηση της αερόβιας γλυκόλυσης και αυξημένη ανάγκη αποκατάστασης των ενδιάμεσων μορίων της γλυκόλυσης μέσω της διάσπασης της φρουκτόζης και μέσω ενεργοποίησης της γλουταμινόλυσης και των αντιδράσεων που έπονται αυτής (βλ. Ενότητα 4.4.1.).

Η μεταβολομική ανάλυση έδειξε ότι τα κύτταρα που είχαν εκπλυθεί με φυσιολογικό ορό (NaCl), έστω και σε ένα βήμα έκπλυσης, παρουσίασαν αύξηση στα επίπεδα δύο χαρακτηριστικών μεταβολιτών, την 6-φωσφορική γλυκόζη και την 6-φωσφορική φρουκτόζη σε σύγκριση με αυτά που είχαν εκπλυθεί με PBS. Η αύξηση αυτών των δυο μεταβολιτών όταν συνδυάζεται και με μειωμένα επίπεδα γλυκόζης υποδεικνύει ότι τα κύτταρα επάγουν άλλους μηχανισμούς παραγωγής των ενδιάμεσων της γλυκόλυσης για την τροφοδότηση της αερόβιας γλυκόλυσης, σε ανταπόκριση στο στρες. Μία εναλλακτική πηγή παραγωγής των ενδιάμεσων της γλυκόλυσης είναι η αποδόμηση μορίων-αποθηκών ενέργειας μέσω της διαδικασίας της γλυκογονόλυσης (βλ. Ενότητα 4.4.1.).

#### ➤ Μεταβολομική ανάλυση των τεσσάρων πρωτοκόλλων

Από αυτήν την ανάλυση ανακτήθηκαν 75 ολόκληρα χρωματογραφήματα για το πρώτο και το δεύτερο πρωτόκολλο και 70 για το πρωτόκολλο 3A και 3B. Η κανονικοποίηση των χρωματογραφημάτων έγινε ως προς δυο παραμέτρους. Το 1<sup>ο</sup> είδος κανονικοποίησης έγινε ως προς τις πειραματικές αποκλίσεις που εισάγονται λόγω της παραγωγής των μεταβολιτών. Η παραγωγή είναι η προσθήκη χημικών παραγόντων (π.χ. υδροχλωρική μεθοξυ-αμίνη) οι οποίοι αντιδρούν με τις χημικές ομάδες των μεταβολιτών (π.χ. κετονομάδα) μετατρέποντας τους μεταβολίτες σε πτητικά και θερμικά σταθερά μόρια. Η δεύτερη έγινε ως προς τα εσωτερικά πρότυπα μεταβολιτών που χρησιμοποιήθηκαν, την ριβιτόλη και την πλήρως σημασμένη στον άνθρακα 13 γλυκόζη ( $[U^{13}\text{-C}]$  γλυκόζη). Η ριβιτόλη εισάγεται στο μίγμα των μεταβολιτών κατά το στάδιο της απομόνωσής τους και θα πρέπει να είναι ανάλογη της ποσότητας της αρχικής μάζας του δείγματος. Για να ελεγχθεί η επαναληψιμότητα των μετρήσεων χρησιμοποιήθηκε η  $[U^{13}\text{-C}]$  γλυκόζη. Η γλυκόζη ως σάκχαρο, με κετονομάδα στη χημική της δομή, την καθιστά ικανή να δημιουργήσει δύο γεωμετρικά ισομερή κατά τη διαδικασία της παραγωγής (μετά την αντίδραση με την υδροχλωρική μεθοξυ-αμίνη), ο λόγος των οποίων πρέπει να παραμένει σταθερός και κοντά στο 5. Η ποσότητα της  $[U^{13}\text{-C}]$  γλυκόζης διαιρείται με το εμβαδόν κάθε κορυφής

στο εκάστοτε μεταβολικό πρότυπο, για να προκύψει το σχετικό εμβαδό κορυφής (Relative Peak Area). Η κανονικοποίηση αυτή γίνεται για να απομακρυνθούν τυχόν διαφοροποιήσεις που μπορεί να προκύψουν στη σχετική ποσότητα των δειγμάτων λόγω σφαλμάτων μεταξύ των διαφορετικών μετρήσεων ώστε τα μεταβολικά πρότυπα να είναι άμεσα συγκρίσιμα μεταξύ τους (Kanani & Klara 2007). Έτσι, τα τελικά κανονικοποιημένα μεταβολικά πρότυπα για το 1<sup>ο</sup> και 2<sup>ο</sup> πρωτόκολλο ήταν 71 και περιείχαν 60 μεταβολίτες, ενώ για το πρωτόκολλο 3A και 3B ήταν 65 και περιείχαν 125 μεταβολίτες. Από τους παραπάνω μεταβολίτες είχαν αφαιρεθεί το φωσφορικό και το πυροφωσφορικό οξύ λόγω της χρήσης του διαλύματος φωσφορικών ιόντων σε ορισμένα δείγματα, αθροίστηκαν η γλουταμίνη και το γλουταμικό οξύ για το λόγο που αναφέρθηκε παραπάνω και η γλυκόζη υπολογίστηκε ως άθροισμα του μεγαλύτερου παραγώγου της και των δύο κορυφών της γλυκοπυρανόζης, δηλαδή, των δυο κυκλικών ισομερών της γλυκόζης.

- Σύγκριση αριθμού και συγκέντρωσης μεταβολιτών των πρωτοκόλλων 1 και 2 με τα πρωτόκολλα 3A και 3B

Οι μεταβολίτες που ταυτοποιήθηκαν στη μεταβολομική ανάλυση των πρωτοκόλλων 1 και 2 ήταν σχεδόν οι μισοί (60) σε σύγκριση με αυτούς των πρωτοκόλλων 3A και 3B (125). Οι συγκεντρώσεις των μεταβολιτών, στα πρώτα δυο πρωτόκολλα εμφανίζονται σε αρκετά μειωμένα επίπεδα σε σχέση με τα άλλα δυο. Σημαντικές διαφορές βρέθηκαν στους μεταβολίτες που υπάρχουν σε αφθονία στο θρεπτικό μέσο ανάπτυξης, όπως είναι η γλυκόζη και τα αμινοξέα λευκίνη, μεθειονίνη, λυσίνη, γλουταμίνη, φαινυλαλανίνη, τυροσίνη, θρεονίνη και τρυπτοφάνη. Αυτοί οι μεταβολίτες ήταν αναμενόμενο να εμφανίζουν υψηλή συγκέντρωση σε όλα τα πρωτόκολλα διότι, αφού υπήρχαν στο θρεπτικό μέσο, θα είχαν πιθανώς εισέλθει με διάχυση ήδη στα κύτταρα. Η μείωση της συγκέντρωσης αυτών των μεταβολιτών στα δύο πρώτα πρωτόκολλα υποδηλώνει μια κατάσταση «λιμοκτονίας» και έρχεται σε αντίθεση με την αναμενόμενη φυσιολογία των κυττάρων που αναπτύσσονται σε ένα πολύ πλούσιο θρεπτικό μέσο ανάπτυξης. Στα πρωτόκολλα αυτά, τα κύτταρα επιδέχονται αρκετά στάδια καταπόνησης λόγω της μηχανικής αποκόλλησης από το τρυβλίο καλλιέργειας και των πολλαπλών βημάτων πλύσεων και φυγοκεντρήσεων. Η μεγάλη διάρκεια αυτών των βημάτων πιθανώς οδηγεί σε διάχυση ενός μέρους των μεταβολιτών αυτών, μέσω της κυτταροπλασματικής μεμβράνης, στα διαλύματα έκπλυσης, λόγω του μικρού μοριακού τους βάρους (βλ. συζήτηση). Ο συνδυασμός της έλλειψης γλυκόζης και των μηχανικών καταπονήσεων πιθανώς οδηγεί τα κύτταρα να επάγουν μηχανισμούς προσαρμογής στις καινούργιες συνθήκες, όπως η αποδόμηση μορίων-αποθηκών ενέργειας ώστε να παράγουν τα απαραίτητα θρεπτικά συστατικά για την ανάπτυξή τους (βλ. συζήτηση). Επιπλέον, η έλλειψη γλυκόζης οδηγεί τα κύτταρα να στραφούν στην οξείδωση των λιπαρών οξέων και στην αποικοδόμηση των αμινοξέων. Τα προϊόντα αυτών των διαδικασιών είναι ενδιάμεσα μόρια του κύκλου του κιτρικού οξέος και της γλυκόλυσης που μπορεί να οδηγήσουν στην παραγωγή κιτρικού οξέος (βλ. Ενότητα 4.4.1.). Το κιτρικό οξύ παρατηρήθηκε σε αυξημένα επίπεδα στο πρωτόκολλο 2 και σε σχετικά αυξημένα επίπεδα στο πρωτόκολλο 1 σε σχέση με αυτά των υπόλοιπων μεταβολιτών του πρωτοκόλλου αυτού. Σε γενικότερο πλαίσιο, το πρωτόκολλο 2 εμφάνισε υψηλότερες συγκεντρώσεις μεταβολιτών σε σχέση με το 1 αλλά χαμηλότερες σε σύγκριση με το

3A και 3B. Η γλυκόζη εμφανίζει ιδιαίτερα χαμηλή συγκέντρωση στο πρωτόκολλο 1, η ανεπάρκεια της οποίας έχει συνδεθεί και με απόπτωση των κυττάρων σε πολύ σύντομο χρονικό διάστημα (Altman & Rathmell 2012). Τα μεταβολικά πρότυπα που συλλέχθηκαν με τα πρωτόκολλα 1 και 2 παρουσίασαν μεγαλύτερη διακύμανση μεταξύ των μετρήσεων των δειγμάτων της εκάστοτε επανάληψης σε σχέση με τα άλλα δυο πρωτόκολλα υποδηλώνοντας μη ικανοποιητικό βαθμό επαναληψιμότητας. Τα παραπάνω αποτελέσματα υποδεικνύουν ότι οι τρόποι επεξεργασίας και συλλογής των δυο πρώτων πρωτοκόλλων δεν είναι ιδιαίτερα αποτελεσματικοί.

➤ Σύγκριση πρωτοκόλλων 3A και 3B

Όπως προαναφέρθηκε, τα πρωτόκολλα 3A και 3B εμφάνισαν το διπλάσιο αριθμό και υψηλότερη συγκέντρωση μεταβολιτών σε σύγκριση με τα άλλα πρωτόκολλα γεγονός που υποδεικνύει ότι έχουν πιο ήπια επίδραση στην κυτταρική φυσιολογία. Τα πλεονεκτήματά τους είναι η μικρή χρονική διάρκεια η οποία έχει ως αποτέλεσμα τα κύτταρα να εκτίθενται λιγότερο χρόνο σε συνθήκες έλλειψης θρεπτικών συστατικών. Έτσι, τα κύτταρα δεν χρειάζεται να ενεργοποιήσουν άλλες μεταβολικές διαδικασίες για την παραγωγή των απαραίτητων θρεπτικών συστατικών, όπως στην περίπτωση των δυο άλλων πρωτοκόλλων. Η διαφορά των δυο πρωτοκόλλων είναι μόνο στο χρόνο καθώς η αύξηση της χρονικής διάρκειας στο πρωτόκολλο 3A είχε ως στόχο την αποτελεσματικότερη έκπλυση του θρεπτικού μέσου ανάπτυξης. Ωστόσο, η μεταβολομική ανάλυση του τελικού διαλύματος έκπλυσης από αυτά τα δυο πρωτόκολλα έδειξε ότι η έκπλυση είναι εξίσου αποτελεσματική και στις δυο περιπτώσεις. Παρόλα αυτά, το πρωτόκολλο της χρονικής διάρκειας των 10min (3B) εμφανίζει κάποια μειονεκτήματα. Η σημαντικότερη διαφορά μεταξύ του πρωτοκόλλου 3A (10min) και του 3B (1min) είναι η μειωμένη συγκέντρωση της γλυκόζης κατά πέντε φορές στο πρώτο. Δυο μεταβολίτες που αυξήθηκαν στο πρωτόκολλο 3A ήταν το γαλακτικό οξύ και η αλανίνη τα οποία παράγονται από το πυροσταφυλικό οξύ (βλ. συζήτηση). Στα φυσιολογικά κύτταρα αυτή η διαδικασία λαμβάνει χώρα σε συνθήκες έλλειψης οξυγόνου ενώ στα καρκινικά κύτταρα συμβαίνει ακόμα και όταν τα επίπεδα οξυγόνου είναι φυσιολογικά. Η διαδικασία αυτή στα καρκινικά κύτταρα ονομάζεται αερόβια γλυκόλυση (φαινόμενο “Warburg”) και περιλαμβάνει την κατανάλωση της γλυκόζης με πολύ υψηλούς ρυθμούς, η οποία αντί να οδηγεί σε αύξηση της ενέργειας του κυττάρου μέσω του κύκλου του κιτρικού οξέος, ρέει προς την παραγωγή του γαλακτικού οξέος (Warburg 1956, Bayley & Devilee 2012). Άλλοι μεταβολίτες που βρέθηκαν με αυξημένη συγκέντρωση στο πρωτόκολλο 3A ήταν το φουμαρικό οξύ, το μηλικό οξύ και η ασπαραγίνη, τα οποία παράγονται στις μεταβολικές αντιδράσεις που έπονται της διαδικασίας της γλουταμινόλυσης, δηλαδή, της απαμίνωσης της γλουταμίνης σε γλουταμινικό οξύ (βλ. Ενότητα 4.4.1.). Η γλουταμινόλυση συμβαίνει και σε φυσιολογικά και σε καρκινικά κύτταρα όταν αυτά βιώνουν κάποιου είδους στρες (π.χ. υποξία, έλλειψη γλυκόζης) (Benjamin et al. 2012) (βλ. Ενότητα 4.4.1.). Τα παραπάνω αποτελέσματα υποδεικνύουν ότι τα κύτταρα που συλλέχθηκαν με το πρωτόκολλο 3A ενεργοποιούν μηχανισμούς απόκρισης (π.χ. γλουταμινόλυση) σε συνθήκες έλλειψης θρεπτικών συστατικών (π.χ. γλυκόζη) σε σχέση με αυτά του πρωτοκόλλου 3B.

➤ Σύγκριση κυττάρων που προέρχονται από διαφορετική ανακαλλιέργεια

Μια επιπρόσθετη σύγκριση που πραγματοποιήθηκε ήταν μεταξύ των κυττάρων που προήλθαν από διαφορετικό στάδιο ανακαλλιέργειας από τη στιγμή της απόψυξής τους. Πρέπει να σημειωθεί ότι όταν το στάδιο ανακαλλιέργειας αναφέρεται με τον αριθμό 2 (Πιν. 21) έχουν πραγματοποιηθεί δυο θρυψινοποιήσεις, μία μετά την απόψυξή τους και μία όταν έγινε η επίστρωσή τους στα τρυβλία καλλιέργειας. Αντίστοιχα στα κύτταρα της 8<sup>ης</sup> ανακαλλιέργειας έχουν γίνει 7 θρυψινοποιήσεις μετά την απόψυξή τους και μία κατά την επίστρωσή τους στα τρυβλία. Τα κύτταρα της 2<sup>ης</sup> ανακαλλιέργειας που χρησιμοποιήθηκαν στο πρωτόκολλο 3B σε σύγκριση με αυτά της 8<sup>ης</sup> του ίδιου πρωτοκόλλου παρατηρήθηκε ότι εμφανίζουν υψηλότερα επίπεδα γλυκόζης. Το γεγονός αυτό υποδεικνύει ότι τα κύτταρα της 2<sup>ης</sup> ανακαλλιέργειας δεν εκτίθενται παρατεταμένα σε συνθήκες στρες. Επιπλέον εμφανίζουν αυξημένα επίπεδα γαλακτικού οξέος και χαμηλά επίπεδα του φουμαρικού οξέος, του μηλικού οξέος και της ασπαραγίνης. Τα αποτελέσματα αυτά υποδεικνύουν την αύξηση της αερόβιας γλυκόλυσης που σημαίνει ότι υπάρχει υψηλή συγκέντρωση γλυκόζης. Έτσι δεν ενεργοποιούνται οι αντιδράσεις που έπονται της γλουταμινόλυσης, επειδή τα κύτταρα δεν χρειάζονται πρόσθετη υποστήριξη από άλλους μεταβολικούς μηχανισμούς για την παραγωγή των ενδιάμεσων της γλυκόλυσης (βλ. Ενότητα 4.4.1.). Επίσης, τα κύτταρα αυτά εμφανίζουν κανονικά επίπεδα 6-φωσφορικής γλυκόζης και καθόλου 6-φωσφορική φρουκτόζη το οποίο υποδεικνύει ότι δεν απαιτείται η ενεργοποίηση μεταβολικών μηχανισμών αποκατάστασης (π.χ. γλυκογονόλυση) των ενδιάμεσων της γλυκόλυσης διότι τα κύτταρα δεν εκτίθενται σε συνθήκες έλλειψης γλυκόζης (βλ. Ενότητα 4.4.1.). Τέλος, στα κύτταρα της 2<sup>ης</sup> ανακαλλιέργειας παρατηρείται αύξηση σε κάποια λιπαρά οξέα που δεν εμφανίζονται καθόλου στα κύτταρα της 8<sup>ης</sup> ανακαλλιέργειας, το οποίο υποδεικνύει αυξημένη λιπογένεση η οποία αποτελεί χαρακτηριστικό των καρκινικών κυττάρων. Η αυξημένη λιπογένεση υποδηλώνει αυξημένα επίπεδα γλυκόζης, η οποία μετατρέπεται, μέσω της γλυκόλυσης, σε πυροσταφυλικό οξύ που εισέρχεται στο κύκλο του κιτρικού οξέος. Με αυτόν τον τρόπο παράγεται κιτρικό οξύ το οποίο εξάγεται στο κυτταρόπλασμα και χρησιμοποιείται για τη σύνθεση λιπαρών οξέων (βλ. Ενότητα 4.4.1.). Συμπερασματικά τα κύτταρα της 2<sup>ης</sup> ανακαλλιέργειας εμφάνισαν την καλύτερη αναλογία ως προς το είδος και τις συγκεντρώσεις των μεταβολικών γεγονόσ που τα καθιστά καταλληλότερα για χρήση στα πειράματα που ακολούθησαν.

3.4.1.2. Το πρωτόκολλο 3B, το PBS ως ρυθμιστικό διάλυμα έκπλυσης και τα κύτταρα της 2<sup>ης</sup> ανακαλλιέργειας αποτελούν το βέλτιστο τρόπο επεξεργασίας και συλλογής κυττάρων για ομικές αναλύσεις

Σύμφωνα με τα παραπάνω αποτελέσματα, το πρωτόκολλο 3B θεωρήθηκε ως ο καταλληλότερος τρόπος επεξεργασίας και συλλογής κυττάρων διότι προκαλεί την πιο ήπια επίδραση στην κυτταρική φυσιολογία καθώς τα κύτταρα εκτίθενται το ελάχιστο δυνατό διάστημα σε συνθήκες έλλειψης θρεπτικών συστατικών και δεν υπόκεινται σε συνθήκες στρες επειδή η μηχανική αποκόλληση έπεται της προσθήκης της μεθανόλης και των φυγοκεντρήσεων. Τα τρία διαλύματα έκπλυσης που χρησιμοποιήθηκαν εξασφάλιζαν την πλήρη απομάκρυνση του θρεπτικού μέσου από τα κύτταρα καθώς η

μεταβολική ανάλυση του τελευταίου διαλύματος έκπλυσης έδειξε ότι τα κύτταρα είχαν εκπλυθεί αποτελεσματικά από το θρεπτικό μέσο. Ωστόσο, το καταλληλότερο διάλυμα πλύσης θεωρήθηκε το PBS καθώς το διάλυμα φυσιολογικού ορού (NaCl) υπέβαλε τα κύτταρα σε στρεσογόνες καταστάσεις ενεργοποιώντας επιπρόσθετες μεταβολικές οδούς. Επιπλέον, το διάλυμα DMEM δεν απέδιδε ικανοποιητικό αριθμό μεταβολιτών τόσο σε ποσοτικό όσο και σε ποιοτικό επίπεδο. Τα κύτταρα της 2<sup>ns</sup> ανακαλλιέργειας επιλέχθηκαν διότι εμφάνισαν την πιο αντιπροσωπευτική εικόνα της φυσιολογίας των καρκινικών κυττάρων.

Έχοντας επιλέξει το πρωτόκολλο 3B για την επεξεργασία και τη συλλογή των κυττάρων που προορίζονται για ομικές αναλύσεις, έπρεπε να γίνει η κατάλληλη προσαρμογή αυτού για να χρησιμοποιηθεί στην ανάλυση του υπο-μελέτη συστήματος σε μεταγραφικό, πρωτεϊνικό και μεταβολικό επίπεδο. Συνοψίζοντας τη διαδικασία του πρωτοκόλλου, τα βήματα που περιλαμβάνει είναι: α) η έκπλυση των κυττάρων στο τρυβλίο καλλιέργειας για τρεις συνεχόμενες φορές με PBS σε 1min, β) η προσθήκη μεθανόλης για την καταστολή των ενζυμικών διεργασιών των κυττάρων και την απελευθέρωση των μεταβολιτών, γ) η μηχανική αποκόλληση των κυττάρων με τη χρήση ειδικής ξύστρας και επακόλουθη συλλογή τους. Για τα κύτταρα που προορίζονταν για μεταβολομική ανάλυση, το πρωτόκολλο ακολουθήθηκε όπως περιγράφηκε παραπάνω. Το δεύτερο βήμα προσαρμόστηκε ανάλογα για την ανάκτηση του RNA και των πρωτεϊνών. Όσον αφορά τη μεταγραφωμική ανάλυση, χρησιμοποιήθηκε το διάλυμα λύσης από ειδικό κιτ απομόνωσης RNA, (RNeasy Micro kit, Qiagen) που έχει χρησιμοποιηθεί κατά κόρον για την απομόνωση RNA από κύτταρα και ιστούς που προορίζονται για την ανάλυση του μεταγραφικού προτύπου με τη χρήση των μικροσυστοιχιών (ενδεικτικά: Srinivasan et al. 2016, Adrian et al. 2015, Kamtchueng et al. 2014). Στην πρωτεομική ανάλυση χρησιμοποιήθηκε το διάλυμα αμμωνιακού άλατος TEAB (0.5M tetraethyl-ammonium bromide) και της οργανικής ένωσης SDS (0.05% Sodium dodecyl sulfate) τα οποία συμβάλλουν στην κυτταρική λύση και οδηγούν στην απελευθέρωση μεγάλης ποσότητας πρωτεϊνών σε διαλυτή μορφή. Η επιλογή του SDS βοηθάει στην αύξηση της διαλυτότητας των πρωτεϊνών και στην αποδιάταξή τους ενισχύοντας τη μετέπειτα ενζυματική πέψη αυτών (Ying Zhang et al. 2015). Η χρήση του αμμωνιακού άλατος TEAB διασφαλίζει τη συμβατότητα των πεπτιδίων που προκύπτουν με τα επακόλουθα αντιδραστήρια που χρησιμοποιούνται για τη σήμανση συγκεκριμένων αμινοξέων (σήμανση iTRAQ) (León et al. 2013) ώστε να αποτελεί το καταλληλότερο διάλυμα λύσης για πειράματα ποσοτικής πρωτεομικής μέσω υγρής χρωματογραφίας-φασματομετρίας μάζας (LC-MS ή LC-MS/MS) (Latosinska et al. 2015, Sandberg et al. 2014).

#### 3.4.2. Καταστολή της έκφρασης του γονιδίου FRA10AC1 στα κύτταρα HeLa

Εφόσον έγινε η βελτιστοποίηση του πρωτοκόλλου επεξεργασίας και συλλογής κυττάρων, ακολούθησε η δημιουργία του κυτταρικού μοντέλου *HeLa* με τροποποιημένη την έκφραση του γονιδίου *FRA10AC1*. Λαμβάνοντας υπ' όψιν τα αποτελέσματα της ανάλυσης του δικτύου του σωματίου συναρμογής που υποδεικνύουν την συμμετοχή της πρωτεΐνης στις διαδικασίες επεξεργασίας mRNA, η μελέτη των αλλαγών που επιφέρει η τροποποιημένη έκφραση του *FRA10AC1* στα μόρια που



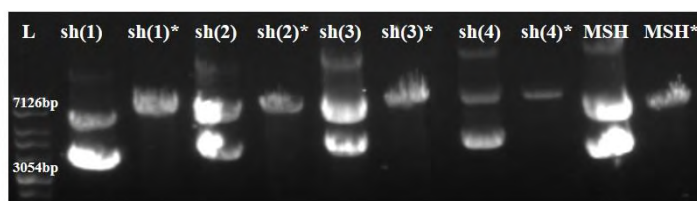
συμμετέχουν σε αυτές τις διαδικασίες θα συνεισέφερε στην κατανόηση των μοριακών μηχανισμών στους οποίους λαμβάνει μέρος η πρωτεΐνη FRA10AC1. Η τροποποίηση της έκφρασης του *FRA10AC1* σε κύτταρα *HeLa* αφορά την δημιουργία μόνιμα διαμολυσμένων κυτταρικών σειρών (stably transfected) στις οποίες πραγματοποιήθηκε καταστολή της έκφρασης του ενδογενούς γονιδίου μέσω τεχνολογίας παρεμβολής RNA (RNAi interference, RNAi).

Για να επιτευχθεί η καταστολή της έκφρασης του γονιδίου *FRA10AC1* στα κύτταρα *HeLa* δημιουργήθηκαν μόνιμα διαμολυσμένες κυτταρικές σειρές με διαγονίδια που φέρουν μόρια shRNA (short-hairpin RNAs), μήκους 59 βάσεων, κλωνοποιημένα στο φορέα pLKO.1-puro (7086bp). Συγκεκριμένα, χρησιμοποιήθηκαν τέσσερα μόρια shRNA με τους κωδικούς Sigma:

- TRCN0000160246: shRNA (1)
- TRCN0000158896: shRNA (2)
- TRCN0000159265: shRNA (3)
- TRCN0000159795: shRNA (4)

Τα ποσοστά αποτελεσματικότητας της αποσιώπησης, με βάση την εταιρία, ήταν 83%, 90%, 95% και 72%, αντίστοιχα. Ως αρνητικός έλεγχος χρησιμοποιήθηκε η πλασμιδιακή κατασκευή (#SHC002, Sigma) που περιέχει shRNA το οποίο δεν στοχεύει σε κανένα γνωστό γονίδιο κανενός είδους αλλά είναι απαραίτητο για τον προσδιορισμό της επίδρασης του μηχανισμού του RNAi στη γονιδιακή έκφραση.

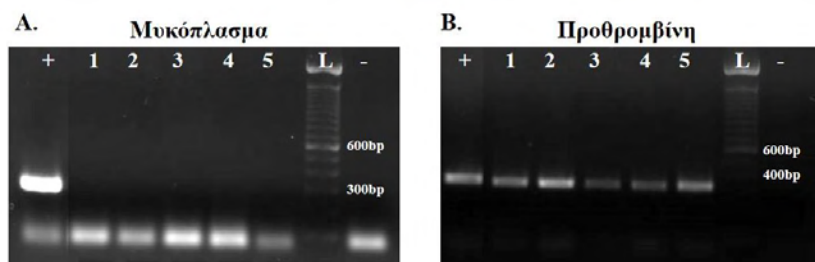
Το πρώτο στάδιο, πριν τη διαδικασία της διαμόλυνσης, ήταν η απομόνωση μεγάλης ποσότητας πλασμιδιακού DNA των τεσσάρων διαγονιδίων και του αρνητικού ελέγχου, έπειτα από το μετασχηματισμό βακτηριακών κυττάρων *DH5a* (Εικ. 59). Ακολούθησε η γραμμοποίησή τους με το ένζυμο περιορισμού *KpnI* (nt 1189) που εντοπίζεται ανοδικά της 3' θέσης LTR του φορέα κλωνοποίησης. Η πέψη σε αυτό το σημείο δεν διαταράσσει την αλληλουχία των shRNA καθώς και άλλα στοιχεία που απαιτούνται για την έκφραση σε κύτταρα θηλαστικών (Εικ. 59). Η συγκέντρωση του DNA των γραμμοποιημένων διαγονιδίων (7145bp) και του διαγονιδίου αρνητικού ελέγχου εκτιμήθηκε με φωτομέτρηση σε μήκος κύματος 260nm, και ήταν 250ng/μl (shRNA\_1), 200ng/μl (shRNA\_2), 250ng/μl (shRNA\_3), 60ng/μl (shRNA\_4) και 200ng/μl (shRNA αρνητικού ελέγχου) (Εικ. 59). Η συνολική ποσότητα των shRNA\_1, 2, 3, 4 και του shRNA αρνητικού ελέγχου ήταν κατ' αντιστοιχία 100μg, 80μg, 100μg, 36μg και 80μg.



**Εικ. 59:** Γραμμοποίηση των τεσσάρων διαγονιδίων και του διαγονιδίου που χρησιμοποιήθηκε ως αρνητικός έλεγχος (MSH) με το ένζυμο περιορισμού *KpnI* (nt 1189). Sh(1), sh(2), sh(3), sh(4), MSH: άκοπο DNA των shRNA\_1, 2, 3, 4 και του shRNA αρνητικού ελέγχου. Σύμβολο \*: γραμμοποιημένα διαγονίδια, L= μάρτυρας μοριακών μεγεθών (Cat. No 15615-016, Invitrogen).

Παράλληλα, έγινε ο καθορισμός της ελάχιστης συγκέντρωσης πουρομυκίνης που είναι ικανή να σκοτώσει, σε διάστημα 2 εβδομάδων, τα κύτταρα *HeLa* που δεν φέρουν ενσωματωμένο στο γονιδιωματικό τους DNA το διαγονίδιο pLKO.1-puro-shRNA. Χρησιμοποιήθηκαν 14 διαφορετικές συγκεντρώσεις πουρομυκίνης (0,01-2μg/ml) στα κύτταρα τα οποία καλλιεργήθηκαν σε τρυβλία 6 θέσεων για δυο εβδομάδες. Σε αυτό το διάστημα η αλλαγή του πλήρους θρεπτικού μέσου με το αντιβιοτικό πουρομυκίνη γινόταν κάθε τρεις ημέρες. Η ελάχιστη συγκέντρωση πουρομυκίνης που επιλέχθηκε, στην οποία πεθαίνουν όλα τα κύτταρα σε διάστημα 10 ημερών, ήταν 400 ng/ml. Ακολούθησε η διαμόλυνση των κυττάρων *HeLa* με 20μg από το κάθε γραμμοποιημένο διαγονίδιο shRNA 1-4 και με το γραμμοποιημένο διαγονίδιο του αρνητικού ελέγχου. Η μέθοδος διαμόλυνσης που χρησιμοποιήθηκε βασίζεται στη συνκατακρήμνιση του DNA με Ca<sub>3</sub>(PO<sub>4</sub>)<sub>2</sub> (Graham & Vander 1973) και έγινε 24h μετά την επίστρωση των τρυβλίων (διαμέτρου 10cm) τα οποία περιείχαν 1,2x10<sup>6</sup> κύτταρα το καθένα. Έπειτα από 48h μετά τη διαμόλυνση πραγματοποιήθηκε θρυψινοποίηση των κυττάρων και επανακαλλιέργειά τους σε τρία τρυβλία με κάλυψη επιφάνειας ~25% ώστε να δράσει η πουρομυκίνη. Μετά από 12h έγινε αλλαγή και προσθήκη θρεπτικού μέσου με 400ng/ml πουρομυκίνη. Κάθε τρεις μέρες γινόταν αλλαγή του θρεπτικού μέσου με το αντιβιοτικό ενώ συνολικά πραγματοποιήθηκαν τέσσερις αλλαγές. Τα κύτταρα *HeLa* που είναι ανθεκτικά στη πουρομυκίνη, ως συνέπεια της ενσωμάτωσης των διαγονιδίων, απλώνονται στην επιφάνεια του τρυβλίου και σχηματίζουν αποικίες. Συνολικά επιλέχθηκαν και αναπτύχθηκαν 17 κλώνοι για τα 4 διαφορετικά διαγονίδια των shRNA και 9 κλώνοι για το διαγονίδιο που χρησιμοποιήθηκε ως αρνητικός έλεγχος.

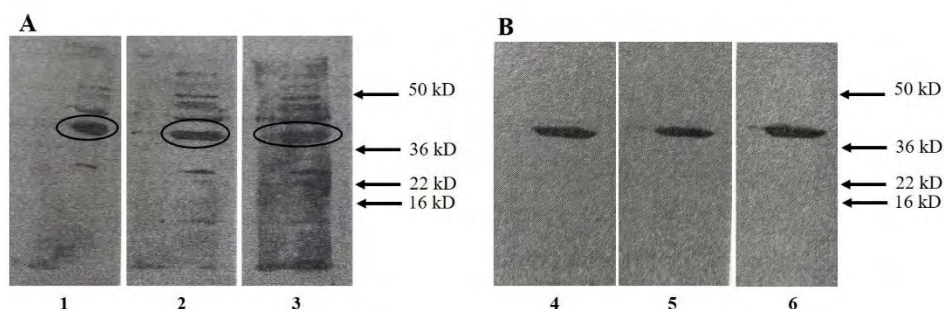
Κατά τη διάρκεια ανάπτυξης των κλώνων αυτών, όπως και σε κάθε περίπτωση καλλιέργειας κυττάρων που πραγματοποιήθηκε στην εργασία, έγινε συλλογή μικρής ποσότητας αυτών κατά την ανακαλιέργεια τους καθώς και πριν την κρυσυντήρηση και μετά την αποψυχή τους, ώστε να πραγματοποιηθεί ο έλεγχός τους για τυχόν παρουσία μόλυνσης από το βακτήριο μυκόπλασμα με PCR. Ειδικότερα, σε κάθε αντίδραση χρησιμοποιούνταν: α) εκκινητές που ενισχύουν μία αλληλουχία 280bp του μυκοπλάσματος, β) εκκινητές για το γονίδιο της προθρομβίνης (350bp) ως εσωτερικός μάρτυρας του πειράματος (βλ. Υλικά & Μέθοδοι) για να διασφαλιστεί η πιστότητα της αντίδρασης, γ) ένα δείγμα κυττάρων που είχε μολυνθεί από μυκόπλασμα ως θετικός μάρτυρας και δ) ένας αρνητικός μάρτυρας. Ενδεικτικά στην Εικόνα 60 φαίνεται ο έλεγχος μυκοπλάσματος για 3 διαφορετικούς κλώνους των διαγονιδίων των shRNA, για έναν του διαγονιδίου που χρησιμοποιήθηκε ως αρνητικός έλεγχος και για μη τροποποιημένα κύτταρα *HeLa*.



**Εικ. 60: Έλεγχος μυκοπλάσματος με PCR.** Φαίνονται τα προϊόντα PCR των κλώνων shRNA(3)\_5, shRNA(4)\_4, shRNA(2)\_3, ενός κλώνου του διαγονιδίου που χρησιμοποιήθηκε ως αρνητικός έλεγχος

και των κυττάρων *HeLa* στις θέσεις 1, 2, 3, 4 και 5, αντίστοιχα, για την αλληλουχία του μυκοπλάσματος (280bp) (A) και για το γονίδιο της προθρομβίνης (350bp) (B). (+), (-): θετικός και αρνητικός μάρτυρας της κάθε αντίδρασης, αντίστοιχα. L: μάρτυρας μοριακών μεγεθών (Cat. No 15628019, Invitrogen).

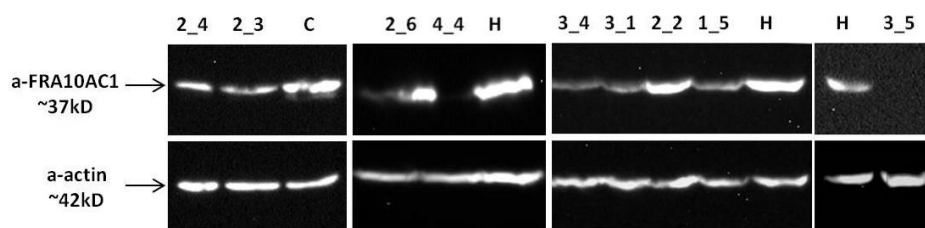
Σε πρώτο στάδιο, έγινε ο έλεγχος της καταστολής του γονιδίου *FRA10AC1* στους 17 κλώνους που αναπτύχθηκαν, με ανάλυση western με το αντίσωμα  $\alpha$ -*FRA10AC1* συγκριτικά με πρωτεϊνικό εκχύλισμα κυττάρων *HeLa* που είχαν διαμολυνθεί με το διαγονίδιο αρνητικού ελέγχου ή με κύτταρα *HeLa* αγρίου τύπου έτσι ώστε να υπολογιστούν τα επίπεδα έκφρασης της ενδογενούς πρωτεΐνης. Για την ανάλυση western χρησιμοποιήθηκαν ~200μg πρωτεϊνικού εκχυλίσματος του κάθε κλώνου και αντίσωμα σε αραιώση 1:500 το οποίο είχε απομονωθεί από πολυκλωνικό ορό κουνελιού (Εικ. 61).



**Εικ. 61: Καθαρισμός αντισώματος  $\alpha$ -*FRA10AC1*.** Ανάλυση western (έκθεση 2min) με διαδοχικές αραιώσεις πολυκλωνικού ορού: 1: 1/1000, 2: 1/600, 3: 1/300 (A), διαδοχικές αραιώσεις καθαρισμένου αντισώματος: 4: 1/100, 5: 1/300, 6: 1/500 (B). Σε κύκλο φαίνεται η ζώνη που αντιστοιχεί στο μοριακό βάρος της *FRA10AC1* (37.5kD).

Για την κανονικοποίηση της ποσότητας των δειγμάτων που εξετάστηκαν, χρησιμοποιήθηκε αντίσωμα έναντι της ακτίνης (αραιώση 1:1000) ως εσωτερικός μάρτυρας του πειράματος.

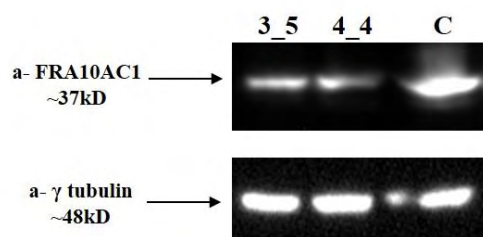
Από τον προκαταρκτικό έλεγχο ταυτοποιήθηκε ότι 9 από τους 17 κλώνους που εμφανίζουν μειωμένη έως σχεδόν μηδενική έκφραση της πρωτεΐνης *FRA10AC1*. Η ποσοτικοποίηση των επιπέδων έκφρασης έγινε έπειτα από κανονικοποίηση της ποσότητας των δειγμάτων που χρησιμοποιήθηκαν με βάση τον εσωτερικό έλεγχο του πειράματος (γονίδιο ακτίνης) με το πρόγραμμα Image J. Με αυτό τον τρόπο, παρατηρήθηκε ότι οι κλώνοι shRNA (3)\_5 και shRNA (4)\_4, σε σύγκριση με το θετικό έλεγχο του πειράματος (*HeLa* αγρίου τύπου), φαίνονται να παρουσιάζουν μηδενική έκφραση της πρωτεΐνης, λαμβάνοντας υπ' όψιν τα όρια ανίχνευσης της ανάλυσης western (Εικ. 62). Όσον αφορά τους υπόλοιπους κλώνους, shRNA (2)\_3, shRNA (2)\_4, shRNA (1)\_5, shRNA (2)\_2, shRNA (3)\_1 και shRNA (3)\_4 εμφάνισαν κατ' αντιστοιχία ποσοστά καταστολής 63%, 33%, 47%, 21%, 54% και 58% με εξαίρεση τον κλώνο shRNA (2)\_6 ο οποίος δεν φαίνεται να καταστέλλει την έκφραση του γονιδίου (Εικ. 62).



**Εικ. 62:** Ανάλυση western των επιπέδων έκφρασης της πρωτεΐνης FRA10AC1 σε κλώνους κυττάρων *HeLa*, μόνιμα διαμολυσμένους με shRNA που στοχεύουν στο γονίδιο *FRA10AC1*. Φαίνονται οι κωδικοί των κλώνων: ο πρώτος αριθμός σε κάθε κλώνο αντιστοιχεί στο shRNA από το οποίο προέκυψε και ο δεύτερος αριθμός στον αύξοντα αριθμό της αποικίας. Η: *HeLa* αγρίου τύπου, C: διαγονίδιο αρνητικού ελέγχου (MSH 2). Στο κάτω μέρος της εικόνας φαίνονται τα επίπεδα έκφρασης της ακτίνης στα αντίστοιχα δείγματα τα οποία χρησιμοποιήθηκαν για την κανονικοποίηση των επιπέδων έκφρασης.

Αξίζει να σημειωθεί ότι οι τέσσερις αποικίες που προέκυψαν από τον κλώνο shRNA (2), ο οποίος με βάση τα δεδομένα της εταιρίας παρουσιάζει ιδιαίτερα υψηλότερα ποσοστά αποτελεσματικότητας (90%) εμφανίζονται, τουλάχιστον με αυτήν την ανάλυση, να αποσιωπούν το γονίδιο σε μικρότερο βαθμό (Εικ. 62). Αντίθετα, η αποικία που προέκυψε από τον κλώνο shRNA (4) εμφανίζεται να καταστέλλει πλήρως το γονίδιο (Εικ. 62) σε αντίθεση με τα χαμηλότερα συγκριτικά επίπεδα καταστολής (72%) που έδινε η εταιρία. Αυτό το αποτέλεσμα πιθανώς σχετίζεται με την επίδραση που μπορεί να έχει η θέση ένθεσης του διαγονιδίου στο γονιδίωμα, η οποία μπορεί να επηρεάζει την έκφρασή του, ή με τον αριθμό αντιγράφων του διαγονιδίου που τελικά ενσωματώνονται.

Με βάση τα αποτελέσματα της ανάλυσης western, επιλέχθηκαν οι δύο κλώνοι shRNA (3)\_5 και shRNA (4)\_4, που παρουσιάζουν πλήρη καταστολή, για επανέλεγχο ώστε να γίνει επιβεβαίωση των αποτελεσμάτων (Εικ. 63). Και στο επαναληπτικό πείραμα, οι δυο κλώνοι εμφάνισαν σχεδόν πλήρη καταστολή, σε ποσοστό ~90%, έπειτα από την κανονικοποίηση των επιπέδων έκφρασής τους.

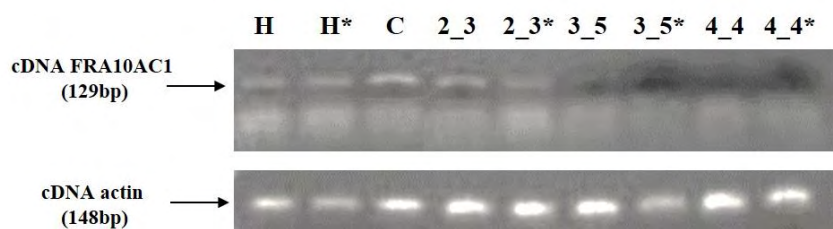


**Εικ. 63:** Ανάλυση western των επιπέδων έκφρασης της πρωτεΐνης FRA10AC1 των κλώνων shRNA(3)\_5 και shRNA(4)\_4. C: διαγονίδιο αρνητικού ελέγχου (MSH 2). Στο κάτω μέρος της εικόνας φαίνονται τα επίπεδα έκφρασης της γ-τουμπουλίνης των 3 δειγμάτων, τα οποία χρησιμοποιήθηκαν για την κανονικοποίηση των επιπέδων έκφρασης.

Βάσει των παραπάνω αποτελεσμάτων, οι κυτταρικοί κλώνοι shRNA (3)\_5, shRNA (4)\_4 που εμφάνισαν πλήρη καταστολή και ο shRNA (2)\_3 που είχε ποσοστό καταστολής 63% επιλέχθηκαν για τα πειράματα των ομικών αναλύσεων σε μεταγραφικό, πρωτεϊνικό και σε μεταβολικό επίπεδο. Η επιλογή και ο έλεγχος πολλών διαφορετικών κλώνων κυττάρων έγινε διότι η ενσωμάτωση των διαγονιδίων στο

γονιδίωμα των κυττάρων γίνεται τυχαία με αποτέλεσμα να υπάρχει η πιθανότητα αυτά να ενσωματωθούν σε οποιοδήποτε γονίδιο ή λειτουργικό στοιχείο του γονιδιώματος. Η ενσωμάτωση του διαγονιδίου που περιέχει το shRNA που στοχεύει στο γονίδιο *FRA10AC1* ή του διαγονιδίου αρνητικού ελέγχου σε οποιαδήποτε λειτουργική αλληλουχία μπορεί να προκαλέσει διάφορες επιδράσεις και μπορεί οι αλλαγές που θα παρατηρηθούν και θα ελεγχθούν μέσω ομικών αναλύσεων να μην προέρχονται από την καταστολή της έκφρασης του *FRA10AC1*. Γι' αυτό το λόγο οι κλώνοι κυττάρων (shRNA (3)\_5, shRNA (4)\_4) που εμφάνισαν παρόμοια ποσοστά καταστολής του γονιδίου ελέγχθηκαν μέσω μεταβολομικής ανάλυσης, ώστε να επιβεβαιωθεί ότι προκαλούν παρόμοιες αλλαγές στην κυτταρική φυσιολογία. Στο ίδιο πλαίσιο, ελέγχθηκαν περισσότεροι από έναν κλώνοι αναφοράς, και πάλι μέσω μεταβολομικής ανάλυσης, ώστε να ελεγχθεί αν προκαλούν παρόμοιες αλλαγές με τα μη τροποποιημένα κύτταρα *HeLa*.

Επιπρόσθετα, σε δεύτερο στάδιο, η επιβεβαίωση της καταστολής της έκφρασης του γονιδίου *FRA10AC1* στους κλώνους shRNA (3)\_5, shRNA (4)\_4 και shRNA (2)\_3 πραγματοποιήθηκε και με ημιποσοτική αντίστροφη μεταγραφή (Reverse Transcription, RT) συζευγμένη με PCR. Η διαδικασία αυτή έγινε από το συνεργαζόμενο εργαστήριο Γενικής Βιολογίας (Επικεφαλής: Καθ. Ν. Μοσχονάς) της Ιατρικής Σχολής του Πανεπιστημίου Πατρών. Στην PCR χρησιμοποιήθηκε ως υπόστρωμα cDNA, που είχε συντεθεί από ολικό RNA των κυττάρων, μέσω της χρήσης κατάλληλων εκκινητών ώστε να παραχθεί ένα προϊόν μήκους 129bp (Εικ. 64). Τα αποτελέσματα της ημιποσοτικής PCR κανονικοποιήθηκαν με βάση τον εσωτερικό έλεγχο του πειράματος (γονίδιο ακτίνης) και συγκρίθηκαν με το πρόγραμμα Image J, ώστε να εξαχθούν συμπεράσματα σχετικά με τα επίπεδα έκφρασης του γονιδίου. Παρατηρήθηκε ότι στους κλώνους shRNA (3)\_5 (αύξοντα αριθμός της αποικίας του αντίστοιχου κλώνου) και shRNA (4)\_4 η έκφραση του γονιδίου φαίνεται να καταστέλλεται πλήρως ή σχεδόν πλήρως, αντίστοιχα, σε σχέση με τους θετικούς ελέγχους του πειράματος (*HeLa* αγρίου τύπου και διαγονίδιο αρνητικού ελέγχου) (Εικ. 64). Όσον αφορά τον κλώνο shRNA (2)\_3, η έκφραση του γονιδίου καταστέλλεται σε ποσοστό 72% (Εικ. 64).



**Εικ. 64: Ημιποσοτικός προσδιορισμός των επιπέδων mRNA του *FRA10AC1* στους κλώνους υπόεκφρασης.** H: *HeLa* αγρίου τύπου C: κλώνος αρνητικός ελέγχου με αύξοντα αριθμό 2 (MSH 2) \*:H δεύτερη ανεξάρτητη επανάληψη κάθε κλώνου. Στο κάτω μέρος της εικόνας φαίνεται το προϊόν της ακτίνης για όλα τα δείγματα (148bp) το οποίο χρησιμοποιήθηκε για την κανονικοποίηση των επιπέδων έκφρασης.

### 3.4.3. Συλλογή δειγμάτων για ομικές αναλύσεις

Σύμφωνα με τα παραπάνω αποτελέσματα οι κυτταρικοί κλώνοι υποέκφρασης του γονιδίου *FRA10AC1* που επιλέχθηκαν για τις ομικές αναλύσεις ήταν οι shRNA (3)\_5 και shRNA (4)\_4 οι οποίοι παρουσίασαν σχεδόν πλήρη καταστολή της έκφρασης του γονιδίου και ο κλώνος shRNA (2)\_3 που εμφάνισε μερική καταστολή σε ποσοστό ~65%.

Το πρώτο βήμα για την προετοιμασία των δειγμάτων για ομικές αναλύσεις ήταν η δημιουργία μιας σειράς κρυσυντηρημένων κυττάρων αυτών των κλώνων ώστε όλα τα κύτταρα που θα χρησιμοποιούνταν για τις ομικές αναλύσεις να προέρχονται από το ίδιο στάδιο ανακαλλιέργειας από τη στιγμή της απόψυξής τους. Όπως προαναφέρθηκε, το βέλτιστο στάδιο ανακαλλιέργειας ήταν το 2<sup>ο</sup> διότι είχε μικρότερη επίδραση στην κυτταρική φυσιολογία και προσομοιάζε σε ικανοποιητικό βαθμό την φυσιολογία των καρκινικών κυττάρων. Έτσι, τα κύτταρα όλων των κλώνων υφίσταντο μία θρυψινοποίηση μετά την απόψυξή και την καλλιέργειά τους και μία κατά την επίστρωσή τους στα τρυβλία καλλιέργειας. Σε κάθε τρυβλίο καλλιέργειας επιστρώθηκαν 10<sup>6</sup> κύτταρα. Η συλλογή των κυττάρων έγινε μετά το πέρας των 48h από την επίστρωσή τους, σύμφωνα με το πρωτόκολλο που επιλέχθηκε και με τις κατάλληλες προσαρμογές αυτού, όπως αναφέρθηκε και παραπάνω.

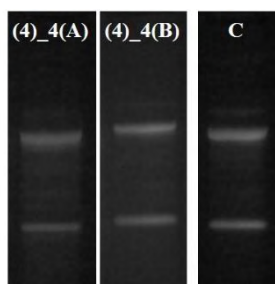
Για τον προσδιορισμό του μεταγραφικού προτύπου χρησιμοποιήθηκαν 3 δείγματα κυττάρων με τροποποιημένη έκφραση του *FRA10AC1*. Συγκεκριμένα:

- δυο βιολογικές επαναλήψεις (Α και Β) του κλώνου υποέκφρασης shRNA(4)\_4, στον οποίο παρατηρείται σχεδόν ολική καταστολή του *FRA10AC1*
- μία επανάληψη του κλώνου αναφοράς για την υποέκφραση του *FRA10AC1* (MSH 2),

Η απομόνωση του ολικού RNA από τα παραπάνω κύτταρα πραγματοποιήθηκε με κατάλληλο κιτ απομόνωσης, όπως περιγράφηκε παραπάνω. Τα δείγματα αποθηκεύτηκαν στους -80°C και μεταφέρθηκαν με ξηρό πάγο στο συνεργαζόμενο εργαστήριο Γενικής Βιολογίας (Επικεφαλής: Καθ. Ν. Μοσχονάς), της Ιατρικής Σχολής του Πανεπιστημίου Πατρών. Ακολούθησε ο έλεγχος των παρασκευασμάτων RNA ως προς την ποιότητά τους, δηλαδή την ακεραιότητά τους, με ηλεκτροφόρηση σε 2% πήκτωμα αγαρόζης, στο οποίο εμφανίζονται οι ζώνες που αντιστοιχούν στα ριβοσωμικά RNA των δυο υπομονάδων του ριβοσώματος (Εικ. 65). Επιπλέον, ελέγχθηκαν ως προς την ποσότητα και την καθαρότητά τους με μέτρηση της οπτικής πυκνότητας και του λόγου καθαρότητας (λόγος 260/280), τα επιθυμητά όρια του οποίου κυμαίνονται στις τιμές 1,8-2, σε φασματοφωτόμετρο (Πιν. 22). Στη συνέχεια τα δείγματα στάλθηκαν σε ξηρό πάγο στο συνεργαζόμενο εργαστήριο του Δρ. Bicciato (Center for Genome Research, Πανεπιστήμιο Modena and Reggio Emilia, Ιταλία) όπου και επιβεβαιώθηκε ότι τα δείγματα RNA ήταν ικανοποιητικής ποιότητας και ποσότητας.

**Πιν. 22: Ποιοτικός και ποσοτικός έλεγχος του RNA των δειγμάτων που προορίζονται για μεταγραφωμική ανάλυση.** Δυο τεχνικές επαναλήψεις (Δείγμα A και B) του κλώνου υποέκφρασης shRNA(4)\_4 και του αντίστοιχου δυο κλώνου αναφοράς (MSH 2).

Κωδικός Δείγματος	ng/μl	Συνολική ποσότητα RNA (μg)	Λόγος 260/280
shRNA(4)_4 (Δείγμα A)	540.5	27	2.02
shRNA(4)_4 (Δείγμα B)	371.4	18,5	2.01
MSH 2	134.54	6.7	1.98



**Εικ. 65: Ηλεκτροφόρηση ολικού RNA των δυο τεχνικών επαναλήψεων (A) και (B) του κλώνου υποέκφρασης (4\_4), και του αρνητικού ελέγχου του, C (MSH 2).**

Για την ανάκτηση του πρωτεϊνικού προτύπου χρησιμοποιήθηκαν 8 δείγματα (τρυβλία) για καθένα από τους παρακάτω 5 κλώνους:

- Κύτταρα *HeLa* που δεν είχαν τροποποιηθεί
- Ο κλώνος ολικής υποέκφρασης του *FRA10AC1*, shRNA(4)\_4
- Ο κλώνος ολικής υποέκφρασης του *FRA10AC1*, shRNA(3)\_5
- Ο κλώνος μερικής υποέκφρασης του *FRA10AC1*, shRNA(2)\_3
- Ο κλώνος αναφοράς για το σύστημα υποέκφρασης (MSH 2)

Τα 8 δείγματα (τρυβλία) κάθε κλώνου ενώθηκαν ανά 2, και τα τελικά 4 δείγματα του κάθε κλώνου αποθηκεύτηκαν στους  $-80^{\circ}\text{C}$  και στάλθηκαν με ξηρό πάγο στο συνεργαζόμενο εργαστήριο του Δρ. S. Garbis (Proteomics Center, Medical School, University of Southampton, U.K.). Εκεί, τα 4 δείγματα ενώθηκαν ανά 2 ώστε να μειωθεί η διαφοροποίηση μεταξύ των τρυβλίων, και το καθένα από τα 2 τελικά δείγματα του κάθε κλώνου περιείχε δείγματα κυττάρων από τέσσερα τρυβλία. Στο συνεργαζόμενο εργαστήριο έγινε η απομόνωση του πρωτεϊνικού περιεχομένου των 10 δειγμάτων, 2 για καθένα από τους 5 κλώνους. Η ποσοτικοποίηση του πρωτεϊνικού περιεχομένου έγινε με φασματοσκοπία υπερύθρου (Infrared spectroscopy, IR) (Garbis et al., 2008).

Για την μέτρηση του μεταβολικού προτύπου χρησιμοποιήθηκαν 6 δείγματα (τρυβλία) για τους κλώνους υποέκφρασης shRNA(4)\_4, shRNA(3)\_5, shRNA(2)\_3 και για τον κλώνο αναφοράς του συστήματος υποέκφρασης (MSH 2) και 4 δείγματα (τρυβλία) για τα κύτταρα *HeLa* που δεν είχαν τροποποιηθεί. Η μηχανική αποκόλληση και η συλλογή των κυττάρων από κάθε τρυβλίο έγινε με την προσθήκη μεθανόλης. Τα δείγματα αποθηκεύτηκαν στους  $-80^{\circ}\text{C}$  και μεταφέρθηκαν με ξηρό πάγο στο συνεργαζόμενο εργαστήριο της Δρ. Μ. Κλάπα (Μεταβολικής Μηχανικής και Βιολογίας Συστημάτων, Ινστιτούτο Χημικών Μηχανικών Επιστημών, Ίδρυμα Έρευνας

και Τεχνολογίας, FORTH/ICE-HT, Πάτρα). Παράλληλα με την καλλιέργεια των 2 (από τα 6) τρυβλίων (δειγμάτων) των κλώνων shRNA(4)\_4, shRNA(3)\_5, shRNA(3)\_5 και MSH 2 χρησιμοποιήθηκε ένα επιπλέον τρυβλίο για κάθε κλώνο από αυτούς με ίδιο αριθμό κυττάρων με τα εξεταζόμενα δείγματα. Το τρυβλίο αυτό χρησιμοποιήθηκε για τη μέτρηση του πρωτεϊνικού περιεχομένου με τη μέθοδο Bradford (Bradford, 1976) (Πιν. 23). Οι τιμές της ποσότητας πρωτεΐνης των υπόλοιπων 4 δειγμάτων αυτών των κλώνων και των 4 δειγμάτων των κυττάρων *HeLa* ήταν κοινές με αυτές του πρωτεϊνικού προτύπου (πραγματοποιήθηκαν στο εργαστήριο του Δρ. S. Garbis) διότι αυτές οι βιολογικές επαναλήψεις του μεταβολικού έγιναν ταυτόχρονα με αυτές του πρωτεϊνικού προτύπου.

**Πιν. 23: Μετρήσεις πρωτεϊνικού περιεχομένου των 2 δειγμάτων των κλώνων shRNA(4)\_4, shRNA(3)\_5, shRNA(3)\_5 και MSH 2.**

	mg πρωτεΐνης
shRNA(2)_3	2,60
shRNA(3)_5	3,10
shRNA(4)_4	2,34
MSH 2	3,03

#### 3.4.4. Εφαρμογή ομικών συνδυαστικών αναλύσεων του κυτταρικού μοντέλου με τροποποιημένη την έκφραση του FRA10AC1

##### 3.4.4.1. Ανάλυση μεταγραφικού προτύπου

Τα επίπεδα έκφρασης των γονιδίων που επηρεάζονται από την υποέκφραση του *FRA10AC1* στο κυτταρικό μοντέλο που δημιουργήθηκε, προσδιορίστηκαν με ανάλυση μικροσυστοιχιών. Στην ανάλυση αυτή χρησιμοποιήθηκαν δυο βιολογικές επαναλήψεις για τον κλώνο υποέκφρασης shRNA(4)\_4 και μια για τον κλώνο αναφοράς (MSH 2), όπως αναλύθηκε και παραπάνω. Η μέτρηση του μεταγραφικού προτύπου των τριών δειγμάτων πραγματοποιήθηκε στο συνεργαζόμενο εργαστήριο του Δρ. Bicciato (Center for Genome Research, Πανεπιστήμιο Modena and Reggio Emilia, Ιταλία) με τη χρήση μικροσυστοιχιών Gene Chip Human Genome U133 Plus 2.0 Array σύμφωνα με το προτυποποιημένο πρωτόκολλο της εταιρείας Affymetrix.

Η ανάλυση των μεταγραφικών προτύπων έγινε τόσο από τον Δρ. Bicciato όσο και από την Δρ. Μ. Κλάπα στο συνεργαζόμενο εργαστήριο Μεταβολικής Μηχανικής και Βιολογίας Συστημάτων, Ινστιτούτο Χημικών Μηχανικών Επιστημών (Ιδρυμα Έρευνας και Τεχνολογίας, FORTH/ICE-HT, Πάτρα). Σε πρώτο στάδιο, η ανάλυση περιελάμβανε: α) την μέτρηση του μέσου όρου έντασης του σήματος κάθε συνόλου ανιχνευτών που αντιστοιχούν στα μετάγραφα ενός γονιδίου, σε κάθε μία από τις συστοιχίες που χρησιμοποιήθηκαν. Ο μέγιστος δυνατός αριθμός του συνόλου ανιχνευτών (probe set) που έχει η εταιρία Affymetrix στη συγκεκριμένη μικροσυστοιχία είναι 16 ολιγονουκλεοτιδικοί ανιχνευτές (Chalifa-Caspi et al. 2004). Κατά τη χρήση των μικροσυστοιχιών αυτών γίνεται φιλτράρισμα ορισμένων probes (π.χ. αυτά που περιλαμβάνονται σε δύο ή περισσότερα probe sets, Ferrari et al. 2007) και ο αριθμός του τελικού συνόλου ανιχνευτών συνήθως είναι τουλάχιστον 11,



αντιπροσωπευτικός αριθμός probe sets, όπως προτείνεται και από την εταιρία και β) την κανονικοποίηση των μεταγραφικών προτύπων των διαφορετικών συστοιχιών με κατάλληλους αλγορίθμους ώστε να αποφευχθούν τυχόν αποκλίσεις που μπορεί να προέκυπταν από πολύ μεγάλες διαφορές στο επίπεδο έκφρασης των γονιδίων στις διαφορετικές καταστάσεις που εξετάζονται (π.χ. στους διαφορετικούς κλώνους με τροποποιημένη και μη έκφραση του γονιδίου *FRA10AC1*) και από διαφορές που προκύπτουν κατά τη διάρκεια της διεκπεραίωσης του πειράματος (Hartemink et al. 2001). Στο δεύτερο στάδιο, εφόσον φιλτραρίστηκαν κάποια σύνολα ανιχνευτών, προσδιορίστηκε η αφθονία (υπολογισμός έντασης του σήματος του συνόλου ανιχνευτών) των εκάστοτε μεταγράφων και έγινε αντιστοίχιση με τους κωδικούς νουκλεοτιδικής αλληλουχίας (Refseq ID) χρησιμοποιώντας την πιο πρόσφατη επισημείωση (annotation) του ανθρώπινου γονιδιώματος. Το επόμενο στάδιο ήταν η στατιστική ανάλυση των τελικών κανονικοποιημένων μεταγραφικών προτύπων. Το τελευταίο βήμα της ανάλυσης περιλάμβανε τη μέτρηση της διαφορικής έκφρασης των μεταγράφων μέσω του υπολογισμού του δυαδικού λογαρίθμου του λόγου ( $\log_2$  ratio) της μέσης τιμής έκφρασης του συνόλου των ανιχνευτών (probe set) που αντιστοιχούν σε κάθε μετάγραφο στον κλώνο υποέκφρασης shRNA(4)\_4 του *FRA10AC1* ως προς τον κλώνο αναφοράς υποέκφρασης (MSH 2).

Σύμφωνα με το όριο σημαντικότητας που επιλέχθηκε, ως μετάγραφα με σημαντική υπερέκφραση (θετικά σημαντικά) θεωρήθηκαν αυτά που παρουσιάζουν τουλάχιστον τριπλάσια αύξηση στην έκφρασή τους στην υποέκφραση του *FRA10AC1* σε σχέση με τον κλώνο αναφοράς (control), δηλαδή ο  $\log_2$  του λόγου της έκφρασής τους είναι μεγαλύτερος ή ίσος του 1,5 ( $\log_2 \geq 1,5$ ). Με τον ίδιο τρόπο, τα μετάγραφα με σημαντική υποέκφραση (αρνητικά σημαντικά) θεωρήθηκαν αυτά που παρουσιάζουν τουλάχιστον τριπλάσια μείωση στην έκφρασή τους στην υποέκφραση του *FRA10AC1* σε σχέση με τον κλώνο αναφοράς (control), δηλαδή ο  $\log_2$  του λόγου της έκφρασής τους είναι μικρότερος ή ίσος του -1,5 ( $\log_2 \leq -1,5$ ). Επίσης, είχε ληφθεί υπόψη η τιμή FDR (False Discovery Rate,  $q < 0,05$ ), η οποία χρησιμοποιείται γενικά στις μικροσυστοιχίες, διότι είναι απαραίτητη όταν γίνεται σύγκριση της μέσης τιμής έκφρασης πολλαπλών συνόλων ανιχνευτών που αφορούν συγκεκριμένα μετάγραφα ώστε να μην συμπεριληφθούν ψευδώς θετικά αποτελέσματα (Cheng & Pounds 2007). Κατά τον υπολογισμό της τιμής (q) αξιολογούνται (αποδοχή ή απόρριψη) οι μηδενικές υποθέσεις (null hypothesis), δηλαδή, οι υποθέσεις στις οποίες δεν υπάρχει καμία στατιστικά σημαντική διαφορά μεταξύ συγκεκριμένων συνόλων ανιχνευτών (probe set) στις δυο καταστάσεις που συγκρίνονται (Benjamini & Hochberg 1995).

#### 3.4.4.2. Ανάλυση πρωτεϊνικού προτύπου

Για την ταυτοποίηση και τον έλεγχο της ποσότητας των πρωτεϊνών σε συνθήκες τροποποιημένης έκφρασης του *FRA10AC1*, χρησιμοποιήθηκε χρωματογραφία προσρόφησης και φασματομετρία μάζας. Η μέτρηση του πρωτεϊνικού προτύπου πραγματοποιήθηκε από το συνεργαζόμενο εργαστήριο του Dr. Garbis στο Proteomics Center, Medical School, University of Southampton, U.K. σε 10 δείγματα, 2 βιολογικές επαναλήψεις (δείγματα) από τον κάθε κλώνο υποέκφρασης του *FRA10AC1*

(shRNA(4)\_4, shRNA(3)\_5, shRNA(2)\_3), 2 από τον κλώνο αναφοράς για το σύστημα υποέκφρασης (MSH 2) και 2 από κύτταρα *HeLa* που δεν είχαν τροποποιηθεί.

Πριν τη μέτρηση του πρωτεϊνικού προτύπου έγινε, στο συνεργαζόμενο εργαστήριο, η επεξεργασία των δειγμάτων (αναγωγή δισουλφιδικών δεσμών, αλκυλίωση και πρωτεόλυση των πρωτεϊνών) και η σήμανση των πρωτεϊνών με ετικέτες iTRAQ, βάσει του πρωτοκόλλου της εταιρείας ABSciex (<http://www.biotech.cornell.edu>, San Hose, CA, USA). Η σήμανση iTRAQ επιτρέπει την ταυτόχρονη ποσοτικοποίηση οκτώ πρωτεϊνικών προτύπων (Aggarwal et al. 2006). Σε κάθε μέτρηση, στην τελευταία θέση του συστήματος iTRAQ, χρησιμοποιήθηκε ένα μίγμα από τα δείγματα *HeLa* (mix *HeLa*) των δύο επαναλήψεων για να είναι δυνατή η κανονικοποίηση και η σύγκριση των δυο μετρήσεων. Η μέτρηση του πρωτεϊνικού προτύπου έγινε με τεχνικές χρωματογραφίας αντίστροφης φάσης (Reversed-Phase, RP) υψηλού pH χρησιμοποιώντας τη χρωματογραφική στήλη Waters, XBridge C8 column (150 x 3mm, 3.5µm particle) στο χρωματογράφο υγρών υψηλής απόδοσης (HPLC) UltiMate (LC Packings, Amsterdam, NL). Κάθε τμήμα του πρωτεϊνικού εκχυλίσματος αναλύθηκε με φασματομετρία μάζας ανασχηματισμού Fourier (FT-MS) με ιονισμό νανοψεκασμού (nanospray ionization) (UltiMate 3000 UHPLC-LTQ-Velos ProOrbitrap Elite, ThermoScientific, Bremen, DE). Τα ανακατασκευασμένα χρωματογραφήματα που ανακτήθηκαν από τα φάσματα μάζας, αναλύθηκαν από το ίδιο συνεργαζόμενο εργαστήριο χρησιμοποιώντας το κατάλληλο λογισμικό με τη χρήση συγκεκριμένων παραμέτρων.

Ακολούθησε η κανονικοποίηση των πρωτεϊνικών προτύπων και υπολογίστηκε ο δυαδικός λογάριθμος ( $\log_2$ ) του λόγου αφθονίας των πρωτεϊνών, σε κάθε μέτρηση, για τα εξής ζεύγη δειγμάτων: α) κλώνοι υποέκφρασης shRNA(4)\_4, shRNA(3)\_5, shRNA(2)\_3, ως προς τον κλώνο αναφοράς για το σύστημα υποέκφρασης (MSH 2) και β) κλώνος αναφοράς για το σύστημα υποέκφρασης ως προς το δείγμα των κυττάρων *HeLa* και ως προς το μίγμα των 2 δειγμάτων των κυττάρων *HeLa* (mix *HeLa*). Κατά την κανονικοποίηση πραγματοποιήθηκε ο υπολογισμός των παραπάνω λόγων αφθονίας των αντίστοιχων πρωτεϊνών σε κάθε δείγμα ως προς το δείγμα του μίγματος των κυττάρων *HeLa* (mix *HeLa*) της ίδιας μέτρησης iTRAQ (π.χ. shRNA(4)\_4/ MSH 2 ως προς mix *HeLa*).

Στη συνέχεια έγινε η αξιολόγηση των δεδομένων που προέκυψαν, δηλαδή το φιλτράρισμα των δεδομένων, βάσει δύο κριτηρίων. Το πρώτο σχετιζόταν με την τιμή του δυαδικού λογαρίθμου ( $\log_2$ ) του λόγου αφθονίας της εκάστοτε πρωτεΐνης στις 2 βιολογικές επαναλήψεις (δείγματα) των κυττάρων *HeLa* ως προς το κοινό δείγμα αναφοράς (mix *HeLa*). Όταν ο λόγος της αφθονίας των πρωτεϊνών ήταν μεγαλύτερος από 1.25 (αύξηση μεγαλύτερη κατά τουλάχιστον ~2,4 φορές) ή μικρότερος από 0.75 (μείωση κατά τουλάχιστον ~1,7 φορές), τότε οι αντίστοιχες πρωτεΐνες απομακρύνονταν. Το δεύτερο ήταν σχετικό με την τιμή της γεωμετρικής τυπικής απόκλισης (geometric standard deviation), δηλαδή της διαφοράς, από το μέσο όρο, των λόγων αφθονίας της εκάστοτε πρωτεΐνης στις 2 βιολογικές επαναλήψεις (δείγματα) των κυττάρων *HeLa* ως προς το κοινό δείγμα αναφοράς. Οι πρωτεΐνες που εμφάνιζαν διαφορά μεγαλύτερη από 1,25 απομακρύνονταν.

Για την εύρεση των πρωτεϊνών που παρουσιάζουν στατιστικά σημαντική αλλαγή (είτε θετική είτε αρνητική) στους εξεταζόμενους τροποποιημένους κλώνους ως προς τους κλώνους αναφοράς της κάθε κατάστασης, χρησιμοποιήθηκε η ανάλυση σημαντικότητας μικροσυστοιχιών (SAM). Η ανάλυση αυτή υπολογίζει τη παρατηρούμενη τιμή (observed score) της διαφοράς της αφθονίας μίας πρωτεΐνης σε δυο διαφορετικές καταστάσεις (π.χ. κλώνος υπερέκφρασης/αντίστοιχος κλώνος αναφοράς) και την αναμενόμενη τιμή (expected score) που προκύπτει από την τυχαία διακύμανση (random fluctuation), όταν δεν υπάρχει διαφορά στην αφθονία της πρωτεΐνης μεταξύ των δυο καταστάσεων. Όταν η διαφορά μεταξύ των δυο αυτών τιμών είναι μικρότερη από μία τιμή-κατώφλι, η οποία εξαρτάται από τον αριθμό των ψευδώς θετικών αποτελεσμάτων, τότε η πρωτεΐνη χαρακτηρίζεται ότι παρουσιάζει στατιστικά σημαντική αλλαγή (Tusher et al. 2001).

#### 3.4.4.3. Ανάλυση μεταβολικού προτύπου

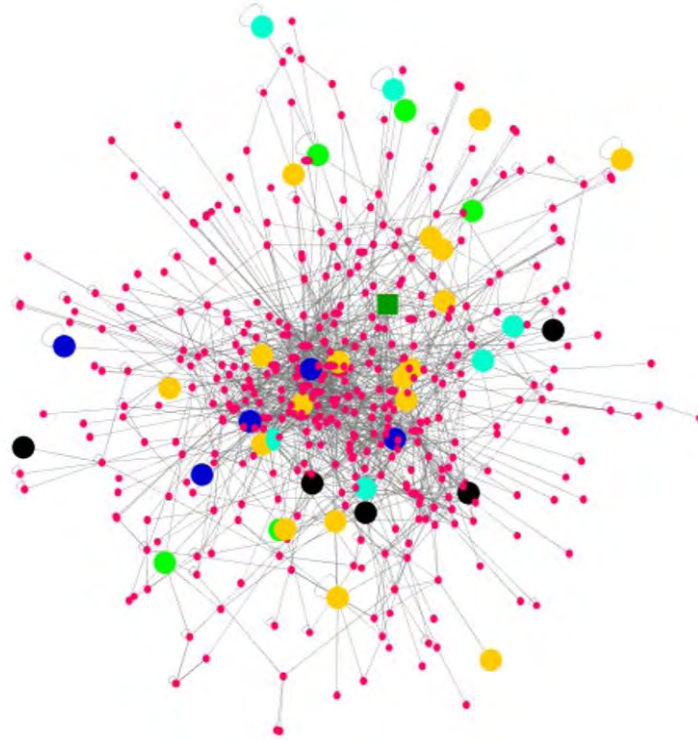
Για την μέτρηση του μεταβολικού προτύπου χρησιμοποιήθηκαν 6 δείγματα (βιολογικές επαναλήψεις) για καθένα από τους κλώνους υποέκφρασης shRNA(4)\_4, shRNA(3)\_5, shRNA(3)\_5 και για τον κλώνο αναφοράς (MSH 2) και 4 δείγματα (βιολογικές επαναλήψεις) για τα κύτταρα *HeLa* που δεν είχαν τροποποιηθεί.

Η μέτρηση του μεταβολικού προτύπου έγινε από την Δρ. Μ. Κλάπα στο συνεργαζόμενο εργαστήριο Μεταβολικής Μηχανικής και Βιολογίας Συστημάτων, Ινστιτούτο Χημικών Μηχανικών Επιστημών (Ίδρυμα Έρευνας και Τεχνολογίας, FORTH/ICE-HT, Πάτρα) με αέρια χρωματογραφία-φασματομετρία μάζας. Το μεταβολικό πρότυπο κάθε δείγματος μετρήθηκε τουλάχιστον τρεις φορές (3 τεχνικές επαναλήψεις) σε συνεχόμενες εκχύσεις, ώστε να πιστοποιηθεί ότι οι πειραματικές συνθήκες ποσοτικοποίησης μεταβολικού προτύπου παρέμεναν σταθερές. Η ταυτοποίηση των μεταβολιτών έγινε από το συνεργαζόμενο εργαστήριο χρησιμοποιώντας την βάση κορυφών NIST (<http://www.nist.gov/pml/data/asd.cfm>), τη βάση Human Metabolome (<http://www.hmdb.ca/>) και την εκτεταμένη βιβλιοθήκη κορυφών του συνεργαζόμενου εργαστηρίου, ενώ η ποσοτικοποίηση βασίστηκε στην επιφάνεια κορυφής του χαρακτηριστικού ιόντος κάθε παράγωγου μεταβολίτη. Στη συνέχεια πραγματοποιήθηκε η κανονικοποίηση και το φιλτράρισμα των μεταβολικών δεδομένων που ανακτήθηκαν για όλα τα δείγματα με βάση την υπολογιστική πλατφόρμα M-IOLITE (Maga-Nteve & Klara 2016) από το συνεργαζόμενο εργαστήριο. Η κανονικοποίηση είναι απαραίτητη για την εξάλειψη των διαφορών που προκύπτουν από πειραματικά σφάλματα με αποτέλεσμα την άμεση σύγκριση των δεδομένων μεταξύ τους. Ένας τύπος κανονικοποίησης αφορά την πραγματική συγκέντρωση των μεταβολιτών σε ένα δείγμα σε σύγκριση με την ποσότητα που μετρείται τελικά από το φασματοφωτόμετρο μάζας. Αυτό επιτυγχάνεται μέσω της εισαγωγής κατάλληλης ποσότητας εσωτερικών προτύπων (βλ. Ενότητα 3.4.1.1.), βάσει του πρωτεϊνικού περιεχομένου των δειγμάτων, ώστε τα μεταβολικά πρότυπα να είναι άμεσα συγκρίσιμα μεταξύ τους (Kanani & Klara 2007). Όταν πρόκειται για αέρια χρωματογραφία-φασματομετρία μάζας, πραγματοποιείται ένας δεύτερος τύπος κανονικοποίησης μέσω της παραγωγής των μεταβολιτών (βλ. Ενότητα 3.4.1.1.). Για να γίνει η ποσοτικοποίηση των μεταβολικών προτύπων θα πρέπει η ποσότητα των

παραγώγων που μετρούνται τελικά να είναι άμεσα ανάλογη με την αρχική συγκέντρωση του μεταβολίτη στο κάθε βιολογικό δείγμα. Μια τεχνική διόρθωσης των μεταβολικών δεδομένων έχει αναπτυχθεί από το συνεργαζόμενο εργαστήριο και ομαδοποιεί τους μεταβολίτες σε τρεις διαφορετικές κατηγορίες ανάλογα με τη χημική τους δομή και τα παράγωγα που σχηματίζουν, με στόχο τον ακριβή προσδιορισμό των πειραματικών αποκλίσεων λόγω της παραγωγίσης (Kanani & Klara 2007, Kanani et al. 2008). Το μέσο μεταβολικό πρότυπο κάθε κλώνου υπολογίστηκε και κανονικοποιήθηκε ως προς το μέσο μεταβολικό πρότυπο των κυττάρων *HeLa*. Η στατιστική ανάλυση που πραγματοποιήθηκε ήταν η ανάλυση σημαντικότητας μικροσυστοιχιών (SAM) η οποία χρησιμοποιείται για να ελεγχθούν ποια μόρια έχουν σημαντικά διαφορετική συγκέντρωση (αυξημένη ή μειωμένη) σε μια κατάσταση φυσιολογίας σε σύγκριση με κάποια άλλη (βλ. Ενότητα 3.4.4.2.). Έτσι, προσδιορίστηκαν οι μεταβολίτες που διαφοροποιούν την ομάδα των κυττάρων *HeLa* και του κλώνου αναφοράς της κατάστασης υποέκφρασης (MSH 2) από την ομάδα των κλώνων shRNA(4)\_4 και shRNA(3)\_5.

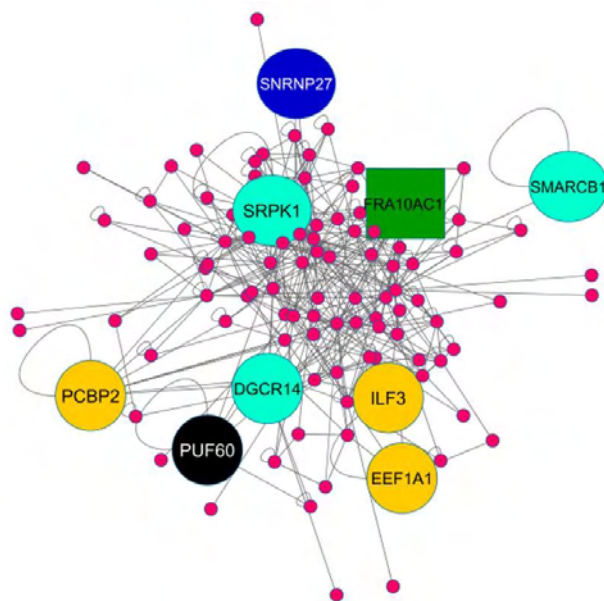
### **3.5. Πρωτεΐνες του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής με τροποποιημένη έκφραση σε πρωτεϊνικό και μεταγραφικό επίπεδο, μετά την καταστολή της έκφρασης του γονιδίου *FRA10AC1***

Χρησιμοποιώντας τα αποτελέσματα που προέκυψαν από τις ομικές αναλύσεις του κυτταρικού μοντέλου με τροποποιημένη την έκφραση του γονιδίου *FRA10AC1*, ακολούθησε η αντιστοίχιση των πρωτεϊνών του συνόλου δεδομένων, του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής και της γειτονιάς της *FRA10AC1* εντός του δικτύου με τα μόρια που εμφάνισαν αλλαγές στην έκφρασή τους σε μεταγραφικό και πρωτεϊνικό επίπεδο. Η σύγκριση αυτή έδειξε ότι 56 πρωτεΐνες του συνόλου δεδομένων του σωματίου συναρμογής εμφάνισαν αλλαγές στην έκφρασή τους, 19 σε πρωτεϊνικό, 29 μόρια σε μεταγραφικό επίπεδο, και 8 ήταν κοινά και στα δυο επίπεδα. Οι 39 από τις 56 πρωτεΐνες ενσωματώθηκαν στο δίκτυο αλληλεπιδράσεων του σωματίου συναρμογής (Εικ. 69). Ειδικότερα, στα πλαίσια του δικτύου, στα μόρια στα οποία παρατηρήθηκε αύξηση της έκφρασής τους σε επίπεδο μεταγράφου ήταν 18 (PCBP2, ILF3, DDX17, WBP11, CDC5L, DIDO1, RBMX, HNRNPU, EEF1A1, EXOSC4, NUP210, PIK3R3, SERBP1, EIF3L, RBBP4, CHD4, EHMT2, HIST1H2AM) και σε επίπεδο πρωτεΐνης 6 (SRPK1, DGCR14, SMARCB1, SMARCE1, VPRBP, HSPA1A) (Εικ. 66). Τα μόρια στα οποία παρατηρήθηκε μείωση της έκφρασής τους ήταν σε επίπεδο μεταγράφου 5 (PQBP1, GPKOW, SRSF4, SYNE1, SNRNP27) και σε επίπεδο πρωτεΐνης 5 (HSP90AA1, PPIL3, RBBP7, ANXA1, JUP) (Εικ. 66). Τα μόρια, η έκφραση των οποίων άλλαξε μετά την καταστολή της έκφρασης του *FRA10AC1* και σε επίπεδο μεταγράφου και πρωτεΐνης ήταν 5 (PUF60, BAG2, GEMIN5, SMARCC1, DHX30) (Εικ. 66).



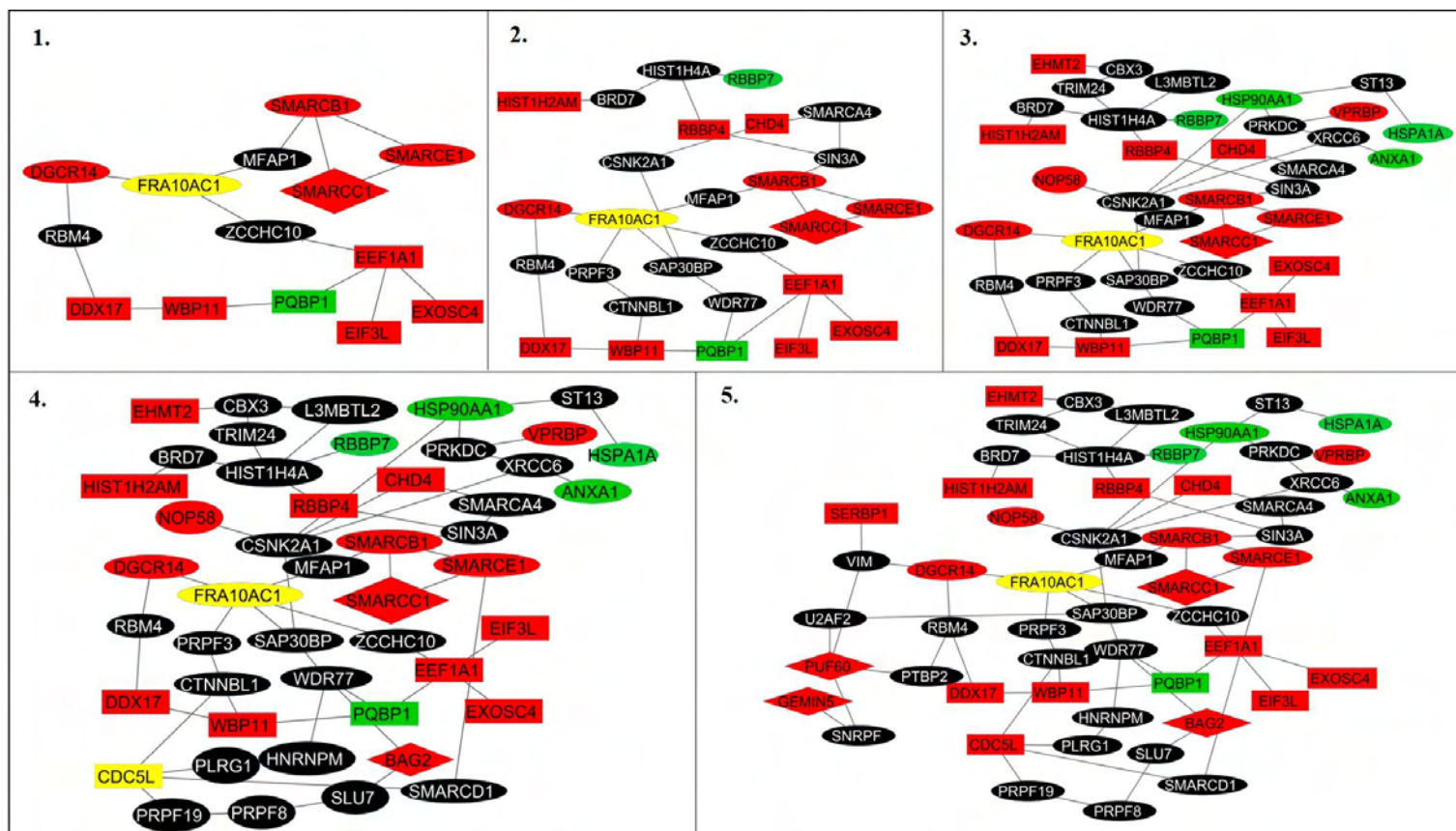
**Εικ. 66:** Πρωτεΐνες του δικτύου αλληλεπιδράσεων του σωματίου συναρμογής με τροποποιημένη έκφραση μετά την καταστολή της έκφρασης του *FRA10AC1*. Φαίνονται με μεγαλύτερου μεγέθους κυκλικό σχήμα οι πρωτεΐνες στις οποίες παρατηρήθηκε αύξηση και μείωση της έκφρασής τους σε μεταγραφικό (κίτρινο και μπλε χρώμα, αντίστοιχα) και σε πρωτεϊνικό επίπεδο (γαλάζιο και πράσινο χρώμα, αντίστοιχα). Οι κοινές πρωτεΐνες μεταξύ των δυο προτύπων που παρουσίασαν αυξημένη έκφραση απεικονίζονται με μαύρο χρώμα. Η *FRA10AC1* φαίνεται με πράσινο χρώμα σε σχήμα τετραγώνου.

Από τις 39 πρωτεΐνες του δικτύου που παρουσίασαν αλλαγές στην έκφρασή τους μετά την καταστολή της έκφρασης του γονιδίου *FRA10AC1*, η μία (*DGCR14*) αποτελεί πρώτο «γείτονα» της *FRA10AC1* και άλλες 7 είναι δεύτεροι «γείτονες» της (Εικ. 67). Στους 3 από τους 8 «γείτονες» παρατηρήθηκε αύξηση της έκφρασής τους σε μεταγραφικό επίπεδο (*EEF1A1*, *PCBP2*, *ILF3*) και σε άλλους 3 σε πρωτεϊνικό (*SMARCB1*, *SRPK1*, *DGCR14*) (Εικ. 67). Τα επίπεδα έκφρασης του γονιδίου που κωδικοποιεί την πρωτεΐνη *SNRNP27* μειώθηκαν σε επίπεδο μεταγράφου (Εικ. 67). Η έκφραση ενός δεύτερου «γείτονα» της *FRA10AC1* (*PUF60*) αυξήθηκε τόσο σε μεταγραφικό όσο και πρωτεϊνικό επίπεδο (Εικ. 67).

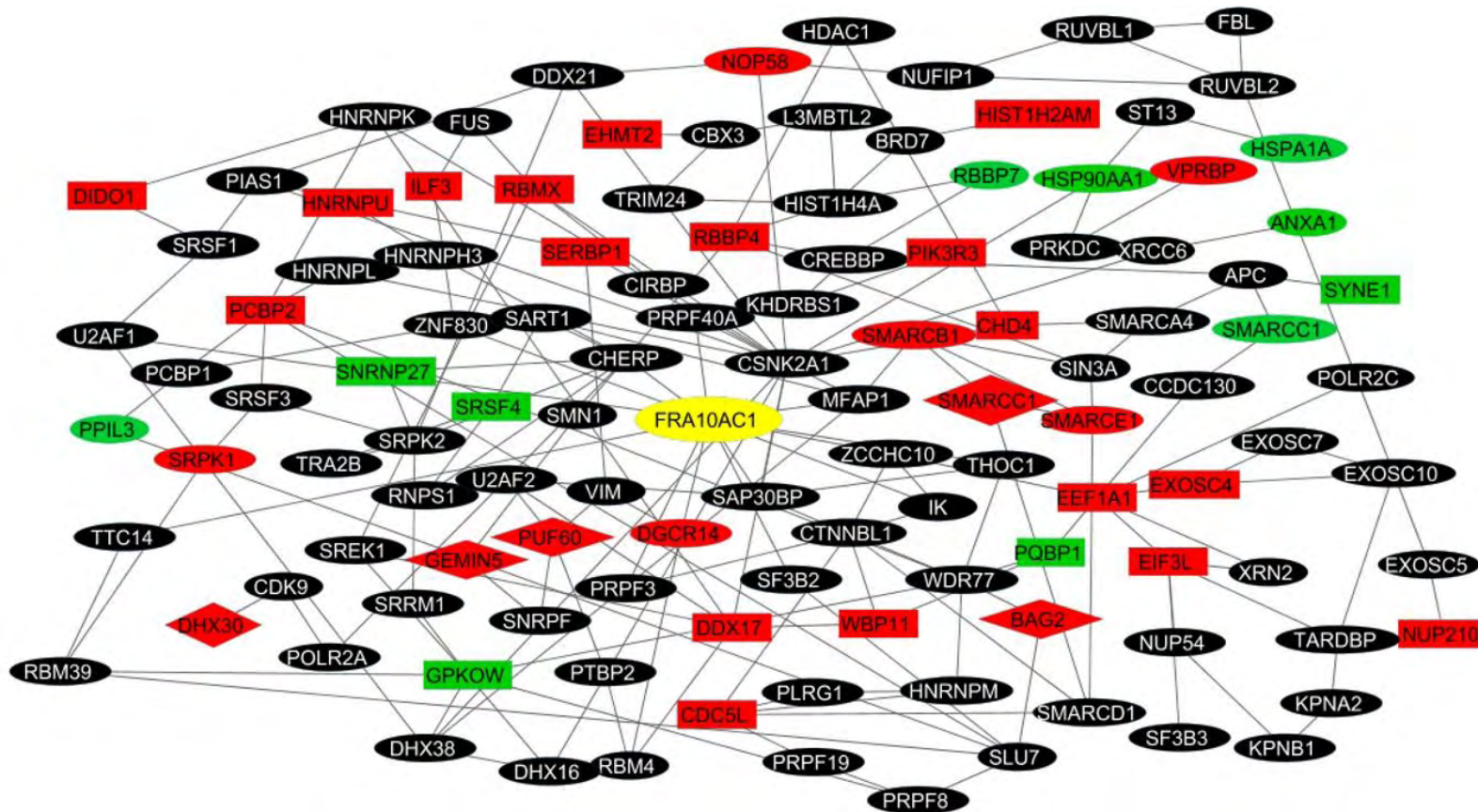


**Εικ. 67: Πρωτεΐνες της «γειτονιάς» της FRA10AC1 με τροποποιημένη έκφραση μετά την καταστολή της έκφρασής της.** Οι πρωτεΐνες στις οποίες παρατηρήθηκε αύξηση και μείωση της έκφρασής τους σε μεταγραφικό επίπεδο φαίνονται με κίτρινο και μπλε χρώμα, αντίστοιχα, και σε αυτές που παρατηρήθηκε αύξηση σε πρωτεϊνικό επίπεδο με γαλάζιο χρώμα. Η κοινή πρωτεΐνη μεταξύ των δυο προτύπων που παρουσίασε αυξημένη έκφραση απεικονίζεται με μαύρο χρώμα. Η FRA10AC1 φαίνεται με πράσινο χρώμα σε σχήμα τετραγώνου.

Στη συνέχεια, για να ελεγχθεί εάν η πρωτεΐνη FRA10AC1 αλληλεπιδρά μέσω πρωτεϊνών του δικτύου που δεν παρουσίασαν αλλαγή στην έκφρασή τους, εκτός από τις 8 πρωτεΐνες που βρίσκονται στη «γειτονιά» της, και με τα υπόλοιπα 31 μόρια που παρουσίασαν αύξηση ή μείωση στα επίπεδα έκφρασής τους είτε σε πρωτεϊνικό είτε σε μεταγραφικό επίπεδο, διερευνήθηκε το μικρότερο δυνατό «μονοπάτι» αλληλεπιδράσεων που συνδέει τη FRA10AC1 με τις πρωτεΐνες αυτές (Εικ. 68, 69). Παρατηρήθηκε ότι η FRA10AC1 αλληλεπιδρά, με τη μεσολάβηση άλλων πρωτεϊνών, με όλες τις 39 πρωτεΐνες που επηρεάστηκαν μετά την καταστολή της έκφρασής της (Εικ. 68, 69). Οι περισσότερες πρωτεΐνες που τοποθετούνται στη «κοντινή γειτονιά» της εμφάνισαν κατά κύριο λόγο αύξηση της έκφρασής τους. Αντίθετα, αυτές που παρουσίασαν μείωση, οι οποίες ήταν λιγότερες σε αριθμό, αλληλεπιδρούσαν με τη FRA10AC1 με τη μεσολάβηση μεγαλύτερου αριθμού πρωτεϊνών. Στις Εικόνες 68 και 69 φαίνονται κατ' αντιστοιχία 5 επιμέρους δίκτυα που δημιουργήθηκαν, ξεκινώντας από τη FRA10AC1 και ενσωματώντας σταδιακά τα μικρότερα δυνατά μονοπάτια (αλληλεπιδράσεις) που τη «συνδέουν» με τις 25 από τις 39 πρωτεΐνες του δικτύου του σωματίου συναρμογής που παρουσίασαν αλλαγή στην έκφρασή τους μετά την καταστολή της έκφρασης της FRA10AC1, και ένα τελικό δίκτυο που περιλαμβάνει τη «σύνδεση» της FRA10AC1 με όλες τις 39 πρωτεΐνες.



**Εικ. 68:** Πρωτεΐνες με τροποποιημένη έκφραση μετά την καταστολή της έκφρασης του γονιδίου *FRA10AC1* στην ευρύτερη «γειτονιά» της *FRA10AC1* εντός του δικτύου του σωματίου συναρμογής. Φαίνονται τα 5 επιμέρους δίκτυα που προέκυψαν μετά τη διαδοχική ενσωμάτωση των αλληλεπιδράσεων που «συνδέουν» τη *FRA10AC1* με τα 25 μόρια που παρουσίασαν αλλαγές στην έκφρασή τους σε μεταγραφικό και πρωτεϊνικό επίπεδο. Με κόκκινο ορθογώνιο και κυκλικό σχήμα φαίνονται τα μόρια που εμφάνισαν αυξημένη έκφραση σε επίπεδο μεταγράφου και πρωτεΐνης, αντίστοιχα, ενώ με πράσινο ορθογώνιο και κυκλικό σχήμα αυτά που παρουσίασαν μειωμένη έκφραση κατ' αντιστοιχία στα δυο επίπεδα. Με κόκκινο ρομβοειδές σχήμα απεικονίζονται οι κοινές πρωτεΐνες μεταξύ των δυο επιπέδων, η έκφραση των οποίων αυξήθηκε, και με κίτρινο κύκλο η *FRA10AC1*. Με μαύρο χρώμα φαίνονται οι πρωτεΐνες που δεν παρουσίασαν τροποποιημένη έκφραση.



Εικ. 69: Δίκτυο της ευρύτερης «γειτονιάς» της FRA10AC1 εντός του σωματίου συναρμογής με όλες τις 39 πρωτεΐνες που παρουσίασαν τροποποιημένη έκφραση μετά την καταστολή της έκφρασης του γονιδίου *FRA10AC1*. Ακολουθείται ο ίδιος χρωματικός κώδικας της Εικ. 68.

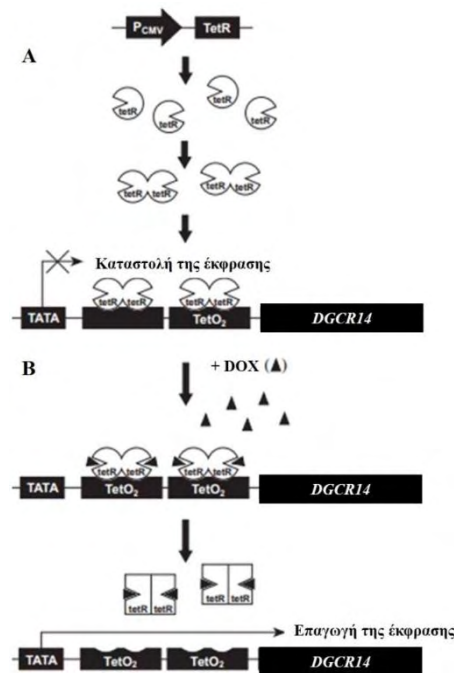


### 3.6. Δημιουργία κυτταρικής σειράς με επαγωγίμη έκφραση της πρωτεΐνης DGCR14

Βάσει της καλά μελετημένης αλληλεπίδρασης μεταξύ των FRA10AC1 και DGCR14 και των αποτελεσμάτων των ομικών αναλύσεων δημιουργήθηκε ένα κυτταρικό μοντέλο *HeLa* που υπερεκφράζει επαγωγίμη την πρωτεΐνη DGCR14. Το μοντέλο αυτό θα μπορούσε να χρησιμοποιηθεί για μια ολιστική διερεύνηση των αλλαγών που επιφέρει η υπερέκφραση του γονιδίου *DGCR14* είτε σε μεταγραφικό είτε σε πρωτεϊνικό ή μεταβολικό επίπεδο μέσω της χρήσης συνδυαστικών ομικών αναλύσεων. Μέσω αυτής της προσέγγισης θα μπορούσε να επιτευχθεί αφενός η μελέτη του λειτουργικού ρόλου της πρωτεΐνης DGCR14 και αφετέρου, εφόσον γίνεται προσομοίωση της αντίθετης κατάστασης που εξετάστηκε στην παρούσα εργασία, ο έλεγχος των επιδράσεων της υπερέκφρασης του DGCR14 στα επίπεδα έκφρασης του γονιδίου *FRA10AC1* ώστε να ενισχυθούν οι ενδείξεις που υπάρχουν για τη λειτουργική σχέση των δύο πρωτεϊνών.

Η δημιουργία της σταθερά διαμολυσμένης κυτταρικής σειράς *HeLa* περιλάμβανε το πλήρους μήκους cDNA του γονιδίου *DGCR14* του ανθρώπου υπό τον έλεγχο ενός συστήματος επαγωγίμης έκφρασης. Αυτό το σύστημα παρέχει τη δυνατότητα ενεργοποίησης της μεταγραφής ενός γονιδίου με την προσθήκη του αντιβιοτικού τετρακυκλίνη ή άλλων αναλόγων της όπως η δοξυκυκλίνη (Doxycycline, Dox) (Εικ. 70).

Η κυτταρική σειρά *HeLa* που δημιουργήθηκε φέρει ενσωματωμένο στο γονιδίωμα το πλασμίδιο pcDNA6/TR που κωδικοποιεί τον καταστολέα της τετρακυκλίνης (tetR) υπό την επίδραση του ισχυρού υποκινητή CMV, και το πλασμίδιο pcDNA4/TO στο οποίο έχει υποκλωνοποιηθεί το πλήρους μήκους cDNA του γονιδίου *DGCR14*. Το δεύτερο πλασμίδιο φέρει ρυθμιστικές αλληλουχίες στις οποίες προσδένεται ο tetR. Έτσι, παρουσία του TetR, η έκφραση του γονιδίου *DGCR14* καταστέλλεται. Αντίθετα, όταν προστεθεί εξωγενώς τετρακυκλίνη, ή κάποιο ανάλογό της (π.χ. Dox), αυτή δεσμεύει τον καταστολέα και έτσι επιτυγχάνεται η έκφραση του γονιδίου *DGCR14* (Εικ. 70).



**Εικ. 70: Επαγωγίμο σύστημα έκφρασης του γονιδίου *DGCR14*** (προσαρμοσμένη από το πρωτόκολλο της εταιρίας Invitrogen, [https://tools.thermofisher.com/content/sfs/manuals/trexssystem\\_man.pdf](https://tools.thermofisher.com/content/sfs/manuals/trexssystem_man.pdf)). Απουσία τετρακυκλίνης, ο TetR συνδέεται στις αλληλουχίες του χειριστή TetO<sub>2</sub>, εμποδίζοντας τη σύνδεση της πολυμεράσης του RNA καταστέλλοντας την έκφραση του γονιδίου *DGCR14* (A). Όταν προστεθεί το ανάλογο της τετρακυκλίνης, Dox, αυτό συνδέεται με τον TetR εμποδίζοντας τη σύνδεσή του με τον χειριστή TetO<sub>2</sub>. Έτσι, η πολυμεράση του RNA συνδέεται στον υποκινητή και πραγματοποιείται η έκφραση του *DGCR14* (B).

### 3.6.1. Δημιουργία σταθερά διαμολυσμένης κυτταρικής σειράς *HeLa* που εκφράζει τον καταστολέα της τετρακυκλίνης

Για τη δημιουργία του επαγωγίμου συστήματος έκφρασης του γονιδίου *DGCR14*, αρχικά πραγματοποιήθηκε η δημιουργία μιας κυτταρικής σειράς *HeLa* που εκφράζει το γονίδιο του καταστολέα της τετρακυκλίνης, TetR. Σε πρώτο στάδιο ελέγχθηκε η ελάχιστη συγκέντρωση του αντιβιοτικού βλαστισιδίνη (γονίδιο ανθεκτικότητας του φορέα σε κύτταρα θηλαστικών) στην οποία πεθαίνουν όλα τα κύτταρα. Χρησιμοποιήθηκαν τρεις διαφορετικές συγκεντρώσεις βλαστισιδίνης (10, 15 και 20 µg/ml) σε κύτταρα *HeLa* τα οποία καλλιεργήθηκαν σε τρυβλία 6 θέσεων για δυο εβδομάδες. Σε αυτό το διάστημα η αλλαγή του πλήρους θρεπτικού μέσου με το αντιβιοτικό βλαστισιδίνη γινόταν κάθε τρεις ημέρες. Η συγκέντρωση βλαστισιδίνης που επιλέχθηκε ήταν 10 µg/ml. Ακολούθησε η διαμόλυνση των κυττάρων *HeLa* με τον γραμμικό πλασμιδιακό φορέα pcDNA6/TR, έπειτα από την πέψη του με το ένζυμο περιορισμού *AhdI*, που τέμνει στη θέση 5745bp, στο γονίδιο της αμικιλίνης. Η διαμόλυνση έγινε με συνκατακρήμνιση του DNA με Ca<sub>3</sub>(PO<sub>4</sub>)<sub>2</sub>, 24h μετά την επίστρωση των δυο τρυβλίων (διαμέτρου 10cm) τα οποία περιείχαν 1,2x10<sup>6</sup> κύτταρα το καθένα. Η ποσότητα του γραμμοποιημένου DNA που χρησιμοποιήθηκε για κάθε τρυβλίο ήταν 20µg. Η διαδικασία που ακολουθήθηκε ήταν ίδια με αυτή που αναφέρθηκε στην Ενότητα 3.4.2.. Συνολικά επιλέχθηκαν και αναπτύχθηκαν 5 κλώνοι (#1, 3, 4, 5, 24).

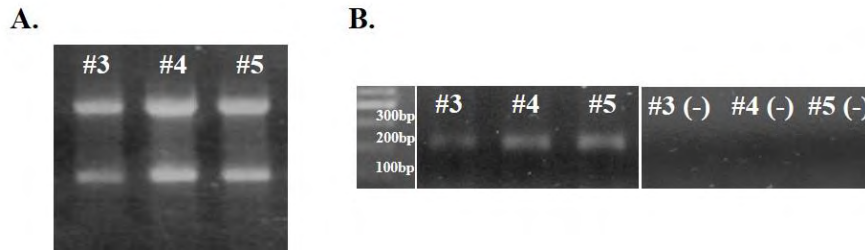
Για την επιβεβαίωση της έκφρασης του γονιδίου *TetR* εφαρμόστηκε RT-PCR σε τρεις από τους παραπάνω κλώνους. Έτσι, σε πρώτο στάδιο απομονώθηκε ολικό RNA από τους κλώνους #3, 4 και 5), μέσω του αντιδραστηρίου TRI (βλ. Υλικά & Μέθοδοι), και η ποιότητά του ελέγχθηκε με ηλεκτροφόρηση σε πήκτωμα αγαρόζης 1,5% (Εικ. 71 Α). Στην ηλεκτροφόρηση διακρίνονται οι δύο χαρακτηριστικές ζώνες του ριβοσωμικού RNA, γεγονός που υποδεικνύει ότι το RNA ήταν ακέραιο και ότι μπορούσε να χρησιμοποιηθεί περαιτέρω για RT-PCR (Εικ. 71 Α). Στη συνέχεια ακολούθησε η επεξεργασία του RNA με το ένζυμο DNAση διότι κατά τη διαδικασία απομόνωσης του RNA, συν-απομονώνεται και κάποια ποσότητα γενωμικού DNA. Αυτή η επεξεργασία είναι απαραίτητη επειδή το γενωμικό DNA θα μπορούσε να δώσει ψευδώς θετικό σήμα κατά την PCR που εφαρμόστηκε στο προϊόν της αντίστροφης μεταγραφής. Η ποσοτικοποίηση του RNA έγινε με φωτομέτρηση και οι αντίστοιχες τιμές φαίνονται στον Πίνακα 24. Η ποσότητα του RNA των τριών κλώνων ήταν ιδιαίτερα ικανοποιητική και η καθαρότητα του (λόγος 260nm/280nm) κυμαινόταν στα επιθυμητά όρια (1,8-2) (Πιν. 24).

**Πιν. 24: Ποσοτικοποίηση του RNA των κλώνων #3, #4 και #5, μετά από την επεξεργασία με DNAση**

Κλώνοι	Συγκέντρωση (ng/μl)	Συνολική συγκέντρωση RNA (μg)	Λόγος 260nm/280nm
#3	1433,7	57,3	1,81
#4	2138,7	85,5	1,84
#5	1614,7	64,5	1,79

Στο επόμενο στάδιο πραγματοποιήθηκε η σύνθεση cDNA με αντίστροφη μεταγραφή με υπόστρωμα το ολικό RNA και με χρήση τυχαίων εξαμερών εκκινητών. Ακολούθησε PCR με ζεύγος εκκινητών (F1, R1) για το γονίδιο *TetR*. Η βέλτιστη θερμοκρασία υβριδοποίησης των εκκινητών προσδιορίστηκε με PCR διαβάθμισης θερμοκρασίας, με υπόστρωμα πλασμιδιακό DNA του φορέα pcDNA/6TR που περιέχει το γονίδιο *TetR*. Το αποτέλεσμα της PCR ήταν συγκρίσιμο σε όλες τις θερμοκρασίες που ελέγχθηκαν (55-62°C), οπότε επιλέχθηκε η υψηλότερη θερμοκρασία καθώς αντιστοιχεί στις πιο αυστηρές συνθήκες.

Στη συνέχεια πραγματοποιήθηκε PCR με υπόστρωμα τα προϊόντα της αντίστροφης μεταγραφής (Εικ. 71 Β). Τα προϊόντα της PCR είχαν μέγεθος 209bp. Ως αρνητικός έλεγχος χρησιμοποιήθηκε για κάθε κλώνο το αντίστοιχο RNA, το οποίο ενισχύθηκε παράλληλα (Εικ. 71 Β). Από το αποτέλεσμα αποδεικνύεται η έκφραση του καταστολέα της τετρακυκλίνης σε όλους τους κλώνους (Εικ. 71 Β).



**Εικ. 71:** Ηλεκτροφόρηση ολικού RNA από τους κλώνους κυττάρων *HeLa* #3, 4 και 5 (A) και προϊόντα PCR του cDNA του γονιδίου *TetR* για τους κλώνους αυτούς (B). Οι δύο έντονες ζώνες που φαίνονται σε κάθε δείγμα αντιστοιχούν στο ριβοσωμικό RNA (A). Φαίνονται τα προϊόντα PCR (209bp) και οι αντίστοιχοι αρνητικοί έλεγχοι του πειράματος (RNA). Αριστερά απεικονίζεται ο δείκτης μοριακών μεγεθών (Cat. No 15628019, Thermo scientific) (B).

Ο κλώνος #5 φαίνεται να εμφανίζει υψηλότερα επίπεδα έκφρασης του καταστολέα της τετρακυκλίνης (*TetR*), αν θεωρήσουμε ότι η απόδοση της αντίστροφης μεταγραφής και της PCR είναι συγκρίσιμη για τους τρεις κλώνους. Έτσι, ο κλώνος αυτός χρησιμοποιήθηκε για τη δημιουργία της κυτταρικής σειράς *HeLa* με επαγωγίμη έκφραση του γονιδίου *DGCR14*.

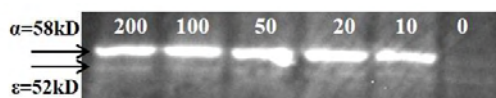
### 3.6.2. Δημιουργία κυτταρικής σειράς με επαγωγίμη έκφραση του γονιδίου *DGCR14*

Ο κλώνος #5 που εκφράζει τον *TetR* διαμολύνθηκε με το διαγονίδιο *pcDNA4/TO-DGCR14* που γραμμοποιήθηκε με το ένζυμο *PvuI* το οποίο δεν διαταράσσει ρυθμιστικά στοιχεία έκφρασης του πλασμιδιακού φορέα. Η διαμόλυνση, που έγινε με συνκατακρήμιση του DNA με  $Ca_3(PO_4)_2$ , πραγματοποιήθηκε 24h μετά την επίστρωση δυο τρυβλίων (διαμέτρου 10cm) τα οποία περιείχαν  $1,2 \times 10^6$  κύτταρα το καθένα. Η ποσότητα του DNA που χρησιμοποιήθηκε για κάθε τρυβλίο ήταν 20μg. Η διαδικασία που ακολουθήθηκε ήταν ίδια με αυτή που αναφέρθηκε στην Ενότητα 3.4.2. η οποία διέφερε μόνο στο στάδιο της θρυψινοποίησης των κυττάρων, κατά την οποία τα κύτταρα κάθε τρυβλίου αραιώθηκαν σε 4 τρυβλία. Στα τρυβλία προστέθηκαν διαφορετικές συγκεντρώσεις του αντιβιοτικού ζεοσίνη (350 μg/ml, 500 μg/ml, 650 μg/ml και 800 μg/ml) που είχαν επιλεγθεί (βλ. Υλικά & Μέθοδοι). Το πλήρες θρεπτικό μέσο στο οποίο καλλιεργήθηκαν τα κύτταρα περιείχε επιπλέον και 10 μg/ml βλαστισιδίνη (γονίδιο ανθεκτικότητας του φορέα *pcDNA6/TR*) διότι ο κλώνος #5 που επιλέχθηκε μπορεί να μην ήταν 100% αμιγής, δηλαδή μπορεί να υπήρχαν και κάποια μη διαμολυσμένα κύτταρα τα οποία θα έπρεπε να μην επιβιώσουν. Η συλλογή και η ανάπτυξη των αποικιών πραγματοποιήθηκε με το ίδιο τρόπο που αναφέρθηκε και στην Ενότητα 3.4.2.. Συνολικά συλλέχθηκαν, από όλα τα τρυβλία που περιείχαν τις τέσσερις διαφορετικές συγκεντρώσεις του αντιβιοτικού ζεοσίνη, 40 αποικίες.

Για τον έλεγχο της επαγωγίμης υπερέκφρασης του γονιδίου *DGCR14* στις παραπάνω αποικίες χρησιμοποιήθηκε το αντιβιοτικό δοξυκυκλίνη, σε συγκέντρωση 200 ng/ml για 24 ώρες. Ο έλεγχος των επιπέδων έκφρασης τόσο της ανασυνδυασμένης όσο και της ενδογενούς πρωτεΐνης *DGCR14* έγινε με ανάλυση Western (δεν παρατίθεται αποτελέσματα). Από τους παραπάνω κλώνους, δυο (#9 και #37) εμφάνισαν υψηλότερα ποσοστά υπερέκφρασης της πρωτεΐνης *DGCR14*. Οι κλώνοι αυτοί είχαν επιλεγθεί από το τρυβλίο στο οποίο τα κύτταρα είχαν καλλιεργηθεί σε συγκέντρωση 650μg/ml του

αντιβιοτικού ζεοσίνη. Συγκριτικά, υψηλότερα επίπεδα υπερέκφρασης εμφάνισε ο κλώνος #37 ο οποίος μελετήθηκε περαιτέρω για τον προσδιορισμό της κατάλληλης συγκέντρωσης δοξυκυκλίνης και του χρόνου επαγωγής της έκφρασης.

Για να προσδιοριστεί η ελάχιστη συγκέντρωση δοξυκυκλίνης που θα μπορεί να επάγει τα μέγιστα επίπεδα έκφρασης της ανασυνδυσμένης DGCR14, αποφεύγοντας το κορεσμό του συστήματος υπερέκφρασης και την άσκοπη κατανάλωση του αντιβιοτικού, πραγματοποιήθηκε μια προκαταρκτική δοκιμή με πέντε διαφορετικές συγκεντρώσεις Dox της τάξης των 10ng/ml, 20ng/ml, 50ng/ml, 100ng/ml και 200ng/ml στον ίδιο χρόνο επαγωγής της έκφρασης των 24 ωρών (Εικ. 72). Ο έλεγχος της έκφρασης πραγματοποιήθηκε με ανάλυση Western, με τη χρήση αντισώματος  $\alpha$ -DGCR14 (Εικ. 72). Συγκεκριμένα χρησιμοποιήθηκε ορός που προέρχεται από κουνέλι ανοσοποιημένο στην πρωτεΐνη DGCR14 σε αραιώση 1:10000. Η ενδογενής πρωτεΐνη DGCR14 έχει μοριακό βάρος 52kD ενώ το μέγεθος της ανασυνδυσμένης αυξάνεται κατά 6kD λόγω των επιτόπων (Xpress και His) και της ακολουθίας που βρίσκεται αμινοτελικά και κωδικοποιείται από τον φορέα pcDNA 3.1 His από τον οποίο αποκόπηκε το cDNA του γονιδίου *DGCR14* για τη μετέπειτα υποκλώνοποίησή του στο φορέα pcDNA4/TO.



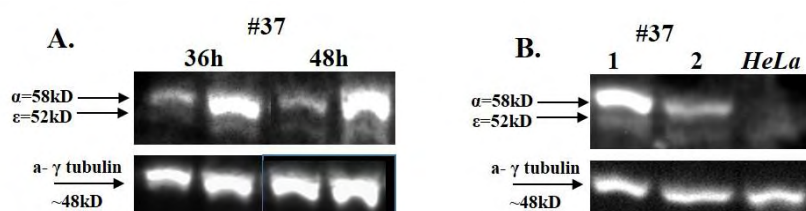
**Εικ. 72: Ανάλυση western των επιπέδων υπερέκφρασης της ανασυνδυσμένης πρωτεΐνης DGCR14 έπειτα από επαγωγή της έκφρασης με δοξυκυκλίνη για 24h.** Χορήγηση διαφορετικών συγκεντρώσεων δοξυκυκλίνης 0, 10, 20, 50, 100 και 200ng/ml στον κλώνο #37 για 24h. α: ανασυνδυσμένη πρωτεΐνη DGCR14, ε: ενδογενής πρωτεΐνη.

Από την ανάλυση western, και συγκρίνοντας τα επίπεδα έκφρασης της ανασυνδυσμένης και της ενδογενούς πρωτεΐνης (Εικ. 72) με το πρόγραμμα Image J, φάνηκε πως η υπερέκφραση της ανασυνδυσμένης πρωτεΐνης DGCR14 ήταν τόσο υψηλή που δεν ήταν ανιχνεύσιμη η ενδογενής (Εικ. 72). Ωστόσο, οι διαφορετικές συγκεντρώσεις Dox δεν οδηγούσαν σε κάποια σαφή διαφορά στα επίπεδα υπερέκφρασή της. Αυτό πιθανόν οφείλεται στο γεγονός ότι οι συνιστώμενες συγκεντρώσεις της δοξυκυκλίνης ήταν τόσο μεγάλες ώστε να προκαλούν κορεσμό του συστήματος υπερέκφρασης.

Γι' αυτό το λόγο έγινε η δοκιμή αρκετά μικρότερων συγκεντρώσεων δοξυκυκλίνης (0,5ng/ml και 0,001ng/ml) καθώς και διαφορετικών χρόνων επαγωγής (36 και 48h). Κάνοντας την ποσοτικοποίηση των επιπέδων υπερέκφρασης της πρωτεΐνης *DGCR14* σε σχέση με την ενδογενή πρωτεΐνη, με τον τρόπο που αναφέρθηκε παραπάνω, παρατηρήθηκε ότι η ανασυνδυσμένη πρωτεΐνη υπερεκφράζεται κατά 2,4 φορές στα κύτταρα στα οποία προστέθηκε 0,5ng/ml δοξυκυκλίνης για 48 ώρες και κατά 4 φορές στις 36 ώρες (Εικ. 73 Α). Από την ανάλυση Western φάνηκε πως η υπερέκφραση είναι μεγαλύτερη όταν χορηγείται δοξυκυκλίνη σε μεγαλύτερη συγκέντρωση (0,5ng/ml) και σε μικρότερο χρόνο. Η αύξηση της έκφρασης σε μικρότερο χρόνο πιθανώς οφείλεται

στο γεγονός ότι, με το πέρασμα του χρόνου, η δοξυκυκλίνη δεσμεύεται από τον καταστολέα και πιθανώς μειώνεται η σταθερότητά της με αποτέλεσμα τα μόρια της δοξυκυκλίνης που απομένουν να μην είναι ικανά να καταστείλουν την έκφραση του καταστολέα. Με αυτόν τον τρόπο, τα μόρια του καταστολέα είναι ελεύθερα και εμποδίζουν την έκφραση του γονιδίου.

Για τον έλεγχο της υπερέκφρασης σε μεγαλύτερες συγκεντρώσεις της δοξυκυκλίνης αλλά σε μικρότερο χρόνο επαγωγής, δοκιμάστηκε μία ενδιάμεση, από τις δυο συγκεντρώσεις που αναφέρθηκαν παραπάνω, δηλαδή 0,05ng/ml σε σύγκριση με την 0,001ng/ml που είχε χρησιμοποιηθεί και στην προηγούμενη δοκιμή. Το χρονικό διάστημα επαγωγής ήταν 36 ώρες. Συγκρίνοντας τα επίπεδα υπερέκφρασης της ανασυνδυασμένης πρωτεΐνης DGCR14 σε σχέση με την ενδογενή στις δυο διαφορετικές συγκεντρώσεις παρατηρήθηκε ότι η υπερέκφραση είναι μεγαλύτερη κατά 2,3 φορές όταν χορηγείται δοξυκυκλίνη σε μεγαλύτερη συγκέντρωση (0,05ng/ml) (Εικ. 73 B). Τα κύτταρα *HeLa* χρησιμοποιούνται σαν αρνητικός μάρτυρας και φαίνεται να εκφράζουν μόνο την ενδογενή πρωτεΐνη DGCR14 (Εικ. 73 B).



**Εικ. 73: Ανάλυση western των επιπέδων υπερέκφρασης της ανασυνδυασμένης πρωτεΐνης DGCR14 έπειτα από επαγωγή της έκφρασης με δοξυκυκλίνη.** Χορήγηση των συγκεντρώσεων δοξυκυκλίνης 0,001ng/ml (1) και 0,5ng/ml (2) για 36 και 48 ώρες (A) και: 0,001ng/ml (1) και 0,5ng/ml (2) για 36 ώρες (B). #37: κλώνος με επαγωγίμη έκφραση του γονιδίου *DGCR14*. α: ανασυνδυασμένη πρωτεΐνη DGCR14, ε: ενδογενής πρωτεΐνη. Τα αγρίου τύπου κύτταρα *HeLa* και η γ-τουμπουλίνη χρησιμοποιήθηκαν ως αρνητικός και εσωτερικός μάρτυρας (α-γ-τουμπουλίνης 1:2500), αντίστοιχα.

## **4. ΣΥΖΗΤΗΣΗ**

### **4.1. Το δίκτυο της FRA10AC1 περιέχει άμεσες και αξιόπιστες πρωτεϊνικές αλληλεπιδράσεις που υποδεικνύουν την συμμετοχή της στις διαδικασίες επεξεργασίας του mRNA**

Το δίκτυο πρωτεϊνικών αλληλεπιδράσεων της πρωτεΐνης FRA10AC1, περιλαμβάνει αποκλειστικά άμεσες αλληλεπιδράσεις. Έως τώρα υπάρχουν πληροφορίες για την ένταξη της FRA10AC1 εντός πρωτεϊνικών συμπλόκων οι οποίες έχουν προκύψει: α) από μια πρόσφατη μελέτη μεγάλης κλίμακας (Huttlin et al. 2017) που περιλαμβάνει πειράματα χρωματογραφίας συγγένειας με πρωτεΐνες-«δολώματα» σε συνδυασμό με φασματομετρία μάζας με την οποία η FRA10AC1 ταυτοποιήθηκε ως συστατικό 7 διαφορετικών πρωτεϊνικών συμπλόκων εκ των οποίων τα 2 είναι εμπλουτισμένα με πρωτεΐνες που σχετίζονται με την συναρμογή και την μεταγραφή (βλ. Ενότητα 3.1.3.), και β) από βιοχημικές μελέτες κατά τις οποίες η FRA10AC1 ταυτοποιήθηκε ως συστατικό του σωματίου συναρμογής και των υποσυμπλόκων του, B *act*, C και P (βλ. Ενότητα 1.3.1.). Η ταυτοποίηση της FRA10AC1 στα παραπάνω σύμπλοκα υποδεικνύει την ύπαρξη έμμεσων αλληλεπιδράσεων μεταξύ αυτής και των υπολοίπων συστατικών των συμπλόκων και την εντάσσει σε ευρείες βιολογικές διαδικασίες, όπως για παράδειγμα, στην επεξεργασία του mRNA. Ωστόσο, για την αποσαφήνιση συγκεκριμένου βιολογικού ρόλου της FRA10AC1 εντός των ευρύτερων βιολογικών διαδικασιών που έχουν υποδειχθεί με βάση τις πληροφορίες της ανάλυσης των πρωτεϊνικών συμπλόκων που συμμετέχει, απαιτείται ο εντοπισμός των πρωτεϊνών με τις οποίες αλληλεπιδρά άμεσα, δηλαδή των πρώτων «γειτόνων» της. Εφόσον αυτές οι πρωτεΐνες είναι λειτουργικά χαρακτηρισμένες, μπορούν να συνεισφέρουν στην εξαγωγή βιολογικών συμπερασμάτων σχετικά με τον ρόλο της FRA10AC1, να την εντάξουν σε μία πιο ειδική διεργασία στο πλαίσιο, για παράδειγμα, της επεξεργασίας του mRNA και να υποδείξουν εστιασμένες πειραματικές διαδικασίες επιβεβαίωσης υποθέσεων. Για αυτούς τους λόγους ανακτήθηκαν αποκλειστικά οι άμεσες αλληλεπιδράσεις της FRA10AC1 από βάσεις δεδομένων που στηρίζονται στη βιβλιογραφία τόσο στον άνθρωπο όσο και σε οργανισμούς μοντέλα. Συγκεκριμένα οι αλληλεπιδράσεις ορίστηκαν ως άμεσες βάσει της πειραματικής διαδικασίας που εφαρμόστηκε και προήλθαν από πειραματικές μελέτες τόσο μεγάλης όσο και μικρής κλίμακας (βλ. Ενότητα 3.1.2.). Έτσι, ταυτοποιήθηκαν 27 άμεσα αλληλεπιδρώντα μόρια υποδεικνύοντας μια πρωτεΐνη με υψηλή συνδεσιμότητα.

Ωστόσο, τα πειράματα μεγάλης κλίμακας έχουν κάποιους περιορισμούς (Snider et al. 2015, Yu & Finley 2009), όπως είναι ο αυξημένος αριθμός των ψευδώς θετικών αποτελεσμάτων τα οποία προκύπτουν λόγω του μεγάλου αριθμού των εξεταζόμενων πρωτεϊνών και των εγγενών χαρακτηριστικών των μεθόδων (βλ. Ενότητα 1.5.1.3.). Γι' αυτό το λόγο, τέθηκαν ορισμένα κριτήρια αξιολόγησης των αλληλεπιδράσεων και των αντίστοιχων πρωτεϊνών τα οποία αποτελούν ένα ισχυρό εργαλείο πιστοποίησης των αποτελεσμάτων από μελέτες μεγάλης κλίμακας (Hayes et al. 2016, Figeys 2008) (βλ. Ενότητα 1.5.1.3.).

Ειδικότερα, η αξιοπιστία των αλληλεπιδράσεων μεταξύ της FRA10AC1 και των πρώτων «γειτόνων» της αλλά και μεταξύ των αλληλεπιδρώντων μορίων της ελέγχθηκε βάσει του αριθμού και της κλίμακας των πειραμάτων τα οποία υποστηρίζουν την

εκάστοτε αλληλεπίδραση και της εξελικτικής συντήρησής τους (βλ. Ενότητα 3.1.2.). Αναλυτικότερα, παρατηρήθηκε ότι η αξιοπιστία των αλληλεπιδράσεων του δικτύου είναι σχετικά αυξημένη, καθώς το 40% αυτών υποστηρίζονται από πειραματικές μελέτες μικρής κλίμακας (π.χ. συν-ανοσοκατακρήμιση απομονωμένων ανασυνδυασμένων πρωτεϊνών). Δυο αλληλεπιδράσεις (FRA10AC1-DGCR14, FRA10AC1-SF3B2) ταυτοποιήθηκαν και με συμπληρωματικές πειραματικές προσεγγίσεις, τη συν-κατακρήμιση και τη δοκιμή με το σύστημα δυο υβριδίων του σακχαρομύκητα για μόνο ένα ζεύγος πρωτεϊνών. Μία από αυτές τις αλληλεπιδράσεις (FRA10AC1-DGCR14) ταυτοποιήθηκε και στο πλαίσιο συμπλόκου. Αυτό το σύμπλοκο που συμμετείχε η FRA10AC1, αλληλεπιδρώντας έμμεσα με την DGCR14, ήταν εμπλουτισμένο με πρωτεΐνες που συμμετέχουν στις διαδικασίες επεξεργασίας του mRNA και κατά κύριο λόγο στη συναρμογή. Αξιοσημείωτο είναι ότι, για πρώτη φορά, η FRA10AC1 βρέθηκε να εντάσσεται σε ένα σύμπλοκο με άλλες πρωτεΐνες του σωματίου συναρμογής σε διαφορετική κυτταρική σειρά (HEK293) από τα κύτταρα *HeLa* στα οποία έχει ταυτοποιηθεί ως συστατικό του σωματίου συναρμογής και των υποσυμπλόκων του. Την αξιοπιστία της τελευταίας αλληλεπίδρασης ενισχύει η ταυτοποίησή της και σε οργανισμούς-μοντέλα (βλ. Ενότητα 3.1.2.).

Η αξιολόγηση των αλληλεπιδρώντων μορίων έγινε μέσω της εφαρμογής τριών κριτηρίων: την ένταξη σε συγκεκριμένο πρωτεϊνικό σύμπλοκο, τη λειτουργική κατηγοριοποίηση και τον υποκυτταρικό εντοπισμό τους. Σύμφωνα με αυτά, περισσότερα από τα μισά αλληλεπιδρώντα μόρια (15) έχουν ταυτοποιηθεί ως συστατικά του μεγαλύτερου ριβονουκλεοπρωτεϊνικού συμπλόκου, του σωματίου συναρμογής, σε πρωτεομικές μελέτες απομόνωσης των συστατικών του. Επιπρόσθετα, 13 από τα 15 παραπάνω αλληλεπιδρώντα μόρια (εκτός των TTC14, ZCCHC10), και δυο επιπλέον, που δεν αποτελούν συστατικά του σωματίου συναρμογής (HABP4 και EEF1D), συμμετέχουν σε διαδικασίες επεξεργασίας του mRNA, κατά κύριο λόγο στη συναρμογή, και σε μικρότερο ποσοστό στη μεταγραφή και στην έξοδο του mRNA από τον πυρήνα (βλ. Ενότητα 3.1.2.). Για τους μηχανισμούς αυτούς είναι γνωστό ότι υπάρχει λειτουργική σύνδεση μεταξύ τους, η οποία είναι απαραίτητη για τη ρύθμιση της γονιδιακής έκφρασης (Montes et al. 2012). Όλα τα αλληλεπιδρώντα μόρια που αναφέρθηκαν παραπάνω εντοπίζονται αποκλειστικά στον πυρήνα ή/και σε αυτόν.

Για τα υπόλοιπα αλληλεπιδρώντα μόρια της FRA10AC1 δεν ήταν δυνατόν να γίνει η ομαδοποίησή τους βάσει της συμμετοχής τους σε ένα λειτουργικά μελετημένο πρωτεϊνικό σύμπλοκο ή σε μία κοινή βιολογική διαδικασία. Ειδικότερα, δυο πρωτεΐνες συμμετέχουν σε σηματοδοτικά μονοπάτια (MAP3K12, GLRX3), μία στο μονοπάτι αποικοδόμησης της πρωτεϊνικής κινάσης C μέσω ουβικουιλίνης (TRIM41), δύο στην προσθήκη ή αφαίρεση μετα-μεταφραστικών τροποποιήσεων (ADPRHL2, VPS29), δύο σε λειτουργίες του κυτταροσκελετού (CCDC155, GPD2) και τρεις για τις οποίες δεν υπάρχει σχετική πληροφορία με το λειτουργικό τους ρόλο (C7orf25, PLPPR4, GATD3A). Οι πρωτεΐνες MAP3K12, GLRX3, TRIM41, ADPRHL2, C7orf25 εντοπίζονται στον πυρήνα ή/και σε αυτόν. Για τις υπόλοιπες είτε δεν υπάρχει διαθέσιμη πληροφορία είτε δεν εμφανίζουν πυρηνικό εντοπισμό.

Έτσι, το ανακατασκευασμένο δίκτυο της FRA10AC1 το οποίο περιέχει 28 πρωτεΐνες και 55 άμεσες πρωτεϊνικές αλληλεπιδράσεις (βλ. Εικ. 32) χαρακτηρίζεται από



αυξημένη αξιοπιστία καθώς περίπου οι μισές (48%) από τις αλληλεπιδράσεις της FRA10AC1 και το 32% των αλληλεπιδράσεων μεταξύ των πρώτων «γειτόνων» της υποστηρίζονται από επιπρόσθετες μελέτες μικρής κλίμακας. Ακόμη, το 81% των αλληλεπιδρώντων μορίων εμφανίζουν πυρηνικό εντοπισμό γεγονός το οποίο συνάδει με την υποκυτταρική τοπολογία της FRA10AC1.

Περισσότερα από τα μισά αλληλεπιδρώντα μόρια (55%) της FRA10AC1 έχουν ταυτοποιηθεί ως συστατικά του σωματίου συναρμογής και συμμετέχουν σε διαδικασίες επεξεργασίας του mRNA, κατά κύριο λόγο στη συναρμογή, γεγονός που υποδεικνύει την ένταξη της πρωτεΐνης FRA10AC1 στο μοριακό μηχανισμό της συναρμογής του mRNA ή, με την ευρύτερη έννοια, στην επεξεργασία του mRNA. Το συμπέρασμα αυτό έρχεται σε συμφωνία με τις διαθέσιμες πληροφορίες που υπάρχουν για την πρωτεΐνη FRA10AC1, δηλαδή: α) τη συμμετοχή του ορθολόγου της στον μονοκύτταρο οργανισμό *Chlamydomonas reinhardtii* στην αναγνώριση των θέσεων συναρμογής, β) το γεγονός ότι τα ορθόλογα μόρια της πρωτεΐνης έχουν ταυτοποιηθεί αποκλειστικά σε ευκαρυωτικούς οργανισμούς στους οποίους υφίσταται η συναρμογή, γ) τη βιοχημική απομόνωσή της από κύτταρα *HeLa* ως συστατικό των υποσυμπλόκων του σωματίου συναρμογής, που σχηματίζονται κατά τη διάρκεια των δύο καταλυτικών βημάτων της διαδικασίας, B *act* (Bessonov et al. 2010), C (Bessonov et al. 2008, Bessonov et al. 2010, Pagan et al. 2013, Schmidt et al. 2014) και του υποσυμπλόκου P που έπεται αυτών (Pagan et al. 2013), και δ) την ταυτοποίησή της σε 2 πρωτεϊνικά σύμπλοκα κυττάρων HEK293, τα οποία είναι εμπλουτισμένα με πρωτεΐνες που συμμετέχουν στην επεξεργασία του mRNA, και κατά κύριο λόγο στη συναρμογή και στη μεταγραφή (Huttlin et al. 2017).

Για να εξαχθούν συμπεράσματα ως προς τη βιολογική διαδικασία στην οποία θα μπορούσε επιπλέον να συμμετέχει η FRA10AC1 βάσει των υπολοίπων αλληλεπιδρώντων μορίων, τα οποία δεν μπορούν να ομαδοποιηθούν με βάση μία κοινή λειτουργική διαδικασία ή ένα κοινό πρωτεϊνικό σύμπλοκο, θα πρέπει να αξιολογηθούν και οι αλλαγές που προκάλεσε η τροποποίηση της έκφρασης του γονιδίου *FRA10AC1* σε αυτά τα αλληλεπιδρώντα μόρια σε μεταγραφικό και πρωτεϊνικό επίπεδο (βλ. συζήτηση παρακάτω).

#### 4.1.1. Η χρήση της μεθόδου FRET για τον έλεγχο των αλληλεπιδράσεων της FRA10AC1 με πρωτεΐνες που συμμετέχουν στις διαδικασίες επεξεργασίας του mRNA

Όπως προαναφέρθηκε, μέσω της διερεύνησης του λειτουργικού ρόλου των πρώτων «γειτόνων» της FRA10AC1, η πλειοψηφία αυτών συμμετέχουν στις διαδικασίες επεξεργασίας του mRNA και κυρίως στη συναρμογή. Όπως είναι γνωστό υπάρχουν πολλές διαφορετικές μέθοδοι για την ταυτοποίηση των πρωτεϊνικών αλληλεπιδράσεων. Για να ελεγχθεί περαιτέρω η συμμετοχή της FRA10AC1 στην επεξεργασία του mRNA εξετάστηκαν οι αλληλεπιδράσεις της με 3 πρωτεΐνες (MFAP1, SAP30BP, THOC1) που αποτελούν συστατικά του σωματίου συναρμογής και συμμετέχουν στην συναρμογή, στη μεταγραφή και στην έξοδο του mRNA από τον πυρήνα, αντίστοιχα, μέσω της χρήσης της μεθόδου Μεταφοράς Ενέργειας Φθορισμού μέσω Συντονισμού (FRET). Η επιλογή αυτής της μεθόδου έναντι άλλων έγινε διότι μέσω αυτής μπορούν να ταυτοποιηθούν άμεσες αλληλεπιδράσεις *in vivo* και ακόμη, να

ελεγχθεί σε ποιο υποκυτταρικό διαμέρισμα λαμβάνει χώρα η αλληλεπίδραση (Bastiaens & Squire 1999). Αντίθετα, άλλες μέθοδοι, όπως η ανοσοκατακρήμνιση ή χρωματογραφία συγγένειας, αν και παρέχουν πληροφορίες για την πιθανότητα δύο πρωτεΐνες να αλληλεπιδρούν *in vivo*, χρησιμοποιούν κυτταρικά εκχυλίσματα που προσομοιάζουν εν μέρει τις φυσιολογικές κυτταρικές συνθήκες. Επιπλέον, οι δοκιμές συνκατακρήμνισης (pull down) μέσω της χρήσης καθαρισμένων και ανασυνδυασμένων πρωτεϊνών οι οποίες ελέγχουν άμεσες αλληλεπιδράσεις αντιπροσωπεύουν μια *in vitro* κατάσταση η οποία απέχει από τις φυσιολογικές συνθήκες των κυττάρων. Η ανάλυση FRET και για τις 3 αλληλεπιδράσεις έδειξε ότι αυτές οι πρωτεΐνες δεν αλληλεπιδρούν. Ωστόσο, θα μπορούσε να υπάρχει αλληλεπίδραση *in vivo* παρόλο που δεν ταυτοποιήθηκε με αυτήν την μέθοδο. Η απουσία αλληλεπίδρασης πιθανώς μπορεί να εξηγηθεί από το μεγάλο μέγεθος των επιτόπων το οποίο μπορεί να αλλάζει την δομή των χμαιρικών πρωτεϊνών με τέτοιο τρόπο ώστε τα δίπολα του μορίου-δότη (η FRA10AC1 σε σύντηξη με τον επίτοπο EGFP) να μην βρίσκονται στον σωστό προσανατολισμό, δηλαδή να μην είναι παράλληλα, με τα δίπολα του μορίου-δέκτη (οι 3 τρεις «γείτονες» της FRA10AC1 σε σύντηξη με τον επίτοπο mCherry) (Sun et al. 2012). Επιπλέον, ένας επιπρόσθετος λόγος που έχει διατυπωθεί είναι ότι, παρόλο που οι δυο εξεταζόμενες χμαιρικές πρωτεΐνες υπερεκφράζονται υπό τη δράση ισχυρών υποκινητών στα κύτταρα, τα υψηλά επίπεδα παραγωγής τους μπορεί να οδηγήσουν στη δημιουργία συσσωματωμάτων των πρωτεϊνών με αποτέλεσμα οι χμαιρικές πρωτεΐνες να μην είναι λειτουργικές όπως οι ενδογενείς (Sun et al. 2012). Οι παραπάνω λόγοι μπορεί να εξηγήσουν την απουσία αλληλεπίδρασης βάσει της φύσης της μεθόδου, ωστόσο, λαμβάνοντας υπόψη τις πληροφορίες που προαναφέρθηκαν για τις 3 εξεταζόμενες πρωτεΐνες (βλ. Ενότητα 3.3.2.) θα μπορούσε να υποθεθεί ότι η FRA10AC1 εμφανίζει λιγότερες πιθανότητες να αλληλεπιδρά με την SAP30BP καθώς αυτές οι δυο έχουν ταυτοποιηθεί σε διαφορετικά υποσύμπλοκα του σωματίου συναρμογής. Για να ελεγχθούν με εναλλακτικούς τρόπους οι παραπάνω αλληλεπιδράσεις θα μπορούσαν να χρησιμοποιηθούν *in vitro* πειραματικές μέθοδοι (π.χ. δοκιμή συν-κατακρήμνισης).

#### **4.2. *In silico* ανακατασκευή του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής**

Η συμμετοχή του ορθόλογου μορίου της FRA10AC1 στο μονοκύτταρο οργανισμό *Chlamydomonas reinhardtii* στην αναγνώριση των θέσεων συναρμογής υποδεικνύει τη συμμετοχή της πρωτεΐνης του ανθρώπου στη διαδικασία της συναρμογής. Επιπλέον, το γεγονός ότι: α) η πλειοψηφία των αλληλεπιδρώντων μορίων της πρωτεΐνης FRA10AC1 αποτελούν συστατικά του σωματίου συναρμογής β) η επανειλημμένη ταυτοποίηση της FRA10AC1 σε βιοχημικές μελέτες απομόνωσης των συμπλόκων του σωματίου συναρμογής, υποδεικνύει την συμμετοχή της στη διαδικασία της συναρμογής, γ) τα αλληλεπιδρώντα μόριά της, τα οποία αποτελούν συστατικά του σωματίου, συμμετέχουν στη συναρμογή και σε σχετιζόμενες με αυτή διαδικασίες, όπως οι SF3B2, U2AF1 (συναρμογή), οι NKAP, SAP30BP (μεταγραφή) και η πρωτεΐνη THOC1 (έξοδος από τον πυρήνα) και δ) η αυξημένη αξιοπιστία των αλληλεπιδράσεων

της με παράγοντες συναρμογής (SF3B2, CHERP, IK, DGCR14) και των αλληλεπιδράσεων μεταξύ των πρώτων «γειτόνων» της που αποτελούν παράγοντες συναρμογής (U2AF1, PRPF3, PRPF40A) ή μεταγραφής (ZNF830, NKAP), υποδεικνύει την συμμετοχή της στις διαδικασίες επεξεργασίας του πρόδρομου mRNA και πιθανώς, λαμβάνοντας υπόψη την αφθονία των «γειτόνων» της, στη σύνδεση της συναρμογής με άλλες διαδικασίες της επεξεργασίας του mRNA, δρώντας ως μόριο-«προσαρμογέας».

Με στόχο λοιπόν την ένταξη της FRA10AC1 σε μια συγκεκριμένη βιολογική διαδικασία εντός του σωματίου συναρμογής πραγματοποιήθηκε η *in silico* ανακατασκευή του δικτύου των άμεσων πρωτεϊνικών αλληλεπιδράσεων ολόκληρου του σωματίου συναρμογής και των επιμέρους υποσυμπλόκων του. Μέσω της ανασύστασης αυτού του δικτύου θα μπορούσε να γίνει προσδιορισμός της τοπολογίας της πρωτεΐνης FRA10AC1 στο ευρύτερο σωματίο συναρμογής και διερεύνηση των βιολογικών διαδικασιών στις οποίες συμμετέχουν οι δεύτεροι «γείτονές» της εντός αυτού, ώστε να συλλεχθούν πληροφορίες για τον πιθανό λειτουργικό της ρόλο εντός του πλαισίου της επεξεργασίας του mRNA.

#### 4.2.1. Οι 630 πρωτεΐνες του σωματίου συναρμογής αποτελούν ένα πλήρες, αξιόπιστο και μη επικαλυπτόμενο σύνολο δεδομένων

Όπως έχει ήδη αναφερθεί, το σωματίο συναρμογής, αυτή η πολύπλοκη και δυναμική ριβονουκλεοπρωτεϊνική μηχανή, περιλαμβάνει εκατοντάδες πρωτεΐνες καθώς απαιτείται εξαιρετική ακρίβεια για την αναγνώριση των θέσεων συναρμογής, την αποκοπή των ιντρονίων και την εναλλακτική συναρμογή (Chen & Moore 2014). Δεδομένου του μεγάλου αριθμού των πρωτεϊνών που συμμετέχουν, για την δημιουργία του συνόλου των συστατικών του σωματίου συναρμογής πραγματοποιήθηκε η αναζήτηση και η εξονυχιστική διερεύνηση τόσο της Spliceosome Database, η οποία αποτελεί πηγή πληροφοριών για τις πρωτεΐνες που έχουν απομονωθεί ως συστατικά του σε βιοχημικές μελέτες, όσο και σχετικών βιβλιογραφικών αναφορών ώστε να βρεθούν όλες οι ερευνητικές εργασίες που αφορούν την σύστασή του (βλ. Ενότητα 3.2.1.1.). Εξετάζοντας λοιπόν και τις 17 διαθέσιμες ερευνητικές εργασίες δημιουργήθηκε ένα πλήρες σύνολο δεδομένων 630 πρωτεϊνών και αποφεύχθηκε η μεροληψία υπέρ κάποιων αναφορών.

Αντίθετα, οι εργασίες που μελετούν τις αλληλεπιδράσεις εντός του σωματίου συναρμογής δεν εμφανίζουν, κατά κανόνα, πληρότητα ως προς τις πηγές άντλησης των δεδομένων τους. Ειδικότερα, οι Hegele et al. (2012) που μελετούν τις φυσικές αλληλεπιδράσεις στο πλαίσιο δικτύων στηρίζονται σε μόλις δύο (Wahl et al. 2009, Jurica & Moore 2003) ερευνητικές εργασίες και το σύνολο δεδομένων τους περιλαμβάνει 244 πρωτεΐνες. Αντίστοιχα, οι Papasaikas et al. (2015), που προσδιορίζουν τις λειτουργικές συνδέσεις μεταξύ πρωτεϊνών του σωματίου βάσει της επίδρασής τους στην εναλλακτική συναρμογή στο πλαίσιο ρυθμιστικών δικτύων, αντλούν δεδομένα μόνο από την εργασία των Wahl et al. (2009) και έτσι η μελέτη αφορά 196 πρωτεΐνες. Οι Pires et al. (2015), παρόλο που δεν αναφέρουν την πηγή άντλησης των δεδομένων τους, δημιουργούν δίκτυο φυσικών αλληλεπιδράσεων με 103 πρωτεΐνες-συστατικά του σωματίου συναρμογής. Σημαντικά μεγαλύτερο αριθμό

συστατικών του σωματίου συναρμογής μελετούν οι Akerman et al. (2015) που εξετάζουν τις φυσικές αλληλεπιδράσεις 630 πρωτεϊνών με *in silico* προσεγγίσεις. Το σύνολο δεδομένων τους προέκυψε από δυο ερευνητικές εργασίες (Hegele et al. 2012, Chen et al. 2007) και δυο βάσεις δεδομένων (Spliceosome Database, KEGG PATHWAY Database). Ωστόσο, συγκρίνοντας τις πρωτεΐνες της Spliceosome Database που ανακτήθηκαν στο πλαίσιο της εργασίας των Akerman et al. (2015) με τις 514 πρωτεΐνες που χρησιμοποιήθηκαν στην παρούσα εργασία και οι οποίες αντιστοιχούν σε όλα τα δεδομένα της βάσης, διαπιστώθηκε ότι στη συγκεκριμένη ερευνητική εργασία δεν συμπεριελήφθησαν όλες, αλλά το 70% αυτών. Επιπλέον, οι Akerman et al. (2015) χρησιμοποιούν στοιχεία που προέρχονται από τη βάση δεδομένων KEGG PATHWAY η οποία αφορά κυρίως την περιγραφή βιολογικών μονοπατιών, και πιθανώς δεν συνάδει με την ανάκτηση πληροφοριών για τα συστατικά του σωματίου συναρμογής. Ειδικότερα, ενσωματώνουν στο σύνολο δεδομένων τους τις πρωτεΐνες τριών μονοπατιών, της συναρμογής, της μεταφοράς RNA και του ελέγχου της ακεραιότητας του RNA. Στη βάση KEGG, κατά την περιγραφή αυτών των μονοπατιών, γίνονται παραπομπές σε πολύ περιορισμένο αριθμό βιβλιογραφικών αναφορών που αφορούν αυτά τα τρία μονοπάτια. Για παράδειγμα, στο μονοπάτι της συναρμογής, γίνεται αναφορά μόνο σε μία ερευνητική εργασία απομόνωσης των συστατικών του σωματίου συναρμογής. Επίσης, στο μονοπάτι της μεταφοράς RNA περιλαμβάνονται όλες οι πρωτεΐνες που συμμετέχουν στην μεταφορά όλων των ειδών RNA και όχι μόνο του mRNA. Στη εργασία των Akerman et al. (2015) δεν εφαρμόστηκε κάποιο κριτήριο επιλογής πρωτεϊνών μεταφοράς αποκλειστικά του mRNA παρόλο τον μεγάλο αριθμό πρωτεϊνών της γενικότερης κατηγορίας μεταφοράς RNA (148). Ακόμη, στο μονοπάτι ελέγχου ακεραιότητας RNA (mRNA surveillance pathway) γίνεται αναφορά στους μηχανισμούς αποικοδόμησης που λαμβάνουν χώρα αποκλειστικά στο κυτταρόπλασμα (π.χ. αποικοδόμηση του mRNA όταν υπάρχει πρόωρο κωδικόνιο λήξης) και παραλείπονται μηχανισμοί που ελέγχουν το mRNA μετά το πέρας της μεταγραφής, στον πυρήνα (π.χ. εξώσωμα). Γι' αυτούς τους λόγους, η βάση KEGG δεν μπορεί να θεωρηθεί κατάλληλη επιλογή για τη συλλογή των συστατικών του σωματίου συναρμογής. Συνολικά, η σύγκριση του συνόλου των δεδομένων των Akerman et al. (2015) με το αντίστοιχο της παρούσας εργασίας έδειξε ότι, παρόλο που ο συνολικός αριθμός των πρωτεϊνών είναι συγκρίσιμος, υπάρχει μικρή αλληλοεπικάλυψη, της τάξεως του 50%. Συμπερασματικά, ο τρόπος ανάκτησης των συστατικών του σωματίου συναρμογής που ακολουθήθηκε στην παρούσα εργασία οδήγησε σε ένα πλήρες σύνολο δεδομένων που περιλαμβάνει αποκλειστικά πρωτεΐνες του σωματίου συναρμογής καθώς οι πηγές άντλησης των δεδομένων (Spliceosome Database και βιβλιογραφικές αναφορές) ήταν οι κατάλληλες.

Επιπρόσθετα, το πλήρες σύνολο δεδομένων των συστατικών του σωματίου συναρμογής χαρακτηρίζεται από αυξημένη εγκυρότητα. Μια παράμετρος που ενίσχυσε την πιστότητα του συνόλου δεδομένων ήταν η χρήση ενός κωδικού κανονικοποίησης για κάθε αναφερόμενο γονίδιο και την αντίστοιχη πρωτεΐνη. Ειδικότερα, κατά τον ενδελεχή έλεγχο των 17 ερευνητικών εργασιών που χρησιμοποιήθηκαν ως πηγές άντλησης δεδομένων παρατηρήθηκε η ύπαρξη διαφορετικών ονομάτων για την ίδια πρωτεΐνη ή τις ισομορφές μιας πρωτεΐνης, για το ίδιο γονίδιο ή για τα εναλλακτικά

μετάγραφα ενός γονιδίου (βλ. Ενότητα 3.2.1.1.). Η άμεση καταγραφή όλων αυτών θα είχε ως αποτέλεσμα το αυξημένο ποσοστό αλληλοεπικάλυψης των δεδομένων και κατ' επέκταση την ψευδή αύξηση του αριθμού των συστατικών του σωματίου συναρμογής. Γι' αυτό το λόγο και για να επιτευχθεί η σύγκριση του πλήθους των δεδομένων, με βάση το γεγονός ότι σε καθεμία από τις 17 ερευνητικές εργασίες ταυτοποιήθηκαν από 100 (π.χ. Behzadnia et al. 2007, Pagan et al. 2013, Zhou et al. 2002) μέχρι και σχεδόν 300 πρωτεΐνες (Makarov et al. 2012, Bessonov et al. 2008, Rappsilber et al. 2002), πραγματοποιήθηκε η αναγωγή όλων των δεδομένων σε έναν κωδικό κανονικοποίησης για κάθε αναφερόμενο γονίδιο, τον κωδικό Gene ID και τον κωδικό UniProt ID για την αντίστοιχη πρωτεΐνη (βλ. Ενότητα 3.2.1.1.). Η παραπάνω διαδικασία έγινε στο ίδιο πλαίσιο με αυτήν που ακολουθήθηκε στη μεταβάση PICKLE στην οποία το σύνολο δεδομένων πρωτεϊνικών αλληλεπιδράσεων είναι κανονικοποιημένο ως προς τον κωδικό UniProt ID για κάθε πρωτεΐνη και ως προς τον κωδικό Gene ID για κάθε γονίδιο. Η κανονικοποίηση της PICKLE έγινε για να αποφευχθούν τυχόν λάθη και επαναλήψεις καθώς οι πρωτογενείς βάσεις από τις οποίες αντλεί πληροφορίες αναφέρονται στα αλληλεπιδρώντα μόρια χρησιμοποιώντας κωδικούς διαφορετικών επιπέδων (Klara et al. 2013). Για παράδειγμα, η HPRD χρησιμοποιεί κωδικούς νουκλεοτιδικής αλληλουχίας, η BioGRID κωδικούς γονιδίου και οι υπόλοιπες τρεις κωδικούς πρωτεΐνης.

Μια επιπρόσθετη παράμετρος που ενίσχυσε την πιστότητα αυτού του συνόλου των δεδομένων ήταν ο καθαρισμός του από πρωτεΐνες που θεωρούνται προσμίξεις (ψευδώς θετικά αποτελέσματα) της διαδικασίας απομόνωσης των συστατικών με χρωματογραφία συγγένειας σε συνδυασμό με φασματομετρία μάζας (Sahni et al. 2015, Berger et al. 2013). Πρέπει να σημειωθεί ότι στις ερευνητικές εργασίες που προήλθαν από το εργαστήριο του Δρ. Luhrmann είχε πραγματοποιηθεί η απομάκρυνση ορισμένων πρωτεϊνών που θεωρήθηκαν ψευδώς θετικά αποτελέσματα, όπως οι ριβοσωμικές πρωτεΐνες, οι τουμπουλίνες και οι πρωτεΐνες του εξωσώματος ο ρόλος των οποίων, μέχρι πρόσφατα, ήταν γνωστός μόνο στο κυτταρόπλασμα. Αν και οι ριβοσωμικές πρωτεΐνες απομακρύνθηκαν και σε αυτή την ερευνητική εργασία λόγω της εδραιωμένης λειτουργίας τους στο κυτταρόπλασμα, ορισμένες τουμπουλίνες συμπεριλήφθηκαν, όπως και όλες οι πρωτεΐνες του εξωσώματος. Ειδικότερα, ορισμένες τουμπουλίνες εμφανίζουν πυρηνικό εντοπισμό ή συσχέτισμό με κάποια πυρηνική λειτουργία (π.χ. μετάγραφή) (βλ. Ενότητα 3.2.1.1.). Όσον αφορά το εξώσωμα, ο ρόλος του είναι γνωστός στον έλεγχο της ακεραιότητας των RNA και στον πυρήνα. Συγκεκριμένα, το πυρηνικό εξώσωμα αποικοδομεί τα mRNA αλλά και άλλα είδη RNA στα οποία δεν γίνεται σωστά ο τερματισμός της μεταγραφής, ο σχηματισμός του 3' άκρου (πολυαδενυλίωση) και η αφαίρεση των ιντρονίων (Kilchert et al. 2016, Pan et al. 2015). Ωστόσο, απομακρύνθηκαν ορισμένες πρωτεΐνες που είχαν εδραιωμένη λειτουργία εκτός πυρήνα καθώς και ένα γονίδιο που μετέπειτα χαρακτηρίστηκε ως ψευδογονίδιο (βλ. Ενότητα 3.2.1.1.).

Η αυξημένη αξιοπιστία του συνόλου δεδομένων του σωματίου συναρμογής ενισχύεται και από τον αριθμό των ερευνητικών εργασιών στις οποίες ταυτοποιήθηκαν αυτές οι 630 πρωτεΐνες καθώς και από το λειτουργικό τους ρόλο. Ειδικότερα, το 62% των πρωτεϊνών-UniProt IDs είχαν απομονωθεί σε περισσότερες από μία πρωτεομικές

μελέτες. Για τις υπόλοιπες πρωτεΐνες, η πλειοψηφία των οποίων είχε ταυτοποιηθεί στο υποσύμπλοκο E (Makarou et al. 2012) ή, σε μικρότερο βαθμό, σε μελέτες ολόκληρου του σωματίου συναρμογής πραγματοποιήθηκε η περαιτέρω αξιολόγησή τους μέσω της εύρεσης των όρων γονιδιακής οντολογίας που συμμετέχουν. Το 60% αυτών (αν εξαιρεθούν οι πρωτεΐνες άγνωστης λειτουργίας) συμμετέχουν στη διαδικασία της συναρμογής και σε σχετιζόμενες με αυτή λειτουργίες. Συνεπώς, οι πρωτεΐνες αυτές δεν θα μπορούσαν να αποτελούν ψευδώς θετικά αποτελέσματα της βιοχημικής μεθόδου απομόνωσης των συστατικών του σωματίου, αλλά, μπορεί να ταυτοποιήθηκαν σε μία μόνο ερευνητική εργασία λόγω της παροδικής τους συνεισφοράς ή της ασθενούς σύνδεσής τους σε ένα ή πολλά διαφορετικά στάδια της συναρμολόγησης. Έτσι, η χαμηλή αφθονία τους μπορεί να μην επιτρέπει την απομόνωσή τους με τις βιοχημικές μεθόδους που χρησιμοποιήθηκαν στις διαφορετικές εργασίες αλλά να απαιτεί τη χρήση μεθόδων με μεγαλύτερη ευαισθησία.

#### 4.2.2. Ογδόντα πρωτεΐνες είναι παρούσες καθ' όλη τη διάρκεια συναρμολόγησης του σωματίου συναρμογής

Παρόλο που ο αριθμός των πρωτεϊνών οι οποίες συμμετέχουν στο μηχανισμό της συναρμογής είναι υψηλός, ένα υποσύνολο αυτών φαίνεται να διαδραματίζουν καθοριστικό ρόλο, καθώς εμφανίζονται καθ' όλη τη διάρκεια της συναρμολόγησης του σωματίου συναρμογής και παρουσιάζουν υψηλή αφθονία είτε σε όλα είτε σε συγκεκριμένα υποσύμπλοκα. Σε αντίστοιχες ερευνητικές εργασίες, οι πρωτεΐνες αυτές έχουν χαρακτηριστεί ως «persistent spliceosome» εξαιτίας της καθολικής τους παρουσίας στα επιμέρους υποσύμπλοκα (Papasaikas et al. 2015). Ο περιορισμός που συναντάται είναι ότι ο χαρακτηρισμός αυτός βασίστηκε σε περιορισμένο αριθμό βιβλιογραφικών αναφορών (Wahl et al. 2009, Zhou et al. 2002), γεγονός που μπορεί να υποδεικνύει κάποια μεροληψία.

Γι' αυτό το λόγο πραγματοποιήθηκε η λεπτομερής εξέταση των 12 ερευνητικών εργασιών που περιλάμβαναν βιοχημικά πειράματα, κατά τα οποία απομονώνονταν τα συστατικά των υποσυμπλόκων. Σε πρώτο στάδιο βρέθηκαν οι πρωτεΐνες που ήταν παρούσες σε όλα τα υποσύμπλοκα A, B, B act και C, με εξαίρεση το E και το P. Τα δύο υποσύμπλοκα εξαιρέθηκαν διότι δεν υπήρχε ικανοποιητικός αριθμός πειραματικών μελετών για την αξιολόγηση των αποτελεσμάτων και επιπλέον επειδή το υποσύμπλοκο P αποτελεί υποσύνολο του C. Σε δεύτερο στάδιο, επιλέχθηκαν πέντε ερευνητικές εργασίες οι οποίες προσέφεραν πληροφορίες για την αφθονία των πρωτεϊνών στα υποσύμπλοκα. Κάποιες από αυτές προσδιόριζαν την αφθονία των πρωτεϊνών (π.χ. υψηλή ή χαμηλή αφθονία) ενώ κάποιες ανέφεραν τον αριθμό των πεπτιδίων. Στις τελευταίες ακολούθησε περαιτέρω αξιολόγηση των δεδομένων αυτών, με τον τρόπο που αναλύθηκε εκτενώς στην Ενότητα 3.2.1.1.4. των αποτελεσμάτων.

Το σύνολο λοιπόν των 80 πρωτεϊνών περιείχε: α) καλά χαρακτηρισμένους παράγοντες συναρμογής με υψηλή αφθονία σε όλα τα υποσύμπλοκα, όπως οι πρωτεΐνες Sm και Lsm (Beggs 2005, Urlaub 2001), β) πρωτεΐνες των συμπλόκων SF3a και SF3b που βρέθηκαν σε υψηλή αφθονία στα πρώτα τρία υποσύμπλοκα ή μόνο στο A, που εξηγείται από το βιολογικό τους ρόλο, καθώς αποτελούν συστατικά του U2snRNP και συνεισφέρουν τόσο στη συναρμολόγηση των αρχικών υποσυμπλόκων

του σωματίου συναρμογής όσο και στην ακριβή αναγνώριση των θέσεων συναρμογής (Lin & Xu, 2012, Rakesh et al. 2016), γ) συστατικά του συμπλόκου Prp19/CDC5L σε υψηλή αφθονία στα τρία τελευταία υποσύμπλοκα, γεγονός που ήταν αναμενόμενο εξαιτίας του καλά χαρακτηρισμένου ρόλου του συμπλόκου αυτού στα δυο καταλυτικά βήματα της συναρμογής, καθώς ρυθμίζει το σχηματισμό και την εξέλιξη των απαραίτητων αναδιατάξεων που απαιτείται να συμβούν στα στάδια αυτά (Chan & Cheng 2005, Chanarat & Sträber 2013, Hogg et al. 2010), δ) πρωτεΐνες SR και hnRNPs σε υψηλή, μέτρια ή χαμηλή αφθονία σε διάφορα υποσύμπλοκα, που συμμετέχουν στη ρύθμιση της συναρμογής με ποικίλους τρόπους (Busch & Hertel 2012) και ε) πρωτεΐνες οι οποίες συμμετέχουν στη συναρμογή και επιπρόσθετα σε σχετιζόμενες διαδικασίες (π.χ. μεταγραφή, έξοδος του mRNA από τον πυρήνα, αποικοδόμηση του mRNA), η παρουσία των οποίων μπορεί να εξηγηθεί βάσει της σημασίας του ρόλου τους καθ' όλη τη διάρκεια της συναρμογής (Πιν. 7, Ενότητα 3.2.1.1.4.).

Συγκρίνοντας τον αριθμό των 80 πρωτεϊνών που εμφάνιζαν καθολική παρουσία με τον αντίστοιχο αριθμό των πρωτεϊνών που αποτελούν το «persistent spliceosome» στην εργασία των Papasaïkas et al. (2015), παρατηρείται ότι ο αριθμός των πρωτεϊνών που βρέθηκαν στην παρούσα εργασία ήταν τριπλάσιος, γεγονός που δείχνει τη σημασία της συλλογής, της σύγκρισης και της αξιολόγησης όλων των διαθέσιμων πληροφοριών που υπάρχουν σχετικά με το πρωτόμα του σωματίου συναρμογής.

Όσον αφορά την αφθονία των 80 πρωτεϊνών, παρατηρήθηκε ότι το 47% αυτών (38 πρωτεΐνες) εμφανίζει υψηλή αφθονία στα υποσύμπλοκα A, B, B *act* και C (12 πρωτεΐνες) ή στα τρία τουλάχιστον από αυτά (26 πρωτεΐνες). Το αποτέλεσμα αυτό έρχεται σε συμφωνία με αντίστοιχες ερευνητικές εργασίες (Hegele et al. 2012, Papasaïkas et al. 2015) οι οποίες χαρακτηρίζουν και αυτές το ~60% των 38 αυτών πρωτεϊνών (22 πρωτεΐνες) ως «core», δηλαδή πρωτεΐνες υψηλής αφθονίας. Οι Hegele et al. (2012) βασίστηκαν σε μία ερευνητική εργασία (Agafonov et al. 2011) η οποία ήταν κοινή με μία από αυτές που χρησιμοποιήθηκαν στην διατριβή. Επιπλέον, οι Papasaïkas et al. (2015) στηρίχτηκαν σε δυο ερευνητικές εργασίες, εκ των οποίων η μία (Zhou et al. 2002) αφορούσε την ταυτοποίηση των συστατικών ολόκληρου του σωματίου συναρμογής και γι' αυτό δεν επιλέχθηκε στη διατριβή, και η άλλη (Wahl et al. 2009) ήταν μία ερευνητική εργασία ανασκόπησης (review) του σωματίου συναρμογής που περιλάμβανε πληροφορίες για την αφθονία των συστατικών του βάσει 2 εργασιών (Bessonov et al. 2008, Behzadnia et al. 2007), οι οποίες χρησιμοποιήθηκαν στη διατριβή. Η ύπαρξη αυτών των κοινών εργασιών μπορεί να δικαιολογήσει εν μέρει το ποσοστό των κοινών πρωτεϊνών που χαρακτηρίστηκαν ως υψηλής αφθονίας.

Η σημασία της συνεισφοράς αυτών των 80 πρωτεϊνών στη διαδικασία της συναρμογής επιβεβαιώθηκε και στα πλαίσια του δικτύου του σωματίου συναρμογής καθώς όλες ενσωματώθηκαν στο δίκτυο, με τοπολογία στις κεντρικές περιοχές του (Εικ. 40). Αυτές οι 80 πρωτεΐνες βρέθηκαν ότι αναπτύσσονται μεταξύ τους 156 αλληλεπιδράσεις, οι οποίες, όπως ήταν αναμενόμενο, διατηρούνται και στα τέσσερα υποσύμπλοκα διότι οι πρωτεΐνες που τις αποτελούν έχουν απομονωθεί ως συστατικά των υποσυμπλόκων αυτών. Παρατηρήθηκε ότι οι πρωτεΐνες υψηλής αφθονίας είχαν την τάση να αλληλεπιδρούν μεταξύ τους. Ένα χαρακτηριστικό παράδειγμα ήταν οι πρωτεΐνες Sm, οι αλληλεπιδράσεις των οποίων επιβεβαιώνουν τη σημασία τους σε όλα

τα στάδια της διαδικασίας. Επιπλέον, οι πρωτεΐνες των συμπλόκων SF3a και SF3b, οι οποίες εμφάνισαν υψηλή αφθονία στα υποσύμπλοκα A, B και B *act*, αλληλεπιδρούσαν τόσο μεταξύ τους όσο και με συστατικά του συμπλόκου U4/U6.U5 (PRPF8, PRPF6) ή του Prp19/CDC5L (PRPF19, CDC5L) που παρουσίαζαν επίσης υψηλή αφθονία στα υποσύμπλοκα B, B *act* και C. Οι αλληλεπιδράσεις μεταξύ των πρωτεϊνών των συμπλόκων SF3a και SF3b επιβεβαιώνουν τη συνεργασία τους για την αναγνώριση των θέσεων συναρμογής που συμβαίνει στα αρχικά στάδια συναρμολόγησης του σωματίου συναρμογής, και οι αλληλεπιδράσεις αυτών με πρωτεΐνες που διαδραματίζουν ενεργό ρόλο στα καταλυτικά βήματα της συναρμογής ενισχύει τη συνεισφορά των συμπλόκων SF3a και SF3b και στα τελικά στάδια συναρμολόγησης του σωματίου. Όσον αφορά τις πρωτεΐνες χαμηλής αφθονίας, είχαν και αυτές την τάση να αλληλεπιδρούν μεταξύ τους. Ωστόσο, δεν απουσίαζαν και οι αλληλεπιδράσεις μεταξύ πρωτεϊνών διαφορετικής αφθονίας. Ένα χαρακτηριστικό παράδειγμα ήταν οι πρωτεΐνες NCBP2, NCBP1 και ALY/REF. Οι δύο από αυτές (NCBP2, NCBP1) εμφάνιζαν υψηλή αφθονία σε όλα τα υποσύμπλοκα ή στα B *act* και C, αντίστοιχα, ενώ η τρίτη (ALY/REF) εμφάνιζε μέτρια αφθονία στο A και χαμηλή στα υπόλοιπα. Οι πρωτεΐνες NCBP2, NCBP1 αποτελούν συστατικά ενός συμπλόκου το οποίο συνδέεται με την 5' καλύπτρα του mRNA (Mazza et al. 2001). Η NCBP1 έχει δειχθεί ότι συνδέεται με την NCBP2 και την ALY/REF (παράγοντας που συμμετέχει στην έξοδο του mRNA από τον πυρήνα) στα αρχικά στάδια συναρμολόγησης του σωματίου συναρμογής (Cheng et al. 2006, Gebhardt et al. 2015). Η υψηλή αφθονία των πρωτεϊνών NCBP2, NCBP1 πιθανώς υποδεικνύει ότι αυτές οι πρωτεΐνες λειτουργούν ως μόρια-προσαρμογείς που στρατολογούν πρωτεΐνες, όπως την ALY/REF, κατά τη διάρκεια συναρμολόγησης του σωματίου συναρμογής ώστε να συνδέσουν τη συναρμογή με άλλες συναφείς διαδικασίες (π.χ. έξοδο του mRNA από τον πυρήνα). Συμπερασματικά παρατηρήθηκε ότι υπάρχει η τάση, και όχι ο κανόνας, οι πρωτεΐνες που εμφάνιζαν την ίδια αναλογία στην αφθονία να αλληλεπιδρούν μεταξύ τους (~55% των αλληλεπιδράσεων). Αν και φαίνεται ότι οι αλληλεπιδράσεις μεταξύ πρωτεϊνών που εμφανίζουν την ίδια αναλογία στην αφθονία τείνουν να αποτελούν τον κανόνα, οι περιπτώσεις αλληλεπιδράσεων μεταξύ πρωτεϊνών διαφορετικής αφθονίας δεν είναι αμελητέες. Η τελευταία περίπτωση μπορεί να υποδηλώνει ότι υπάρχουν κάποιες πρωτεΐνες υψηλής αφθονίας που έχουν βασικό ρόλο στη συναρμογή και συνδέονται με άλλες που εμφανίζουν δευτερεύοντα ή βοηθητικό ρόλο στη διαδικασία γεγονός που πιθανώς εξηγεί και τη χαμηλή τους αφθονία. Η ύπαρξη τέτοιου είδους αλληλεπιδράσεων συνάδει και με την πολυπλοκότητα της συναρμολόγησης του σωματίου συναρμογής το οποίο είναι το μεγαλύτερο ριβονουκλεοπρωτεϊνικό σύμπλοκο και, όπως είναι αναμενόμενο, σε αυτό περιλαμβάνονται πρωτεΐνες που έχουν είτε βασικούς είτε δευτερεύοντες ρόλους ώστε η συντονισμένη συνεργασία τους να συνεισφέρει στη διεκπεραίωση της διαδικασίας με εξαιρετική ακρίβεια.

4.2.3. Το δίκτυο πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής και των υποσυμπλόκων του περιέχει αποκλειστικά άμεσες αλληλεπιδράσεις και είναι ιδιαίτερα αξιόπιστο



Το ολοκληρωμένο και υψηλά αξιόπιστο σύνολο δεδομένων των συστατικών του σωματίου συναρμογής χρησιμοποιήθηκε στην *in silico* ανακατασκευή του δικτύου πρωτεϊνικών αλληλεπιδράσεων ολόκληρου του σωματίου συναρμογής καθώς και των επιμέρους υποσυμπλόκων του. Η αναζήτηση των αποκλειστικά άμεσων αλληλεπιδράσεων εντός του σωματίου συναρμογής πραγματοποιήθηκε με την ανάκτηση δεδομένων τόσο από την μετα-βάση PICKLE, που περιλαμβάνει δεδομένα από τις πέντε κύριες βάσεις που στηρίζονται στη βιβλιογραφία στον άνθρωπο, όσο και από δυο βάσεις δεδομένων στους οργανισμούς-μοντέλα *C. elegans* και *D. melanogaster* (βλ. Ενότητα 3.2.1.2.). Η επιλογή της ένταξης, στο δίκτυο, μόνο των άμεσων αλληλεπιδράσεων συνάδει με την επιλογή αυτής της κατηγορίας των αλληλεπιδράσεων στο πλαίσιο ανακατασκευής του δικτύου της FRA10AC1 (βλ. Ενότητα 4.1.) Με στόχο λοιπόν τον προσδιορισμό συγκεκριμένου λειτουργικού ρόλου της FRA10AC1 εντός της διαδικασίας της συναρμογής, εξετάστηκαν μόνο οι άμεσες αλληλεπιδράσεις μεταξύ των συστατικών του σωματίου συναρμογής ώστε να διερευνηθεί ο βιολογικός ρόλος των πρωτεϊνών που βρίσκονταν στην κοντινή «γειτονιά» της FRA10AC1 στο διευρυμένο δίκτυο του σωματίου συναρμογής.

Η επιλογή αλληλεπιδράσεων από δυο διαφορετικές κλάσεις της μετα-βάσης PICKLE εξασφάλιζε την ενσωμάτωση όλων των διαθέσιμων πληροφοριών για τις αλληλεπιδράσεις του σωματίου συναρμογής. Η πρώτη κλάση της μετα-βάσης περιλάμβανε μόνο τις άμεσες αλληλεπιδράσεις από τις 5 πρωτογενείς βάσεις δεδομένων και η δεύτερη κλάση περιείχε, βάσει συγκεκριμένων κριτηρίων, έμμεσες αλληλεπιδράσεις για τις οποίες όμως κρίθηκε απαραίτητη η περαιτέρω αξιολόγησή τους διότι ορισμένες ερευνητικές εργασίες μπορεί να περιέγραφαν και άμεσες αλληλεπιδράσεις (βλ. παρακάτω). Όσον αφορά την πρώτη κλάση, μία από τις μεθόδους ανίχνευσης αλληλεπιδράσεων χαρακτηρίζεται ως «ανασύσταση συμπλόκου» (reconstituted complex), και χαρακτηρίζει 483 αλληλεπιδράσεις. Ωστόσο, δεν ήταν ξεκάθαρο αν αυτή η μέθοδος αφορούσε μόνο άμεσες αλληλεπιδράσεις διότι περιγράφει την ταυτοποίηση αλληλεπιδράσεων μεταξύ μιας πρωτεΐνης-δολώματος και πρωτεϊνών που προέρχονται από κυτταρικό εκχύλισμα, οι οποίες είναι έμμεσες αλληλεπιδράσεις ή αλληλεπιδράσεις μεταξύ ανασυνδυασμένων και καθαρισμένων πρωτεϊνών που είναι άμεσες (βλ. Ενότητα 3.2.1.2.). Γι' αυτό το λόγο έγινε περαιτέρω διερεύνηση των βιβλιογραφικών αναφορών που περιέγραφαν αυτές τις αλληλεπιδράσεις, η οποία οδήγησε στην απομάκρυνση μεγάλου ποσοστού αλληλεπιδράσεων (~60%) που υφίσταντο μεταξύ μιας πρωτεΐνης-δολώματος και πρωτεϊνών ενός κυτταρικού εκχυλίσματος. Η επιλογή της δεύτερης κλάσης η οποία περιέχει έμμεσες αλληλεπιδράσεις που ταυτοποιήθηκαν σε τουλάχιστον μία μελέτη μικρής κλίμακας, πραγματοποιήθηκε διότι περιλάμβανε αλληλεπιδράσεις που είχαν ταυτοποιηθεί με πειράματα συν-ανοσοκατακρήμνισης (anti-tag ή anti bait co-immunoprecipitation). Τα πειράματα αυτά μπορεί να περιγράφουν και άμεσες αλληλεπιδράσεις, καθώς σε αυτή τη μέθοδο μπορεί να εξετάζονται και αλληλεπιδράσεις μεταξύ ανασυνδυασμένων και καθαρισμένων πρωτεϊνών εκτός από τις αλληλεπιδράσεις μίας πρωτεΐνης-στόχου με πρωτεΐνες που προέρχονται από κυτταρικό εκχύλισμα, γι' αυτό και κρίθηκε απαραίτητο να γίνει η περαιτέρω διερεύνησή τους. Η επιλογή της δεύτερης κλάσης αποσκοπούσε στην ενσωμάτωση όλων των διαθέσιμων πληροφοριών, αποφεύγοντας

την παράλειψη άμεσων αλληλεπιδράσεων που μπορεί να μετακινήθηκαν στη δεύτερη κλάση της μεταβάσης λόγω του τύπου του πειράματος που απέδιδε η εκάστοτε πρωτογενής βάση. Το ποσοστό αυτών των αλληλεπιδράσεων της δεύτερης κλάσης ήταν ~2%. Επιπλέον, στις κατηγορίες των ερευνητικών εργασιών που εξετάστηκαν, παρατηρήθηκαν αρκετές περιπτώσεις αλληλεπιδράσεων που ήταν καταχωρημένες στις βάσεις (~5%) οι οποίες δεν είχαν ταυτοποιηθεί στην αντίστοιχη πειραματική μελέτη. Αυτό οφείλεται στο γεγονός ότι η εισαγωγή των δεδομένων στις βάσεις πραγματοποιείται κυρίως μέσω της μελέτης των περιλήψεων και όχι ολόκληρων των ερευνητικών εργασιών με τη χρήση ενός ειδικού λεξιλογίου διερεύνησης (λέξεις-κλειδιά) (Huang et al. 2008, Zhou & He 2008).

Τέλος, δεδομένου του μικρού ποσοστού αλληλοεπικάλυψης μεταξύ των βάσεων δεδομένων (βλ. Ενότητα 1.5.1.3.), πλέον επιβεβαιώνεται η αναγκαιότητα άντλησης πληροφοριών για πρωτεϊνικές αλληλεπιδράσεις από διαφορετικές βάσεις δεδομένων. Το χαρακτηριστικό αυτό παρατηρήθηκε τόσο στην παρούσα εργασία, όσο και στο πλαίσιο της *in silico* ανακατασκευής ολόκληρου του δικτύου πρωτεϊνικών αλληλεπιδράσεων στον άνθρωπο (Klara et al. 2013). Η αλληλοεπικάλυψη μεταξύ των δεδομένων που προέκυψαν από τις 5 βάσεις του ανθρώπου ήταν αξιοσημείωτα μικρή για τις πρωτεϊνικές αλληλεπιδράσεις, και μηδενική για τις ερευνητικές εργασίες (βλ. Ενότητα 3.2.1.2.1.). Το ίδιο συνέβη και στην περίπτωση ολόκληρου του «interactome» του ανθρώπου στο οποίο βρέθηκαν μόνο 8 κοινές βιβλιογραφικές αναφορές μεταξύ των 5 βάσεων δεδομένων (Klara et al. 2013) που χρησιμοποιεί η PICKLE. Η μεταβάση PICKLE που χρησιμοποιήθηκε στην παρούσα εργασία, εκτός του ότι ενσωματώνει πληροφορίες από τις μεγαλύτερες και ευρέως χρησιμοποιούμενες πρωτογενείς βάσεις δεδομένων, χαρακτηρίζεται και από ιδιότητες που της προσδίδουν αυξημένη αξιοπιστία ως προς την ανάκτηση, την ενσωμάτωση και τη διαχείριση των πρωτογενών δεδομένων. Αυτές είναι: α) η κανονικοποίηση των κωδικών με τους οποίους είναι καταχωρημένες οι πρωτεΐνες στις πρωτογενείς βάσεις δεδομένων σε έναν κοινό κωδικό (identifier) ώστε να αποφευχθεί η αλληλοεπικάλυψη των καταχωρήσεων, δηλαδή, η επανάληψη ίδιων αλληλεπιδράσεων, β) η ταξινόμηση των αλληλεπιδράσεων σε άμεσες και έμμεσες αλληλεπιδράσεις καταγράφοντας τον τύπο του πειράματος που αναφέρεται στις πρωτογενείς βάσεις δεδομένων, γ) ο διαχωρισμός των αλληλεπιδράσεων με βάση την κλίμακα του πειράματος (μικρής, μεσαίας και μεγάλης κλίμακας) και δ) η απόδοση μίας τιμής αξιοπιστίας ανάλογα με τον αριθμό και το είδος των πειραμάτων που περιγράφουν μία αλληλεπίδραση.

Από τα παραπάνω γίνεται κατανοητό ότι για την ενσωμάτωση όλων των διαθέσιμων πληροφοριών για τις πρωτεϊνικές αλληλεπιδράσεις του σωματίου συναρμογής είναι απαραίτητη η χρήση μιας μετα-βάσης που περιλαμβάνει δεδομένα πρωτεϊνικών αλληλεπιδράσεων από διαφορετικές πρωτογενείς πηγές. Σε αντίστοιχες πειραματικές μελέτες, στις οποίες κατασκευάζεται το δίκτυο αλληλεπιδράσεων του σωματίου συναρμογής, δεν χρησιμοποιείται κάποια μετα-βάση (Akerman et al. 2015, Papasaikas et al. 2015, Pires et al. 2015) με εξαίρεση τους Hegele et al. (2012) οι οποίοι συγκρίνουν τις αλληλεπιδράσεις που βρέθηκαν πειραματικά με αυτές που είναι ενσωματωμένες στη μετα-βάση ConsensusPathDB. Η μετα-βάση αυτή δεν ενσωματώνει δεδομένα μόνο για πρωτεϊνικές αλληλεπιδράσεις αλλά και για γενετικές αλληλεπιδράσεις,

αλληλεπιδράσεις φαρμάκου-στόχου, και για βιοχημικά, σηματοδοτικά και μεταβολικά μονοπάτια γεγονός που δεν την κατατάσσει σε μια στοχευμένη μετα-βάση μόνο για πρωτεϊνικές αλληλεπιδράσεις όπως είναι η μετα-βάση PICKLE.

Το τελικό σύνολο δεδομένων περιλάμβανε 1617 αποκλειστικά άμεσες αλληλεπιδράσεις μεταξύ 472 πρωτεϊνών, οι οποίες αναφέρονταν σε 573 ερευνητικές εργασίες. Ο αριθμός των εργασιών αυξήθηκε στις 997 όταν ενσωματώθηκαν και πειράματα τα οποία αφορούσαν την ταυτοποίηση έμμεσων αλληλεπιδράσεων των 472 πρωτεϊνών. Το σύνολο αυτό ωστόσο εμπλουτίστηκε με 109 αλληλεπιδράσεις, οι 105 από τις οποίες προήλθαν αποκλειστικά από τους οργανισμούς-μοντέλα, 14 πρωτεΐνες και 6 ερευνητικές εργασίες, σύμφωνα με τα δεδομένα που αφορούν οργανισμούς-μοντέλα και η συλλογή τους έγινε με βάση την ιδέα των ιντερόλογων (interologs) (Nguyen et al. 2013).

Η αξιοπιστία του τελικού συνόλου ενισχύθηκε από το γεγονός ότι ένα σχετικά αυξημένο ποσοστό (~31%) των αλληλεπιδράσεων, λαμβάνοντας υπόψη και τις εργασίες που αφορούσαν τις έμμεσες αλληλεπιδράσεις, περιγράφονταν σε περισσότερες από μία ερευνητικές εργασίες, η κλίμακα των οποίων κυμαινόταν από 2 έως 44. Αυτό σημαίνει ότι ένας σημαντικός αριθμός αλληλεπιδράσεων είχε ταυτοποιηθεί με εναλλακτικούς τύπους πειραματικής μεθόδου. Η ύπαρξη διαφορετικών πειραμάτων για την ταυτοποίηση μιας αλληλεπίδρασης έρχεται σε συμφωνία με το γεγονός ότι τα αποτελέσματα που προκύπτουν από μελέτες μεγάλης κλίμακας απαιτούν-για την αύξηση της αξιοπιστίας τους-την ύπαρξη διαφορετικών πειραμάτων είτε μικρής ή μεγάλης κλίμακας (Hayes et al. 2016). Επιπλέον, ένα μικρό αλλά όχι αμελητέο ποσοστό πρωτεϊνικών αλληλεπιδράσεων (3%) ήταν εξελικτικά συντηρημένες και στον άνθρωπο.

Στο τελικό ανακατασκευασμένο δίκτυο ενσωματώθηκε: α) το 94% των δεδομένων πρωτεϊνικών αλληλεπιδράσεων (δηλαδή 1686 αλληλεπιδράσεις μεταξύ 444 πρωτεϊνών-κόμβων, με τις 173 από αυτές να σχηματίζουν και ομοδιμερή) και β) το 70% των πρωτεϊνών του συνόλου δεδομένων του σωματίου συναρμογής (Εικ. 43). Παρατηρήθηκε ότι το μεγαλύτερο ποσοστό των πρωτεϊνών που δεν συμπεριελήφθησαν στο δίκτυο είχαν ταυτοποιηθεί σε μία μόνο ερευνητική εργασία. Ειδικότερα είχαν ταυτοποιηθεί είτε στο υποσύμπλοκο E είτε σε μελέτες απομόνωσης ολόκληρου του σωματίου συναρμογής ή ήταν πρωτεΐνες άγνωστης λειτουργίας. Το αποτέλεσμα αυτό υποδεικνύει την πιθανή μη συνεισφορά τους στη διαδικασία της συναρμογής και την ανάγκη επιπρόσθετων πειραματικών μελετών για την πιστοποίηση της συμμετοχής τους.

Επιπλέον, και στα ανακατασκευασμένα δίκτυα των υποσυμπλόκων του σωματίου συναρμογής, το μεγαλύτερο ποσοστό των πρωτεϊνών (>70%) που είχαν ταυτοποιηθεί ως συστατικά τους ενσωματώθηκε στα τελικά δίκτυα (Εικ. 44 & 45) γεγονός που αυξάνει την αξιοπιστία του συνόλου δεδομένων των συστατικών των επιμέρους υποσυμπλόκων. Ο αριθμός των αλληλεπιδράσεων των επιμέρους δικτύων αυξάνει αναλογικά με τον αριθμό των συστατικών των επιμέρους υποσυμπλόκων. Ειδικότερα τα δίκτυα των υποσυμπλόκων B, B act και C παρουσιάζουν αρκετά υψηλό αριθμό αλληλεπιδράσεων (748, 467 και 734, αντίστοιχα), σε συμφωνία με τον μεγάλο αριθμό των συστατικών τους, 321, 228 και 313, αντίστοιχα. Ο μεγάλος αριθμός των

αλληλεπιδράσεων σε αυτά τα υποσύμπλοκα πιθανώς εξηγείται από το βασικό τους ρόλο στα δυο καταλυτικά βήματα της συναρμογής. Ο χαμηλότερος αριθμός των συστατικών και των αλληλεπιδράσεων του *B act* μπορεί να εξηγηθεί από το μικρότερο αριθμό δημοσιεύσεων οι οποίες αφορούν την απομόνωση των συστατικών του. Εναλλακτικά μπορεί να εξηγηθεί από το γεγονός ότι το στάδιο αυτό περιλαμβάνει αρκετές δυναμικές αλληλεπιδράσεις που είναι παροδικές λόγω των ποικίλων αναδιατάξεων που συμβαίνουν στο συγκεκριμένο βήμα (βλ. Ενότητα 1.4.). Έτσι μπορεί να μην έχει βρεθεί έως τώρα όλο το φάσμα των αλληλεπιδράσεων δεδομένης της μικρής διάρκειας αυτού του σταδίου και της έλλειψης σταθερότητας του. Το υποσύμπλοκο P εμφανίζει μικρό αριθμό συστατικών (116 πρωτεΐνες) και αλληλεπιδράσεων σε σχέση με τα υπόλοιπα υποσύμπλοκα, ο οποίος μπορεί να εξηγηθεί από: i) την ύπαρξη μίας ερευνητικής εργασίας για την απομόνωση των συστατικών του και ii) το γεγονός ότι σε αυτό το στάδιο έχει ολοκληρωθεί η διαδικασία της συναρμογής και ακολουθεί ουσιαστικά η αποσυναρμολόγηση του σωματίου συναρμογής. Για τα δίκτυα των αρχικών υποσυμπλόκων παρατηρείται ότι οι πρωτεΐνες και οι αλληλεπιδράσεις του E είναι περισσότερες από αυτές του A το οποίο πιθανώς συμβαίνει διότι το υποσύμπλοκο E περιέχει αρκετούς παράγοντες αναδιαμόρφωσης της χρωματίνης και μεταγραφικούς παράγοντες. Αυτοί οι παράγοντες μπορεί να είναι απαραίτητοι για να στρατολογήσουν τις πρωτεΐνες που συμμετέχουν στα επόμενα στάδια και έπειτα να απομακρύνονται. Έτσι πιθανώς μειώνεται ο αριθμός των αλληλεπιδράσεων στο υποσύμπλοκο A.

#### 4.2.4. Ο προσδιορισμός των μοναδικών πρωτεϊνικών αλληλεπιδράσεων στα διαδοχικά στάδια συναρμολόγησης του σωματίου συναρμογής δίνει μια προκαταρκτική εικόνα της δυναμικής του

Μια προσπάθεια ανάκτησης της δυναμικής του σωματίου συναρμογής πραγματοποιήθηκε μέσω της κατασκευής των δικτύων πρωτεϊνικών αλληλεπιδράσεων των υποσυμπλόκων του. Με αυτόν τον τρόπο δίνεται η δυνατότητα να βρεθούν οι παροδικές αλληλεπιδράσεις σε συγκεκριμένα στάδια συναρμολόγησης του σωματίου συναρμογής καθώς επίσης και το μοτίβο αλλαγής των αλληλεπιδράσεων κατά τη διάρκεια της διαδικασίας, δηλαδή τον τρόπο με τον οποίο οι πρωτεΐνες αλλάζουν «γείτονες». Στο πλαίσιο αυτό βρέθηκαν οι μοναδικές πρωτεϊνικές αλληλεπιδράσεις στα υποσύμπλοκα A, B, *B act* και C με βάση την πρωτεϊνική σύσταση του κάθε υποσυμπλόκου. Ο αριθμός των αλληλεπιδράσεων αυτών ήταν μικρός (25, 70, 29 και 104, αντίστοιχα) σε σχέση με τις συνολικές αλληλεπιδράσεις του κάθε υποσυμπλόκου (βλ. Ενότητα 3.2.2.). Το αποτέλεσμα αυτό επιβεβαιώνει την κοινά αποδεκτή άποψη που υπάρχει για τα διάφορα κυτταρικά σύμπλοκα, δηλαδή την ύπαρξη σταθερών κυτταρικών αλληλεπιδράσεων που διατηρούνται στο χρόνο για την αποφυγή διαταραχών των βασικών λειτουργιών του κυττάρου και παροδικών αλληλεπιδράσεων που συνεισφέρουν στην «πλαστικότητα» των διαφορετικών αντιδράσεων (Ahmed & Xing 2009, Li et al. 2014, Ou-Yang et al. 2014).

Μελετώντας τις αλληλεπιδράσεις των δικτύων των υποσυμπλόκων και των μοναδικών αλληλεπιδράσεων που λαμβάνουν χώρα σε αυτά, ορισμένα προκαταρκτικά συμπεράσματα μπορούσαν να εξαχθούν για το πρότυπο αλλαγής τους κατά τη διάρκεια

συναρμολόγησης του σωματίου συναρμογής. Μια αρχική παρατήρηση ήταν ότι οι αλληλεπιδράσεις μεταξύ των συστατικών του εξωσώματος εμφανίστηκαν μόνο στο υποσύμπλοκο B, με δεδομένο βέβαια ότι τα συστατικά του απομονώθηκαν μόνο σε αυτό το σύμπλοκο. Μέχρι τώρα ο μηχανισμός σύνδεσης του εξωσώματος με τη συναρμογή δεν είναι ξεκάθαρος, αν και έχει δείχθει ότι το σύμπλοκο TRAMP (βασικό συστατικό του εξωσώματος, Callahan & Butler 2010) και γενικότερα το εξώσωμα είναι υπεύθυνα για τον έλεγχο της σωστής λήξης της μεταγραφής και της σωστής εξέλιξης της συναρμογής (Kilchert et al. 2016, Kong et al. 2014). Έτσι, η στρατολόγηση των συστατικών του εξωσώματος μπορεί να απαιτείται να γίνει πριν από τα καταλυτικά βήματα της συναρμογής που λαμβάνουν χώρα στα υποσύμπλοκα B *act* και C, τα οποία συμβαίνουν ταυτόχρονα με την επιμήκυνση της μεταγραφής, ώστε οι πρωτεΐνες του εξωσώματος να έχουν τη δυνατότητα να «επιβλέψουν» αν η μεταγραφή τερματιστεί σωστά και αν η αποκοπή των ιντρονίων γίνει με τον σωστό τρόπο. Η υπόθεση αυτή συνάδει με την στρατολόγηση των πρωτεϊνών και των αλληλεπιδράσεων αυτού του συμπλόκου κατά τον σχηματισμό του υποσυμπλόκου B.

Η διερεύνηση της αλλαγής των «γείτονων» των πρωτεϊνών στα διαφορετικά υποσύμπλοκα του σωματίου συναρμογής στηρίχτηκε στην πρωτεϊνική σύσταση αυτών. Δηλαδή, εάν υπήρχε μια αλληλεπίδραση μεταξύ δυο πρωτεϊνών, αυτή θα εντασσόταν σε εκείνο το υποσύμπλοκο στο οποίο και οι δυο πρωτεΐνες του ζεύγους είχαν ταυτοποιηθεί ως συστατικά του. Αν μία από τις δυο δεν αποτελούσε συστατικό ενός συγκεκριμένου υποσυμπλόκου τότε δε θα υφίσταται αλληλεπίδραση στο υποσύμπλοκο αυτό. Βάσει αυτής της προϋπόθεσης, ένα παράδειγμα αποτελούν η U2AF1 και η SF3B2, δυο πρώτοι «γείτονες» της FRA10AC1, οι οποίες διαδραματίζουν βασικό ρόλο στην αναγνώριση της 3' θέσης συναρμογής, στην στρατολόγηση του U2snRNP στο πρόδρομο μετάγραφο καθώς και στα καταλυτικά βήματα της συναρμογής (Voith von Voithenberg et al. 2016, Lardelli et al. 2010). Η συνεισφορά της U2AF1 στην αναγνώριση της 3' θέσης συναρμογής και στο 1<sup>ο</sup> καταλυτικό βήμα της συναρμογής επιβεβαιώνεται από τον τρόπο με τον οποίο αλλάζουν οι «γείτονες» της. Ειδικότερα, στα υποσύμπλοκα A και B, η U2AF1 αλληλεπιδρά με συστατικά του U2snRNP (SF3B1, SF3B2) και με πρωτεΐνες σχετιζόμενες με αυτό (U2AF2). Στη συνέχεια, στα υποσύμπλοκα B και B *act*, αλληλεπιδρά με πρωτεΐνες του συμπλόκου U4/U6.U5 (π.χ. PRPF4) και μόνο στο B *act* παρουσιάζει μια αλληλεπίδραση με την ελικάση RNA, DHX38 (συνεισφέρει σημαντικά στις αναδιατάξεις που λαμβάνουν χώρα στα καταλυτικά βήματα της συναρμογής, Cordin & Beggs 2013, Tseng et al. 2011) και με την FRA10AC1. Οι μοναδικές αλληλεπιδράσεις της U2AF1 με την DHX38 και FRA10AC1 στο B *act* μπορεί να υποδεικνύουν ότι αυτή η πρωτεΐνη αποτελεί παράγοντα στρατολόγησης πρωτεϊνών οι οποίες συμμετέχουν στα επόμενα βήματα της συναρμογής, με βάση και πάλι την παραδοχή ότι αυτές οι πρωτεΐνες έχουν απομονωθεί μαζί μόνο σε αυτό το υποσύμπλοκο. Σε παρόμοιο πλαίσιο, στα αρχικά υποσύμπλοκα, η SF3B2 αλληλεπιδρά με πρωτεΐνες του συμπλόκου U2snRNP (SF3B1, SF3B2, U2AF1). Στη συνέχεια, από το υποσύμπλοκο B ή B *act* και έπειτα, αλληλεπιδρά με ελικάσες RNA οι οποίες συνεισφέρουν στην αποσυναρμολόγηση (unwinding) του συμπλόκου U4/U6 (π.χ. SNRNP200, DHX16, Cordin & Beggs 2013), με συστατικά του U4/U6.U5 (π.χ. PRPF8) καθώς και με παράγοντες του δεύτερου καταλυτικού

βήματος (SLU7). Επιπλέον, από το υποσύμπλοκο B και τα άλλα υποσύμπλοκα που έπονται αυτού, η SF3B2 συνδέεται με παράγοντες που συμμετέχουν στην έξοδο του mRNA (π.χ. THOC1) από τον πυρήνα και στην αποικοδόμηση αυτού όταν δεν είναι κατάλληλο για να μεταφραστεί (RBM8A). Το γεγονός αυτό υποδεικνύει ότι η πρωτεΐνη αυτή μπορεί να συνδέει τη συναρμογή με άλλες διαδικασίες της επεξεργασίας του RNA.

Επιπρόσθετα, οι αλληλεπιδράσεις ενός πρώτου «γείτονα» της FRA10AC1 (MFAP1) αλλάζουν κατά τη διάρκεια συναρμολόγησης του σωματίου συναρμογής γεγονός το οποίο συνάδει με το λειτουργικό του ρόλο. Η MFAP1 συμμετέχει στη μετάβαση από το σύμπλοκο B στο σύμπλοκο B *act* του σωματίου συναρμογής για την πραγματοποίηση του πρώτου καταλυτικού βήματος (Andersen et al. 2008). Οι αλληλεπιδράσεις της στο σύμπλοκο A είναι ελάχιστες (4) ενώ πληθαίνουν από το B ή B *act* και στα ακόλουθα υποσύμπλοκα. Ειδικότερα στο B ή B *act* και στα άλλα υποσύμπλοκα που ακολουθούν, αλληλεπιδρά με συστατικά του συμπλόκου U4/U6.U5 (π.χ. SART1) και με ελικάσες RNA (DHX38, DHX8) οι οποίες συνεισφέρουν στις αναδιατάξεις των δυο καταλυτικών βημάτων. Οι αλληλεπιδράσεις αυτές θα μπορούσαν να επιβεβαιώσουν τη συνεισφορά της MFAP1 στην μετατροπή του υποσυμπλόκου B σε B *act* ώστε να πραγματοποιηθεί η πρώτη αντίδραση μετεστεροποίησης.

Τα παραπάνω παραδείγματα αποδίδουν μια προκαταρκτική εικόνα της δυναμικής των αλληλεπιδράσεων κατά τη διάρκεια συναρμολόγησης του σωματίου συναρμογής βάσει των πληροφοριών που υπάρχουν για τη σύσταση των διαφορετικών υποσυμπλόκων. Για την περαιτέρω ανάλυση της δυναμικής αυτών των αλληλεπιδράσεων θα πρέπει να εφαρμοστούν εναλλακτικές πειραματικές προσεγγίσεις (βλ. Ενότητα 4.3.3.).

#### 4.2.5. Ο αυτοματοποιημένος προσδιορισμός των οντολογικών κατηγοριών (GO terms) των πρωτεϊνών απαιτεί επιπρόσθετη αξιολόγηση

Το 96% των πρωτεϊνών-συστατικών του σωματίου συναρμογής κατατάσσεται σε 19 όρους γονιδιακής οντολογίας βιολογικής διαδικασίας (GO terms), σύμφωνα με τη βάση δεδομένων DAVID και UniProt και σε συνδυασμό με εξονυχιστική εξέταση της σχετικής βιβλιογραφίας για καθεμία από τις 630 πρωτεΐνες του συνόλου δεδομένων (βλ. Ενότητα 3.2.1.1.1.), με εξαίρεση αυτών που ήταν άγνωστης λειτουργίας. Οι βάσεις DAVID και UniProt χρησιμοποιήθηκαν επειδή παρουσιάζουν ορισμένα πλεονεκτήματα (βλ. παρακάτω). Η εξονυχιστική εξέταση της σχετικής βιβλιογραφίας για καθεμία από τις 630 πρωτεΐνες του συνόλου δεδομένων, παρόλο που ήταν χρονοβόρα, ήταν απαραίτητη καθώς σε κάποιες περιπτώσεις τα ανακτηθέντα δεδομένα για την οντολογία των γονιδίων (GO terms) εμφανίζουν λάθη. Το αυξημένο ποσοστό λάθους έχει επιβεβαιωθεί και από αρκετές μελέτες που ελέγχουν την ορθότητα και την εγκυρότητα των οντολογικών κατηγοριών (Jones et al. 2007, Koskinen et al. 2015, Schnoes et al. 2013, Škunca et al. 2012, Yu et al. 2017).

Η DAVID ανακτά τις πληροφορίες για τους όρους γονιδιακής οντολογίας από τρεις βάσεις δεδομένων οι οποίες όμως έχουν κοινό σημείο αναφοράς, δηλαδή, αποδίδουν οντολογικούς όρους σύμφωνα με τη σύμπραξη γονιδιακής οντολογίας (Gene ontology consortium), η οποία αποτελεί τη συνήθη επιλογή από τα μέλη της επιστημονικής

κοινότητας. Το Gene ontology consortium χρησιμοποιεί τρεις μεθόδους απόδοσης GO terms στα γονίδια (βλ. Υλικά & Μέθοδοι). Το ποσοστό των οντολογικών κατηγοριών που προσδιορίζονται μέσω της προσωπικής αξιολόγησης των βιβλιογραφικών αναφορών από ειδικούς επιμελητές (1<sup>η</sup> μέθοδος) είναι πάρα πολύ μικρό, της τάξεως του 0,6% (The Gene Ontology Consortium 2013, 2015). Ωστόσο, το ποσοστό λάθους είναι αυξημένο καθώς κυμαίνεται στο 28% με 30% (Jones et al. 2007, Schnoes et al. 2013), γεγονός το οποίο μπορεί να οφείλεται στην χρήση, κατά κύριο λόγο, πειραμάτων μεγάλης κλίμακας για τον προσδιορισμό του λειτουργικού ρόλου των πρωτεϊνών. Το μεγαλύτερο ποσοστό των GO terms (99,4%) προσδιορίζεται μέσω των άλλων δυο μεθόδων οι οποίες χρησιμοποιούν αυτοματοποιημένους τρόπους ανάκτησης GO terms με την εποπτεία επιμελητών (2<sup>η</sup> μέθοδο) ή χωρίς αυτήν (3<sup>η</sup> μέθοδο). Το ποσοστό λάθους σε αυτές τις περιπτώσεις ανέρχεται στο 49% (Jones et al. 2007, Škunca et al. 2012). Ένα δεύτερο πλεονέκτημα της DAVID είναι ότι αποδίδει έναν βαθμό σημαντικότητας σε κάθε οντολογικό όρο βάσει του εμπλουτισμού του στα γονίδια που αναζητήθηκαν και παρέχει τη δυνατότητα ομαδοποίησης παρόμοιων όρων γονιδιακής οντολογίας οι οποίοι περιλαμβάνουν κοινά γονίδια. Τέλος, η βάση αυτή θεωρήθηκε επαρκώς ενημερωμένη καθώς η τελευταία ενημέρωσή της έγινε τον Οκτώβριο του 2016 (έκδοση 6.8).

Το πλεονέκτημα της βάσης δεδομένων UniProt, η οποία ανήκει στη σύμπραξη γονιδιακής οντολογίας (Gene ontology consortium), είναι ότι διαθέτει συμπληρωματικές ομάδες επιμελητών οι οποίες κάνουν επιπρόσθετη αξιολόγηση των οντολογικών κατηγοριών που ήδη υπάρχουν (Barrell et al. 2009, Dimmer et al. 2012). Ωστόσο, και σε αυτή τη βάση, αν και γίνεται επιπρόσθετη αξιολόγηση, συνεχίζουν να εμφανίζονται κάποιοι περιορισμοί και ασυνέπειες στον τρόπο απόδοσης των GO terms. Ειδικότερα, όσον αφορά το σύνολο των οντολογικών όρων που προσδιορίζονται μέσω προσωπικής αξιολόγησης των βιβλιογραφικών αναφορών από επιμελητές (manual curation) το οποίο αποτελεί το 0,75% των συνολικών GO terms της βάσης, οι περιορισμοί αφορούν το είδος των βιβλιογραφικών αναφορών. Για παράδειγμα, η λειτουργική κατηγοριοποίηση του 25% των πρωτεϊνών της UniProt που αφορά την απόδοση των GO terms έχει στηριχθεί σε πειράματα μεγάλης κλίμακας (αναφέρονται σε περισσότερες από 100 πρωτεΐνες) τα οποία αποτελούν το 0,14% των συνολικών βιβλιογραφικών αναφορών που εξετάζονται από τους επιμελητές. Αντίθετα, τα πειράματα που αναφέρονται σε μία πρωτεΐνη και αυτά που είναι μικρής κλίμακας (2-9 πρωτεΐνες) αποτελούν το 96% των βιβλιογραφικών αναφορών και έχουν χρησιμοποιηθεί για τον προσδιορισμό του λειτουργικού ρόλου του 53% των πρωτεϊνών της UniProt που αφορά την απόδοση των GO terms (Schnoes et al. 2013). Η ύπαρξη λοιπόν μεγάλου ποσοστού πειραμάτων μεγάλης κλίμακας είναι υπεύθυνη για την απόδοση ίδιων GO term σε μεγάλο αριθμό πρωτεϊνών γεγονός το οποίο αυξάνει την πιθανότητα για την μη εγκυρότητα των αποτελεσμάτων (Schnoes et al. 2013).

Το αυξημένο ποσοστό λάθους στις βάσεις δεδομένων προσδιορισμού γονιδιακής οντολογίας επιβεβαιώθηκε και στην παρούσα εργασία. Ειδικότερα παρατηρήθηκε ότι ένα σχετικά υψηλό ποσοστό πρωτεϊνών (30%) συμμετείχε σε διαφορετικές λειτουργικές διαδικασίες ή δεν σχετιζόταν με το συγκεκριμένο όρο γονιδιακής οντολογίας που ήταν καταχωρημένος στη βάση δεδομένων DAVID. Μια συχνή

περίπτωση λανθασμένης απόδοσης GO term ήταν ότι σε αρκετές πρωτεΐνες αποδίδονταν ο οντολογικός όρος της συναρμογής βάσει ερευνητικών εργασιών οι οποίες αφορούσαν την απομόνωση των πρωτεϊνών αυτών ως συστατικά του σωματίου συναρμογής και όχι βάσει της τεκμηριωμένης λειτουργίας τους στη διαδικασία της συναρμογής. Επιπρόσθετα, ένα ποσοστό της τάξεως του 8% ήταν πολυλειτουργικές πρωτεΐνες. Η επιλογή της Uniprot βοήθησε διότι μέσω αυτής μειώθηκε κατά 36% το ποσοστό των πρωτεϊνών που δεν τους αποδίδονταν οντολογικοί όροι (52 πρωτεΐνες).

Η συνήθης διαδικασία σε αντίστοιχες ερευνητικές εργασίες που αφορούν το σωματίο συναρμογής είναι η ανάκτηση πληροφοριών είτε από βάσεις δεδομένων που αναφέρουν όρους γονιδιακής οντολογίας, είναι ευρέως χρησιμοποιούμενες και θεωρούνται έγκυρες (NCBI και Uniprot) (Papasaikas et al. 2015) είτε από συνδυασμό δυο βάσεων (π.χ. Spliceosome Database, KEGG PATHWAY) που όμως εμφανίζουν κάποιους περιορισμούς (Akerman et al. 2015). Ειδικότερα, στη δεύτερη περίπτωση, η Spliceosome Database δεν περιέχει όρους γονιδιακής οντολογίας και κάνει κατηγοριοποίηση των συστατικών της βάσει των συμπλόκων στα οποία έχουν ταυτοποιηθεί και βάσει των χαρακτηριστικών περιοχών της δομής τους (motifs) ή λειτουργικών περιοχών (domains) που διαθέτουν (Cvitkovic & Jurica 2013). Στην KEGG PATHWAY database ο ίδιος οντολογικός όρος αποδίδεται σε όλες τις πρωτεΐνες ενός συγκεκριμένου μονοπατιού. Για παράδειγμα, στο μονοπάτι της συναρμογής (ko03041) υπάρχει παραπομπή στον όρο γονιδιακής οντολογίας κυτταρικού διαμερίσματος «spliceosomal complex». Όλες οι πρωτεΐνες που ανήκουν σε αυτό το μονοπάτι χαρακτηρίζονται με αυτόν τον όρο οντολογίας. Άρα αυτή η βάση καθίσταται ακατάλληλη για τη λειτουργική κατηγοριοποίηση των πρωτεϊνών. Στην εργασία των Akerman et al. (2015) δε χρησιμοποιούνται οι επίσημοι οντολογικοί όροι (GO terms) αλλά πραγματοποιείται διαφορετικό είδος λειτουργικής κατηγοριοποίησης σύμφωνα με την ένταξη των πρωτεϊνών σε διαφορετικά σύμπλοκα και διαφορετικά βιολογικά μονοπάτια.

Έτσι, γίνεται κατανοητό ότι η περαιτέρω επεξεργασία αυτού του είδους των δεδομένων είναι απολύτως απαραίτητη για την εξαγωγή έγκυρων συμπερασμάτων σχετικά με την λειτουργία των πρωτεϊνών, εφόσον βέβαια αυτό είναι εφικτό, ανάλογα με τον όγκο των δεδομένων που εξετάζεται κάθε φορά. Ειδικότερα, στο ερώτημα της παρούσας εργασίας, η παραπάνω διαδικασία ήταν απαραίτητη διότι η ανακατασκευή του δικτύου του σωματίου συναρμογής είχε ως στόχο την εύρεση συγκεκριμένου βιολογικού ρόλου αλλά και πολλαπλών ρόλων της καθεμίας από τις πρωτεΐνες που βρίσκονταν στην ευρύτερη «γειτονιά» της FRA10AC1 ώστε να μπορούν να εξαχθούν ορθά βιολογικά συμπεράσματα για συγκεκριμένη βιολογική λειτουργία της FRA10AC1 εντός της διαδικασίας της συναρμογής.

#### 4.2.6. Ο προσδιορισμός των GO terms των πρωτεϊνών του σωματίου συναρμογής εμπλουτίζει το σύνολο δεδομένων με πρωτεΐνες που συμμετέχουν και σε άλλες διαδικασίες της επεξεργασίας και μεταφοράς του mRNA, πλην της συναρμογής

Για τις 540 από τις 630 πρωτεΐνες του συνόλου δεδομένων του σωματίου συναρμογής, υπάρχουν πληροφορίες στη βιβλιογραφία για το λειτουργικό τους ρόλο. Αναλύοντας τα αποτελέσματα της λειτουργικής κατηγοριοποίησης των πρωτεϊνών-



συστατικών του σωματίου συναρμογής, παρατηρείται ότι σχεδόν οι μισές από αυτές (44,6%) συμμετέχουν στη διαδικασία της συναρμογής και της μεταγραφής/ρύθμισης της μεταγραφής. Πρόκειται για κάτι αναμενόμενο καθώς οι δυο αυτές κυτταρικές διεργασίες είναι γνωστό ότι λαμβάνουν χώρα ταυτόχρονα (Alpert et al. 2017, Saldi et al. 2016), όπως έχει αναλυθεί και στην Ενότητα 1.4.2. της εισαγωγής. Το αντίστοιχο ποσοστό ξεπερνάει το 50% αν ληφθεί υπόψη και ο αριθμός των παραγόντων που συμμετέχουν στην αναδιαμόρφωση της χρωματίνης. Οι παράγοντες αυτοί συνεισφέρουν στη ρύθμιση της μεταγραφής (Venkatesh & Workman 2015), στη συναρμολόγηση των αρχικών υποσυμπλόκων του σωματίου συναρμογής και στη ρύθμιση της συναρμογής με έμμεσο τρόπο μέσω της μεταγραφής (de Almeida & Carmo-Fonseca 2012). Πρόκειται δηλαδή για τρεις μηχανισμούς οι οποίοι συνδέονται άρρηκτα μεταξύ τους (de Almeida & Carmo-Fonseca 2012, Allemand et al. 2008) (βλ. Ενότητα 1.4.2.).

Επιπρόσθετα, σημαντικό ποσοστό αντιπροσώπευσης παρουσιάζουν οι οντολογικές κατηγορίες της προσθήκης καλύπτρας στο 5' άκρο του πρόδρομου μεταγράφου και της επεξεργασίας του 3' άκρου αυτού (~2,6%), της εξόδου του ώριμου μεταγράφου από τον πυρήνα (4,1%) καθώς και της αποικοδόμησης των μη κατάλληλων μεταγράφων για τη μετάφρασή τους σε πρωτεΐνες (3,6%). Το αποτέλεσμα αυτό θεωρείται επίσης αναμενόμενο, διότι είναι γνωστό ότι το mRNA υφίσταται, εκτός της συναρμογής, και άλλα στάδια επεξεργασίας μέχρι την έξοδό του στο κυτταρόπλασμα για να μεταφραστεί. Όλοι οι παράγοντες που συμμετέχουν στην επεξεργασία και μεταφορά του mRNA στρατολογούνται σταδιακά σχηματίζοντας το αγγελιοφόρο ριβονουκλεοπρωτεϊνικό σωματίδιο mRNP. Στην οντολογική κατηγορία της εξόδου του mRNA από τον πυρήνα ενσωματώθηκαν, μεταξύ άλλων, το ετεροδιμερές σύμπλοκο NXF1–P15, οι 9 από τις 13 πρωτεΐνες του συμπλόκου TREX (transcription-export complex) και οι 8 από τις 9 πρωτεΐνες του συμπλόκου EJC (exon junction complex). Τα συστατικά του EJC ενσωματώθηκαν επιπλέον στην οντολογική κατηγορία της αποικοδόμησης των ακατάλληλων μεταγράφων για μετάφραση. Αυτή η κατηγορία περιείχε, μεταξύ άλλων, και 10 από τις 11 πρωτεΐνες του πυρηνικού εξωσώματος. Το σύμπλοκο TREX (Barbosa et al. 2012) αναγνωρίζει τα μόρια mRNA και σε συνεργασία με το σύμπλοκο NXF1–P15 τα μεταφέρει στις νουκλεοπορίνες ώστε να εξαχθούν από τον πυρήνα (Katahira 2015, Elbarbary & Maquat 2016). Το σύμπλοκο EJC, που διαδραματίζει καθοριστικό ρόλο στη λειτουργική σύνδεση της συναρμογής τόσο με την έξοδο των mRNA από τον πυρήνα όσο και με την αποικοδόμηση των mRNA που δεν είναι κατάλληλα για να μεταφραστούν, στρατολογείται εκατέρωθεν των ορίων των εξονίων. Όσον αφορά τον πρώτο ρόλο, το σύμπλοκο EJC περιέχει πρωτεΐνες οι οποίες αλληλεπιδρούν με παράγοντες εξόδου των μεταγράφων (π.χ. συστατικά του συμπλόκου TREX) συμβάλλοντας στην έξοδό τους στο κυτταρόπλασμα. Σχετικά με την αποικοδόμηση των mRNA που διαθέτουν πρόωρο κωδικόνιο λήξης, τα συστατικά του EJC «συνοδεύουν» το μετάγραφο έως το κυτταρόπλασμα και δρουν ως «αναγνωριστικά» μόρια για πρωτεΐνες που αλληλεπιδρούν με το ριβόσωμα προκαλώντας την παύση της μετάφρασης στο σημείο που υπάρχει λάθος και την έναρξη της αποικοδόμησης του mRNA (Chuang et al. 2015). Ένα άλλο σύμπλοκο που ελέγχει την ακεραιότητα του mRNA αλλά και άλλων ειδών RNA στον πυρήνα, είναι

το πυρηνικό εξώσωμα το οποίο αποικοδομεί τα RNA στα οποία δεν γίνεται σωστά ο τερματισμός της μεταγραφής, η πολυαδενυλίωση και η αφαίρεση των ιντρονίων (Kilchert et al. 2016, Pan et al. 2015).

Τέλος, δυο οντολογικές κατηγορίες που εμφανίζουν σχετικά αυξημένο ποσοστό είναι αυτές της μετάφρασης (4%) και της μίτωσης (4%). Από το σύνολο των 31 πρωτεϊνών που κατηγοριοποιήθηκαν με βάση την ανάλυση γονιδιακής οντολογίας (GO analysis) ότι έχουν λειτουργικό ρόλο στην μετάφραση, οι 22 έχουν εντοπιστεί στον πυρήνα είτε με βάση ερευνητικές δημοσιεύσεις ή με δεδομένα της βάσης δεδομένων Human Protein Atlas. Για τις 5 από αυτές είναι γνωστός ο πολυλειτουργικός τους ρόλος. Ειδικότερα αυτές οι πρωτεΐνες διαδραματίζουν βασικό ρόλο και στη μετάφραση καθώς δρουν ως «αναγνωριστικά» μόρια για παράγοντες που συμμετέχουν στη μετάφραση των mRNAs (Michlewski et al. 2008) και παράλληλα έχουν ρόλο και στην συναρμογή. Δύο από αυτές αποτελούν συστατικά του συμπλόκου EJC (POLDIP3, CASC3) και «συνοδεύουν» το μετάγραφο στο κυτταρόπλασμα και εκεί: α) στρατολογούν κινάσες που φωσφορυλιώνουν τις ριβοσωμικές υπομονάδες και παράγοντες μετάφρασης, ενεργοποιώντας την έναρξη της μετάφρασης (Ma et al. 2008) ή β) συνδέονται με υπομονάδες του συμπλόκου έναρξης eIF3 της μετάφρασης το οποίο στρατολογεί τη μικρή ριβοσωμική υπομονάδα και συνεισφέρει στην αναγνώριση των κωδικονίων έναρξης (Chazal et al. 2013). Οι υπόλοιπες τρεις (SRSF1, hnRNPk, RBM4) είναι παράγοντες συναρμογής και εναλλακτικής συναρμογής, «συνοδεύουν» το μετάγραφο στο κυτταρόπλασμα όπου συνεισφέρουν στη ρύθμιση της έναρξης της μετάφρασης μέσω της στρατολόγησης παραγόντων έναρξης της μετάφρασης (π.χ. πρωτεΐνες του συμπλόκου eIF4) οι οποίοι συνδέονται είτε στην 5' καλύπτρα του mRNA είτε σε άλλες περιοχές του mRNA (π.χ. 3' μη μεταφραζόμενη περιοχή) (Bomszyk et al. 2004, Habelhah et al. 2001, Michlewski et al. 2008). Συνεπώς για τις υπόλοιπες πυρηνικές πρωτεΐνες που δεν έχουν γνωστό ρόλο στην επεξεργασία του mRNA, μπορεί να υποθεθεί είτε ότι οι πρωτεΐνες αυτές έχουν πιθανόν κάποια πυρηνική λειτουργία η οποία δεν έχει ταυτοποιηθεί ακόμα ή ότι παράγουν εναλλακτικές ισομορφές με πυρηνική λειτουργία. Οι ερευνητικές εργασίες που χρησιμοποιήθηκαν για την ανακατασκευή του πρωτεόματος του σωματίου συναρμογής δεν περιέχουν πληροφορίες για τις πρωτεϊνικές ισομορφές. Ένα παράδειγμα που υποδεικνύει ότι η υπόθεση πυρηνικών εναλλακτικών ισομορφών είναι πιθανή, είναι η περίπτωση του παράγοντα επιμήκυνσης της μετάφρασης eEF1A που ταυτοποιήθηκε ως συστατικό του σωματίου συναρμογής. Υπάρχουν δυο εναλλακτικές πρωτεϊνικές ισομορφές, η eEF1A2 η οποία λειτουργεί ως παράγοντας επιμήκυνσης της μετάφρασης και η eEF1A1 που συμμετέχει στη μεταγραφή γονιδίων σε απόκριση στο θερμικό σοκ (Vera et al. 2014). Στα πειράματα απομόνωσης των συστατικών του σωματίου συναρμογής δεν προσδιορίζεται ποια ισομορφή απομονώνεται ενώ παράλληλα η ανάλυση γονιδιακής οντολογίας (GO analysis) που εφαρμόζεται αφορά κατά κανόνα την κύρια πρωτεϊνική ισομορφή. Είναι γνωστό ότι οι εναλλακτικές ισομορφές μιας πρωτεΐνης μπορεί να εμφανίζουν διαφορετική υποκυτταρική τοπολογία (Nakao et al. 2005) ενώ μία πρόσφατη ερευνητική εργασία (Yang et al. 2016) που αφορούσε περίπου 250 πρωτεϊνικές ισομορφές 160 γονιδίων έδειξε ότι τουλάχιστον το 50% των ισομορφών παρουσιάζουν διαφορετικό πρότυπο στις αλληλεπιδράσεις τους σε σχέση με τις κύριες

ισομορφές τους, δηλαδή συνδέονται με πρωτεΐνες που συμμετέχουν σε διαφορετικές βιολογικές διεργασίες. Οι υπόλοιπες 9 πρωτεΐνες που δεν εμφανίζουν πυρηνική τοπολογία (π.χ. EEF2, EIF2AK2, DHX29, αμινοακυλο-tRNA συνθετάση της αργινίνης- RARS) μπορεί να απομονώθηκαν ως συστατικά του σωματίου συναρμογής διότι δημιουργούν εκ νέου αλληλεπιδράσεις, κατά τη διαδικασία της *in vitro* συναρμογής ή της ανοσοκατακρήμνισης που πραγματοποιείται για την απομόνωση του σωματίου, με τις πολυλειτουργικές πρωτεΐνες που συμμετέχουν τόσο στη συναρμογή όσο και στη μετάφραση. Αυτή η υπόθεση εκ πρώτης όψεως δεν συνάδει με την χρήση πυρηνικού εκχυλίσματος στις αντίστοιχες εργασίες ωστόσο, κατά κανόνα τα πυρηνικά εκχυλίσματα δεν είναι απόλυτα απαλλαγμένα από κυτταροπλασματικά συστατικά αλλά είναι απλώς αρκετά εμπλουτισμένα με συστατικά του πυρήνα. Αυτός είναι ίσως ο λόγος που για τις συγκεκριμένες 9 πρωτεΐνες ταυτοποιήθηκε μικρός αριθμός πεπτιδίων.

Αντίστοιχα, από τις 30 πρωτεΐνες που κατηγοριοποιήθηκαν στην οντολογική κατηγορία της μίτωσης, το 27% αυτών (9) είναι πολυλειτουργικές δηλαδή συμμετέχουν στη συναρμογή, στη μεταγραφή ή στην αναδιαμόρφωση της χρωματίνης αλλά έχουν και ενεργό ρόλο στη διαδικασία της μίτωσης και πιθανώς απομονώθηκαν ως συστατικά του σωματίου λόγω του ρόλου τους στην επεξεργασία του mRNA. Ειδικότερα, οι πρωτεΐνες ZNF830, NCL, ERH, SSRP1 (μεταγραφή) και HNRNPU (συναρμογή/εναλλακτική συναρμογή, μεταγραφή) συμμετέχουν στη στοίχιση και στο διαχωρισμό των χρωμοσωμάτων (Vandormael-Pournin et al. 2005), στην σύνδεση των μικροσωληνίσκων με τον κινητοχώρο των χρωμοσωμάτων (Fujimura et al. 2012, Ma et al. 2011), στον πολυμερισμό των μικροσωληνίσκων (Zeng et al. 2010) και στην οργάνωση της μιτωτικής ατράκτου (Ma et al. 2007, 2011), ενώ αυξάνουν τη διάρκεια της διαδικασίας της μίτωσης (Ma et al. 2007, Ma et al. 2011). Επιπλέον, η πρωτεΐνη RBMX (συναρμογή/εναλλακτική συναρμογή) αποτελεί ρυθμιστικό μόριο κατά τη σύναψη των αδελφών χρωματίδων (sister chromatid cohesion) (Matsunaga et al. 2012) και η IK (συναρμογή) συμμετέχει στην ενεργοποίηση του σημείου ελέγχου οργάνωσης της ατράκτου (spindle assembly checkpoint) το οποίο είναι απαραίτητο για τον έλεγχο της ακεραιότητας της μιτωτικής ατράκτου και της σύνδεσης των αδελφών χρωματίδων σε αυτή (Yeh et al. 2012). Τέλος, δυο πρωτεΐνες PPP1R10 και RUVBL1 (μεταγραφή και αναδιαμόρφωση της χρωματίνης, αντίστοιχα) συμμετέχουν στην αποσυσπείρωση των χρωμοσωμάτων κατά το τέλος της μίτωσης (Landsverk et al. 2005, Magalska et al. 2014). Επιπλέον, από τις υπόλοιπες 21 πρωτεΐνες οι οποίες με βάση την ανάλυση GO συμμετέχουν μόνο στη διαδικασία της μίτωσης, οι 17 έχουν εντοπιστεί και στον πυρήνα οπότε, με το ίδιο σκεπτικό που αναλύθηκε παραπάνω, μπορεί να έχουν κάποιο επιπλέον, ως τώρα άγνωστο, λειτουργικό ρόλο και στην επεξεργασία του mRNA. Εναλλακτικά, για τις 4 πρωτεΐνες που δεν εμφανίζουν πυρηνικό εντοπισμό, μπορεί να υποθεθεί ότι απομονώθηκαν ως συστατικά του σωματίου λόγω εκ νέου αλληλεπίδρασής τους, με τις πολυλειτουργικές πρωτεΐνες που συμμετέχουν τόσο στη συναρμογή, στη μεταγραφή ή στην αναδιαμόρφωση της χρωματίνης όσο και στη μίτωση. Αυτές οι νέες αλληλεπιδράσεις μπορούν να αιτιολογηθούν είτε με βάση τον λόγο που προαναφέρθηκε, δηλαδή την καθαρότητα του πυρηνικού εκχυλίσματος, ή με το γεγονός ότι ένα ποσοστό των κυττάρων από τα οποία απομονώθηκε εκχύλισμα

μπορεί να ήταν σε διαδικασία μίτωσης καθώς δεν αναφέρεται να χρησιμοποιούνται συγχρονισμένες καλλιέργειες. Και στην περίπτωση των 4 πρωτεϊνών που εντάχθηκαν στην οντολογική κατηγορία της μίτωσης, ο αριθμός των πεπτιδίων που ταυτοποιήθηκαν ήταν πολύ μικρός.

Από τα παραπάνω παρατηρείται ότι αρκετές πρωτεΐνες μπορεί να ανήκουν σε παραπάνω από μία οντολογικές κατηγορίες, δηλαδή να είναι πολυλειτουργικές. Ειδικότερα, πρωτεΐνες που είναι παράγοντες συναρμογής μπορεί να συμμετέχουν στη διαδικασία της μετάφρασης, της μίτωσης, της εξόδου του mRNA και της αποικοδόμησης των ακατάλληλων μεταγράφων. Ακόμη, παράγοντες μεταγραφής μπορεί να παρουσιάζουν λειτουργικό ρόλο στη μίτωση ή τη μετάφραση και κάποιοι παράγοντες αναδιαμόρφωσης της χρωματίνης μπορεί να συμμετέχουν στη μίτωση. Το ποσοστό των πολυλειτουργικών πρωτεϊνών στο σύνολο των συστατικών του σωματίου συναρμογής είναι ιδιαίτερα αυξημένο (24%). Η πλειοψηφία αυτών των πρωτεϊνών συμμετείχαν στη διαδικασία της συναρμογής και της μεταγραφής ή της μεταγραφής και της οργάνωσης της χρωματίνης (βλ. Ενότητα 3.2.1.1.1.). Αυτό μπορεί να εξηγηθεί από την ύπαρξη διαφορετικών λειτουργικών μοτίβων (domains) και χαρακτηριστικών περιοχών (motifs) που μπορεί να προσδίδουν στις πρωτεΐνες πολλαπλούς λειτουργικούς ρόλους ή διαφορετική τοπολογία μέσα στο κύτταρο, αντίστοιχα (Zaretsky & Wreschner 2008). Επιπλέον, είναι γνωστό ότι σχεδόν όλο το ποσοστό των γονιδίων του ανθρώπου (95%) υφίσταται εναλλακτική συναρμογή με αποτέλεσμα να παράγεται ποικιλία πρωτεϊνικών ισομορφών που μπορεί να συμμετέχουν σε διαφορετικές κυτταρικές διεργασίες (Gerstein et al. 2014, Lu et al. 2012), όπως στην περίπτωση του γονιδίου *eEF1A*.

#### 4.2.7. Τα ποσοστά ενσωμάτωσης των πρωτεϊνών του δικτύου στα GO terms συνάδουν με τα αντίστοιχα του συνόλου δεδομένων του σωματίου συναρμογής, ενώ μειώνεται το ποσοστό των πρωτεϊνών που δεν έχουν χαρακτηρισμένο λειτουργικό ρόλο

Αξιοσημείωτη ήταν η ενσωμάτωση στο δίκτυο, του μεγαλύτερου μέρους των πρωτεϊνών των 19 οντολογικών κατηγοριών στις οποίες κατατάχθηκαν τα συστατικά του σωματίου συναρμογής. Ενδεικτικά, ενσωματώθηκε το 89% της οντολογικής κατηγορίας της συναρμογής, το 64% της μεταγραφής/ρύθμισης της μεταγραφής, το 77% της οργάνωσης και αναδιαμόρφωσης της χρωματίνης, και κατ' αντιστοιχία το 81% και 83% των παραγόντων που συμμετείχαν στην έξοδο του mRNA από τον πυρήνα και στην αποικοδόμηση μη κατάλληλων μεταγράφων για μετάφραση (Εικ. 42). Το σχετικά χαμηλό ποσοστό των πρωτεϊνών που συμμετέχουν στη μεταγραφή οφείλεται στο γεγονός ότι το 24% των πρωτεϊνών που δεν ενσωματώθηκαν στο δίκτυο (186) συμμετείχαν στην οντολογική κατηγορία της μεταγραφής. Το αποτέλεσμα αυτό συνάδει με τον εξίσου αυξημένο αριθμό πρωτεϊνών που δεν περιλαμβάνονταν στο δίκτυο και είχαν απομονωθεί ως συστατικά του συμπλόκου E, το οποίο είναι εμπλουτισμένο με μεταγραφικούς παράγοντες και πρωτεΐνες που ρυθμίζουν τη μεταγραφή. Επιπλέον, το 31% των πρωτεϊνών οι οποίες δεν συμπεριλήφθηκαν στο δίκτυο ήταν άγνωστης λειτουργίας γεγονός που εξηγεί το μικρό ποσοστό (7,4%) αυτών των πρωτεϊνών στο πλαίσιο του δικτύου. Τέλος, στο σύνολο των πρωτεϊνών που δεν ενσωματώθηκαν στο δίκτυο, η οντολογική κατηγορία της συναρμογής εμφάνιζε πολύ

χαμηλή αντιπροσώπευση (10%), γεγονός που ενισχύει τον εμπλουτισμό του ανακατασκευασμένου δικτύου με παράγοντες συναρμογής ή με παράγοντες που συμβάλλουν στη ρύθμιση της διαδικασίας.

Ο εμπλουτισμός σε παράγοντες συναρμογής φαίνεται και κατά την οπτικοποίηση του δικτύου (Εικ. 40). Το δίκτυο που σχημάτισαν μόνο οι πρωτεΐνες που συμμετέχουν στη συναρμογή αποτελεί τη «ραχοκοκαλιά» του συνολικού δικτύου (~41% του συνόλου των πρωτεϊνών του δικτύου) (Εικ. 40). Αξίζει να σημειωθεί ότι και η πλειοψηφία των υπολοίπων πρωτεϊνών εντάχθηκε σε συναφείς λειτουργικά διαδικασίες με τη συναρμογή. Ενδεικτικά το 22% στην μεταγραφή/ρύθμιση της μεταγραφής, περίπου το 10% στην αναδιαμόρφωση της χρωματίνης, το 4,9% στην έξοδο του mRNA από τον πυρήνα και το 4,3% στην αποικοδόμηση ακατάλληλων μεταγράφων για μετάφραση (Εικ. 40). Έτσι επιβεβαιώνεται και μέσω της ανακατασκευής του δικτύου του σωματίου συναρμογής, η καλά εδραιωμένη σχέση της συναρμογής με άλλες διαδικασίες επεξεργασίας του RNA.

Το ποσοστό των πολυλειτουργικών πρωτεϊνών ήταν αυξημένο (22%) και στα πλαίσια του δικτύου. Η πλειοψηφία των πρωτεϊνών αυτής της κατηγορίας (80%) ενσωματώθηκε στο δίκτυο του σωματίου συναρμογής και το μεγαλύτερο ποσοστό αυτών συμμετείχαν κυρίως στη συναρμογή και τη μεταγραφή/ρύθμιση της μεταγραφής ή στην οργάνωση/αναδιαμόρφωση της χρωματίνης και τη μεταγραφή/ρύθμιση της μεταγραφής στα πλαίσια του συνόλου δεδομένων του σωματίου συναρμογής.

#### 4.2.8. Η λειτουργική κατηγοριοποίηση των πρωτεϊνών των υποσυμπλόκων του σωματίου συναρμογής και των αντίστοιχων δικτύων τους συνάδει με την πορεία συναρμολόγησής τους στο πρόδρομο μετάγραφο

Στο ίδιο πλαίσιο, πραγματοποιήθηκε η λειτουργική κατηγοριοποίηση των πρωτεϊνών που αποτελούσαν συστατικά ενός ή περισσότερων υπο-συμπλόκων (87% του συνόλου των πρωτεϊνών) για κάθε υποσύμπλοκο του σωματίου συναρμογής. Η πορεία εκπροσώπησης της εκάστοτε οντολογικής κατηγορίας έδειξε να συνάδει με τα διαδοχικά στάδια συναρμολόγησης του σωματίου συναρμογής (βλ. Ενότητα 3.2.1.1.3.). Ειδικότερα, το αυξημένο ποσοστό των παραγόντων αναδιαμόρφωσης της χρωματίνης στο σύμπλοκο E έρχεται σε συμφωνία με την αλληλεπίδραση των πρωτεϊνών αυτών με παράγοντες των snRNPs του σωματίου συναρμογής (π.χ. U2snRNP), οι οποίοι συμμετέχουν στα αρχικά στάδια συναρμολόγησης του σωματίου (Kfir et al. 2015, Haque & Oberdoerffer, 2014) (βλ. Ενότητα 1.4.2.). Επιπλέον, οι πρωτεΐνες που συμμετείχαν στη διαδικασία της συναρμογής βρέθηκαν να αποτελούν την πλειοψηφία στα υποσύμπλοκα B, B *act* και C, κάτι που θεωρείται αναμενόμενο, καθώς σε αυτά τα στάδια λαμβάνουν χώρα οι δυο αντιδράσεις μετεστεροποίησης, οι οποίες αποτελούν τα καταλυτικά βήματα του μηχανισμού και οδηγούν στην τελική αποκοπή των ιντρονίων (Matera & Wang 2014, Tseng & Cheng 2008). Ακόμη, σημαντικό ήταν το ποσοστό εμπλουτισμού των συγκεκριμένων υποσυμπλόκων σε παράγοντες που συμμετείχαν τόσο στην έξοδο των mRNAs από τον πυρήνα όσο και στην αποικοδόμηση αυτών, εφόσον δεν είναι κατάλληλα για να μεταφραστούν. Οι παράγοντες αυτοί στρατολογούνται κατά τη διάρκεια της συναρμογής και αλληλεπιδρούν με πρωτεΐνες που συμμετέχουν στην καθοδήγηση του mRNA προς

τους πυρηνικούς πόρους και στον έλεγχο της ακεραιότητας του στο κυτταρόπλασμα για την επακόλουθη αναγνώρισή του από τα ριβοσώματα (Katahira 2015, Barbosa et al. 2012, Chuang et al. 2015). Τέλος, στα πλαίσια των υποσυμπλόκων, επιβεβαιώνεται για ακόμα μια φορά η εδραιωμένη λειτουργική σχέση της συναρμογής με τη μεταγραφή, καθώς παρατηρείται ότι το ποσοστό των παραγόντων στην οντολογική κατηγορία της μεταγραφής δεν αλλάζει σημαντικά σε κάποιο υποσύμπλοκο, αλλά παραμένει σχετικά αυξημένο σε όλα τα στάδια (~15-20%). Για τις υπόλοιπες 81 πρωτεΐνες που δεν είχαν ταυτοποιηθεί ως συστατικά κάποιου υποσυμπλόκου, η ένταξή τους στο σύνολο δεδομένων του σωματίου συναρμογής ενισχύθηκε από το μεγάλο ποσοστό συμμετοχής τους (53%) στις οντολογικές κατηγορίες της συναρμογής και των στενά λειτουργικά σχετιζόμενων κατηγοριών (π.χ. μεταγραφή-ρύθμιση της μεταγραφής, έξοδος του mRNA από τον πυρήνα).

Και στο πλαίσιο των επιμέρους δικτύων, το ποσοστό αλλαγής εκπροσώπησης των κύριων οντολογικών κατηγοριών στα επιμέρους δίκτυα των υποσυμπλόκων έρχεται σε πλήρη συμφωνία με τις διαδικασίες που λαμβάνουν χώρα κατά τα διαδοχικά στάδια συναρμολόγησης του σωματίου συναρμογής. Η μόνη διαφορά που παρατηρήθηκε στα πλαίσια των δικτύων είναι ο επιπρόσθετος εμπλουτισμός σε παράγοντες συναρμογής, ειδικότερα των υποσυμπλόκων B, B *act* και C, πιθανώς λόγω της σημασίας της συνεισφοράς αυτών των παραγόντων στα δυο καταλυτικά βήματα της συναρμογής, τα οποία συμβαίνουν κατά το σχηματισμό αυτών των συμπλόκων. Ωστόσο, η αύξηση της οντολογικής κατηγορίας της συναρμογής στα δίκτυα αυτά μπορεί να είναι αυξημένη διότι μειώθηκαν άλλες κατηγορίες (π.χ. πρωτεΐνες άγνωστης λειτουργίας).

#### 4.2.9. Η ανάλυση των τοπολογικών παραμέτρων του δικτύου συνεισφέρει στον προσδιορισμό των ιδιοτήτων και της αρχιτεκτονικής του

Το «τοπολογικό τοπίο» των αλληλεπιδράσεων ενός δικτύου αποτελεί μια πλούσια πηγή πληροφοριών για την εύρεση κοινών λειτουργικών ρόλων των πρωτεϊνών-κόμβων που συμμετέχουν σε αυτές (Yu et al. 2013). Σύμφωνα με την προσέγγιση αυτή μπορεί να εξαχθούν συμπεράσματα για τις ιδιότητες των πρωτεϊνών άγνωστης λειτουργίας. Αυτό μπορεί να επιτευχθεί μέσω της άντλησης στοιχείων από «γειτονικές» πρωτεΐνες, οι οποίες είναι μελετημένες, δεδομένου ότι πρωτεΐνες που συν-εντοπίζονται και έχουν παρόμοιες τοπολογικές ιδιότητες σε ένα δίκτυο πρωτεϊνικών αλληλεπιδράσεων είναι πιθανό να σχετίζονται λειτουργικά (Yu et al. 2013, Cho et al. 2016).

Ορισμένες από τις σημαντικότερες ιδιότητες του δικτύου είναι η πυκνότητά του (network density), η κατανομή της συνδεσιμότητας των πρωτεϊνών (node degree distribution), ο συντελεστής ομαδοποίησης (clustering coefficient), η δυνατότητα επικοινωνίας δυο πρωτεϊνών με το μικρότερο δυνατό μονοπάτι (closeness centrality) και η ικανότητα δράσης μιας πρωτεΐνης ως «γέφυρα» στις συνδέσεις των υπολοίπων πρωτεϊνών του δικτύου (betweenness centrality). Όλες αυτές μπορούν να χρησιμοποιηθούν για την ανάκτηση σχετικών πληροφοριών με τη συνδεσιμότητα και τη θέση-τοπολογία («wiring») μίας πρωτεΐνης σε ένα δίκτυο (Davis et al. 2015).

Η μελέτη αυτών των παραμέτρων έχει χρησιμοποιηθεί από πολλούς ερευνητές και έχει οδηγήσει στην εξαγωγή συμπερασμάτων σε ποικίλα βιολογικά ερωτήματα.

Ενδεικτικά, αξίζει να αναφερθούν η συσχέτιση των πρωτεϊνών που βρίσκονται στο κέντρο ενός δικτύου με την κρίσιμη βιολογική τους σημασία στη διαδικασία που μελετάται, και η σύνδεση των περιφερειακών πρωτεϊνών με δευτερεύοντες ρόλους στις κυτταρικές διεργασίες (Acencio & Lemke 2009, da Silva et al. 2008, Estrada 2006, Jeong et al. 2001). Επιπλέον, σε δίκτυα πρωτεϊνικών αλληλεπιδράσεων τα οποία περιλαμβάνουν πρωτεΐνες που σχετίζονται με νευροεκφυλιστικές ασθένειες (π.χ. Alzheimer), συγκεκριμένες τοπολογικές παράμετροι (π.χ. shortest path length) χρησιμοποιούνται ως μέσο πρόβλεψης υποψήφιων γονιδίων για τις ασθένειες αυτές (Goñi et al. 2008, Krauthammer et al. 2004, Liu et al. 2012). Επίσης, έχει δειχθεί ότι πρωτεΐνες που σχετίζονται με διάφορες μορφές καρκίνου έχουν την τάση να εμφανίζουν υψηλές τιμές των παραμέτρων «betweenness centrality», «node degree» και χαμηλές τιμές των παραμέτρων «clustering coefficient» και «shortest path length» σε σύγκριση με πρωτεΐνες που δεν σχετίζονται με κάποιο είδος καρκίνου (Safari-Alighiarloo et al. 2014, Sun & Zhao 2010). Στο δίκτυο πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής επαληθεύονται τα παραπάνω αποτελέσματα καθώς πρωτεΐνες που έχουν συσχετιστεί με καρκίνο (π.χ. SRPK2, RBM10, U2AF1, CDC5L, SRPK1) εμφανίζουν υψηλό αριθμό αλληλεπιδράσεων (13-33), υψηλές τιμές «betweenness centrality» και χαμηλές τιμές «clustering coefficient».

Στο πλαίσιο του δικτύου του σωματίου συναρμογής, τα αποτελέσματα που προκύπτουν από την ανάλυσή του οδηγούν στην εξαγωγή συμπερασμάτων για την «αρχιτεκτονική» και τις ιδιότητές του. Αρχικά, αν και η πυκνότητα του δικτύου είναι σχετικά χαμηλή (0,015), η κατά μέσο όρο ελάχιστη απόσταση ανάμεσα σε οποιοδήποτε ζεύγος πρωτεϊνών του δικτύου είναι μικρή (characteristic path length=3,527) και το μεγαλύτερο μονοπάτι που συνδέει δυο πρωτεΐνες του δικτύου είναι σχετικά μικρό (network diameter=8) (βλ. Ενότητα 3.2.4.). Τα αποτελέσματα αυτά, σε συνδυασμό με την υψηλή τιμή (96%) της παραμέτρου «shortest path metric», αν ληφθούν υπόψη τα ετεροδιμερή ή ομοδιμερή που δεν συμπεριελήφθησαν στο ενιαίο δίκτυο, υποδεικνύουν ένα καλά συνδεδεμένο δίκτυο. Η καλή συνδεσιμότητα των πρωτεϊνών του σωματίου συναρμογής υποδεικνύει ότι όλες οι πρωτεΐνες συνεργάζονται μεταξύ τους ώστε οι διαδικασίες της επεξεργασίας του mRNA να λάβουν χώρα συντονισμένα. Δηλαδή, πρωτεΐνες που συμμετέχουν στα διαφορετικά στάδια συναρμολόγησης του σωματίου συναρμογής συνδέονται με πρωτεΐνες που λαμβάνουν μέρος και σε άλλες διαδικασίες της επεξεργασίας του mRNA, μέσω εναλλακτικών αλληλεπιδράσεων, ακολουθώντας διαφορετικά μονοπάτια. Αυτό έχει ως αποτέλεσμα η συναρμογή να πραγματοποιείται συντονισμένα και ταυτόχρονα με τη μεταγραφή και να αποτελεί το συνδεδετικό κρίκο με διαδικασίες που έπονται αυτής, όπως η έξοδος του mRNA από τον πυρήνα και ο έλεγχος της ακεραιότητας αυτού είτε στον πυρήνα είτε στο κυτταρόπλασμα.

Η ομοιομορφία της συνδεσιμότητας του δικτύου υποδηλώνεται από την τιμή της παραμέτρου «network centralization» (0,159), η οποία δείχνει ότι οι πρωτεΐνες-κόμβοι του δικτύου είναι ομοιόμορφα κατανομημένες τόσο στο κέντρο, όσο και στην περιφέρειά του. Το δίκτυο είναι κατά κάποιο τρόπο οργανωμένο σε «στρώματα», δηλαδή στην περιφέρειά του είναι συγκεντρωμένη η πλειοψηφία των κόμβων που συνδέονται με μία μόνο πρωτεΐνη ή με ένα μικρό αριθμό πρωτεϊνών, ενώ στο κέντρο βρίσκονται οι πρωτεΐνες-κόμβοι που παρουσιάζουν υψηλή συνδεσιμότητα (hubs) (Εικ.

40). Αυτή η κατανομή των κόμβων του δικτύου μπορεί να υποδηλώνει ότι πρωτεΐνες που διαδραματίζουν βασικό ρόλο στη διαδικασία της συναρμογής βρίσκονται στο κέντρο του δικτύου και εμφανίζουν συνδέσεις με πολλές πρωτεΐνες. Αντίθετα, κόμβοι που τοποθετούνται στην περιφέρεια του δικτύου μπορεί να είναι βοηθητικές πρωτεΐνες, με δευτερεύοντα ρόλο στη συναρμογή, οι οποίες μπορεί να συνδέονται παροδικά σε συγκεκριμένα στάδια της διαδικασίας, αν ληφθεί υπόψη και ο μικρός αριθμός των αλληλεπιδράσεών τους.

Σύμφωνα με την παράμετρο κατανομής του αριθμού των αλληλεπιδράσεων (node degree distribution) το δίκτυο φαίνεται να ακολουθεί την κατανομή «power law», καθώς λίγες πρωτεΐνες-κόμβοι εμφανίζουν πολλές συνδέσεις (hub). Η πρωτεΐνη-hub με τις περισσότερες συνδέσεις η CSNK2A1, έχει 79 αλληλεπιδράσεις, ενώ το μεγαλύτερο ποσοστό των πρωτεϊνών (73%) αλληλεπιδρούν με 1-10 πρωτεΐνες. Οι υπόλοιπες είναι μέτριας συνδεσιμότητας με αριθμό αλληλεπιδράσεων από 11 έως 27. Τα αποτελέσματα αυτά υποδεικνύουν ένα δίκτυο το οποίο εμφανίζει ετερογένεια στις συνδέσεις του, μια ιδιότητα που αποτελεί χαρακτηριστικό των βιολογικών δικτύων (Barabasi et al. 2009, Barabási & Oltvai 2004). Αυτή η ετερογένεια των αλληλεπιδράσεων των συστατικών του σωματίου συναρμογής υποδεικνύει ότι το 27% των πρωτεϊνών (118 πρωτεΐνες) που εμφανίζει υψηλή και μέτρια συνδεσιμότητα, διαδραματίζουν βασικό ρόλο καθ' όλη τη διάρκεια της συναρμογής, αν ληφθεί υπόψη ότι αρκετές από αυτές (37 πρωτεΐνες) είναι παρούσες στα υποσύμπλοκα A, B, B act και C. Η πλειοψηφία όμως των πρωτεϊνών παρουσιάζουν χαμηλή συνδεσιμότητα, το οποίο μπορεί να υποδεικνύει ότι οι πρωτεΐνες αυτές συμμετέχουν σε συγκεκριμένα στάδια της διαδικασίας, επιτελώντας μια συγκεκριμένη λειτουργία. Το γεγονός αυτό συνάδει με τη «φύση» της διαδικασίας της συναρμογής, δηλαδή, τη διαδοχική συναρμολόγηση των υποσυμπλόκων του σωματίου συναρμογής στο πρόδρομο μετάγραφο, η οποία πιθανώς απαιτεί τη στρατολόγηση και τη συνεργασία διαφορετικών πρωτεϊνών ή ομάδων πρωτεϊνών στα διάφορα στάδια της διαδικασίας. Επίσης η ύπαρξη πολλών πρωτεϊνών χαμηλής συνδεσιμότητας μπορεί να έρχεται σε συμφωνία με τη συνεργασία της συναρμογής και με άλλες διαδικασίες της επεξεργασίας του mRNA. Δηλαδή, πρωτεΐνες μπορεί να στρατολογούνται σε συγκεκριμένα στάδια της συναρμολόγησης του σωματίου συναρμογής και να λειτουργούν ως μόρια-«προσαρμογείς» τα οποία βοηθούν να συνεχιστεί η επεξεργασία του mRNA μετά το πέρας της συναρμογής.

Οι πρωτεΐνες που αποτέλεσαν «hubs» του δικτύου χαρακτηρίστηκαν από υψηλές τιμές των παραμέτρων «betweenness centrality» και «closeness centrality». Αυτές που παρουσιάζουν υψηλές τιμές στην παράμετρο «betweenness centrality» υποδεικνύουν τη δράση των πρωτεϊνών-hub ως «γέφυρες», οι οποίες συνδέουν ομάδες που παρουσιάζουν παρόμοια κυτταρική λειτουργία (Girvan & Newman 2002). Αυτές που παρουσιάζουν υψηλές τιμές στην παράμετρο «closeness centrality» υποδεικνύουν τη βιολογική σημασία τους στη διεκπεραίωση της διεργασίας στην οποία συμμετέχουν (Jalili et al. 2016, Tew et al. 2007). Με αυτόν τον τρόπο η λειτουργία ενός παράγοντα συναρμογής μπορεί να προβλεφθεί από τη συνδεσιμότητά του μέσα στο δίκτυο, όπως έχει προταθεί και από τους Akerman et al. (2015). Ειδικότερα, στην εργασία αυτή, οι παράγοντες που αποτελούν θετικούς ρυθμιστές της συναρμογής (πρωτεΐνες SR) παρουσίασαν υψηλή συνδεσιμότητα και υψηλή τιμή «betweenness centrality» ενώ αυτοί



που αποτελούν αρνητικούς ρυθμιστές της συναρμογής (πρωτεΐνες hnRNP) εμφάνισαν χαμηλή συνδεσιμότητα και χαμηλή τιμή «betweenness centrality». Ωστόσο, η ομαδοποίηση αυτή δεν παρατηρήθηκε στο δίκτυο του σωματίου συναρμογής που ανακατασκευάστηκε στη διατριβή, καθώς οι πρωτεΐνες που εμφάνισαν υψηλή συνδεσιμότητα και τις υψηλότερες τιμές «betweenness centrality» και «closeness centrality» παρατηρήθηκε ότι συμμετέχουν στη συναρμογή/εναλλακτική συναρμογή ή στη μεταγραφή χωρίς να μπορεί να γίνει περαιτέρω ομαδοποίηση.

➤ Όλες οι πρωτεΐνες-hubs του δικτύου συμμετέχουν στη διαδικασία της συναρμογής και σε συναφείς λειτουργικά διαδικασίες

Πέντε από τις πρωτεΐνες-hubs του δικτύου συμμετέχουν αποκλειστικά στη διαδικασία της συναρμογής (SMN1, U2AF2, PRPF40A, SRPK1, SRPK2) και δυο στην εναλλακτική συναρμογή (RBM10, CHERP). Οι υπόλοιπες συμμετέχουν στη συναρμογή και τη μεταγραφή/ρύθμιση της μεταγραφής (RBM39) ή και στα δυο με επιπρόσθετη συμμετοχή στην έξοδο του mRNA από τον πυρήνα (CSNK2A1) και στη μετάφραση (HNRNPK). Παρατηρείται ότι οι πρωτεΐνες αυτές δικαιολογημένα αποτελούν hubs του δικτύου διότι εμφανίζουν βασικό ρόλο στη διαδικασία της συναρμογής. Για παράδειγμα, η SMN1 συμμετέχει στη συναρμολόγηση των snRNPs (Paushkin et al. 2002), η U2AF2 είναι απαραίτητη για την αναγνώριση των θέσεων συναρμογής (Agrawal et al. 2016) και η κινάση CSNK2A1 συνεισφέρει σε κάθε βήμα συναρμολόγησης και αποσυναρμολόγησης του σωματίου συναρμογής φωσφορυλιώνοντας πρωτεΐνες που στρατολογούνται στα διαφορετικά στάδια (Bian et al. 2013).

➤ Η ενσωμάτωση της κινάσης CSNK2A1, του μεγαλύτερου «hub», απαιτεί περαιτέρω διερεύνηση

Ένα ερώτημα που διερευνήθηκε είναι εάν σε ένα δίκτυο θα πρέπει να ενσωματώνονται οι πρωτεΐνες που είναι υπεύθυνες για μετα-μεταφραστικές τροποποιήσεις. Όσον αφορά το σωματίο συναρμογής, η σημασία των μετα-μεταφραστικών τροποποιήσεων στη δυναμική του είναι ιδιαίτερα σημαντική (McKay & Johnson 2010). Συγκεκριμένα, η φωσφορυλίωση των πρωτεϊνών ρυθμίζει τις δυναμικές αλληλεπιδράσεις που συμβαίνουν κατά τη διάρκεια της συναρμολόγησης, κατάλυσης και αποσυναρμολόγησης του σωματίου συναρμογής (McKay & Johnson 2010), όπως έχει δειχθεί και για την CSNK2A1. Χαρακτηριστικά παραδείγματα για τη συνεισφορά της φωσφορυλίωσης στην καθοδήγηση της δυναμικής του σωματίου είναι η επανειλημμένη ταυτοποίηση των πρωτεϊνών SR ως υποστρώματα φωσφορυλίωσης, και η επακόλουθη συνεισφορά της φωσφορυλίωσής τους στις δυναμικές ανακατατάξεις του σωματίου (Fluhr 2008; Lipp et al. 2015, Tenenbaum & Aguirre-Ghiso 2005, Zhou & Fu 2013). Αρκετές από τις πρωτεΐνες SR φωσφορυλιώνονται από τις κινάσες SRPK1 και SRPK2, οι οποίες αποτελούν «hubs» του συγκεκριμένου δικτύου (Aubol et al. 2016, Bullock & Oltean 2017, Mathew et al. 2008). Άλλα παραδείγματα της σημασίας της φωσφορυλίωσης είναι οι πρωτεΐνες SF3B1 και PRPF8. Τα διαδοχικά στάδια φωσφορυλίωσης και αποφωσφορυλίωσης της SF3B1 συμβάλλουν στα καταλυτικά βήματα της συναρμογής (Boudrez et al. 2002, De Graaf et al. 2006). Η ρύθμιση της PRPF8 μέσω της συντονισμένης φωσφορυλίωσης και

ουβικουτινιλίωσης οδηγεί σε ένα πλήθος διαδοχικών και συντονισμένων αναδιατάξεων RNA–RNA και RNA–πρωτεΐνης (Bellare et al. 2008, Grainger 2005).

Άρα, σε κάθε δίκτυο που κατασκευάζεται θα πρέπει να ελέγχεται η σημασία του λειτουργικού ρόλου των πρωτεϊνών οι οποίες είναι υπεύθυνες για τις μετα-μεταφραστικές τροποποιήσεις. Για παράδειγμα, στο συγκεκριμένο δίκτυο, η CSNK2A1 είναι μια πρωτεΐνη η οποία δεν συμμετέχει μόνο σε μία διαδικασία αλλά σε τρεις (συναρμογή, μεταγραφή, έξοδος από τον πυρήνα), γεγονός το οποίο υποδεικνύει τον βασικό λειτουργικό της ρόλο. Επιπλέον, πρέπει να ελέγχεται κατά πόσο η προσθήκη των πρωτεϊνών που είναι υπεύθυνες για τις μετα-μεταφραστικές τροποποιήσεις «παραμορφώνουν» το δίκτυο που δημιουργείται. Για παράδειγμα, η προσθήκη της ουβικουτίνης σε αυτό το δίκτυο θα συγκέντρωνε τις περισσότερες αλληλεπιδράσεις το οποίο βέβαια δεν σημαίνει ότι διαδραματίζει το σημαντικότερο ρόλο αλλά αποτελεί «hub» λόγω της φύσης της αλληλεπίδρασης, δηλαδή πρέπει να αλληλεπιδρά με τις άλλες πρωτεΐνες για να προσθέσει τα μόρια ουβικουτίνης. Η απόρριψη της ενσωμάτωσης των αλληλεπιδράσεων αυτών στο δίκτυο του σωματίου συναρμογής έρχεται σε συμφωνία και με άλλες μελέτες κατασκευής πρωτεϊνικών δικτύων οι οποίες επιλέγουν να απομακρύνουν την ουβικουτίνη λόγω του τεράστιου αριθμού των αλληλεπιδράσεών της (Pellegrini et al. 2016, Zhu et al. 2011). Στα πλαίσια κατασκευής ολόκληρου του «interactome» του ανθρώπου (Klara et al. 2013), παρόλο που η ουβικουτίνη ενσωματώνεται στο δίκτυο και συγκεντρώνει περίπου 1000 αλληλεπιδράσεις, έχει δειχθεί ότι η ενσωμάτωση αυτών των PPI αλλάζει σε μεγάλο βαθμό την «αρχιτεκτονική» του δικτύου και έχει προταθεί ότι θα ήταν δόκιμο οι αλληλεπιδράσεις της να λαμβάνονται υπόψη στα πλαίσια δικτύων που αφορούν τις μετα-μεταφραστικές τροποποιήσεις των πρωτεϊνών (Klara et al. 2013, Du et al. 2011, Venancio et al. 2009).

➤ Η αφαίρεση πρωτεϊνών-hub και της FRA10AC1 δεν επηρεάζει την «στιβαρότητα» του δικτύου

Το δίκτυο που κατασκευάστηκε υποστηρίζεται από αυξημένη συνδεσιμότητα (μεγάλο αριθμό PPIs) των πρωτεϊνών-hubs με αυτές που εμφανίζουν χαμηλό αριθμό «γειτόνων». Επίσης υποστηρίζεται από μειωμένη συνδεσιμότητα (μικρό αριθμό PPIs) μεταξύ των πρωτεϊνών-«hubs» του δικτύου, μια ιδιότητα γνωστή ως «dissortativity» (Newman 2002, Pavlopoulos et al. 2011). Η ιδιότητα αυτή αυξάνει την «στιβαρότητα» (robustness) των βιολογικών δικτύων (Maslov & Sneppen 2002), τα καθιστά δηλαδή «ανθεκτικά» στην απώλεια πρωτεϊνών-κόμβων γενικά, αλλά συγχρόνως εξαρτώμενα σε μεγάλο βαθμό από την απομάκρυνση των ίδιων των πρωτεϊνών-«hubs» (Jeong et al. 2001, Khuri & Wuchty 2015, Memmott et al. 2004). Η αφαίρεση πρωτεϊνών υψηλής συνδεσιμότητας έχει δειχθεί ότι οδηγεί σε μεγάλη αύξηση της διαμέτρου του δικτύου, δηλαδή, σε αύξηση της απόστασης (αριθμός μονοπατιών) μεταξύ των πρωτεϊνών του δικτύου και στην διάσπαση του δικτύου σε μικρότερες ομάδες (Jeong et al. 2001, (Albert et al. 2000, Friedel & Zimmer 2007). Ωστόσο, η απομάκρυνση των πρωτεϊνών-κόμβων με χαμηλή συνδεσιμότητα μπορεί να έχει ποικίλες επιδράσεις στη «στιβαρότητα» του δικτύου. Οι επιδράσεις αυτές μπορεί να αυξάνονται εάν απομακρύνονται πρωτεΐνες που συνδέονται με πρωτεΐνες-«hubs» (Pires et al. 2015). Παρ' όλα αυτά, μέχρι τώρα, δεν υπάρχουν εργασίες που μελετούν τις επιδράσεις της

απομάκρυνσης πρωτεϊνών που συνδέονται με «hubs», τα οποία θεωρούνται υψηλής σημασίας λόγω του μεγάλου αριθμού PPI που εμφανίζουν. Ωστόσο, έχει δειχθεί ότι η απομάκρυνση πρωτεϊνών χαμηλής συνδεσιμότητας (ανεξάρτητα αν συνδέονται με hub ή όχι) από πρωτεϊνικά δίκτυα μπορεί να επηρεάσει τη «στιβαρότητα» του δικτύου (Tew et al. 2007).

Μια άλλη ιδιότητα των δικτύων που συνεισφέρει στη «στιβαρότητά» του είναι η ύπαρξη μικρού μήκους μονοπατιών, χαρακτηριστικό και του δικτύου του σωματίου συναρμογής, η οποία αποτελεί χαρακτηριστικό των δικτύων με ιδιότητες «small-world». Σε ένα τέτοιου είδους δίκτυο, η κατά μέσο όρο απόσταση όλων των ζευγών των κόμβων των δικτύων είναι μικρή (Telesford et al. 2011, Xu et al. 2011, Zhang Z. & Zhang J. 2009), μια ιδιότητα που συνεισφέρει στη διατήρηση της ακεραιότητας του δικτύου έναντι διαταραχών οι οποίες μπορούν να συμβούν μέσω της απομάκρυνσης κάποιου κόμβου ή αλληλεπίδρασης τα οποία συμμετέχουν σε οποιοδήποτε από αυτά τα μικρά μονοπάτια του δικτύου (Telesford et al. 2011, Xu et al. 2011, Zhang Z. & Zhang J. 2009).

Σε μια προκαταρκτική προσπάθεια ελέγχου της «στιβαρότητας» του δικτύου του σωματίου συναρμογής έπειτα από την αφαίρεση πρωτεϊνών-«hubs», απομακρύνθηκαν οι δυο πρωτεΐνες με τις υψηλότερες συνδέσεις (CSNK2A1 και HNRNPK) (βλ. Ενότητα 3.2.4.2.). Αν και θα ήταν αναμενόμενο η απομάκρυνση πρωτεϊνών υψηλής συνδεσιμότητας να οδηγήσει στην αύξηση της διαμέτρου του δικτύου και στην διάσπαση αυτού σε μικρότερες ομάδες, στο δίκτυο που κατασκευάστηκε στην εργασία αυτή, δεν παρατηρήθηκαν τέτοιου είδους αλλαγές, με δεδομένο ότι η διάμετρος διατηρήθηκε σταθερή και μόνο ένας μικρός αριθμός κόμβων απομακρύνθηκε από το ενιαίο δίκτυο (βλ. Ενότητα 3.2.4.2.). Το γεγονός αυτό υποδεικνύει ότι το δίκτυο του σωματίου συναρμογής είναι «ανθεκτικό» σε τυχόν διαταραχές και διατηρεί την καλή συνδεσιμότητα μεταξύ των πρωτεϊνών του, αν ληφθεί υπόψη και η μικρή τιμή της ακτίνας του δικτύου (1) έπειτα από την αφαίρεση της πρωτεΐνης με τις περισσότερες συνδέσεις (CSNK2A1, 79 συνδέσεις). Επιπλέον, ένα ακόμη στοιχείο που υποδεικνύει την «ανθεκτικότητα» του δικτύου του σωματίου συναρμογής σε τυχόν διαταραχές είναι η ύπαρξη μικρών μονοπατιών μεταξύ των πρωτεϊνών-κόμβων του (characteristic path length=3,527).

Όσον αφορά την πρωτεΐνη FRA10AC1, θα μπορούσε να υποθεθεί, σύμφωνα με τις σχετικά υψηλές τιμές των «betweenness centrality» και «closeness centrality» που εμφανίζει, και την παρουσία των 8 από τις 10 πρωτεΐνες-hubs στους πρώτους και δεύτερους γείτονές της, ότι εμφανίζει σημαντικό ρόλο στην αρχιτεκτονική του δικτύου και κατ' επέκταση στην επεξεργασία του mRNA. Ωστόσο, η στοχευμένη απουσία της τόσο από ολόκληρο το δίκτυο όσο και από συγκεκριμένα σύμπλοκα δεν επιφέρει αλλαγές στην αρχιτεκτονική του δικτύου (βλ. Ενότητα 3.2.4.2.), γεγονός το οποίο θεωρείται αναμενόμενο αν ληφθεί υπόψη ότι η απομάκρυνση ακόμα και των πιο υψηλά συνδεδεμένων πρωτεϊνών είχε ελάχιστες επιδράσεις στις τοπολογικές παραμέτρους του δικτύου του σωματίου συναρμογής.

Συμπερασματικά, ο παραπάνω έλεγχος της «στιβαρότητας» του δικτύου του σωματίου συναρμογής υποδεικνύει ένα δίκτυο το οποίο διατηρεί την «ακεραιότητα» του έπειτα και από την απομάκρυνση των πρωτεϊνών υψηλής συνδεσιμότητας.

#### 4.2.10. Η εύρεση συμπλεγμάτων πρωτεϊνών (clusters) συνεισφέρει στον προσδιορισμό της λειτουργίας πρωτεϊνών με μη χαρακτηρισμένο βιολογικό ρόλο

Μεταξύ των διαφόρων ειδών προσεγγίσεων που έχουν προταθεί για τον καθορισμό της λειτουργίας άγνωστων πρωτεϊνών βάσει των δικτύων PPI, οι πιο δημοφιλείς είναι: α) η προσέγγιση που είναι «προσανατολισμένη από τους γείτονες» (Neighborhood-oriented) (Saha et al. 2014, Shen et al. 2016), κατά την οποία μελετώνται είτε οι άμεσοι είτε οι έμμεσοι (δεύτεροι) «γείτονες» της πρωτεΐνης ή οι κοινοί «γείτονες» που μοιράζεται με τις υπόλοιπες ώστε να προβλεφθεί η λειτουργία της με βάση τις λειτουργίες των «γειτόνων» της (Chua et al. 2006, Nabieva et al. 2005, Schwikowski et al. 2000, Sharan et al. 2007) και β) η προσέγγιση η οποία στηρίζεται στην ανάκτηση πληροφοριών από καλά συνδεδεμένα μεταξύ τους συμπλέγματα πρωτεϊνών (clusters), αν ληφθεί υπόψη ότι οι υψηλά συνδεδεμένες πρωτεΐνες συχνά συμμετέχουν σε κοινές βιολογικές λειτουργίες (Shen et al. 2016, Lin et al. 2015).

Εφαρμόζοντας τη δεύτερη προσέγγιση στο δίκτυο PPI του σωματίου συναρμογής βρέθηκαν οκτώ στατιστικά σημαντικά συμπλέγματα πρωτεϊνών {π.χ. πρωτεΐνες Sm και Lsm, πρωτεΐνες που περιέχουν μοτίβο σερίνης-θρεονίνης (πρωτεΐνες SR), παράγοντες αναδιαμόρφωσης της χρωματίνης, πρωτεΐνες του εξωσώματος, (βλ. Ενότητα 3.2.4.3.)}, τα οποία επιβεβαιώνουν τη θεωρία ότι πρωτεΐνες με κοινό βιολογικό ρόλο στο κύτταρο είναι πυκνά συνδεδεμένες μεταξύ τους και αραιά συνδεδεμένες με τις υπόλοιπες. Ο αριθμός των συμπλεγμάτων που βρέθηκαν είναι σχετικά μικρός αναλογικά με τον συνολικό αριθμό των πρωτεϊνών αλλά, αν ληφθεί υπόψη η σχετικά χαμηλή τιμή των συντελεστών ομαδοποίησης (clustering coefficient) και της κατανομής μέσου όρου των συντελεστών ομαδοποίησης (average clustering coefficient) που δεν υποδεικνύει ένα δίκτυο με αυξημένη ικανότητα σχηματισμού συμπλεγμάτων, τότε είναι μεγάλος. Ο αριθμός των συμπλεγμάτων αναμένεται να αυξάνεται όταν εισάγονται στο δίκτυο και οι έμμεσες αλληλεπιδράσεις καθώς η εισαγωγή τους πιθανώς δημιουργεί ένα δίκτυο που προσομοιάζει τα σύμπλοκα που σχηματίζονται *in vivo*, με αποτέλεσμα να σχηματίζονται περισσότερα συμπλέγματα που περιέχουν μεγαλύτερο αριθμό πρωτεϊνών. Αυτό παρατηρήθηκε και στα πλαίσια του δικτύου συναρμογής (δεν παρουσιάζεται στα αποτελέσματα), δηλαδή όταν συμπεριελήφθησαν στο δίκτυο και οι έμμεσες αλληλεπιδράσεις αυξήθηκε τόσο ο αριθμός των συμπλεγμάτων (13 συμπλέγματα) όσο και των πρωτεϊνών (σε 4 από αυτά σχεδόν διπλασιάστηκε). Η αύξηση της ικανότητας ομαδοποίησης των πρωτεϊνών ενός δικτύου υποδηλώνει ότι αυτό είναι «ανθεκτικό» στη διάδοση τυχόν διαταραχών (Stelling et al. 2004). Η σχετικά χαμηλή τιμή των παραμέτρων αυτών στο δίκτυο του spliceosome μπορεί πιθανώς να οφείλεται στην ύπαρξη αρκετών «hubs» καθώς όσο μεγαλώνει ο αριθμός των αλληλεπιδράσεων συγκεκριμένων πρωτεϊνών, μειώνεται συνολικά η ικανότητα των πρωτεϊνών να ομαδοποιούνται («average clustering coefficient») λόγω της προϋπόθεσης ύπαρξης μεγάλου αριθμού αλληλεπιδράσεων μεταξύ των πρώτων γειτόνων των «hubs» που πρέπει να λάβουν χώρα. Επιπλέον, οι πρωτεΐνες-«hubs» έχουν μειωμένη πιθανότητα να μοιράζονται κοινούς γείτονες με τις υπόλοιπες πρωτεΐνες του δικτύου (topological coefficient) διότι ο μικρός αριθμός

αλληλεπιδράσεων των «γειτόνων» τους έχει ως αποτέλεσμα τη μειωμένη πιθανότητα εμφάνισης πολλών κοινών «γειτόνων» μεταξύ τους.

Στα συμπλέγματα του δικτύου που δημιουργήθηκαν οι περισσότερες πρωτεΐνες έχουν χαρακτηρισμένη κυτταρική λειτουργία. Βάσει της δεύτερης προσέγγισης που προαναφέρθηκε παραπάνω για τον καθορισμό της λειτουργίας άγνωστων πρωτεϊνών, σε ένα από αυτά, που περιλαμβάνει πρωτεΐνες του εξωσώματος, θα μπορούσε να γίνει πρόβλεψη για μία πρωτεΐνη που δεν έχει γνωστή λειτουργία (PBRM1). Θα μπορούσε να υποθεθεί ότι αυτή η πρωτεΐνη σχετίζεται με την αποικοδόμηση μη κατάλληλων μεταγράφων, αν ληφθεί υπόψη ότι όλες οι υπόλοιπες πρωτεΐνες του συμπλέγματος αποτελούν συστατικά του εξωσώματος. Επίσης, στο σύμπλεγμα των 10 πρωτεϊνών που είναι εμπλουτισμένο με μόρια που σχετίζονται με την προσκόλληση μορίων (5 πρωτεΐνες), υπάρχουν 2 πρωτεΐνες οι οποίες δεν έχουν γνωστό λειτουργικό ρόλο (CCDC130, SRP68). Έτσι επειδή όλες οι πρωτεΐνες του συμπλέγματος συμμετέχουν στην προσκόλληση μορίων είναι λογικό να γίνει η υπόθεση ότι και οι 2 πρωτεΐνες άγνωστης λειτουργίας συνεισφέρουν σε αυτή τη διαδικασία. Ωστόσο, διερευνώντας τις επιπρόσθετες λειτουργίες των πρωτεϊνών του συμπλέγματος και των ισομορφών τους παρατηρήθηκε ότι υπάρχουν πρωτεϊνικές ισομορφές με επιπρόσθετες πυρηνικές λειτουργίες καθώς και πολυλειτουργικές πρωτεΐνες. Ειδικότερα, μια πρωτεϊνική ισομορφή (PKP1b) του γονιδίου *PKP1* που σχετίζεται με την προσκόλληση μορίων εντοπίζεται αποκλειστικά στον πυρήνα και υπάρχουν ενδείξεις για τη συμμετοχή της στη ρύθμιση της μεταγραφής (Schmidt et al. 1997, Sobolik-Delmaire et al. 2010). Μία ακόμη πρωτεΐνη που συμμετέχει στην προσκόλληση μορίων (JUP, Huang et al. 2008) είναι πολυλειτουργική και εμφανίζει ενεργό ρόλο και στη μεταγραφή (Maeda et al. 2004). Οι υπόλοιπες 3 πρωτεΐνες που σχετίζονται με την προσκόλληση μορίων (PKP2, DSP, DSG1) έχουν διαφορετικές ισομορφές (Brennan et al. 2004, Cabral et al. 2010, Hatzfeld 2007) οι οποίες, αν και δεν έχουν γνωστό λειτουργικό ρόλο, μπορεί να συμμετέχουν σε κάποια πυρηνική λειτουργία, όπως συμβαίνει και με την πλακοφιλίνη 1 (PKP1b). Επιπλέον, μια ισομορφή (EEF1A1) του γονιδίου *EEF1A* έχει συσχετιστεί με τη μεταγραφή των γονιδίων σε απόκριση στο θερμικό σοκ (Vera et al. 2014) και η πρωτεΐνη KIN η οποία έχει γνωστό ρόλο στην αντιγραφή του DNA (Miccoli et al. 2003), έχει δειχθεί στους προκαρυώτες ότι σχετίζεται με τη ρύθμιση της μεταγραφής (Timchenko et al. 1996, Spurio et al. 1992). Με βάση τα δεδομένα αυτά θα μπορούσε να εξαχθεί το συμπέρασμα ότι οι 2 πρωτεΐνες άγνωστης λειτουργίας αυτού του συμπλέγματος μπορεί να συμμετέχουν στη μεταγραφή-ρύθμιση της μεταγραφής.

Όσον αφορά την πρωτεΐνη FRA10AC1, αυτή τοποθετήθηκε σε σύμπλεγμα πρωτεϊνών στα όρια της στατιστικής σημαντικότητας, το οποίο περιείχε όλους τους πρώτους «γείτονες» της, εκτός από δύο. Οι υπόλοιπες πρωτεΐνες συμπεριλαμβάνονταν στους δεύτερους «γείτονες» της FRA10AC1. Το σύμπλεγμα αυτό ήταν εμπλουτισμένο με παράγοντες που συμμετείχαν στη διαδικασία της συναρμογής (~53%). Ειδικότερα παρατηρήθηκαν αρκετές ελικάσες RNA (π.χ. DHX16, DHX38), οι οποίες συμμετέχουν στα καταλυτικά βήματα της συναρμογής (Gencheva et al. 2010), καθώς και παράγοντες αυτών των βημάτων (π.χ. SLU7) (Hegele et al. 2012). Επιπλέον βρέθηκαν πρωτεΐνες οι οποίες συμμετέχουν μεν στη διαδικασία της συναρμογής αλλά ταυτόχρονα ελέγχουν την ακεραιότητα των μεταγράφων συγκρατώντας στον πυρήνα τα μετάγραφα στα

οποία δεν έχει γίνει σωστά η συναρμογή {SNIP1, RBMX2, συστατικά του συμπλόκου συναρμογής και «συγκράτησης» (retention & splicing complex, RES) (Wysoczanski et al., 2014, Wysoczanski & Zweckstetter 2016). Σύμφωνα με τα παραπάνω θα μπορούσε να γίνει η υπόθεση ότι η FRA10AC1 καθώς και οι άλλες 6 πρωτεΐνες του συμπλέγματος που δεν έχουν γνωστό λειτουργικό ρόλο συνεισφέρουν στους μηχανισμούς επεξεργασίας του mRNA.

#### 4.2.11. Μεγάλος αριθμός πρωτεϊνών του δικτύου του σωματίου συναρμογής και της «γειτονιάς» της FRA10AC1 συσχετίζονται με νοσήματα

Δυο κατηγορίες ασθενειών οι οποίες βρέθηκαν σε αυξημένο ποσοστό στο δίκτυο του σωματίου συναρμογής είναι οι νευρολογικές και αναπτυξιακές ασθένειες, 27% και 36% του συνόλου των μενδελικών νοσημάτων, αντίστοιχα (βλ. Ενότητα 3.2.5.). Σε πολλές περιπτώσεις, έχει παρατηρηθεί ότι την αιτία αυτού του είδους των ασθενειών αποτελεί το αυξημένο ποσοστό της εσφαλμένης συναρμογής υπεύθυνων γονιδίων γι' αυτές τις ασθένειες, λόγω μεταλλάξεων που συμβαίνουν σε παράγοντες *trans* της συναρμογής (Vuong et al. 2016). Το γεγονός αυτό μπορεί να οφείλεται στην εκτεταμένη χρήση της εναλλακτικής συναρμογής στο νευρικό σύστημα των θηλαστικών, προκειμένου να παραχθούν συγκεκριμένες πρωτεϊνικές ισόμορφες, οι οποίες επηρεάζουν όλα τα στάδια της ανάπτυξης και λειτουργίας των νευρώνων (Vuong et al. 2016, Raj & Blencowe 2015, Zheng & Black 2013). Μια κατηγορία πρωτεϊνών, βλάβες των οποίων προκαλούν τέτοιου είδους νοσήματα είναι οι πρωτεΐνες που συνδέονται με το RNA (Nussbacher et al. 2015), οι οποίες διαδραματίζουν σημαντικό ρόλο στη ρύθμιση της συναρμογής και συναντώνται σε ποσοστό μεγαλύτερο του 70% στο δίκτυο του σωματίου συναρμογής. Για παράδειγμα, μεταλλάξεις των γονιδίων *TARDBP*, *HNRNPA1*, *FUS*, τα οποία περιλαμβάνονται και στο δίκτυο του σωματίου συναρμογής, προκαλούν μενδελικά νευρολογικά και αναπτυξιακά νοσήματα όπως η αμυοτροφική πλάγια σκλήρυνση (amyotrophic lateral sclerosis, ALS) και η εκφύλιση του μετωπιαίου και του κροταφικού λοβού (frontotemporal lobar degeneration, FTL) (Highley et al. 2014, Kim et al. 2013, Scotti & Swanson 2016, Zhou et al. 2014). Ειδικότερα οι μεταλλάξεις κέρδους λειτουργίας αυτών των γονιδίων προκαλούν συσσωματώματα πρωτεϊνών (πλάκες αμυλοειδούς-amyloids) ή οδηγούν σε λανθασμένη τοπολογία των αντίστοιχων γονιδιακών προϊόντων. Με αυτό τον τρόπο διαταράσσεται η ομοιοστάση των πρωτεϊνών και η συναρμογή. Για παράδειγμα, στην περίπτωση της *TARDBP* παρατηρείται η πλήρης απουσία της από τον πυρήνα, εξαιτίας της δημιουργίας κυτταροπλασματικών συσσωματωμάτων, με αποτέλεσμα τη διαταραχή του μηχανισμού της συναρμογής.

Ένα άλλο καλά μελετημένο παράδειγμα πρωτεΐνης που προκαλεί αναπτυξιακό γενετικό νόσημα (μυϊκή ατροφία του νωτιαίου μυελού, Spinal Muscular Atrophy) (Jangi et al. 2017) (βλ. Ενότητα 1.5.2.) είναι η *SMN1* η οποία αποτελεί «hub» του δικτύου του σωματίου συναρμογής. Η έλλειψη της πρωτεΐνης *SMN1* οδηγεί στην αποτυχία συναρμολόγησης των snRNPs, το οποίο υποδεικνύει ότι έχει βασικό ρόλο στη διαδικασία της συναρμογής (Cooper et al. 2009). Η εκδήλωση των φαινοτυπικών χαρακτηριστικών της ασθένειας στους κινητικούς νευρώνες μπορεί να οφείλεται στην ύπαρξη ιστοειδικών παραγόντων οι οποίοι λειτουργούν σε συνεργασία με την *SMN1*

και καθορίζουν τη σύσταση των snRNPs. Έτσι, η έλλειψη της SMN1 οδηγεί σε αλλαγή της σύστασής τους με ιστοειδικό τρόπο με αποτέλεσμα να αλλάζουν και οι διαταραχές στη διαδικασία της συναρμογής (Jangi et al. 2017, Cooper et al. 2009).

Όσον αφορά τα υπόλοιπα «hubs», τρία από αυτά (SRPK1, SRPK2, RBM10) έχουν συσχετιστεί με διάφορα είδη καρκίνου. Ειδικότερα η SRPK1 λειτουργεί ως ογκοκατασταλτικό γονίδιο και η υπερέκφρασή της έχει παρατηρηθεί σε διάφορους τύπους καρκίνου (π.χ. καρκίνο του μαστού, καρκίνο του πνεύμονα) (Li et al. 2014, Wang et al. 2014). Ομοίως, και η SRPK2 υπερεκφράζεται σε αυτούς τους τύπους καρκίνου (Gout et al. 2012). Επιπρόσθετα, έχει αναφερθεί συσχέτιση της SRPK2 με τη νόσο Alzheimer (Hong et al. 2012). Για την RBM10 έχει δειχθεί ότι σωματικές μεταλλάξεις του γονιδίου της έχουν σχετιστεί με αδενοκαρκίνωμα του πνεύμονα (Imielinski et al. 2012) ή καρκίνο του παγκρέατος (Furukawa et al. 2011).

Όσον αφορά τη «γειτονιά» της FRA10AC1, παρατηρήθηκε ότι είναι εμπλουτισμένη με πρωτεΐνες (PRPF3, PRPF4, PRPF6, PRPF8, PRPF31 και SNRNP200) μεταλλάξεις των οποίων προκαλούν μελαγχρωστική αμφιβληστροειδοπάθεια, μια κληρονομική μορφή τύφλωσης, που οφείλεται στον εκφυλισμό του αμφιβληστροειδή (Liu & Zack 2013) (βλ. Ενότητα 1.5.2.). Ειδικότερα μία από αυτές τις πρωτεΐνες (PRPF3) αποτελεί πρώτο «γείτονα» της FRA10AC1, γεγονός που μπορεί πιθανώς να υποδεικνύει τη συσχέτιση και της FRA10AC1 με αυτήν τη διαταραχή, δεδομένου ότι όταν μια πρωτεΐνη που είναι γνωστό ότι προκαλεί ή σχετίζεται με ένα γενετικό νόσημα, τότε οι πρωτεΐνες του δικτύου που ενώνονται μαζί της με τον μικρότερο αριθμό βημάτων (shortest path length) μπορεί να αποτελέσουν υποψήφια γονίδια που προκαλούν ή σχετίζονται με την ίδια γενετική ασθένεια ή την ίδια κατηγορία ασθενειών (Liu et al. 2012, Krauthammer et al. 2004, Goñi et al. 2008)

Επιπλέον στους πρώτους και δεύτερους «γείτονες της FRA10AC1 υπάρχουν αρκετές πρωτεΐνες (SF3B1, U2AF1, U2AF2 και SRSF2), στα γονίδια των οποίων έχουν ταυτοποιηθεί σωματικές μεταλλάξεις που σχετίζονται με διάφορα είδη καρκίνου (π.χ. μυελοειδείς νεοπλασίες, χρόνια μυελοκυτταρική λευχαιμία (Yoshida & Ogawa 2014, Malcovati et al. 2015) (βλ. Ενότητα 1.5.2.). Οι σωματικές μεταλλάξεις των γονιδίων *U2AF1* και *SRSF2* σε αιμοποιητικά κύτταρα προκαλούν αλλαγές στην αναγνώριση της 3' θέσης συναρμογής των γονιδίων-στόχων που σχετίζονται με την αιμοποίηση, με αποτέλεσμα τη λανθασμένη συναρμογή εκατοντάδων μεταγράφων και κατ' επέκταση βλάβες στη διαφοροποίηση των αιμοποιητικών κυττάρων (Ilgan et al. 2015, Komeno et al. 2015, Shirai et al. 2015, Zhang et al. 2015). Ειδικότερα, μεταλλάξεις του γονιδίου *U2AF1*, το οποίο αποτελεί πρώτο «γείτονα» της FRA10AC1, φαίνεται να καταστέλλουν τον κυτταρικό πολλαπλασιασμό και την κυτταρική διαφοροποίηση και να επάγουν τον προγραμματισμένο κυτταρικό θάνατο γεγονός το οποίο συνάδει με το κλινικό χαρακτηριστικό αυτών των ειδών καρκίνου, τη δριμεία κυτταροπενία, δηλαδή την αναποτελεσματική αιμοποίηση η οποία συνοδεύεται από αυξημένη ενεργότητα του μηχανισμού της απόπτωσης (Yoshida et al. 2011). Όσον αφορά την SF3B1, η οποία είναι υπεύθυνη για την πιστότητα της αναγνώρισης της θέσης διακλάδωσης στο 3' άκρο του ιντρονίου (Corrionero et al. 2011), έχει δειχθεί ότι, σωματικές μεταλλάξεις απώλειας λειτουργίας του γονιδίου *SF3B1* προκαλούν διαταραχές στη διαδικασία της συναρμογής, δηλαδή, δημιουργείται εναλλακτική 3' θέση συναρμογής στα γονίδια που

εξετάστηκαν στις αντίστοιχες μελέτες (Quesada et al. 2011). Οι τέσσερις αυτές πρωτεΐνες διαδραματίζουν καίριο ρόλο στα αρχικά στάδια της συναρμογής και όλες σχετίζονται κυρίως με αιμοποιητικούς καρκίνους, αν και οι ίδιες σωματικές μεταλλάξεις των γονιδίων *U2AF1* και *SF3B1* έχουν βρεθεί και σε άλλους τύπους καρκίνου (π.χ. καρκίνος του μαστού, του πνεύμονα, του παγκρέατος). Έτσι ενισχύεται η άποψη ότι πρωτεΐνες με παρόμοιο λειτουργικό ρόλο σχετίζονται με παρόμοιες ή ίδιες διαταραχές, με αποτέλεσμα να μπορεί να ενταχθεί και η πρωτεΐνη FRA10AC1 σε αυτήν την κατηγορία, δεδομένου του σκεπτικού που αναλύθηκε παραπάνω. Ένας τρόπος για να ελεγχθεί αυτό είναι η χρήση μεθόδων ταυτοποίησης μεταλλάξεων και αλληλούχισης (π.χ. exome sequencing) του γονιδίου *FRA10AC1* σε άτομα που εμφανίζουν αυτά τα είδη καρκίνου στον συγκεκριμένο ιστό.

Τέλος, η «γειτονιά» της FRA10AC1 τόσο εντός όσο και εκτός του σωματίου συναρμογής είναι εμπλουτισμένη σε πρωτεΐνες οι οποίες έχουν συσχετιστεί με νευροεκφυλιστικά νοσήματα. Ειδικότερα οι δεύτεροι γείτονες της, SNRNP70 και SRPK2, έχουν συσχετιστεί με τη νόσο Alzheimer (Bai et al. 2013, Hong et al. 2012), και οι πρώτοι «γείτονες» της εκτός του σωματίου συναρμογής, MAP3K12, VPS29 και EEF1D, τόσο με την ασθένεια Parkinson όσο και με τη νόσο Alzheimer (Berman et al. 2015, Schulte et al. 2014, Tedeschi & Bradke 2013). Αυτές οι παρατηρήσεις σε συνδυασμό με μια πρόσφατη πειραματική μελέτη στην οποία βρέθηκε συσχετισμός πολυμορφισμών του γονιδίου *FRA10AC1* με τη νόσο Alzheimer (Li et al. 2015) αποτελούν ισχυρή ένδειξη για τη συσχέτιση της FRA10AC1 με νοσήματα νευροεκφυλιστικής φύσεως.

Πέρα από την προσέγγιση που προαναφέρθηκε και αφορά τη διερεύνηση της «γειτονιάς» των πρωτεϊνών του δικτύου του σωματίου συναρμογής που προκαλούν ή σχετίζονται με γενετικά νοσήματα ώστε να προβλεφθούν πιθανά υποψήφια γονίδια για τα ίδια νοσήματα, εφαρμόστηκε και η επέκταση αυτής, η οποία αφορά την πρόβλεψη υποψήφιων γονιδίων για γενετικά νοσήματα, διερευνώντας τη «γειτονιά» πρωτεϊνών που έχουν τον ίδιο βιολογικό ρόλο και προκαλούν ή σχετίζονται με το ίδιο γενετικό νόσημα εντός των συμπλεγμάτων (clusters) του δικτύου. Ειδικότερα, στο 5<sup>ο</sup> cluster του δικτύου βρέθηκαν μεταλλάξεις 6 γονιδίων που προκαλούν το μενδελικό σύνδρομο Coffin-Siris και κωδικοποιούν πρωτεΐνες του συμπλόκου SWI/SNF, και στο 8<sup>ο</sup> cluster βρέθηκαν μεταλλάξεις 3 γονιδίων που προκαλούν μια μενδελική μορφή καρδιοπάθειας (ARVC) και κωδικοποιούν πρωτεΐνες οι οποίες συμμετέχουν στην προσκόλληση μορίων. Σχετικά με το 5<sup>ο</sup> cluster, στη «γειτονιά» των 6 πρωτεϊνών του συμπλόκου SWI/SNF υπάρχουν 3 πρωτεΐνες (SMARCC1, SMARCC2, SMARCD1) οι οποίες συμμετέχουν και αυτές στην οργάνωση και αναδιαμόρφωση της χρωματίνης καθώς αποτελούν συστατικά του συμπλόκου SWI/SNF. Έτσι, τα γονίδια που κωδικοποιούν αυτές τις 3 πρωτεΐνες θα μπορούσαν να αποτελέσουν πιθανά υποψήφια για το σύνδρομο Coffin-Siris. Στο ίδιο πλαίσιο, στο 8<sup>ο</sup> cluster, 2 γονίδια που κωδικοποιούν τις πρωτεΐνες DSG1 και PKP1, οι οποίες συμμετέχουν και αυτές όπως και οι «γείτονες» τους (JUP, PKP2, DSP) στην προσκόλληση μορίων και κυττάρων, θα μπορούσαν να αποτελέσουν πιθανά υποψήφια γονίδια για τη μενδελική μορφή καρδιοπάθειας η οποία οφείλεται σε μεταλλάξεις γονιδίων που κωδικοποιούν αυτές οι 3 πρωτεΐνες-«γείτονες».



### 4.3. Περιορισμοί και προοπτικές του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής

Παρά την πληθώρα των πληροφοριών που μπορούν να ανακτηθούν από το δίκτυο πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής του ανθρώπου, υπάρχουν ορισμένοι περιορισμοί που συναντώνται. Γενικότερα, στα δίκτυα PPI του ανθρώπου αυτοί οι περιορισμοί είναι μεγαλύτεροι σε σχέση με άλλα είδη οργανισμών, διότι οι κυτταρικοί μηχανισμοί και ο τρόπος σύνδεσης μεταξύ αυτών χαρακτηρίζεται από ιδιαίτερη πολυπλοκότητα λόγω της ποικιλίας των κυτταρικών τύπων, της ύπαρξης πολλών διαφορετικών ισομορφών των πρωτεϊνών, και του αυξημένου αριθμού των πιθανών μετα-μεταφραστικών τροποποιήσεων των πρωτεϊνών που συμμετέχουν στις αλληλεπιδράσεις (Figeys 2008). Οι δυο πρώτοι λόγοι της αυξημένης πολυπλοκότητας των κυτταρικών μηχανισμών του ανθρώπου μπορεί να οδηγήσουν στην διαφοροποίηση των πρωτεϊνικών αλληλεπιδράσεων μεταξύ των διαφορετικών ιστών του ανθρώπου (tissue-specific interactions), και ο τελευταίος μπορεί να επιφέρει αλλαγές στην δυναμική των αλληλεπιδράσεων των πρωτεϊνών οι οποίες μπορεί να αλλάζουν «γειτόνες» κατά τη διάρκεια του χρόνου που λαμβάνει χώρα μια κυτταρική διαδικασία. Όπως έχει ήδη αναφερθεί, η διαδικασία της συναρμογής είναι μια δυναμική διαδικασία, καθώς κατά τη διάρκεια της συναρμολόγησης του σωματίου συναρμογής αλλάζει η σύσταση των διαφορετικών υποσυμπλόκων σε πρωτεΐνες, αλληλεπιδράσεις μεταξύ πρωτεϊνών, μορίων RNA και μεταξύ πρωτεϊνών και μορίων RNA. Στην εργασία αυτή πραγματοποιήθηκε η ανασύσταση του πρωτεόματος του σωματίου συναρμογής και του δικτύου των αλληλεπιδράσεών του, το οποίο περιλάμβανε αποκλειστικά πρωτεΐνες και όχι μόρια RNA. Επιπλέον, σε αυτήν την εργασία έγινε μια προκαταρκτική προσπάθεια εξέτασης της δυναμικής των αλληλεπιδράσεων κατά τη διάρκεια συναρμολόγησης των υποσυμπλόκων του, η οποία στηρίχτηκε αποκλειστικά στην πρωτεϊνική σύστασή τους, καθώς υπάρχουν πολύ περιορισμένες πληροφορίες για μεταβολές αλληλεπιδράσεων των αντίστοιχων πρωτεϊνών κατά τη διάρκεια της διαδικασίας λόγω μετα-μεταφραστικών τροποποιήσεων. Στις επόμενες ενότητες αναλύονται οι περιορισμοί που συναντώνται στο δίκτυο του σωματίου συναρμογής και οι προοπτικές που μπορεί να υπάρξουν.

#### 4.3.1. Η εισαγωγή των αλληλεπιδράσεων των πρωτεϊνικών ισομορφών πιθανώς θα τροποποιούσε το δίκτυο

Ένας περιορισμός που συναντάται γενικότερα στα δίκτυα πρωτεϊνικών αλληλεπιδράσεων, συμπεριλαμβανομένου του δικτύου του σωματίου συναρμογής, είναι ότι δεν λαμβάνονται υπόψη οι ισομορφές των πρωτεϊνών και κατ' επέκταση οι αλληλεπιδράσεις τους. Όπως είναι γνωστό, περισσότερο από το 90% των γονιδίων του ανθρώπου που κωδικοποιούν πρωτεΐνες υφίστανται εναλλακτική συναρμογή (Gerstein et al. 2014, Lu et al. 2012), με αποτέλεσμα οι πρωτεϊνικές ισομορφές στον άνθρωπο να είναι τουλάχιστον 58% αυξημένες σε σχέση με τη *D. melanogaster*, 44% σε σχέση με τον *C. elegans* και 46% σε σχέση με το *Gallus gallus domesticus* (όρνιθα) (Blakeley et al. 2010). Ο μηχανισμός της εναλλακτικής συναρμογής μπορεί να τροποποιήσει τις πρωτεϊνικές αλληλεπιδράσεις αλλάζοντας την πρωτεϊνική ακολουθία και τη σύσταση των πρωτεϊνών σε περιοχές-domains οι οποίες μπορεί επιπλέον να αλλάζουν και την

τοπολογία των πρωτεϊνών (π.χ. πυρήνας-κυτταρόπλασμα) (Ellis et al. 2012). Έως τώρα ο αριθμός των πληροφοριών για τις αλληλεπιδράσεις μεταξύ των πρωτεϊνικών ισομορφών είναι περιορισμένος. Σε μια πρόσφατη ερευνητική μελέτη (Yang et al. 2016), ελέγχθηκαν οι αλληλεπιδράσεις 248 ισομορφών που κωδικοποιούνται από 161 γονίδια του ανθρώπου και δείχθηκε ότι τα πρότυπα των αλληλεπιδράσεων των εναλλακτικών ισομορφών που παράγονται από ένα γονίδιο, διαφέρουν σε πολύ μεγάλο ποσοστό (>50%). Το αντίστοιχο δίκτυο αλληλεπιδράσεων όλων των πρωτεϊνικών ισομορφών των 161 γονιδίων περιέχει τον τριπλάσιο αριθμό αλληλεπιδράσεων από αυτό που περιλαμβάνει μόνο τις αλληλεπιδράσεις των κύριων ισομορφών. Καμία βάση δεδομένων δεν διαθέτει τέτοιου είδους στοιχεία, με εξαίρεση την IntAct που περιλαμβάνει μόνο 116 αλληλεπιδράσεις μεταξύ πρωτεϊνικών ισομορφών χωρίς να δίνεται η δυνατότητα στο χρήστη να τις δει στο σύνολό τους. Μια σχετικά πρόσφατη μελέτη (Tseng et al. 2015), αφορά την δημιουργία μιας βάσης δεδομένων, η οποία προβλέπει τις αλληλεπιδράσεις των ισομορφών εφόσον ο χρήστης εισάγει τα δεδομένα έκφρασης των εναλλακτικών μεταγράφων που τον ενδιαφέρουν. Ειδικότερα, η βάση υπολογίζει ποια εναλλακτικά μετάγραφα συν-εκφράζονται σε διαφορετικούς ιστούς και σύμφωνα με επιπρόσθετα δεδομένα τα οποία σχετίζονται με τις αλληλεπιδράσεις μεταξύ περιοχών-domains των αντίστοιχων ισομορφών (οι πληροφορίες αυτές προέρχονται από τη βάση δεδομένων iPfam), προβλέπει ποιες από τις εναλλακτικές ισομορφές των πρωτεϊνών είναι πιθανότερο να αλληλεπιδρούν. Όπως προαναφέρθηκε, οι διαφορετικές ισομορφές μπορεί να παρουσιάζουν διαφορετική τοπολογία και κατ'επέκταση να έχουν διαφορετικές αλληλεπιδράσεις, μία περίπτωση που συναντάται στο δίκτυο των άμεσων αλληλεπιδράσεων της FRA10AC1. Ο άμεσος «γείτονας» της, η EEF1D, που δεν αποτελεί συστατικό του σωματίου συναρμογής, παράγει δυο διαφορετικές ισομορφές, την eEF1Bδ, η οποία λειτουργεί ως παράγοντας επιμήκυνσης της μετάφρασης, και την eEF1BδL που λειτουργεί ως μεταγραφικός παράγοντας γονιδίων τα οποία διαθέτουν στοιχεία θερμικού σοκ. Η κύρια λειτουργία του γονιδίου *EEF1D*, που είναι κυτταροπλασματική, δεν συνάδει με την αλληλεπίδραση της αντίστοιχης πρωτεΐνης με τη FRA10AC1, η οποία εμφανίζει αποκλειστικά πυρηνική τοπολογία κατά τη μεσόφαση. Ωστόσο, η ύπαρξη της πυρηνικής ισομορφής, eEF1BδL, πιθανώς υποδεικνύει την αλληλεπίδραση της FRA10AC1 με αυτή την ισομορφή. Η υπόθεση αυτή ενισχύεται και από τις αλληλεπιδράσεις της eEF1BδL με δυο πρωτεΐνες που εμφανίζουν πυρηνική λειτουργία, τους μεταγραφικούς παράγοντες HSF1, ο οποίος ενεργοποιείται στο στρες από υψηλή θερμοκρασία, και NFE2L2 που ρυθμίζει τη μεταγραφή γονιδίων τα οποία έχουν, στους υποκινητές τους, στοιχεία απόκρισης στο οξειδωτικό στρες (Kaitsuka et al. 2011) Αντίθετα, η πλειοψηφία των αλληλεπιδράσεων της ισομορφής eEF1Bδ σχετίζεται με παράγοντες έναρξης (π.χ. EIF1B) ή επιμήκυνσης (π.χ. EEF1A2, EEF1G) της μετάφρασης (Andersen et al. 2001, Sasikumar et al. 2012).

Στο πλαίσιο του δικτύου του σωματίου συναρμογής, εάν υπήρχαν οι πληροφορίες για τις πρωτεϊνικές ισομορφές που ταυτοποιήθηκαν σε κάθε υποσύμπλοκο και για τις αλληλεπιδράσεις των πρωτεϊνικών ισομορφών θα μπορούσαν πιθανώς να συνεισφέρουν στην προσθήκη αλληλεπιδράσεων μεταξύ πρωτεϊνών που δεν αλληλεπιδρούν οι κύριες ισομορφές τους ή αυτές που εξετάζονταν στην κάθε πειραματική δοκιμή ελέγχου των αλληλεπιδράσεων, αλλά και στην αφαίρεση

αλληλεπιδράσεων μεταξύ πρωτεϊνών καθώς μπορεί να μην υφίσταται αλληλεπίδραση μεταξύ αυτών των ισομορφών που είχαν εξεταστεί στα αντίστοιχα πειράματα. Επιπλέον οι παραπάνω πληροφορίες θα μπορούσαν να συμβάλλουν στην αιτιολόγηση της παρουσίας ορισμένων πρωτεϊνών που έχουν γνωστό κυτταροπλασματικό ρόλο στο πρωτόμα και στο δίκτυο του σωματίου συναρμογής καθώς θα υποδείκνυαν κάποια πιθανή πυρηνική λειτουργία των εναλλακτικών ισομορφών τους.

#### 4.3.2. Τα ιστο-ειδικά δίκτυα πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής θα προσομοιάζαν περισσότερο την *in vivo* κατάσταση

Οι ερευνητικές εργασίες που εξετάστηκαν για την ανακατασκευή του πρωτομάτος του σωματίου συναρμογής παρουσιάζουν ένα σημαντικό πλεονέκτημα. Περιλαμβάνουν πειράματα ταυτοποίησης και απομόνωσης των συστατικών του και των υποσυμπλόκων του τα οποία έλαβαν χώρα αποκλειστικά σε έναν κυτταρικό τύπο (*HeLa*). Ωστόσο, η ταυτοποίηση των αλληλεπιδράσεων των συστατικών του, που ανακτήθηκαν από τις βάσεις δεδομένων, πραγματοποιήθηκε σε ποικίλους κυτταρικούς τύπους και με διαφορετικές πειραματικές μεθόδους. Το γεγονός αυτό περιορίζει ως ένα βαθμό τις πληροφορίες διότι κάθε κυτταρικός τύπος εμφανίζει ιστο-ειδικές λειτουργίες, οι οποίες έχουν μεταβαλλόμενο αντίκτυπο στην αλληλεπίδραση μεταξύ των πρωτεϊνών. Έτσι, παρ' όλο που θα ήταν προτιμότερο να επιλέγονται οι μελέτες που έχουν πραγματοποιηθεί στον ίδιο κυτταρικό τύπο, αυτό θα μείωνε σε μεγάλο ποσοστό τα δεδομένα των πρωτεϊνών και των πρωτεϊνικών αλληλεπιδράσεων. Η ιστο-ειδική κατασκευή του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής θα μπορούσε να γίνει βάσει του συνδυασμού των δεδομένων έκφρασης (σε επίπεδο mRNA και πρωτεΐνης) σε διαφορετικούς ιστούς, αντλώντας πληροφορίες από αντίστοιχες βάσεις δεδομένων, για παράδειγμα την GTEx (<https://www.gtexportal.org>, Lonsdale et al. 2013) ή την MERAV (<http://merav.wi.mit.edu/>, Shaul et al. 2016).

#### 4.3.3. Η διερεύνηση της αλλαγής του «προτύπου» των αλληλεπιδράσεων κατά τη διάρκεια συναρμολόγησης του σωματίου συναρμογής θα συνεισέφερε στην ανάκτηση πληροφοριών για τη δυναμική του

Όπως έχει αναφερθεί, η συναρμογή είναι μια δυναμική διαδικασία κατά την οποία οι αλληλεπιδράσεις μεταξύ των μορίων που συμμετέχουν σε αυτή μπορεί να διατηρούνται καθ' όλη τη διάρκεια συναρμολόγησης του σωματίου συναρμογής ή να λαμβάνουν χώρα σε διαφορετικό χρόνο και να συνοδεύονται από την εναλλαγή μορίων RNA ή πρωτεϊνών. Στη διατριβή πραγματοποιήθηκε μια προκαταρκτική προσπάθεια για την ανάκτηση πληροφοριών σχετικά με τη δυναμική των αλληλεπιδράσεων με κριτήριο την πρωτεϊνική σύσταση των υποσυμπλόκων, δηλαδή, μία αλληλεπίδραση θεωρήθηκε ότι υφίσταται σε ένα υποσύμπλοκο εφόσον και οι δυο πρωτεΐνες που συμμετέχουν έχουν ταυτοποιηθεί ως συστατικά του υποσυμπλόκου αυτού. Το κριτήριο αυτό, παρόλο που περιορίζει την ανάκτηση μιας ολοκληρωμένης εικόνας της δυναμικής του σωματίου, χρησιμοποιήθηκε ως το μοναδικό γιατί δεν υπήρχαν αρκετές πληροφορίες για τις αλλαγές των αλληλεπιδράσεων κατά τη διάρκεια της συναρμολόγησης του σωματίου π.χ. λόγω μετα-μεταφραστικών τροποποιήσεων (βλ. Ενότητα 4.2.9.).

Ένας τρόπος πρόβλεψης της δυναμικής πρωτεϊνικών δικτύων, σε επίπεδο μεγάλης κλίμακας, θα μπορούσε να είναι η ενσωμάτωση πληροφοριών στα δίκτυα πρωτεϊνικών αλληλεπιδράσεων των υποσυμπλόκων του σωματίου συναρμογής, σχετικά με τα επίπεδα έκφρασης των συστατικών του είτε σε επίπεδο μεταγράφου είτε πρωτεΐνης. Σε μελέτες που εξετάζουν τη δυναμική των αλληλεπιδράσεων στα διαφορετικά στάδια του κυτταρικού κύκλου δημιουργούν ένα δίκτυο πρωτεϊνικών αλληλεπιδράσεων και στη συνέχεια αξιολογούν τις αλληλεπιδράσεις αυτές ως προς τη δυναμική τους στις διαφορετικές χρονικές στιγμές (Tang et al. 2011, Wang et al. 2013, Zhang et al. 2014). Η αξιολόγηση βασίζεται στα δεδομένα έκφρασης σε επίπεδο μεταγράφου των πρωτεϊνών που συμμετέχουν στις αλληλεπιδράσεις, στις διαφορετικές χρονικές στιγμές του κυτταρικού κύκλου που εξετάζονται. Ειδικότερα υπολογίζουν τα χρονικά σημεία κατά τα οποία ένα μετάγραφο εμφανίζει την υψηλότερη έκφραση. Αν τις ίδιες χρονικές στιγμές και κάποια άλλη πρωτεΐνη, η οποία με βάση το δίκτυο αλληλεπιδρά με την πρώτη, παρουσιάζει επίσης την υψηλότερη έκφρασή της, τότε θεωρούν αυτήν την αλληλεπίδραση αξιόπιστη αλλιώς την καταργούν. Σε παρόμοιο πλαίσιο θα μπορούσαν να αξιολογηθούν οι αλληλεπιδράσεις του κάθε υποσυμπλόκου βάσει της αφθονίας των πρωτεϊνών στα διαφορετικά υποσύμπλοκα, όπως έγινε και στην περίπτωση των 80 πρωτεϊνών που εμφανίζουν καθολική παρουσία κατά τη διάρκεια της συναρμογής. Λαμβάνοντας υπ' όψιν αυτά τα στοιχεία, οι ήδη υπάρχουσες αλληλεπιδράσεις του δικτύου του σωματίου συναρμογής που θεωρήθηκαν ότι λαμβάνουν χώρα καθ' όλη τη διάρκεια της διαδικασίας θα μπορούσαν να καταργηθούν σε ένα ή περισσότερα υποσύμπλοκα και να ανακτηθούν πληροφορίες για τη δυναμική τους.

Ένας εναλλακτικός τρόπος ελέγχου της δυναμικής των αλληλεπιδράσεων είναι η χρήση στοχευμένων πειραμάτων για συγκεκριμένες πρωτεΐνες χρησιμοποιώντας για παράδειγμα συναγωνιστικές δοκιμές συν-ανοσοκατακρήμνισης, όπως πραγματοποιήθηκε και στην εργασία των Hegele et. al. (2012) (βλ. Ενότητα 1.5.3.). Με τις δοκιμές αυτές είναι δυνατόν να εξεταστεί αν για μια πρωτεΐνη, που είναι γνωστό ότι αλληλεπιδρά με δυο άλλες πρωτεΐνες, είναι απαραίτητη, για να αλληλεπιδράσει με μία από αυτές, η παρουσία της τρίτης. Έτσι, σε κάποια υποσύμπλοκα μπορεί να καταργούνταν η αλληλεπίδραση μεταξύ δύο πρωτεϊνών διότι δεν θα μπορούσε να γίνει λόγω της απουσίας της τρίτης πρωτεΐνης.

Ένα άλλο χαρακτηριστικό γενικά των πρωτεϊνικών αλληλεπιδράσεων είναι ότι εξαρτώνται από τις μετα-μεταφραστικές τροποποιήσεις, οι οποίες έχουν την ικανότητα να ρυθμίζουν την ειδικότητα σύνδεσης μιας πρωτεΐνης με ποικίλους «γείτονες» σε διαφορετικές καταστάσεις. Στον άνθρωπο ο αριθμός των πρωτεϊνών που υφίστανται 5 βασικά είδη μετα-μεταφραστικών τροποποιήσεων (ακετυλίωση, γλυκοζυλίωση, μεθυλίωση, φωσφορυλίωση και ουβικουτινιλίωση) είναι κατά περίπου 12 φορές μεγαλύτερος (28894) σε σύγκριση με τους οργανισμούς *D. melanogaster* (2517) και *C. elegans* (2370), και κατά περίπου 7 φορές σε σχέση με τον *R. norvegicus* (αρουραίος, 4684) και τον *S. cerevisiae* (σακχαρομύκητας, 3896) (Duan & Walther 2015). Επιπλέον, αρκετά είδη τέτοιων τροποποιήσεων (π.χ. δημιουργία δισουλφιδικού δεσμού, αποκοπή πεπτιδικού τμήματος, υδροξυλίωση, σουμοϋλίωση) δεν υφίστανται σε λιγότερους πολύπλοκους οργανισμούς όπως αυτοί που προαναφέρθηκαν (Duan & Walther 2015). Μελέτες έχουν δείξει τη σημασία των τροποποιήσεων αυτών στη

διαδικασία της συναρμογής, όπως για παράδειγμα τη φωσφορυλίωση των SR πρωτεϊνών (McKay & Johnson 2010). Αν και μέχρι τώρα τα διαθέσιμα δεδομένα πρωτεϊνικών αλληλεπιδράσεων προέρχονται από μελέτες που δεν εξετάζουν αν οι πρωτεΐνες είναι τροποποιημένες ή όχι, έχουν αναπτυχθεί μέθοδοι για την ανάλυση των αλληλεπιδράσεων που εξαρτώνται από τέτοιου είδους τροποποιήσεις {π.χ. τροποποιημένο σύστημα δυο υβριδίων (Guo et al. 2004), μέθοδοι ταυτοποίησης με φασματομετρία μάζας (Li et al. 2012)}. Η ύπαρξη τέτοιων δεδομένων θα μπορούσε να συμβάλλει στην πιο ολοκληρωμένη απεικόνιση της δυναμικής του σωματίου συναρμογής. Για παράδειγμα, είναι γνωστό ότι τα διαδοχικά στάδια φωσφορυλίωσης και αποφωσφορυλίωσης της SF3B1 συμβάλλουν στα καταλυτικά βήματα της συναρμογής (Boudrez et al. 2002, De Graaf et al. 2006). Έτσι, η ταυτοποίηση των αλληλεπιδράσεων της πρωτεΐνης SF3B1 όταν είναι φωσφορυλιωμένη ή όχι θα μπορούσε να δώσει μια πιο πληροφοριακή εικόνα για τους «γείτονές» της, και πως αυτοί αλλάζουν ανάλογα με τη μορφή της πρωτεΐνης στα δυο καταλυτικά βήματα της συναρμογής. Ένα άλλο παράδειγμα αποτελεί η PRPF8, η συντονισμένη φωσφορυλίωση και ουβικουτινίωση της οποίας οδηγεί σε ένα πλήθος διαδοχικών και συντονισμένων αναδιατάξεων RNA-RNA και RNA-πρωτεΐνης (Grainger & Beggs 2005, Bellare et al. 2008). Γνωρίζοντας λοιπόν σε τι μορφή βρίσκεται μία πρωτεΐνη (π.χ. φωσφορυλιωμένη ή μη) στα διαφορετικά στάδια συναρμολόγησης του σωματίου συναρμογής, θα μπορούσε να ελεγχθεί αν υφίσταντο οι αλληλεπιδράσεις της στα διαφορετικά υποσύμπλοκα. Για παράδειγμα, μπορεί να έχει βρεθεί μια αλληλεπίδραση στα τρία πρώτα υποσύμπλοκα και σε ένα από αυτά, η μία από τις δυο πρωτεΐνες του ζεύγους να είναι σε αποφωσφορυλιωμένη μορφή ώστε να μην αλληλεπιδρά, ενώ στα άλλα δυο να φωσφορυλιώνεται και να αλληλεπιδρά.

#### 4.3.4. Η εισαγωγή των αλληλεπιδράσεων RNA-RNA και RNA-πρωτεϊνών στο δίκτυο του σωματίου συναρμογής θα συνέβαλλε στη δημιουργία ενός πιο ολοκληρωμένου δικτύου

Όλες οι διαδικασίες της επεξεργασίας του RNA, συμπεριλαμβανομένης της συναρμογής, διεκπεραιώνονται με τις συνεχείς αλληλεπιδράσεις πρωτεϊνών και μορίων RNA, είτε των πρόδρομων μεταγράφων είτε των μη κωδικών RNA. Έτσι, ένα ολοκληρωμένο δίκτυο πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής θα μπορούσε να αλλάξει αν προστίθεντο και οι αλληλεπιδράσεις των μορίων RNA. Το επόμενο βήμα που θα μπορούσε να ακολουθηθεί είναι η ενσωμάτωση στο δίκτυο των μορίων RNA (snRNA, mRNA). Μέχρι τώρα υπάρχει μια δημοσιευμένη μελέτη μεγάλης κλίμακας που αναφέρεται στις αλληλεπιδράσεις των U1snRNA και U2snRNA με περίπου 220 πρωτεΐνες του σωματίου συναρμογής (Chu et al. 2015). Σε επίπεδο μικρής κλίμακας, έχουν πραγματοποιηθεί αρκετές μελέτες που ταυτοποίησαν αλληλεπιδράσεις μεταξύ μορίων RNA-RNA και RNA-πρωτεϊνών των συμπλόκων U2snRNP, SF3b, SF3a (Lin & Xu 2012, Wang et al. 2005) και Prp19/CDC5L (Grote et al. 2010) με τα U1, U2, U5 και U6 snRNA και με το πρόδρομο μετάγραφο καθώς και αλληλεπιδράσεις μεταξύ πρωτεϊνών και mRNA κατά τη διάρκεια του 2<sup>ου</sup> καταλυτικού βήματος (Coltri et al. 2011) (βλ. Ενότητα 1.5.3.). Για τη συνολική ανάκτηση όλων των παραπάνω πληροφοριών, σημαντική είναι η συνεισφορά και των

βάσεων δεδομένων, οι οποίες περιέχουν πληροφορίες αλληλεπιδράσεων πρωτεϊνών με μη κωδικά μόρια RNA (Junge et al. 2017). Για παράδειγμα η βάση NPInter (<http://www.bioinfo.org/NPInter/index.htm>, Hao et al. 2016), η οποία στηρίζεται τόσο σε βιβλιογραφικές αναφορές όσο και σε πρόβλεψη αλληλεπιδράσεων, προσφέρει, μεταξύ άλλων δεδομένων, πληροφορίες για τις αλληλεπιδράσεις μεταξύ των snRNA και των πρωτεϊνών του σωματίου συναρμογής. Ενδεικτικά, σε αυτή τη βάση υπάρχουν περίπου 200-300 αλληλεπιδράσεις, από τη βιβλιογραφία, των U1snRNA και U2snRNA με πρωτεΐνες του σωματίου συναρμογής. Έτσι, η αξιοποίηση των διαθέσιμων πληροφοριών θα συνεισέφερε στη διαλεύκανση των αναδιατάξεων που λαμβάνουν χώρα σε όλη τη διάρκεια του σωματίου συναρμογής και ειδικότερα στα καταλυτικά βήματα αυτής.

Εναλλακτικά, θα μπορούσαν να γίνουν επιπρόσθετα στοχευμένα πειράματα, για τον προσδιορισμό των αλληλεπιδράσεων πρωτεϊνών με μόρια RNA σε συγκεκριμένα στάδια του σωματίου συναρμογής μετά από *in vitro* συναρμογή και διακοπή της στα στάδια αυτά. Οι πειραματικές προσεγγίσεις μπορεί να περιλαμβάνουν: α) τη χρήση ανοσοκατακρήμνισης και επακόλουθης ταυτοποίησης των μορίων RNA με αλληλούχηση έπειτα από την ομοιοπολική σύνδεση (crosslinking) πρωτεϊνών και RNA (Hafner et al. 2010, McHugh et al. 2014) β) την κατακρήμνιση των πρωτεϊνών και την επακόλουθη ταυτοποίησή τους με φασματομετρία μάζας έπειτα από την αλληλεπίδρασή τους με συγκεκριμένο μόριο RNA το οποίο ακινητοποιείται σε σφαιρίδια μέσω της χρήσης πρωτεϊνών ή RNA απταμερών (μικρού μήκους αλληλουχία ριβονουκλεοτιδίων που λειτουργεί ως «ετικέτα συγγένειας», π.χ. S1 απταμερή που συνδέονται στη βιοτίνη, Srisawat & Engelke 2001) τα οποία μπορούν να συνδέονται τόσο με το RNA και όσο με σφαιρίδια (π.χ. τα S1 απταμερή συνδέονται σε σφαιρίδια στρεπταβιδίνης) (McHugh et al. 2014, Keene et al. 2006) και γ) την ταυτοποίηση μορίων RNA με αλληλούχηση νέας γενιάς (RNA seq) έπειτα από την προσθήκη ετικετών στο μόριο RNA (RNA tagging) οι οποίες προστίθεντο από κατάλληλο ένζυμο (π.χ. πολυ-U πολυμεράση που προσθέτει ουριδίνες στο RNA) αφού έχει προηγηθεί η σύνδεση αυτού του ενζύμου με την εξεταζόμενη πρωτεΐνη και μόνο εφόσον λάβει χώρα η αλληλεπίδραση της πρωτεΐνης με το RNA (Lapointe et al. 2015).

#### **4.4. Η ανάλυση του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής σε συνδυασμό με τις ομικές αναλύσεις του κυτταρικού μοντέλου με τροποποιημένη έκφραση του γονιδίου *FRA10AC1* παρέχουν σημαντικές ενδείξεις για τη βιολογική λειτουργία της *FRA10AC1***

Σύμφωνα με την ανάλυση του δικτύου πρωτεϊνικών αλληλεπιδράσεων της *FRA10AC1* και του σωματίου συναρμογής που έχει προηγηθεί (βλ. Ενότητα 4.1. & 4.2., υποδεικνύεται ότι ένα δίκτυο πρωτεϊνικών αλληλεπιδράσεων, μεταξύ άλλων, μπορεί να λειτουργήσει ως εργαλείο για την εξαγωγή συμπερασμάτων σχετικά με τη βιολογική λειτουργία των συστατικών του για τα οποία δεν υπάρχουν επαρκείς πληροφορίες, εξετάζοντας τις λειτουργικές διαδικασίες που συμμετέχουν οι «γείτονές» τους. Ακόμη, σε ένα δίκτυο, έπειτα από την απομάκρυνση μιας πρωτεΐνης, μια κατάσταση που προσομοιάζει την καταστολή της έκφρασής της, μπορεί να ελεγχθούν

οι επιδράσεις αυτής, σε ποιοτικό επίπεδο, δηλαδή, στον αριθμό των αλληλεπιδράσεων του δικτύου και στην αρχιτεκτονική του (βλ. Ενότητα 4.2.9.). Ωστόσο, δεν είναι δυνατή η εξέταση των επιδράσεων αυτών σε ποσοτικό επίπεδο, δηλαδή δεν μπορεί να υπολογιστεί το ποσοστό αλλαγής που επιφέρει η απομάκρυνση της πρωτεΐνης στα επίπεδα έκφρασης των υπολοίπων πρωτεϊνών του δικτύου. Ένας από τους εναλλακτικούς τρόπους εξέτασης των επιδράσεων αυτών σε ποσοτικό επίπεδο οι οποίοι συνεισφέρουν, σε συνδυασμό με τα δίκτυα πρωτεϊνικών αλληλεπιδράσεων, στη διερεύνηση της βιολογικής λειτουργίας μίας πρωτεΐνης είναι η εύρεση των ποσοτικών αλλαγών που μπορεί να επιφέρει η τροποποίηση της έκφρασής της στην κυτταρική φυσιολογία, σε πολλαπλά μοριακά επίπεδα, από τα μετάγραφα στις πρωτεΐνες και στους μεταβολίτες, μέσω της χρήσης ομικών τεχνολογιών (βλ. Ενότητα 1.6.). Στην παρούσα εργασία, για να διερευνηθούν οι αλλαγές στην έκφραση των μορίων που προκαλούνται από την τροποποιημένη έκφραση της FRA10AC1 δημιουργήθηκε ένα κυτταρικό μοντέλο *HeLa* στο οποίο πραγματοποιήθηκε καταστολή της έκφρασής της. Οι αλλαγές αυτές ελέγχθηκαν μέσω της εφαρμογής ομικών τεχνολογιών οι οποίες προϋποθέτουν ένα πρωτόκολλο επεξεργασίας και συλλογής κυττάρων που επιφέρει τις λιγότερες αλλαγές στη κυτταρική φυσιολογία (βλ. Ενότητα 4.4.1.). Τα αποτελέσματα των ομικών αναλύσεων σε μεταγραφικό, πρωτεϊνικό και μεταβολικό επίπεδο του κυτταρικού μοντέλου σε συνδυασμό με τις αλληλεπιδράσεις της FRA10AC1 εντός του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής οδήγησαν στην διατύπωση πιθανών μοντέλων λειτουργίας της FRA10AC1 στις διαδικασίες επεξεργασίας του mRNA (βλ. Ενότητα 4.4.3.). Ακόμη, λαμβάνοντας υπ' όψιν τα αποτελέσματα των αναλύσεων αυτών και συνδυαστικά με τις αλληλεπιδράσεις της FRA10AC1 εκτός του σωματίου συναρμογής προτάθηκαν επιπρόσθετα μοντέλα λειτουργίας της πρωτεΐνης σε βιολογικές διαδικασίες πέραν της επεξεργασίας του mRNA, τα οποία χρήζουν περαιτέρω διερεύνηση (βλ. Ενότητα 4.4.2.).

#### 4.4.1. Η εφαρμογή ομικών αναλύσεων συνεισφέρει στην εύρεση των αλλαγών που προκαλούνται μετά την καταστολή της FRA10AC1 στην κυτταρική φυσιολογία

##### ➤ *Βελτιστοποίηση πρωτοκόλλου επεξεργασίας και συλλογής των κυττάρων HeLa*

Για την αξιολόγηση των αποτελεσμάτων των ομικών αναλύσεων του κυτταρικού μοντέλου *HeLa* με τροποποιημένη την έκφραση του γονιδίου *FRA10AC1*, και την εξαγωγή συμπερασμάτων για τις αλλαγές που επιφέρει η τροποποίηση αυτή στην κυτταρική φυσιολογία, έπρεπε να ληφθούν υπόψη δυο σημαντικές παράμετροι: α) η καρκινική φύση της κυτταρικής σειράς που χρησιμοποιήθηκε ως σύστημα μελέτης και β) η μέθοδος που χρησιμοποιήθηκε για την επεξεργασία και τη συλλογή των υπό εξέταση κυττάρων η οποία πρέπει να επιφέρει όσον το δυνατόν λιγότερες αλλαγές στη φυσιολογία τους. Έτσι, το βήμα που προηγήθηκε των ομικών αναλύσεων ήταν η βελτιστοποίηση της μεθόδου επεξεργασίας και συλλογής των κυττάρων *HeLa* η οποία έγινε μέσω της μεταβολομικής ανάλυσης τους.

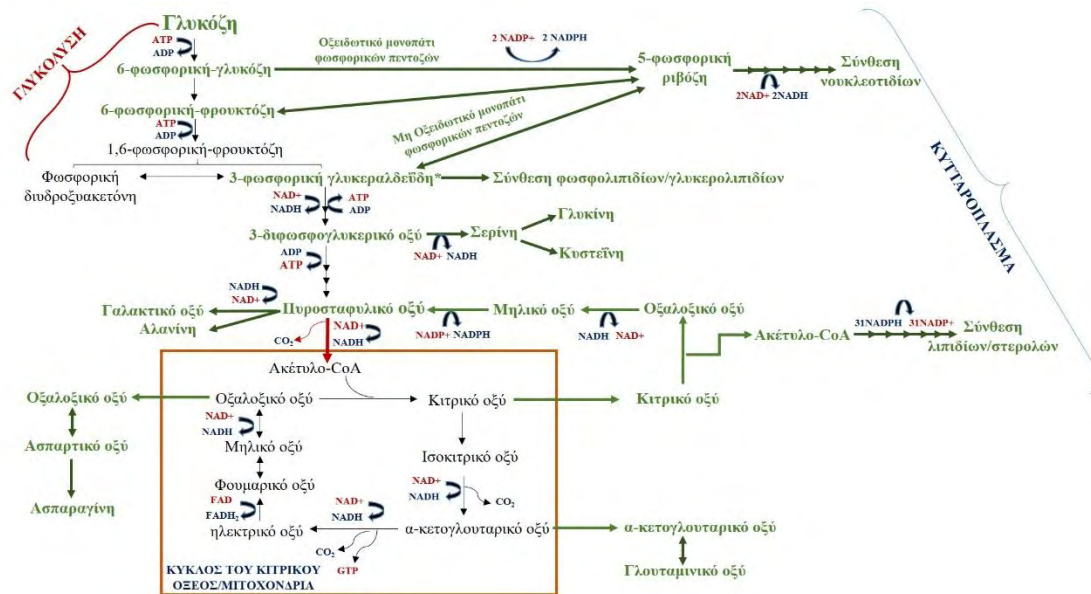
Η φυσιολογία των καρκινικών κυττάρων διαφέρει από αυτήν των φυσιολογικών ως προς την ανάγκη των πρώτων να χρησιμοποιήσουν τις ήδη υπάρχουσες μεταβολικές διαδικασίες για την εξυπηρέτηση ενός και μόνο σκοπού, την επιβίωση και τον

πολλαπλασιασμό τους. Ειδικότερα, τροποποιούν τις μεταβολικές διαδικασίες που υπάρχουν φυσιολογικά στα κύτταρα, δηλαδή, επαναπρογραμματίζουν τον μεταβολισμό τους, ώστε να ευνοούνται οι αναβολικές αντιδράσεις του μεταβολισμού. Με αυτόν τον τρόπο επιτρέπουν την χρήση θρεπτικών συστατικών για να παράγουν νουκλεοτίδια, λιπίδια, πρωτεΐνες και ενέργεια υπό τη μορφή μορίων υψηλής ενέργειας (π.χ. 5'-τριφωσφορική αδενοσίνη-ATP) ώστε να εξασφαλίσουν γρήγορη κυτταρική διαίρεση (De Berardinis & Chandel 2016, Phan et al. 2014). Αντίθετα, τα φυσιολογικά κύτταρα ευνοούν τις καταβολικές αντιδράσεις του μεταβολισμού, που επιτρέπουν τη διάσπαση των θρεπτικών συστατικών για την παραγωγή μορίων ATP μέσω της οξειδωτικής φωσφορυλίωσης, διατηρώντας μια ισορροπία μεταξύ της αποικοδόμησης και της σύνθεσης των υδατανθράκων, των λιπιδίων και των πρωτεϊνών (Mayers et al. 2015).

Η παλαιότερη και η πιο γνωστή μεταβολική αλλαγή (φαινόμενο “Warburg”) στα καρκινικά κύτταρα είναι η κατανάλωση της γλυκόζης με πολύ υψηλούς ρυθμούς, η οποία αντί να οξειδώνεται μέσω του κύκλου του κιτρικού οξέος, ρέει προς την παραγωγή του γαλακτικού οξέος ακόμα και όταν τα επίπεδα οξυγόνου είναι φυσιολογικά, διαδικασία γνωστή ως αερόβια γλυκόλυση (Bayley & Devilee 2012; Warburg 1956) (Εικ. 74). Η επιλογή της γλυκόλυσης και της ροής της προς την παραγωγή γαλακτικού οξέος από τα καρκινικά κύτταρα μπορεί να εξηγηθεί με ποικίλους τρόπους (DeBerardinis & Chandel 2016, Mayers et al. 2015). Ένας από τους αυτούς είναι ότι η γλυκόλυση αποτελεί την πηγή πρόδρομων μορίων για τις αναβολικές διεργασίες του μεταβολισμού που εξυπηρετούν τη βιοσύνθεση των μακρομορίων, μια βασική ανάγκη των καρκινικών κυττάρων (Mayers et al. 2015, Fuster & Esko 2005, Lunt & Vander Heiden 2011). Ειδικότερα, συνεισφέρει (Εικ. 74): α) στη βιοσύνθεση νουκλεοτιδίων μέσω του οξειδωτικού και του μη οξειδωτικού μονοπατιού των φωσφορικών πεντοζών (πάνω δεξιά), β) στη σύνθεση γλυκίνης και σερίνης από το φωσφογλυκερικό οξύ και αλανίνης από το πυροσταφυλικό οξύ (δεξιά και αριστερά στο κέντρο) γ) στη σύνθεση φωσφολιπιδίων και γλυκερολιπιδίων, τα οποία αποτελούν δομικό χαρακτηριστικό των κυτταρικών μεμβρανών, από την 3-φωσφορική γλυκεραλδεΐδη (δεξιά στο κέντρο) και δ) στο βιοσυνθετικό μονοπάτι των εξοζαμινών το οποίο παρέχει τα υποστρώματα για τη γλυκοζυλίωση των πρωτεϊνών και των λιπιδίων. Επιπλέον, το τελευταίο προϊόν της γλυκόλυσης, το πυροσταφυλικό οξύ μετατρέπεται, μέσω του κύκλου του κιτρικού οξέος στα μιτοχόνδρια, σε κιτρικό οξύ, το οποίο μεταφέρεται στο κυτταρόπλασμα και συνεισφέρει στη σύνθεση λιπιδίων και στερολών (δεξιά στο κέντρο) που αποτελούν συστατικά των μεμβρανών των κυττάρων. Ακόμη, τα ενδιάμεσα μόρια του κύκλου του κιτρικού οξέος, ακετογλουταρικό και οξαλοξικό οξύ συμμετέχουν στη σύνθεση του γλουταμικού οξέος, ασπαρτικού οξέος και ασπαραγίνης, αντίστοιχα (Εικ. 74). Τα παραπάνω υποδεικνύουν ότι η γλυκόλυση είναι η βασική πηγή τροφοδότησης των καρκινικών κυττάρων για να αναπτυχθούν και να διαιρεθούν. Ένας επιπρόσθετος λόγος επιλογής της γλυκόλυσης από τα καρκινικά κύτταρα είναι ότι το ποσοστό της παραγόμενης ενέργειας από τη γλυκόλυση, το οποίο είναι χαμηλότερο σε σύγκριση με την οξειδωτική φωσφορυλίωση, μπορεί να ξεπεράσει αυτό της οξειδωτικής φωσφορυλίωσης όταν η



ροή της γλυκόλυσης είναι πολύ μεγάλη (DeBerardinis et al. 2008, Gatenby & Gillies 2004).



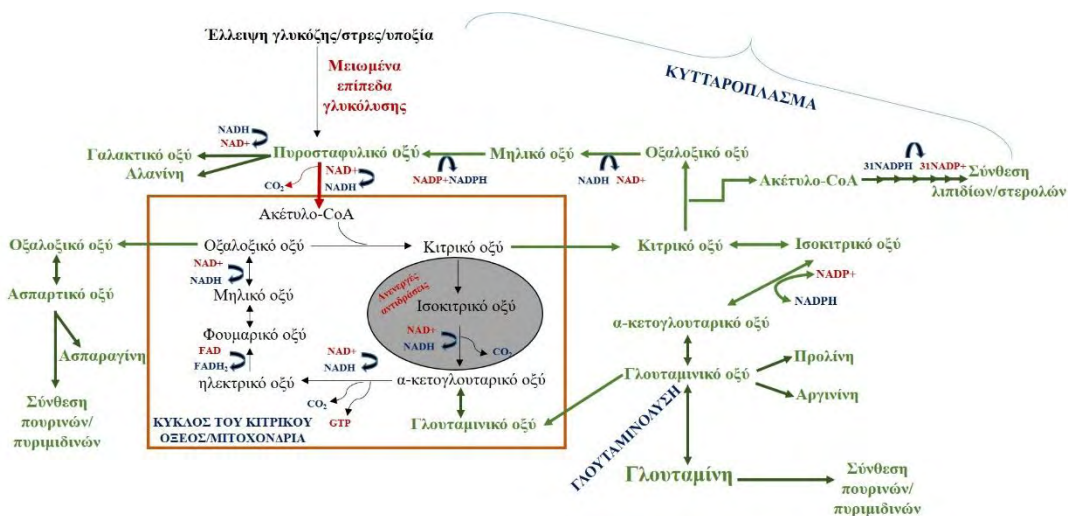
**Εικ. 74: Μεταβολικές προσαρμογές των καρκινικών κυττάρων.** Φαίνονται με πράσινη, έντονη γραφή και πράσινα βέλη οι μεταβολικές διαδικασίες που ενισχύονται στα καρκινικά κύτταρα. \*: σχηματίζονται δυο μόρια 3-φωσφορικής γλυκεραλδεΐδης.

Επιπλέον, η επιλογή της αερόβιας γλυκόλυσης από τα καρκινικά κύτταρα μπορεί να εξηγηθεί, από το γεγονός ότι η αναγωγή του πυροσταφυλικού οξέος σε γαλακτικό επιτρέπει την μετατροπή του νικοτιναμίδου-αδενινου-δινουκλεοτιδίου NADH στην οξειδωμένη μορφή του, το NAD<sup>+</sup> (Εικ. 74 κέντρο αριστερά). Το μόριο αυτό είναι απαραίτητο για τη συνεχή γλυκολιτική ροή καθώς συνεισφέρει στη μετατροπή της 3-φωσφορικής γλυκεραλδεΐδης σε 3-φωσφογλυκερικό οξύ και στη σύνθεση των αμινοξέων και των νουκλεοτιδίων ευνοώντας και πάλι τις αναβολικές διεργασίες του μεταβολισμού (Εικ. 74 κέντρο και πάνω δεξιά). Μέσα από αυτές τις διαδικασίες παράγεται NADH ώστε να επιτυγχάνεται η διατήρηση της οξειδοαναγωγικής ισορροπίας NAD<sup>+</sup>/NADH (Chiarugi et al. 2012). Αυτή είναι απαραίτητη καθώς τα δύο μόρια συμμετέχουν σε σηματοδοτικά μονοπάτια που ρυθμίζουν την έκφραση γονιδίων τα οποία συνεισφέρουν στην κυτταρική ανάπτυξη μέσω της ρύθμισης της προόδου του κυτταρικού κύκλου και στην επιδιόρθωση DNA (Chiarugi et al. 2012).

Μια άλλη μεταβολική αλλαγή των καρκινικών κυττάρων, η οποία ενεργοποιείται σε συνθήκες στρες (Benjamin et al. 2012), έλλειψης γλυκόζης (Wise et al. 2011) και υποξίας (Le et al. 2012) καθώς και σε κύτταρα που φέρουν μεταλλάξεις στα σύμπλοκα πρωτεϊνών της αλυσίδας μεταφοράς ηλεκτρονίων (Mullen et al. 2012), είναι η διαδικασία της γλουταμινόλυσης. Η διαδικασία αυτή λαμβάνει χώρα στο κυτταρόπλασμα και περιλαμβάνει την απαμίνωση της γλουταμίνης σε γλουταμινικό οξύ το οποίο μπορεί να ακολουθήσει δυο δρόμους (Altman et al. 2016, Lunt & Heiden, 2011) (Εικ. 75 κάτω δεξιά). Ο πρώτος οδηγεί είτε στην παραγωγή των λιπιδίων και στερολών είτε στην αερόβια γλυκόλυση (Εικ. 75 δεξιά και στο κέντρο). Ο δεύτερος οδηγεί στην παραγωγή του α-κετογλουταρικού οξέος το οποίο, στα καρκινικά κύτταρα,

δεν παράγεται μέσω του κύκλου του κιτρικού οξέος διότι παραμένουν ανενεργές οι αντιδράσεις από το κιτρικό οξύ έως το α-κετογλουταρικό οξύ, μια διαδικασία γνωστή ως «κομμένος» κύκλος του κιτρικού οξέος (truncated TriCarboxylic Acid, TCA cycle). Η διαδικασία αυτή συμβαίνει στα καρκινικά κύτταρα για να ενισχύσει την παραγωγή των λιπιδίων και των στερολών καθώς και την αερόβια γλυκόλυση, (Lunt & Heiden 2011) (Εικ. 75 κέντρο και δεξιά).

Επιπλέον, μέσω της γλουταμινόλυσης «ευνοείται» η σύνθεση πρωτεϊνών (Εικ. 75 κάτω δεξιά, κάτω και κέντρο αριστερά) (Altman et al. 2016, Hosios et al. 2016, Phang et al. 2015) και η σύνθεση πουρινών και πυριμιδινών ενισχύοντας την κυτταρική ανάπτυξη και διαίρεση των καρκινικών κυττάρων (Εικ. 75 κάτω δεξιά και αριστερά) (Lane & Fan 2015, Sullivan et al. 2015). Ακόμη, η παραγωγή του πυροσταφυλικού οξέος οδηγεί αφενός στην ενίσχυση της αερόβιας γλυκόλυσης και στην αναγέννηση του NAD<sup>+</sup> (Εικ. 75 κέντρο αριστερά), και αφετέρου στην αναγέννηση του NADPH το οποίο είναι απαραίτητο για τη σύνθεση λιπιδίων και στερολών (Εικ. 75 κέντρο δεξιά). Εκτός από τη συνεισφορά της γλουταμινόλυσης στις βιοσυνθετικές διεργασίες των καρκινικών κυττάρων, η συμμετοχή της στην αναπλήρωση των ενδιάμεσων του κύκλου του κιτρικού οξέος οδηγεί στην παραγωγή των μορίων νικοτιναμιδο-αδενινουκλεοτιδίου (NADH) και φλαβινο-αδενινουκλεοτιδίου (FADH<sub>2</sub>), μέσω των οποίων παράγεται ενέργεια (ATP) με τη διαδικασία της οξειδωτικής φωσφορυλίωσης (Εικ. 75 κέντρο πορτοκαλί πλαίσιο).



**Εικ. 75:** Η διαδικασία της γλουταμινόλυσης και οι μεταβολικές αντιδράσεις που έπονται αυτής. Φαίνονται με πράσινη, έντονη γραφή και πράσινα βέλη οι διαδικασίες που ενεργοποιούνται στα καρκινικά κύτταρα όταν ο ρυθμός της γλυκόλυσης είναι μειωμένος.

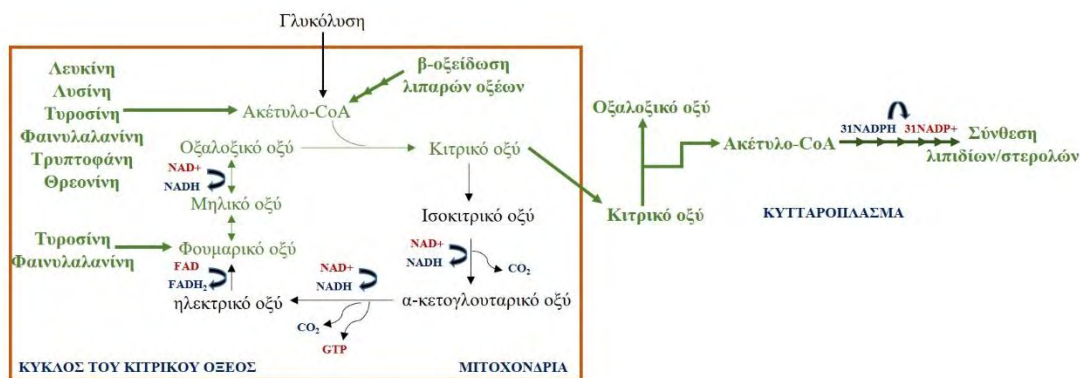
Τα παραπάνω χαρακτηριστικά του καρκινικού μεταβολισμού λήφθησαν υπόψη για την αξιολόγηση της μεταβολομικής ανάλυσης των τεσσάρων εναλλακτικών πρωτοκόλλων που σχεδιάστηκαν για τη συλλογή των κυττάρων *HeLa*. Ένα σημαντικό βήμα της ανάλυσης αυτής είναι ο χειρισμός των κυττάρων πριν και κατά τη συλλογή τους (Zhang et al. 2013). Αυτό είναι απαραίτητο διότι τα κύτταρα πρώτον, πρέπει να διατηρήσουν την ακεραιότητα των μεμβρανών τους ώστε να μπορεί να γίνει η μέτρηση των ενδογενών μεταβολιτών τους αποφεύγοντας την διαρροή αυτών στο θρεπτικό μέσο,

και δεύτερον, πρέπει να μην αναπτύσσονται σε συνθήκες στρες οι οποίες τους οδηγούν να προσαρμόσουν τις μεταβολικές τους διαδικασίες ώστε να μπορούν να επιβιώσουν (Karooro et al. 2015, Muschet et al. 2016). Τα τέσσερα πρωτόκολλα που σχεδιάστηκαν περιλάμβαναν είτε την έκπλυση των κυττάρων στο τρυβλίο καλλιέργειας, την μηχανική αποκόλληση και συλλογή τους στο διάλυμα έκπλυσης (πρωτόκολλο 1), είτε την μηχανική αποκόλληση των κυττάρων στο θρεπτικό μέσο ανάπτυξης, την έκπλυσή τους μέσω φυγοκεντρήσεων, και τη συλλογή τους στο διάλυμα έκπλυσης (πρωτόκολλο 2), ή την έκπλυση των κυττάρων στο τρυβλίο καλλιέργειας σε 10min (πρωτόκολλο 3A) και σε 1min (πρωτόκολλο 3B) και την μηχανική αποκόλληση και συλλογή τους σε κρύο διάλυμα μεθανόλης (πρωτόκολλο 3A και 3B) (βλ. Ενότητα 2.4.4.). Χρησιμοποιήθηκαν τρία διαφορετικά διαλύματα έκπλυσης, το διάλυμα φωσφορικών αλάτων (1xPBS), ο φυσιολογικός ορός (0.9% w/v NaCl) και το θρεπτικό μέσο DMEM, ώστε να ελεγχθεί η αποτελεσματικότητα της έκπλυσης σε κάθε περίπτωση. Επίσης, δοκιμάστηκαν και διαφορετικά στάδια ανακαλλιέργειας για να διαπιστωθεί εάν ο αυξημένος αριθμός των θρυψινοποιήσεων υποβάλλει τα κύτταρα να αναπτυχθούν σε συνθήκες στρες (βλ. Ενότητες 2.4.4., 3.4.1.). Οι διαφορές μεταξύ των διαφορετικών καταστάσεων που χρησιμοποιήθηκαν υποδεικνύουν το βέλτιστο πρωτόκολλο επεξεργασίας και συλλογής των κυττάρων *HeLa*.

- Τα πρωτόκολλα 1 και 2 παρουσιάζουν «φτωχό» μεταβολικό πρότυπο και υποβάλλουν τα κύτταρα σε συνθήκες έλλειψης γλυκόζης και άλλων θρεπτικών συστατικών

Στα πρωτόκολλα 1 και 2, αν και χρησιμοποιήθηκε ο ίδιος τρόπος μηχανικής αποκόλλησης των κυττάρων στα διαλύματα έκπλυσης, οι διαφορές μεταξύ τους έγκεινται: α) στο διάλυμα αποκόλλησης των κυττάρων, το οποίο είναι PBS ή 0.9% w/v NaCl για το πρώτο και DMEM για το δεύτερο, και β) στα βήματα των φυγοκεντρήσεων τα οποία απουσιάζουν από το πρώτο πρωτόκολλο και μειώνουν τη διάρκειά του. Η μεταβολομική ανάλυση έδειξε ότι η χρήση του PBS ως διάλυμα αποκόλλησης των κυττάρων πιθανώς επηρεάζει τις τιμές των φωσφορικών μεταβολιτών καθώς οι μεταβολίτες που έχουν στη δομή τους φωσφορικές ομάδες εμφάνισαν υψηλότερες συγκεντρώσεις στο πρωτόκολλο 1. Σύμφωνα με αυτό, η χρήση του DMEM ήταν αφενός μια εναλλακτική επιλογή για την αποκόλληση των κυττάρων και αφετέρου προσέφερε πιο ευνοϊκές συνθήκες ανάπτυξης στα κύτταρα καθώς οι συγκεντρώσεις όλων των μεταβολιτών ήταν υψηλότερες στο δεύτερο πρωτόκολλο σε σχέση με το πρώτο. Ωστόσο, ο αριθμός και οι συγκεντρώσεις των μεταβολιτών στα δυο αυτά πρωτόκολλα ήταν αρκετά πιο χαμηλοί σε σχέση με τα πρωτόκολλα 3A και 3B. Η διαφορά αυτών των ζευγών πρωτοκόλλων έγκειται στην προσθήκη μεθανόλης. Στα πρωτόκολλα 1 και 2, η διαδικασία της επεξεργασίας και συλλογής κυττάρων είχε διάρκεια 30 και 45min, αντίστοιχα, και στα κύτταρα, αφού είχαν συλλεχθεί σε μορφή πελλέτας, προστίθετο η μεθανόλη ώστε να κατασταλούν οι ενζυμικές αντιδράσεις και να εξαχθούν οι μεταβολίτες. Αντίθετα, στα πρωτόκολλα 3A και 3B, η μεθανόλη προστίθετο στο τρυβλίο καλλιέργειας για την αποκόλληση των κυττάρων και την ταυτόχρονη καταστολή της ενζυμικής δραστηριότητας, έπειτα από τη μικρή χρονική διάρκεια των εκπλύσεων. Οι διαφορές λοιπόν του πρώτου ζεύγους πρωτοκόλλων σε σχέση με τα άλλα δυο, ως προς τον αριθμό των μεταβολιτών, μπορούν να αποδοθούν

στη διαρροή των μεταβολιτών μέσω της κυτταροπλασματικής μεμβράνης λόγω της μεγάλης διάρκειας της επεξεργασίας των κυττάρων μέχρι την καταστολή των ενζυμικών αντιδράσεων. Σε αυτό το χρονικό διάστημα, η σύσταση σε ιόντα των διαλυμάτων έκπλυσης μπορεί να δημιουργεί διαφορετικές ωσμωτικές πιέσεις μεταξύ του εξωκυττάρου και ενδοκυττάρου περιβάλλοντος με αποτέλεσμα τα κύτταρα να υποστούν ωσμωτικό σοκ (Karooro et al. 2017). Επιπλέον, η μεγάλη χρονική διάρκεια των σταδίων έκπλυσης, μηχανικής αποκόλλησης και συλλογής των πρωτοκόλλων 1 και 2 υποβάλλει τα κύτταρα σε συνθήκες έλλειψης γλυκόζης και άλλων θρεπτικών συστατικών ακόμη και στην περίπτωση του δεύτερου πρωτοκόλλου κατά το οποίο η αποκόλληση γίνεται στο θρεπτικό μέσο ανάπτυξης αλλά οι εκπλύσεις έπονται αυτής. Όταν τα κύτταρα υποβάλλονται σε τέτοιες συνθήκες ενεργοποιούν αντίστοιχους μηχανισμούς απόκρισης, ξεκινώντας με την ενεργοποίηση των μεταφορέων γλυκόζης για την εισαγωγή νέων μορίων γλυκόζης και την αποικοδόμηση μορίων-αποθηκών ενέργειας (π.χ. γλυκογόνο) ώστε να παραχθούν εκ νέου ενδιάμεσα μόρια της γλυκόλυσης (Labak et al. 2016, Zois & Harris 2016). Εναλλακτικοί μηχανισμοί είναι η β-οξειδωση των λιπαρών οξέων, η αποικοδόμηση των αμινοξέων και η γλουταμινόλυση, μέσω των οποίων παράγονται ενδιάμεσα μόρια του κύκλου του κιτρικού οξέος και της γλυκόλυσης, που συμβάλλουν στις βιοσυνθετικές διεργασίες των κυττάρων (Owen et al. 2002, Raulien et al. 2017). Τα μεταβολικά πρότυπα των πρωτοκόλλων 1 και 2 χαρακτηρίζονται από μικρές συγκεντρώσεις της γλυκόζης, του πυροσταφυλικού και του γαλακτικού οξέος, σε σύγκριση με τα πρωτόκολλα 3A και 3B, γεγονός που υποδηλώνει ότι σε αυτά τα κύτταρα δεν «υποστηρίζεται» μια χαρακτηριστική διαδικασία του καρκινικού μεταβολισμού, η αερόβια γλυκόλυση. Επιπλέον, η μειωμένη συγκέντρωση συγκεκριμένων αμινοξέων υποδεικνύει την αποικοδόμησή τους για την αναπλήρωση ενδιάμεσων μορίων του κύκλου του κιτρικού οξέος τα οποία οδηγούν στην παραγωγή του κιτρικού οξέος ώστε αυτό να χρησιμοποιηθεί στη λιπογένεση (Beloribi-Djefaflija et al. 2016) (Εικ. 76).



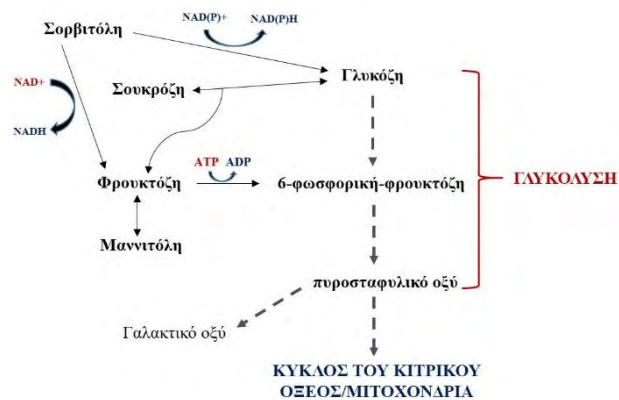
**Εικ. 76: Μεταβολικές αλλαγές που παρατηρούνται στα πρωτόκολλα 1 και 2.** Φαίνονται με πράσινη, έντονη γραφή και πράσινα βέλη οι αντιδράσεις ανάκτησης του κιτρικού οξέος και η επακόλουθη διαδικασία της λιπογένεσης.

Η μειωμένη συγκέντρωση της γλουταμίνης θα μπορούσε να ενισχύσει την παραγωγή του κιτρικού οξέος μέσω της γλουταμινόλυσης, ωστόσο, μόρια τα οποία συμμετέχουν στις αντιδράσεις που έπονται αυτής της διαδικασίας δεν παρουσιάζουν αυξημένες

συγκεντρώσεις, γεγονός που υποδεικνύει τη μη συμμετοχή της γλουταμίνης στην αναπλήρωση αυτών των μορίων. Οι συνθήκες έλλειψης θρεπτικών συστατικών στις οποίες υποβάλλονται τα κύτταρα των πρωτοκόλλων 1 και 2 μπορούν να ωθήσουν τα κύτταρα να στραφούν στην αυτοφαγία, το κύριο μονοπάτι λυσοσωματικής αποικοδόμησης, που ανακυκλώνει το κατεστραμμένο και πιθανόν επιβλαβές κυτταρικό υλικό, μέσω της ενεργοποίησης διαφόρων σηματοδοτικών μονοπατιών ώστε να αναπαραχθούν τα απαραίτητα πρόδρομα μόρια των βιοσυνθετικών διαδικασιών (Altman & Rathmell 2012, Caro-Maldonado 2011). Ωστόσο, αυτός ο μηχανισμός μπορεί να μην είναι επαρκής και τα κύτταρα να οδηγηθούν σε απόπτωση, μέσω της ενεργοποίησης και πάλι σηματοδοτικών μονοπατιών (Altman & Rathmell 2012, Ramírez-Reinado et al. 2013). Σύμφωνα με τα παραπάνω, και λαμβάνοντας υπόψη τον μη ικανοποιητικό βαθμό επαναληψιμότητας των μετρήσεων των επαναλήψεων στα πρωτόκολλα 1 και 2, γίνεται κατανοητό ότι αυτοί οι τρόποι επεξεργασίας και συλλογής κυττάρων δεν είναι αποτελεσματικοί. Τα πρωτόκολλα 3A και 3B εμφανίζουν ένα πλούσιο μεταβολικό πρότυπο καθώς περιλαμβάνουν τους διπλάσιους μεταβολίτες σε σχέση με τα άλλα δυο και δεν φαίνεται να υποβάλλουν τα κύτταρα σε συνθήκες στρες, καθώς η συγκέντρωση των μεταβολιτών τους είναι ιδιαίτερα αυξημένη και επαρκής ώστε τα καρκινικά κύτταρα να μπορούν να ανταποκριθούν στις αυξανόμενες μεταβολικές ανάγκες. Ωστόσο, υπάρχουν κάποιες διαφορές μεταξύ των πρωτοκόλλων 3A και 3B οι οποίες καθιστούν το ένα από αυτά το βέλτιστο για την επεξεργασία και συλλογή των κυττάρων.

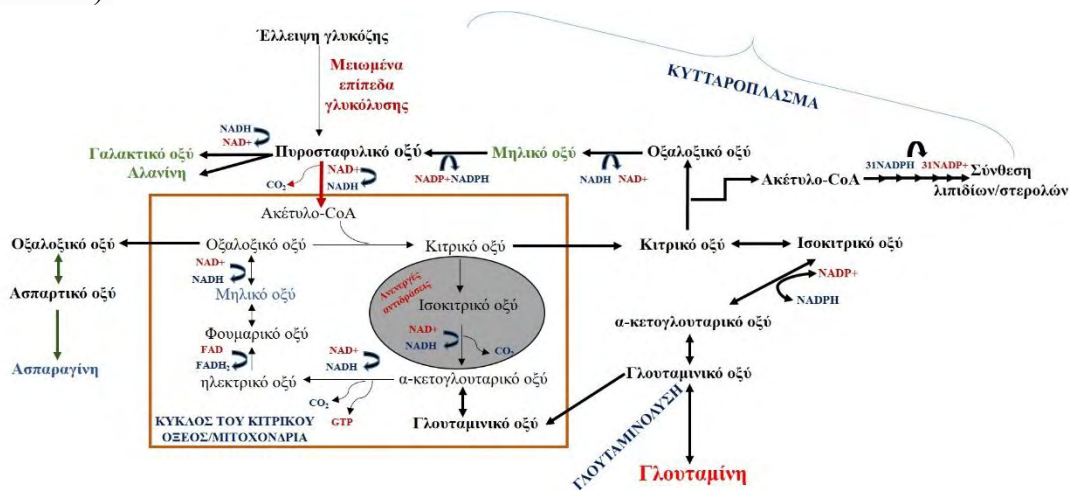
- Στο πρωτόκολλο 3A, σε σύγκριση με το 3B, επάγονται μηχανισμοί απόκρισης σε συνθήκες έλλειψης θρεπτικών συστατικών

Η διαφορά των πρωτοκόλλων 3A και 3B έγκειται μόνο στο χρονικό διάστημα των εκπλύσεων μέχρι την καταστολή της ενζυμικής δραστηριότητας μέσω της προσθήκης μεθανόλης. Η επιλογή του μεγαλύτερου χρονικού διαστήματος στο πρωτόκολλο 3A έγινε για να αυξήσει την αποτελεσματικότητα της έκπλυσης η οποία, ωστόσο, παρατηρήθηκε ότι είναι εξίσου αποτελεσματική και στα δυο πρωτόκολλα. Ωστόσο, η αύξηση του χρόνου επεξεργασίας των κυττάρων του πρωτοκόλλου 3A φαίνεται, από τη μεταβολομική ανάλυση, ότι επηρέασε αρνητικά τα κύτταρα. Ειδικότερα, η γλυκόζη εμφάνισε υποπενταπλάσια συγκέντρωση στο πρωτόκολλο 3A σε σύγκριση με το 3B, γεγονός που υποδεικνύει ότι τα κύτταρα πρέπει να ενεργοποιήσουν μηχανισμούς αποκατάστασης των ενδιάμεσων της γλυκόλυσης και του κύκλου του κιτρικού οξέος ώστε να εξασφαλίσουν την ενίσχυση της αερόβιας γλυκόλυσης και των βιοσυνθετικών διεργασιών. Η ενεργοποίηση αυτών των μηχανισμών αντικατοπτρίζεται στο μεταβολικό πρότυπο του πρωτοκόλλου 3A. Η παρουσία των υδατανθράκων σουκρόζη, μαννιτόλη και σορβιτόλη αποκλειστικά στα κύτταρα του πρωτοκόλλου 3A υποδεικνύει τη χρήση τους ως μόρια τα οποία διασπώνται για να αποκαταστήσουν τα χαμηλά επίπεδα της γλυκόζης και τα ενδιάμεσα μόρια της γλυκόλυσης (Εικ. 77).



**Εικ. 77: Πορεία αποκατάστασης της γλυκόζης.** Φαίνεται η διαδικασία μετατροπής των σακχάρων σουκρόζη, μαννιτόλη και σορβιτόλη σε γλυκόζη και στα ενδιάμεσα μόρια της γλυκόλυσης (6-φωσφορική φρουκτόζη).

Ένας άλλος μηχανισμός προσαρμογής που φαίνεται να ενεργοποιήθηκε στα κύτταρα του πρωτοκόλλου 3Α είναι η διαδικασία της γλουταμινόλυσης καθώς μεταβολίτες που περιέχονται στις αντιδράσεις που έπονται αυτής εμφάνισαν αυξημένη συγκέντρωση. Ειδικότερα, οι υψηλές τιμές του φουμαρικού οξέος, του μηλικού οξέος και της ασπαραγίνης (μέσω του ασπαρτικού οξέος) δείχνουν την ανάγκη αναπλήρωσης των ενδιάμεσων του κύκλου του κιτρικού οξέος ώστε να λάβουν χώρα η οξειδωτική φωσφορυλίωση (μέσω της παραγωγής των NADH και FADH<sub>2</sub>) και η σύνθεση των πρωτεϊνών (Εικ. 78 αριστερά). Επιπλέον, η αύξηση του μηλικού οξέος, της αλανίνης και του γαλακτικού οξέος υποδεικνύει την ενίσχυση της αερόβιας γλυκόλυσης μέσω των αντιδράσεων που έπονται της γλουταμινόλυσης στο κυτταρόπλασμα (Εικ. 78 δεξιά και πάνω).



**Εικ. 78: Μεταβολικές αλλαγές του πρωτοκόλλου 3Α.** Φαίνονται με πράσινη και μπλε γραφή οι μεταβολίτες που αλλάζουν στις δυο διαφορετικές πορείες που έπονται της γλουταμινόλυσης.

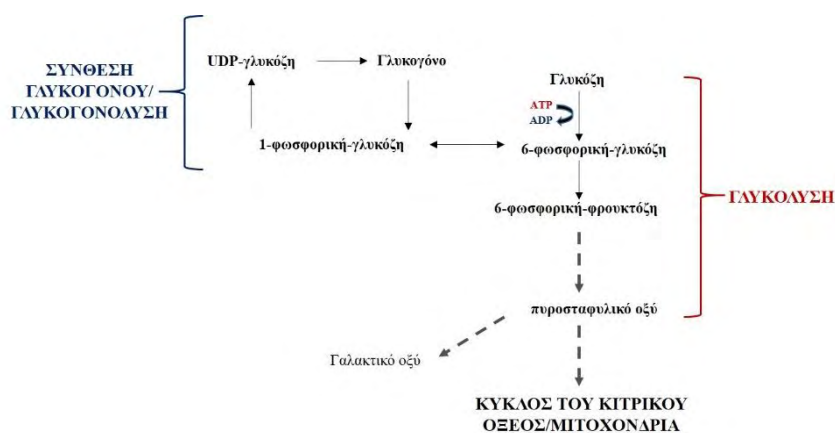
Η υποβολή των κυττάρων του πρωτοκόλλου 3Α σε συνθήκες στρες υποδεικνύεται και από την αύξηση του μεταβολίτη σερμιδίνης, η οποία οδηγεί στην επαγωγή του μηχανισμού της αυτοφαγίας μέσω της τροποποίησης των ιστονών και της επακόλουθης ρύθμισης της έκφρασης των γονιδίων που σχετίζονται με την αυτοφαγία (Eisenberg et

al. 2009). Η σπερμιδίνη ανήκει στην οικογένεια των πολυαμινών οι οποίες λειτουργούν ως μόρια-συνδέτες στο DNA, στο RNA ή στις πρωτεΐνες προκαλώντας τροποποιήσεις στη διαμόρφωσή τους και κατ' επέκταση επηρεάζοντας την έκφραση των γονιδίων και τις αλληλεπιδράσεις μεταξύ των παραπάνω μορίων (π.χ. μεταξύ DNA και πρωτεϊνών) (Murray-Stewart et al. 2016, Thomas & Thomas 2003). Αντίθετα, τα πολύ υψηλά επίπεδα γλυκόζης στο πρωτόκολλο 3B φαίνεται ότι ικανοποιούν τις μεταβολικές ανάγκες των καρκινικών κυττάρων καθώς ενισχύεται η διαδικασία της αερόβιας γλυκόλυσης (λόγω της αυξημένης συγκέντρωσης του γαλακτικού οξέος). Ακόμη, η αυξημένη συγκέντρωση του πυροσταφυλικού οξέος φαίνεται ότι εξυπηρετεί αφενός την αερόβια γλυκόλυση και αφετέρου την ενίσχυση των ενδοκυτταρικών του κύκλου του κιτρικού οξέος μετά την είσοδό του στα μιτοχόνδρια. Παράλληλα, μέσω του πυροσταφυλικού οξέος, φαίνεται να ενισχύονται και οι κυτταροπλασματικές βιοσυνθετικές διεργασίες από την έξοδο του κιτρικού οξέος στο κυτταρόπλασμα καθώς η γλουταμινόλυση και οι αντιδράσεις που έπονται αυτής δεν φαίνεται να ενεργοποιούνται λόγω της μικρής συγκέντρωσης των μεταβολιτών που συμμετέχουν σε αυτές. Επίσης, τα χαμηλά επίπεδα της σπερμιδίνης στο πρωτόκολλο 3B υποδεικνύουν ότι τα κύτταρα δεν επάγουν μηχανισμούς απόκρισης στο στρες διότι δεν χρειάζεται να αναπληρώσουν τα πρόδρομα μόρια των βιοσυνθετικών διεργασιών. Γίνεται κατανοητό λοιπόν ότι το πρωτόκολλο 3B τείνει να αντιπροσωπεύει αποτελεσματικότερα την φυσιολογία των καρκινικών κυττάρων χωρίς να επάγονται μηχανισμοί για την αναπλήρωση των απαραίτητων θρεπτικών συστατικών που αποτελούν βασική ανάγκη των ταχέως πολλαπλασιαζόμενων καρκινικών κυττάρων.

- Το μεταβολικό πρότυπο των κυττάρων που εκπλύθηκαν με PBS, υποδεικνύει τη χρήση του ως το καταλληλότερο διάλυμα έκπλυσης

Το PBS, ως διάλυμα έκπλυσης, ελέγχθηκε διότι χρησιμοποιείται επανειλημμένα στις καλλιέργειες προσκολλημένων κυττάρων για την έκπλυσή τους αλλά και σε αρκετές μελέτες μεταβολομικής ανάλυσης τέτοιου είδους καρκινικών σειρών (Karooro et al. 2015, Dettmer et al. 2011, Lorenz et al. 2011). Η εναλλακτική επιλογή του φυσιολογικού ορού (0,9% NaCl) έγινε διότι το διάλυμα αυτό είναι ισοτονικό, με αποτέλεσμα η σύστασή του σε ιόντα να μην προκαλεί ώσμωση στα κύτταρα και με αυτόν τον τρόπο να μειώνεται η πιθανότητα καταστροφής της ακεραιότητας της κυτταροπλασματικής μεμβράνης και η επακόλουθη διαρροή των ενδογενών μεταβολιτών. Η επιλογή του DMEM πραγματοποιήθηκε για να μειωθεί ο χρόνος έκθεσης των κυττάρων σε συνθήκες έλλειψης θρεπτικών συστατικών κατά τη διάρκεια της έκπλυσής τους. Η μεταβολομική ανάλυση των κυττάρων που εκπλύθηκαν με το DMEM δεν απέδωσε εμπλουτισμένο μεταβολικό πρότυπο σε σύγκριση με τα άλλα δυο. Ακόμη, το DMEM περιέχει πολλούς μεταβολίτες του κεντρικού μεταβολισμού του άνθρακα γεγονός που δυσχεραίνει την ακριβή μέτρηση των αντίστοιχων ενδογενών μεταβολιτών, διότι κατά την έκπλυση των κυττάρων αυτοί οι μεταβολίτες μπορεί να εισέρχονται στα κύτταρα. Αν και το μεταβολικό πρότυπο των κυττάρων που εκπλύθηκαν με τα άλλα δυο διαλύματα υποδεικνύει την ενεργοποίηση μηχανισμών απόκρισης σε συνθήκες έλλειψης γλυκόζης, όπως η γλουταμινόλυση (αύξηση μηλικού και φουμαρικού οξέος) και η διάσπαση της φρουκτόζης (μειωμένη συγκέντρωση) για την αναπλήρωση των ενδοκυτταρικών της γλυκόλυσης, το DMEM δε θεωρήθηκε σωστή

επιλογή διότι επηρεάζει τις μετρήσεις των ενδογενών μεταβολιτών και δεν αποδίδει μεγάλο αριθμό μεταβολιτών. Συγκρίνοντας τα μεταβολικά προφίλ των κυττάρων που εκπλύθηκαν με φυσιολογικό ορό και PBS, παρατηρήθηκε ότι τα πρώτα εμφάνισαν αύξηση των μεταβολιτών 6-φωσφορική γλυκόζη, 6-φωσφορική φρουκτόζη σε συνδυασμό με μειωμένα επίπεδα γλυκόζης. Όταν η γλυκόζη δεν είναι διαθέσιμη, αυτοί οι δυο μεταβολίτες είναι δυνατόν να προκύψουν μέσω της διαδικασίας της γλυκογονόλυσης. Το γλυκογόνο είναι ένα διακλαδισμένο πολυμερές γλυκόζης το οποίο λειτουργεί ως μόριο-αποθήκη ενέργειας και η διάσπασή του ενεργοποιείται μέσω μιας σειράς αντιδράσεων για την τελική παραγωγή της 6-φωσφορικής γλυκόζης (Labak et al. 2016) (Εικ. 79). Το μόριο αυτό εισέρχεται στη γλυκολιτική πορεία, γεγονός που εξηγεί και την αύξηση της 6-φωσφορικής φρουκτόζης.



**Εικ. 79: Διαδικασία σύνθεσης και διάσπασης του γλυκογόνου.** Η διάσπαση του γλυκογόνου συνεισφέρει στην τροφοδότηση των ενδιάμεσων μορίων της γλυκόλυσης (6-φωσφορικής γλυκόζης).

Οι μεταβολίτες (π.χ. 6-φωσφορική γλυκόζη, 6-φωσφορική φρουκτόζη) που περιέχουν στη δομή τους φωσφορικές ομάδες ήταν αναμενόμενο να είναι αυξημένοι στα κύτταρα που εκπλύθηκαν με PBS διότι κατά την έκπλυση των κυττάρων, το διάλυμα αυτό μπορεί να εισέρχεται στα κύτταρα και να επηρεάζει τις μετρήσεις των ενδογενών φωσφορικών μεταβολιτών. Παρ' όλα αυτά, το PBS θεωρήθηκε η αποτελεσματικότερη επιλογή διότι εμφανίζει εμπλουτισμένο μεταβολικό πρότυπο σε σύγκριση με το DMEM και επάγει λιγότερους μηχανισμούς απόκρισης σε συνθήκες έλλειψης γλυκόζης σε σχέση με τον φυσιολογικό ορό.

- Τα κύτταρα που προέρχονται από πρώιμο στάδιο ανακαλλιέργειας αποδίδουν μία αντιπροσωπευτική εικόνα της καρκινικής φυσιολογίας

Η σύγκριση του μεταβολικού προτύπου κυττάρων διαφορετικού σταδίου ανακαλλιέργειας πραγματοποιήθηκε για να διαπιστωθούν τυχόν επιδράσεις του αυξανόμενου αριθμού των θρυψινοποιήσεων στο μεταβολισμό των καρκινικών κυττάρων. Τέτοιου είδους σύγκριση δεν έχει πραγματοποιηθεί έως τώρα σε αντίστοιχες μεταβολομικές αναλύσεις πρωτοκόλλων επεξεργασίας και συλλογής προσκολλημένων καρκινικών κυττάρων (Dettmer et al. 2011, Lorenz et al. 2011, Karoore et al. 2015, Karoore et al. 2017, Khin et al. 2009). Τα χαρακτηριστικά του μεταβολικού προτύπου των κυττάρων της 2<sup>ης</sup> ανακαλλιέργειας που χρησιμοποιήθηκαν στο πρωτόκολλο 3B,



σε σύγκριση με αυτά της  $8^{15}$  του ίδιου πρωτοκόλλου, ήταν τα αυξημένα επίπεδα γλυκόζης και η ενίσχυση της αερόβιας γλυκόλυσης (αύξηση του γαλακτικού οξέος). Επίσης, τα χαμηλά επίπεδα του φουμαρικού οξέος, του μηλικού οξέος και της ασπαραγίνης υποδεικνύουν ότι δεν ενεργοποιείται η γλουταμινόλυση ώστε να αναπληρωθούν τα πρόδρομα μόρια (μηλικό οξύ) που οδηγούν στην παραγωγή του πυροσταφυλικού οξέος ώστε να καλυφθεί η ανάγκη των καρκινικών κυττάρων για αερόβια γλυκόλυση. Επιπλέον, δεν φαίνεται να ενεργοποιούνται επιπρόσθετοι μηχανισμοί απόκρισης στην έλλειψη γλυκόζης, όπως η γλυκογονόλυση, καθώς σε αυτά τα κύτταρα εμφανίζονται φυσιολογικά επίπεδα των ενδιάμεσων της γλυκόλυσης (π.χ. 6-φωσφορικής γλυκόζης). Ακόμη, μόνο το μεταβολικό πρότυπο των κυττάρων της  $2^{15}$  ανακαλλιέργειας φαίνεται να είναι εμπλουτισμένο με λιπαρά οξέα γεγονός που συνάδει με ένα από τα βασικά χαρακτηριστικά του καρκινικού μεταβολισμού, την αυξημένη λιπογένεση (Beloribi-Djefaflija et al. 2016). Γίνεται κατανοητό λοιπόν ότι η καταπόνηση των κυττάρων κατά τη διάρκεια των θρυψινοποιήσεων έχει αντίκτυπο στην καρκινική φυσιολογία γεγονός που υποδεικνύει ότι τα κύτταρα που χρησιμοποιούνται για μεταβολομικές προσεγγίσεις είναι δόκιμο να προέρχονται από πρώιμο στάδιο ανακαλλιέργειας.

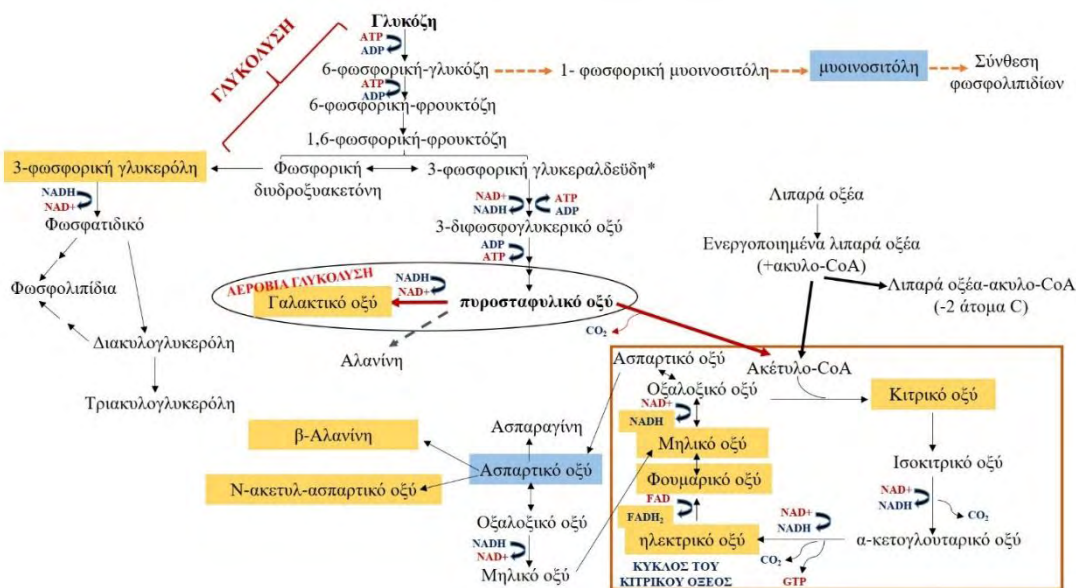
Λαμβάνοντας υπόψη τις συγκρίσεις των μεταβολικών προφίλ των διαφορετικών πρωτοκόλλων, διαλυμάτων έκπλυσης και σταδίων ανακαλλιέργειας, προτείνεται η επιλογή του πρωτοκόλλου 3B ως η βέλτιστη μέθοδος επεξεργασίας και συλλογής των κυττάρων *HeLa*. Η επιλογή της μεθανόλης ως μέσο αποκόλλησης προσκολλημένων κυττάρων έχει προταθεί και σε άλλη μελέτη που αφορά τη συγκεκριμένη κυτταρική σειρά με τη διαφορά ότι τα κύτταρα προορίζονται για υγρή χρωματογραφία σε συνδυασμό με φασματομετρία μάζας (LC/MS) (Khin et al. 2009). Όσον αφορά άλλες καρκινικές σειρές προσκολλημένων κυττάρων και πάλι προτείνεται η χρήση της μεθανόλης ως μέσο αποκόλλησης των κυττάρων είτε σε διαφορετικές συγκεντρώσεις είτε σε συνδυασμό με την προσθήκη οργανικού διαλύματος (π.χ. HEPES) το οποίο συνεισφέρει στην ελαχιστοποίηση της διαρροής των ενδογενών μεταβολιτών (Dettmer et al. 2011, Lorenz et al. 2011, Karoore et al. 2015, Karoore et al. 2017). Η επιλογή του PBS ως διάλυμα έκπλυσης επιβεβαιώνεται και σε αντίστοιχες πειραματικές μελέτες με τη διαφορά ότι ο αριθμός των εκπλύσεων μειώνεται σε 1 ή 2 ώστε να ελαχιστοποιηθεί ο χρόνος επεξεργασίας των κυττάρων (Karoore et al. 2015, Karoore et al. 2017). Ωστόσο, στο πρωτόκολλο 3B που προτείνεται στην διατριβή έχει πραγματοποιηθεί η ελαχιστοποίηση του χρονικού διαστήματος (1min) στο οποίο τα κύτταρα είναι εκτεθειμένα σε συνθήκες έλλειψης θρεπτικών συστατικών έως ότου γίνει η καταστολή της ενζυμικής δραστηριότητας. Επιπλέον, το πρωτόκολλο αυτό εμφανίζει την πιο αντιπροσωπευτική εικόνα των μεταβολικών αναγκών των καρκινικών κυττάρων επιφέροντας τις λιγότερες επιδράσεις στη φυσιολογία τους. Εφόσον επιλέχθηκε το πρωτόκολλο 3B για την επεξεργασία και συλλογή των κυττάρων *HeLa* έγινε και η κατάλληλη προσαρμογή αυτού για να χρησιμοποιηθεί στην ανάλυση του κυτταρικού μοντέλου με τροποποιημένη την έκφραση του γονιδίου *FRA10AC1* σε μεταγραφικό, πρωτεϊνικό και μεταβολικό επίπεδο. Τα διαλύματα και ο αριθμός των εκπλύσεων παρέμειναν ίδια για τα κύτταρα που προορίζονταν για τα τρία είδη ομικών αναλύσεων, ενώ το τελευταίο διάλυμα συλλογής των κυττάρων διαφοροποιήθηκε.

Συγκεκριμένα, στο τελευταίο βήμα χρησιμοποιήθηκε η μεθανόλη για την μεταβολομική ανάλυση, το διάλυμα λύσης του κιτ απομόνωσης του RNA για την μεταγραφωμική ανάλυση, και το διάλυμα αμμωνιακού άλατος TEAB-SDS για την ανάλυση του πρωτεϊνικού προτύπου (βλ. Ενότητα 3.4.1.2.).

Στις επόμενες ενότητες αναλύονται οι μεταβολικές διαδικασίες που επηρεάστηκαν μετά την καταστολή της έκφρασης του γονιδίου *FRA10AC1* καθώς και οι αλλαγές που ταυτοποιήθηκαν σε μεταγραφικό και πρωτεϊνικό επίπεδο. Ακολούθως αναλύονται οι υποθέσεις λειτουργίας της *FRA10AC1* που υποδεικνύονται από το συνδυασμό των αποτελεσμάτων των ομικών αναλύσεων και των πρωτεϊνικών αλληλεπιδράσεων της εντός και εκτός του δικτύου του σωματίου συναρμογής.

➤ *Μεταβολικές διαδικασίες που επηρεάζονται μετά την καταστολή της έκφρασης του γονιδίου FRA10AC1*

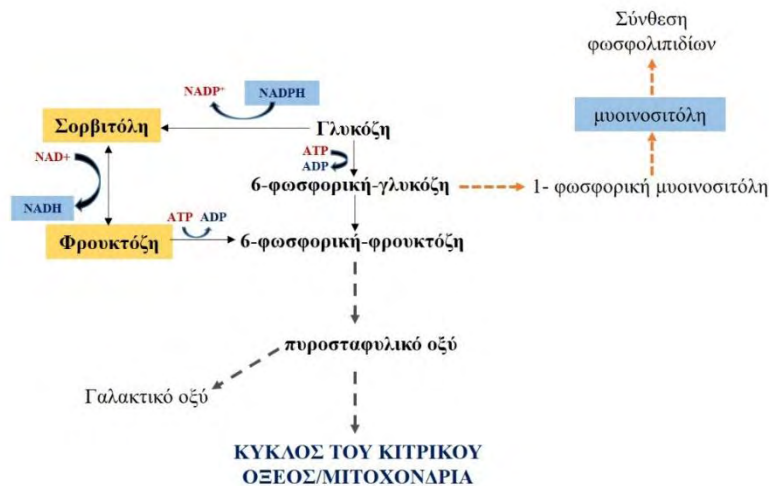
Οι μεταβολίτες που διαφοροποιήθηκαν στους κλώνους ολικής καταστολής shRNA(4)<sub>4</sub> και shRNA(3)<sub>5</sub> σε σχέση με τα κύτταρα *HeLa* και τον κλώνο αναφοράς (MSH 2) ενσωματώθηκαν στα μονοπάτια του κεντρικού μεταβολισμού του άνθρακα. Οι 4 από αυτούς που παρουσίασαν αυξημένη συγκέντρωση αποτελούν ενδιάμεσα μόρια του κύκλου του κιτρικού οξέος, όπως το ηλεκτρικό οξύ, το μηλικό οξύ, το φουμαρικό οξύ και το κιτρικό οξύ. Βάσει του καρκινικού μεταβολισμού των κυττάρων *HeLa*, η αυξημένη συγκέντρωση κιτρικού, ηλεκτρικού και φουμαρικού οξέος υποδεικνύει έναν αυξημένο ρυθμό β-οξειδωσης λιπιδίων η οποία οδηγεί στην παραγωγή ακετύλο-συνενζύμου A (ακετύλο-CoA) για την αναπλήρωση του κύκλου του κιτρικού οξέος (Εικ. 80) και επακόλουθη αυξημένη ενεργότητα των σχετικών διεργασιών των υπεροξειδιοσωμάτων (Wanders et al. 2016). Επιπλέον, η αύξηση των μεταβολιτών της αντίδρασης μετατροπής ηλεκτρικού σε φουμαρικό οξύ που οδηγεί στην απελευθέρωση του φλαβινο-αδενινο-δινουκλεοτιδίου (FADH<sub>2</sub>) και η αύξηση της παραγωγής του μηλικού οξέος που υποδηλώνει την απελευθέρωση του νικοτιναμιδο-αδενινο-δινουκλεοτιδίου (NADH) (Εικ. 80), δυο μόρια τα οποία λειτουργούν ως δότες ηλεκτρονίων, δείχνουν άμεση σύνδεση των αντιδράσεων αυτών με την οξειδωτική φωσφορυλίωση, την κυτταρική αναπνοή και τη μεταφορά ηλεκτρονίων, των οποίων η ενεργότητα φαίνεται να αυξάνεται.



**Εικ. 80: Μεταβολίτες με σημαντικά αυξημένη ή μειωμένη συγκέντρωση στους κλώνους υποέκφρασης του γονιδίου *FRA10AC1* και τα μεταβολικά μονοπάτια στα οποία ενσωματώθηκαν.** Οι μεταβολίτες που παρουσίασαν αυξημένη ή μειωμένη συγκέντρωση απεικονίζονται με κίτρινο ή μπλε πλαίσιο, αντίστοιχα.

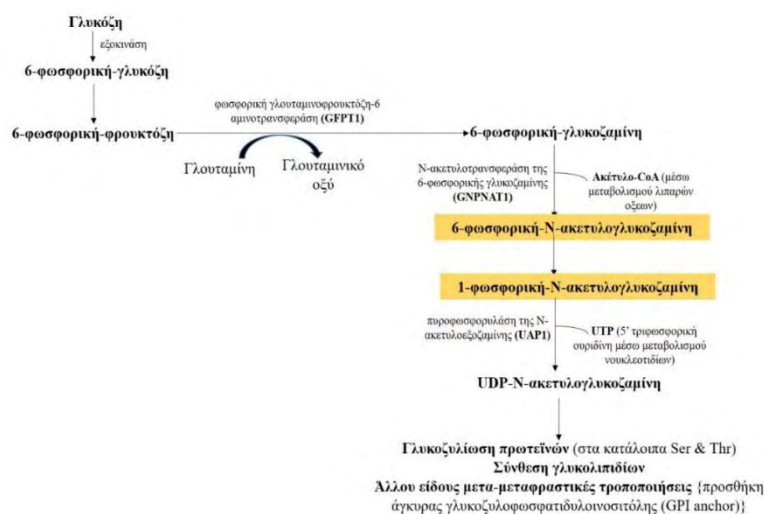
Επίσης, παρατηρείται αύξηση στους μεταβολίτες N-ακετυλ-γλουταμικό οξύ, N-ακετυλ-ασπαρτικό οξύ και β-αλανίνη που αποτελούν κύρια αντιοξειδωτικά μόρια και τα δυο τελευταία παράγονται από το ασπαρτικό οξύ, γεγονός που μπορεί να εξηγήσει τη μείωση της συγκέντρωσης του ασπαρτικού οξέος (Εικ. 80). Η αύξηση αυτών των μεταβολιτών μπορεί να συνδεθεί με την αυξημένη ενεργότητα του κύκλου του κιτρικού οξέος και της οξειδωτικής φωσφορυλίωσης λειτουργώντας ως προστατευτικός μηχανισμός έναντι των μορίων που απελευθερώνονται από τη συνεχή ενεργότητα των διεργασιών αυτών. Δυο άλλοι μεταβολίτες που αυξάνονται είναι η 3-φωσφορική γλυκερόλη (δεύτερη στη σειρά σημαντικότητας), η οποία συμμετέχει στο μεταβολισμό των γλυκερολιπιδίων (Εικ. 80), η βιταμίνη Β6 που αποτελεί συνένζυμο των αντιδράσεων του μεταβολισμού των αμινοξέων, της γλυκόζης και των λιπιδίων (Mooney et al. 2009) και το γαλακτικό οξύ που υποδεικνύει αύξηση της αερόβιας γλυκόλυσης, ένα χαρακτηριστικό του καρκινικού μεταβολισμού (Bayley & Devilee, 2012, Warburg 1956) (Εικ. 80).

Δυο από τους μεταβολίτες που εμφάνισαν την υψηλότερη αυξημένη συγκέντρωση ήταν η σορβιτόλη (πρώτη στη σειρά σημαντικότητας) και η φρουκτόζη (έκτη στη σειρά σημαντικότητας). Αυτοί συμμετέχουν στη μεταβολική οδό των πολυολών κατά την οποία παράγεται φρουκτόζη από γλυκόζη μέσω σορβιτόλης και ενεργοποιείται σε συνθήκες υπεργλυκαιμίας (Kinoshita 1990) (Εικ. 81). Η αύξηση της ενεργότητας του μονοπατιού των πολυολών συνδέεται με μείωση του ανηγμένου νικοτιναμιδο-αδενινω-δινουκλεοτιδίου-φωσφορικού άλατος (NADPH) και του οξειδωμένου νικοτιναμιδο-αδενινω-δινουκλεοτιδίου (NADH)  $NAD^+$  τα οποία χρησιμοποιούνται ως αντιδρώντα στις δυο αντιδράσεις του μονοπατιού (Εικ. 81). Η μείωση της συγκέντρωσης αυτών των συμπαραγόντων έρχεται σε συμφωνία με τη μείωση στη συγκέντρωση της μυοϊνοσιτόλης (Goldfarb et al. 1991).



**Εικ. 81:** Μεταβολίτες που αυξήθηκαν ή μειώθηκαν κατά την καταστολή της έκφρασης του γονιδίου *FRA10AC1* και το μονοπάτι των πολυολών στο οποίο ενσωματώθηκαν. Οι μεταβολίτες που παρουσίασαν αυξημένη συγκέντρωση απεικονίζονται με κίτρινο πλαίσιο, ενώ αυτοί με μειωμένη συγκέντρωση με μπλε πλαίσιο.

Τέλος, ένας μεταβολίτης που παρουσίασε σημαντικά αυξημένη συγκέντρωση και ταυτοποιήθηκε τρίτος στη σειρά σημαντικότητας είναι η N-ακετυλογλυκοζαμίνη η οποία εμπλέκεται στη N-γλυκοζυλίωση των πρωτεϊνών και είναι πιθανό προϊόν μεταβολισμού της γλυκόζης μέσω ενός μονοπατιού που «παρεκκλίνει» της γλυκόλυσης (Santoyo-Ramos et al. 2012) (Εικ. 82). Η πρώτη αντίδραση αυτού του μονοπατιού καταλύεται από το ένζυμο φωσφορική γλουταμινοφρουκτόζη-6 αμινοτρανσφεράση (GFPT1) και συνεχίζει προς την παραγωγή της UDP-N-ακετυλογλυκοζαμίνης, που συμμετέχει στη γλυκοζυλίωση των πρωτεϊνών, μέσω της αντίδρασης της N-ακετυλοτρανσφεράσης της 6-φωσφορικής γλυκοζαμίνης (GNPNAT1) και της αντίδρασης της πυροφωσφορυλάσης της N-ακετυλοεξοζαμίνης (UAP1) (Santoyo-Ramos 2012) (Εικ. 82).



**Εικ. 82:** Μεταβολίτες που αυξήθηκαν στους κλώνους υποέκφρασης του γονιδίου *FRA10AC1* και το μονοπάτι παραγωγής της N-ακετυλογλυκοζαμίνης που είναι υπεύθυνο για τη γλυκοζυλίωση των πρωτεϊνών. Οι μεταβολίτες που παρουσίασαν αυξημένη συγκέντρωση απεικονίζονται με κίτρινο πλαίσιο.

➤ *Επιδράσεις της καταστολής της έκφρασης του γονιδίου FRA10AC1 σε μετάγραφα και πρωτεΐνες που συμμετέχουν στην επεξεργασία του mRNA αλλά και σε άλλες βιολογικές διαδικασίες*

Κατά την καταστολή της έκφρασης της FRA10AC1 παρατηρήθηκε ότι, σε επίπεδο μεταγράφου, 627 γονίδια παρουσίασαν αύξηση της έκφρασής τους ενώ 920 εμφάνισαν μείωση.

Ένα αρκετά αυξημένο ποσοστό των μεταγράφων των γονιδίων που παρουσίασαν στατιστικά σημαντική αύξηση στην έκφρασή τους μετά την καταστολή της έκφρασης του FRA10AC1 συμμετέχουν στην επεξεργασία, συναρμογή και μεταβολισμό του RNA/mRNA, στην μεταγραφή και στη μετάφραση. Μια άλλη κατηγορία μεταγράφων γονιδίων που αυξήθηκε είναι αυτά που συμμετέχουν σε διάφορα σηματοδοτικά μονοπάτια (π.χ. μονοπάτι των διαμεμβρανικών υποδοχέων κινασών τυροσίνης, μονοπάτια των κυττάρων άμυνας, σηματοδότηση μέσω της φωσφατιδυλινοσιτόλης και μέσω υποδοχέων των στεροειδών ορμονών συμπεριλαμβανομένων των θυροειδών ορμονών, σηματοδότηση της ινσουλίνης). Επιπρόσθετα, στατιστικά αυξημένη έκφραση παρουσίασαν μετάγραφα γονιδίων που συμμετέχουν στην αντιγραφή και επιδιόρθωση του DNA, στον κυτταρικό κύκλο, στον κύκλο του τρικαρβοξυλικού οξέος, στην οξειδωτική φωσφορυλίωση, στο μεταβολισμό των λιπιδίων και φωφολιπιδίων και στη βιοσύνθεση του συμπαραγόντα NAD.

Όσον αφορά τα μετάγραφα γονιδίων που μειώθηκε σημαντικά η έκφρασή τους, αυτά συμμετέχουν σε διαδικασίες που σχετίζονται με τη δομή του κυτταροσκελετού, σε διεργασίες του λυσοσώματος, του ενδοπλασματικού δικτύου και του συστήματος Golgi, σε διάφορα σηματοδοτικά μονοπάτια (π.χ. σχετικά με τον κυτταρικό πολλαπλασιασμό, την κυτταρική διαφοροποίηση, την απόπτωση), στην αποικοδόμηση πρωτεϊνών μέσω πρωτεασώματος, στην N-γλυκοζυλίωση πρωτεϊνών και σε άλλα είδη μετα-μεταφραστικών τροποποιήσεων, στη μεταφορά πρωτεϊνών στην κυτταροπλασματική μεμβράνη και στη ρύθμιση των μορίων προσκόλλησης (adhesion molecules) στο κύτταρο, στην πήξη του αίματος και στην ανάπτυξη των αιμοφόρων αγγείων.

Οι πρωτεΐνες που εμφάνισαν αυξημένα επίπεδα έκφρασης ήταν 190, ενώ αυτές που παρουσίασαν μειωμένα επίπεδα ήταν 370. Οι πρωτεΐνες που εμφάνισαν σημαντικά αυξημένη έκφραση μετά την καταστολή της έκφρασης του γονιδίου FRA10AC1 συμμετέχουν στην επεξεργασία των μη κωδικών RNA (tRNA, rRNA), στις διαδικασίες που λαμβάνουν χώρα στο μιτοχόνδριο, στη βιοσύνθεση και στο μεταβολισμό των αμινοξέων, των υδατοδιαλυτών βιταμινών, των πυριμιδινών, της νικοτιναμίδης και του νικοτιναμίδο-αδενινο-δινουκλεοτιδίου (NAD), στη σύνδεση σε νουκλεοτίδια και στο μόριο της τριφωσφορικής αδενοσίνης (ATP), και στη ρύθμιση της απόπτωσης. Οι πρωτεΐνες που παρουσίασαν σημαντικά μειωμένη έκφραση συνεισφέρουν στην οργάνωση του κυτταροσκελετού, στην αρνητική ρύθμιση του πολυμερισμού, του αποπολυμερισμού και του μήκους των ινιδίων της ακτίνης, στη θετική ρύθμιση της μεταφοράς της γλυκόζης, στο μεταβολισμό του φωσφόρου και του φωσφορικού οξέος, στη φωσφορυλίωση των πρωτεϊνών, στην απόπτωση και τη ρύθμισή της, στη ρύθμιση του καταρράκτη αντιδράσεων των κινασών σερίνης/θρεονίνης, στην προσκόλληση

μεταξύ των κυττάρων και στις διεργασίες των δεσμοσωμάτων που συμμετέχουν σε αυτή τη διαδικασία, στη σηματοδότηση (π.χ. συναρμολόγηση σηματοδοτικών συμπλόκων, σηματοδοτικό μονοπάτι μέσω ιντεγκρινών), και στη μεταφορά των πρωτεϊνών μέσω κυτταροπλασματικών κυστιδίων και συγκεκριμένα μεταφορά μέσω κυστιδίων της συσκευής Golgi.

Από το σύνολο των μορίων που εμφάνισαν αυξημένη ή μειωμένη έκφραση, είτε σε επίπεδο μεταγράφου ή πρωτεΐνης, μετά την καταστολή της έκφρασης της FRA10AC1, τα 56 κωδικοποιούν συστατικά του σωματίου συναρμογής (Πιν. 1 Παράρτημα), 39 εκ των οποίων ενσωματώθηκαν στο δίκτυο πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής. Σε επίπεδο μεταγράφου, 29 συστατικά του σωματίου εμφάνισαν αλλαγές στην έκφρασή τους και τα 23 από αυτά ενσωματώθηκαν στο δίκτυο. Σε επίπεδο πρωτεΐνης, 19 μόρια, που παρουσίασαν τροποποιημένη έκφραση, αποτελούσαν συστατικά του σωματίου συναρμογής (Πιν. 1 Παράρτημα), 13 εκ των οποίων ενσωματώθηκαν στο δίκτυο. Υπήρχαν και κοινά μόρια που εμφάνισαν αυξημένη ή μειωμένη έκφραση και σε επίπεδο μεταγράφου και πρωτεΐνης (βλ. επόμενη ενότητα). Οι 39 πρωτεΐνες του σωματίου συναρμογής τοποθετούνται στο δίκτυο πρωτεϊνικών αλληλεπιδράσεων του σωματίου και συνδέονται με την FRA10AC1, μέσω της μεσολάβησης μιας ή περισσότερων αλληλεπιδράσεων (βλ. Ενότητα 3.5.). Η πλειοψηφία των 39 πρωτεϊνών κατηγοριοποιήθηκαν, σύμφωνα με το βιολογικό τους ρόλο, σε τέσσερις λειτουργικές ομάδες: α) στην συναρμογή, και συγκεκριμένα στα καταλυτικά βήματα της διαδικασίας (4 μόρια), β) στην εναλλακτική συναρμογή (6 μόρια), γ) στη μεταγραφή (8 μόρια), και δ) στην αναδιαμόρφωση της χρωματίνης (6 μόρια). Τρία μόρια (DGCR14, PQBP1, DDX17) εκ των 39, συμμετέχουν σε περισσότερες από μία εκ των τεσσάρων λειτουργικών ομάδων και εμφάνισαν αυξημένη έκφραση, απουσία της FRA10AC1. Ειδικότερα, η DGCR14 έχει διττό ρόλο στην μεταγραφή και στη συναρμογή (βλ. Ενότητα 1.3.1.). Η PQBP1 έχει προταθεί ότι συνεισφέρει στη σύνδεση του συμπλόκου PRP19/CDC5L με το τριμερές σύμπλοκο U4/U6.U5 και συγκεκριμένα με πρωτεΐνες του U5 snRNP (Mizuguchi et al. 2014). Επιπρόσθετα παίζει ρόλο και στην εναλλακτική συναρμογή γονιδίων και στη μεταγραφή. Έχει δειχθεί ότι επηρεάζει τη στρατολόγηση του U2 snRNP στις εναλλακτικές θέσεις συναρμογής μέσω της αλληλεπίδρασής της με την πρωτεΐνη SF3B1 η οποία συνδέεται με το mRNA (Wang et al. 2013), και επιπλέον, συνδέεται με το καρβοξυτελικό άκρο της RNA πολυμεράσης II όταν βρίσκεται στη φωσφορυλιωμένη μορφή προκαλώντας την καταστολή της μεταγραφής κάποιων γονιδίων (Okazawa et al. 2002). Η DDX17 συμμετέχει στην εναλλακτική συναρμογή έπειτα από την αλληλεπίδρασή της με τις πρωτεΐνες hnRNP H και hnRNP F τις οποίες βοηθά να στρατολογηθούν στο mRNA, στις θέσεις των εναλλακτικών εξονίων (Dardenne et al. 2014), και στην ρύθμιση της μεταγραφής, μέσω της σύνδεσής της με μεταγραφικούς παράγοντες, ενεργοποιώντας τη μεταγραφή γονιδίων (Germann et al. 2012).

Στην πρώτη λειτουργική ομάδα, αυτή της συναρμογής, 2 μόρια (SNRNP27, GPKOW) παρουσίασαν μειωμένη έκφραση, απουσία της FRA10AC1, και τα άλλα 2 (CDC5L, WBP11) αυξημένη. Η SNRNP27, που αποτελεί δεύτερο «γείτονά» της FRA10AC1, είναι πρωτεΐνη SR η οποία συνδέεται με άλλες πρωτεΐνες αυτής της

κατηγορίας και έχει προταθεί ότι συνεισφέρει στη στρατολόγηση του τριμερούς συμπλόκου U4/U6.U5 στα συστατικά του σωματίου συναρμογής που έχουν ήδη τοποθετηθεί στο mRNA, αντικαθιστώντας την SNRNP70 η οποία βοηθά στη σύνδεση του U1 snRNP στην 5' θέση συναρμογής μέσω αλληλεπιδράσεων με πρωτεΐνες SR στην αρχή της συναρμογής (Fetzer et al. 1997). Η GPKOW συνδέεται με το RNA και έχει προταθεί ότι συμμετέχει στη συναρμογή πριν το πρώτο καταλυτικό βήμα πιθανώς μέσω της αλληλεπίδρασης της με την ελικάση DHX16 χωρίς να είναι γνωστός ο μοριακός μηχανισμός (Aksaas et al. 2011, Zang et al. 2014). Η CDC5L είναι συστατικό του συμπλόκου PRP19/CDC5L το οποίο σταθεροποιεί τη σύνδεση των U5snRNP και U6snRNP με το ήδη συναρμολογημένο σωματίο συναρμογής κατά τη διάρκεια των καταλυτικών βημάτων της συναρμογής (Grote et al. 2010), χωρίς να είναι γνωστός ο ακριβής ρόλος της (Chan et al. 2003, Chanarat & Sträber 2013, Grote et al. 2010). Η WBP11 έχει προταθεί ότι συνεισφέρει, σε συνεργασία με την PQBP1, στη σύνδεση του συμπλόκου PRP19/CDC5L με το τριμερές σύμπλοκο U4/U6.U5 και συγκεκριμένα με πρωτεΐνες του U5 snRNP, καθώς και οι δυο αποτελούν συστατικά του U4/U6.U5 tri-snRNP. Η WBP11 αλληλεπιδρά με την CD2BP2 (U5-52K) και η PQBP1 με την TXNL4A (U5-15K). Η σύνδεση της WBP11 με την PQBP1 έχει δείχθει ότι μειώνει την ικανότητα της τελευταίας να αλληλεπιδράσει με την U5-15K, γεγονός που πιθανώς υποδεικνύει ότι υψηλές συγκεντρώσεις της WBP11 θα μπορούσαν να τροποποιούν τη δράση του σωματίου συναρμογής με αρνητικό τρόπο, δηλαδή να προκαλούν την αποσυναρμολόγηση του PRP19/CDC5L από το U5 snRNP (Mizuguchi et al. 2014).

Στην δεύτερη λειτουργική ομάδα, αυτή της εναλλακτικής συναρμογής, 5 μόρια (RBMX, hnRNP U, PCBP2, SRPK1, PUF60) εμφάνισαν αυξημένα επίπεδα έκφρασης, μετά την καταστολή της FRA10AC1, και ένα (SRSF4) παρουσίασε μειωμένα επίπεδα. Η RBMX συνδέεται σε ένα μοτίβο αλληλουχίας του mRNA (μοτίβο CC(A/C)) που βρίσκεται σε κοντινή απόσταση από τα εναλλακτικά εξόνια συνεισφέροντας στην εναλλακτική συναρμογή γονιδίων (Heinrich et al. 2009). Η hnRNP U ρυθμίζει την εναλλακτική συναρμογή γονιδίων μέσω της σύνδεσής της με παράγοντες συναρμογής που αναγνωρίζουν τις 5' και 3' θέσεις συναρμογής (Xiao et al. 2012). Η SRSF4 είναι μια πρωτεΐνη SR που συμμετέχει, σε συνεργασία με την RBMX (hnRNP G) και άλλες 5 πρωτεΐνες SR και hnRNP, στην εναλλακτική συναρμογή γονιδίων μέσω της σύνδεσής της με αλληλουχίες των ιντρονίων του mRNA (intronic splicing silencer) (Wang et al. 2011). Η PCBP2 (hnRNP E2) συμμετέχει στην εναλλακτική συναρμογή γονιδίων μέσω της σύνδεσής της με αλληλουχίες των ιντρονίων του mRNA (intronic splicing enhancer) και πιθανώς «ανταγωνίζεται» τη δράση της πρωτεΐνης SRSF4 (Broderick et al. 2004, Wang et al. 2011). Η SRPK1 είναι μία κινάση που φωσφορυλιώνει τις πρωτεΐνες SR στα αμινοξικά κατάλοιπα αργινίνη και σερίνη, συνεισφέροντας στην αναγνώριση των 5' και 3' εναλλακτικών ή μη θέσεων συναρμογής (Aubol et al. 2016, Aubol et al. 2013, Jamros et al. 2015, Zhong et al. 2009). Η PUF60, σε συνεργασία με την U2AF1 και U2AF2, συμβάλλει στην αναγνώριση της 3' θέσης συναρμογής (Hastings et al. 2007) και συνεισφέρει στην εναλλακτική συναρμογή γονιδίων έπειτα από τη σύνδεσή της με στοιχεία της αλληλουχίας του mRNA (intronic splicing enhancer) η οποία έχει ως αποτέλεσμα να περιλαμβάνονται κάποια εξόνια (exon inclusion) (Poleev et al. 2000).

Στην τρίτη λειτουργική ομάδα, αυτή της μεταγραφής, 8 μόρια (EIF3L, eEF1A1, ILF3, DIDO1, SERBP1, PIK3R3, VPRBP, EHMT2) εμφάνισαν αυξημένα επίπεδα έκφρασης, απουσία της FRA10AC1, ενώ ένα (JUP) παρουσίασε μειωμένη έκφραση. Η EIF3L αποτελεί συστατικό του συμπλόκου eIF3 που είναι υπεύθυνο για την έναρξη της μετάφρασης (Lee et al. 2015, Masutani et al. 2007) αλλά έχει δειχθεί ότι αλληλεπιδρά (πιθανώς κάποια πρωτεϊνική ισομορφή) με μεταγραφικούς παράγοντες οι οποίοι συνδέονται με τους υποκινητές των γονιδίων που κωδικοποιούν τα ριβοσωμικά RNA συνεισφέροντας στην έναρξη της μεταγραφής τους (Seither et al. 2001, Yuan et al. 2002). Στο ίδιο πλαίσιο, η eEF1A1 που αποτελεί ισομορφή του παράγοντα επιμήκυνσης της μετάφρασης eEF1A έχει δειχθεί ότι συμμετέχει στη μεταγραφή γονιδίων σε απόκριση στο θερμικό σοκ (Vera et al. 2014). Η SERBP1 και η ILF3 είναι συστατικά ενός συμπλόκου που συμμετέχει στη ρύθμιση της μεταγραφής γονιδίων (Mari et al. 2015). Επίσης αυτές οι δυο πρωτεΐνες συνεισφέρουν στη διατήρηση της σταθερότητας των mRNA, σε απόκριση σε διάφορα ερεθίσματα, μέσω της σύνδεσής τους στην 3' μη μεταφραζόμενη περιοχή εμποδίζοντας την αποικοδόμηση του mRNA (Heaton et al. 2001, Masuda et al. 2013, Song et al. 2017, Vumbaca et al. 2008). Η DIDO1 είναι μεταγραφικός παράγοντας που πιθανώς ρυθμίζει την έκφραση των γονιδίων που σχετίζονται με τους αποπτωτικούς μηχανισμούς (Gomes et al. 2002). Η PIK3R3 ενεργοποιεί το σηματοδοτικό μονοπάτι NF-κB (Qingqing Wang et al. 2013) και συγκεκριμένα φωσφορυλιώνει το μεταγραφικό παράγοντα NF-κB, η φωσφορυλίωση του οποίου ρυθμίζει τη μεταγραφή γονιδίων (Ye et al. 2017). Η VPRBP συμμετέχει στη ρύθμιση της μεταγραφής γονιδίων (Kim et al. 2013, Wang et al. 2017) και η JUP ενεργοποιεί τη μεταγραφή γονιδίων στο σηματοδοτικό μονοπάτι Wnt (Maeda et al. 2004). Η EHMT2 είναι μια μεθυλο-τρανσφεράση που μεθυλιώνει την ιστόνη H3 στο κατάλοιπο 9 συνεισφέροντας στη μεταγραφική καταστολή γονιδίων (Tachibana et al. 2002). Αν και κατά κύριο λόγο λειτουργεί ως συν-καταστολέας της μεταγραφής, έχει δειχθεί ότι μπορεί να λειτουργήσει και ως συν-ενεργοποιητής της μεταγραφής συγκεκριμένων κατηγοριών γονιδίων (Ενδεικτικά: Bittencourt et al. 2012, Laumet et al. 2015, Lu et al. 2013, Mozzetta et al. 2015).

Στην τέταρτη λειτουργική ομάδα, αυτή της αναδιαμόρφωσης της χρωματίνης, 5 μόρια (RBBP4, CHD4, SMARCC1, SMARCB1, SMARCE1) εμφάνισαν αυξημένη έκφραση και η RBBP7 παρουσίασε μειωμένη. Τρεις πρωτεΐνες (RBBP4, RBBP7, CHD4) αποτελούν συστατικά του συμπλόκου NuRD (NUcleosome Remodelling and Deacetylase, Denslow & Wade 2007), και τρεις (SMARCC1, SMARCB1, SMARCE1) είναι συστατικά του συμπλόκου SWI/SNF, το οποίο συνεισφέρει κατά κύριο λόγο στην ενεργοποίηση της μεταγραφής (Euskirchen et al. 2011, Tang et al. 2010). Ειδικότερα, οι RBBP4 και RBBP7 συνδέονται με τις ουρές των ιστονών H3 και H4 συμβάλλοντας στην αναδιαμόρφωση της χρωματίνης και στην συναρμολόγηση και αποσυναρμολόγηση των νουκλεοσωμάτων. Επιπλέον συνδέονται και με μεταγραφικούς παράγοντες και συνεισφέρουν, με αυτόν τον τρόπο, στη στρατολόγηση του συμπλόκου NuRD στους υποκινητές των γονιδίων. Η CHD4 συνδέεται με την ιστόνη H3, έχει την ιδιότητα ελικάσης και συμμετέχει στην αναδιαμόρφωση της χρωματίνης (Allen et al. 2013, Millard et al. 2016). Η SMARCB1, η οποία αποτελεί και δεύτερο «γείτονα» της FRA10AC1, και η SMARCC1 συνεισφέρουν στη



διατήρηση της ακεραιότητας του συμπλόκου SWI/SNF, δηλαδή στην προστασία του από την αποικοδόμηση (Sohn et al. 2007, Wang et al. 2017), και η SMARCB1 έχει δειχθεί ότι συνεισφέρει στην ενεργοποίηση της μεταγραφής γονιδίων (Wang et al. 2017, Nakayama et al. 2017) και συγκεκριμένα αυτών που εκφράζονται στους μυς (Joliot et al. 2014). Η SMARCE1 συμμετέχει στη ρύθμιση της μεταγραφής των γονιδίων που σχετίζονται με τον κυτταρικό κύκλο (Hah et al. 2010) και στην καταστολή της έκφρασης γονιδίων που συμμετέχουν στη διαφοροποίηση των T-κυττάρων (Lomeli & Castillo-Robles 2016).

Διερευνώντας τα μόρια που παρουσίασαν τροποποιημένη έκφραση, απουσία της FRA10AC1, στην ευρύτερη «γειτονιά» της εκτός του σωματίου συναρμογής, παρατηρήθηκε ότι η πλειοψηφία αυτών συμμετέχουν σε κοινές βιολογικές διεργασίες (π.χ. μεταγραφή, σηματοδότηση). Δύο πρώτοι «γειτόνες» της, η EEF1D, που παρουσίασε αυξημένη έκφραση, είναι παράγοντας επιμήκυνσης της μετάφρασης αλλά μια πρωτεϊνική της ισομορφή (eEF1BδL) μεταναστεύει στον πυρήνα και λειτουργεί ως μεταγραφικός παράγοντας γονιδίων τα οποία διαθέτουν στοιχεία θερμικού σοκ (heat-shock elements) (Kaitsuka et al. 2011, Kaitsuka & Matsushita 2015). Και η HABP4, η οποία εμφάνισε μειωμένη έκφραση, ρυθμίζει την εναλλακτική συναρμογή γονιδίων, όπως έχει δειχθεί *in vitro* (Bressan et al. 2009), αλλά επιπρόσθετα έχει δειχθεί ότι αλληλεπιδρά με πρωτεΐνες που έχουν λειτουργικό ρόλο στη ρύθμιση της μεταγραφής (Lemos et al. 2003). Εξετάζοντας τις αλλαγές των δευτέρων «γειτόνων» της εκτός του σωματίου συναρμογής παρατηρήθηκαν αλλαγές στην έκφρασή της PRKCA που αποτελεί πρώτο «γείτονα» των EEF1D και HABP4 και της DYNC1H1 που αλληλεπιδρά μόνο με την EEF1D. Ειδικότερα, η PRKCA ανήκει στην οικογένεια των κινασών σερίνης/θρεονίνης και ενεργοποιείται σε απόκριση διαφόρων ερεθισμάτων, συμπεριλαμβανομένων σημάτων που συνδέονται σε υποδοχείς πρωτεϊνών G, κινασών τυροσίνης, αυξητικών παραγόντων και μορίων που σχετίζονται με την προσκόλληση κυττάρων (Konopatskaya & Poole, 2010, Nakashima 2002). Η ενεργοποίηση της κίνησης αυτής οδηγεί στην φωσφορυλίωση διαφόρων υποστρωμάτων, για παράδειγμα μεταγραφικών παραγόντων ώστε να ρυθμιστεί η μεταγραφή γονιδίων, μορίων που σχετίζονται με τη ρύθμιση του κυτταρικού κύκλου ελέγχοντας τον κυτταρικό πολλαπλασιασμό, και μορίων που συμμετέχουν στην προσκόλληση των κυττάρων ή προωθούν το σχηματισμό ελασματοποδίων (lamellipodia) τα οποία λαμβάνουν μέρος στην κυτταρική μετανάστευση (Gould & Newton 2008, Sun & Rotenberg 1999). Η DYNC1H1 αποτελεί την βαριά αλυσίδα της δυνεΐνης, ενός συμπλόκου το οποίο αποτελείται από πολλές υπομονάδες και συμβάλλει στη μεταφορά κυτταρικών συστατικών και δομών κατά μήκος των μικροσωληνίσκων (Schiavo et al. 2013). Ειδικότερα, αυτές οι δομές μπορεί να είναι πρωτεΐνες ή πρωτεϊνικά σύμπλοκα, συσσωματώματα πρωτεϊνών, mRNAs, αλλά ακόμα και μεγάλων δομών όπως του πυρήνα, των κεντροσωμάτων, των χρωμοσωμάτων και της μιτωτικής ατράκτου καθώς η δυνεΐνη συνεισφέρει στην οργάνωση του κυτταροσκελετού κατά τη διάρκεια του κυτταρικού κύκλου (Allan 2011, Raaijmakers & Medema 2014, Vale 2003). Στο νευρικό σύστημα η DYNC1H1 παρουσιάζει ιστοειδικό ρόλο διότι μεταφέρει κυτταρικά συστατικά και δομές κατά μήκος των μικροσωληνίσκων των δενδριτών, και επιπλέον, αποτελεί το μοναδικό σύστημα μεταφοράς προς τα αρνητικά άκρα των

μικροσωληνίσκων (minus ends) του άξονα των νευρικών κυττάρων. Συγκεκριμένα μεταφέρει σηματοδοτικά σύμπλοκα που συνεισφέρουν στην μεταγραφή των γονιδίων και στην ανάπτυξη, πρωτεΐνες που δεν είναι σωστά αναδιπλωμένες και οργανίδια από τις συνάψεις στο σώμα του νευρικού κυττάρου (Schiavo et al. 2013). Οι υπόλοιποι δεύτεροι «γείτονες» της FRA10AC1 εκτός του σωματίου συναρμογής που εμφάνισαν τροποποιημένα επίπεδα έκφρασης ήταν 21. Η ανάλυση γονιδιακής οντολογίας των 21 δεύτερων «γειτόνων» της FRA10AC1 εκτός του σωματίου συναρμογής (η ανάλυση δεν παρουσιάζεται στα αποτελέσματα) έδειξε ότι τρεις συμμετέχουν στην διαδικασία της μεταγραφής, επτά σε σηματοδοτικά μονοπάτια (π.χ. σηματοδοτικό μονοπάτι Wnt, Ras), και τρεις στην οργάνωση του κυτταροσκελετού και στην ρύθμιση του κυτταρικού κύκλου. Επιπλέον, ανάλυση γονιδιακής οντολογίας εφαρμόστηκε (δεν παρουσιάζεται στα αποτελέσματα) και στις πρωτεΐνες που είχαν απομονωθεί στα ίδια 7 πρωτεϊνικά σύμπλοκα με την FRA10AC1 μέσω της χρήσης πειραμάτων χρωματογραφίας συγγένειας (Huttlin et al. 2017) (βλ. Ενότητα 3.1.3.). Παρατηρήθηκε και πάλι ότι οι πρωτεΐνες που εμφάνισαν αλλαγή στην έκφρασή τους (10) κατά την καταστολή της έκφρασης του γονιδίου FRA10AC1 συμμετέχουν στη ρύθμιση της μεταγραφής (1), σε ποικίλα σηματοδοτικά μονοπάτια (4), στην οργάνωση του κυτταροσκελετού και στη ρύθμιση του κυτταρικού κύκλου (2).

➤ *Κοινές βιολογικές διαδικασίες μεταξύ μεταγραφικού και πρωτεϊνικού προτύπου που επηρεάζονται μετά την καταστολή της έκφρασης του FRA10AC1*

Τα μόρια τα οποία εμφάνισαν αύξηση ή μείωση της έκφρασής τους, μετά την καταστολή της έκφρασης του FRA10AC1, τόσο σε επίπεδο μεταγράφου όσο και σε επίπεδο πρωτεΐνης ήταν 82 και 111, αντίστοιχα, αριθμοί που αναλογούν σε ποσοστό 43% και 30% του συνολικού αριθμού των πρωτεϊνών. Από αυτά, τα 8 αποτελούν συστατικά του σωματίου συναρμογής και κατά πλειοψηφία (7) παρουσίασαν αυξημένα επίπεδα έκφρασης ενώ τα 5 ενσωματώθηκαν στο δίκτυο του σωματίου (PUF60, BAG2, GEMIN5, SMARCC1, DHX30). Τα παραπάνω ποσοστά είναι ικανοποιητικά υποδεικνύοντας αυξημένη αξιοπιστία στις αλλαγές που παρατηρούνται σε σημαντικό αριθμό μορίων. Η διερεύνηση των οντολογικών όρων (GO terms) έδειξε ότι η πλειοψηφία των κοινών μορίων συμμετέχουν σε κοινές βιολογικές διαδικασίες. Ειδικότερα, σύμφωνα με την ανάλυση γονιδιακής οντολογίας των 82 κοινών μορίων που παρουσίασαν αυξημένη έκφραση και στα δυο επίπεδα μετά την καταστολή της έκφρασης του FRA10AC1, παρατηρήθηκε ότι 8 μόρια συμμετέχουν στη μεταγραφή, 4 στην επεξεργασία είτε του mRNA είτε των μη κωδικών RNA (tRNA, rRNA), 7 στις διαφορετικές φάσεις και στη ρύθμιση του κυτταρικού κύκλου, 4 στην οργάνωση του κυτταροσκελετού, 7 σε διαφορετικά σηματοδοτικά μονοπάτια, 12 στη μεταφορά μορίων (π.χ. ιόντα, πρωτεΐνες, βιταμίνες), 5 σε διάφορα αναπτυξιακά στάδια, 3 στη ρύθμιση της απόπτωσης, 4 στο μεταβολισμό λιπιδίων και φωσφολιπιδίων, 6 στην οξειδωτική φωσφορυλίωση και 16 στο μεταβολισμό ενώσεων που περιέχουν άζωτο (π.χ. NAD). Οι οντολογικοί όροι που εμφάνισαν τον μεγαλύτερο βαθμό εμπλουτισμού (enrichment score) ήταν η οξειδωτική φωσφορυλίωση και ο μεταβολισμός ενώσεων που περιέχουν άζωτο και ακολούθησαν οι υπόλοιποι. Αυτό μπορεί να εξηγηθεί από τον συνολικό αριθμό των γονιδίων που περιέχει κάθε οντολογικός όρος, δηλαδή, για

παράδειγμα, το GO term της οξειδωτικής φωσφορυλίωσης περιέχει συνολικά 129 γονίδια, με αποτέλεσμα τα 6 γονίδια που ενσωματώνονται σε αυτόν, αν και μικρός αριθμός, να αυξάνουν τον βαθμό εμπλουτισμού αυτού του οντολογικού όρου. Με βάση την ανάλυση γονιδιακής οντολογίας των 111 κοινών μορίων που παρουσίασαν μειωμένη έκφραση μετά την καταστολή της έκφρασης του *FRA10AC1*, παρατηρήθηκε ότι 15 συμμετέχουν σε σηματοδοτικά μονοπάτια (π.χ. NF-κB), 5 στη ρύθμιση μορίων-υποστρωμάτων προσκόλλησης (adhesion molecules) των κυττάρων, 7 στις διαφορετικές φάσεις του κυτταρικού κύκλου, 8 στη μεταφορά μορίων (π.χ. λιπιδίων, πρωτεϊνών), 15 σε διαφορετικά αναπτυξιακά στάδια, 18 στο μεταβολισμό ενώσεων που έχουν άζωτο και φώσφορο, 6 στο μεταβολισμό πουρινών και πυριμιδινών και 3 στο μεταβολισμό λιπιδίων.

Ωστόσο, υπάρχουν αρκετά μόρια που δεν παρουσιάζουν ταυτόσημες αλλαγές στην έκφραση και στα δύο επίπεδα κυτταρικής λειτουργίας, γεγονός που θα μπορούσε να εξηγηθεί με διάφορους τρόπους (Greenbaum et al. 2002, Maier et al. 2009). Ειδικότερα, η ύπαρξη μορίων που παρουσίασαν αυξημένη έκφραση σε επίπεδο μεταγράφου και όχι σε επίπεδο πρωτεΐνης μπορεί να οφείλεται: α) στο σχηματισμό δευτεροταγών δομών μεταξύ των mRNA που υπερεκφράζονται, οι οποίες πιθανώς μπορεί να οδηγήσουν στην αποικοδόμηση των μεταγράφων πριν τη μετάφραση, β) στην ύπαρξη ορισμένων ρυθμιστικών RNA (π.χ. miRNAs) τα οποία επηρεάζουν τη σταθερότητα του mRNA και κατ' επέκταση μπορεί να εμποδίσουν την μετάφρασή του, εξισορροπώντας την υπερέκφραση των μεταγράφων σε επίπεδο πρωτεΐνης και γ) στον μειωμένο χρόνο ημιζωής μιας πρωτεΐνης (protein half-life), δηλαδή, η υπερέκφραση ενός μεταγράφου μπορεί να εξισορροπείται από την άμεση ενεργοποίηση των μηχανισμών αποικοδόμησης της αντίστοιχης πρωτεΐνης. Η μειωμένη έκφραση των μεταγράφων και όχι των αντίστοιχων πρωτεϊνών μπορεί πιθανώς να εξηγηθεί από την αυξημένη αποικοδόμηση του mRNA και την ταυτόχρονη ενεργοποίηση ενός μηχανισμού που αυξάνει την απόδοση της μετάφρασης εξισορροπώντας την υποέκφραση των μεταγράφων σε επίπεδο πρωτεΐνης.

Αντίθετα, η εμφάνιση αυξημένης έκφρασης μορίων σε επίπεδο πρωτεΐνης και όχι σε επίπεδο μεταγράφου μπορεί να οφείλεται στην αυξημένη απόδοση της μετάφρασης, δηλαδή, στη διαθεσιμότητα του κυττάρου σε ριβοσώματα και στο χρόνο στον οποίο τα ριβοσώματα στρατολογούνται στο mRNA ώστε να αρχίσει η μετάφραση. Η μείωση της έκφρασης της πρωτεΐνης και όχι του μεταγράφου μπορεί να εξηγηθεί από την αυξημένη αποικοδόμησή της και την ενεργοποίηση ενός μηχανισμού που μειώνει την απόδοση της μετάφρασης ώστε να εξισορροπηθεί η υποέκφραση της πρωτεΐνης σε επίπεδο μεταγράφου.

➤ *Κοινές βιολογικές διαδικασίες μεταξύ των τριών ομικών προτύπων που επηρεάζονται μετά την καταστολή της έκφρασης του FRA10AC1*

Η σύγκριση μεταξύ των τριών προτύπων είναι δυνατόν να γίνει ως προς τον έλεγχο των μεταβολικών μονοπατιών που μπορεί να επηρεάζονται και στα τρία επίπεδα έπειτα από την καταστολή της έκφρασης του γονιδίου *FRA10AC1*. Ωστόσο, δεν μπορεί να γίνει η σύγκριση μεταξύ μορίων διότι στο μεταγραφικό και πρωτεϊνικό πρότυπο εξετάζονται οι αλλαγές των μεταγράφων των γονιδίων και των αντίστοιχων πρωτεϊνών

ενώ στο μεταβολικό οι αλλαγές των μεταβολιτών που ενσωματώνονται στα μονοπάτια του κεντρικού μεταβολισμού του άνθρακα. Στην προηγούμενη ενότητα, έγινε η σύγκριση μεταξύ του μεταγραφικού και πρωτεϊνικού προτύπου, κατά την οποία βρέθηκε ότι 26 μόρια εμφάνισαν αυξημένη έκφραση και στα δυο επίπεδα μετά την καταστολή της έκφρασης του *FRA10AC1* τα οποία συμμετέχουν στο μεταβολισμό λιπιδίων, φωσφολιπιδίων και ενώσεων που περιέχουν άζωτο (π.χ. NAD) καθώς και στην οξειδωτική φωσφορυλίωση. Στην μεταβολομική ανάλυση βρέθηκε ότι, από το σύνολο των 22 μεταβολιτών που παρουσίασαν αυξημένη συγκέντρωση μετά την καταστολή της έκφρασης του *FRA10AC1*, 4 συμμετέχουν στις αντιδράσεις του κύκλου του κιτρικού οξέος με αποτέλεσμα να αυξάνονται μόρια (NADH, FADH<sub>2</sub>) που συμμετέχουν στην οξειδωτική φωσφορυλίωση προκαλώντας αύξηση της ενεργότητάς της. Επιπλέον, η αύξηση μεταβολιτών που δρουν ως αντιοξειδωτικά μόρια (N-ακετυλ-γλουταμικό οξύ, N-ακετυλ-ασπαρτικό οξύ, β-αλανίνη) υποδεικνύουν την αυξημένη ενεργότητα της οξειδωτικής φωσφορυλίωσης. Ακόμη παρατηρήθηκε η αύξηση δυο μεταβολιτών (3-φωσφορική γλυκερόλη, βιταμίνη B6) που συμμετέχουν στο μεταβολισμό γλυκερολιπιδίων και λιπιδίων. Τα κοινά μόρια μεταξύ μεταγραφικού και πρωτεϊνικού προτύπου που εμφάνισαν μειωμένη έκφραση απουσία της *FRA10AC1*, και σχετίζονται με μεταβολικά μονοπάτια (μεταβολισμός λιπιδίων, ενώσεων που έχουν άζωτο και φώσφορο, πουρινών και πυριμιδινών) ήταν 27. Από τους 2 μεταβολίτες που εμφάνισαν μειωμένη συγκέντρωση απουσία της *FRA10AC1*, ο ένας (μυοϊνοσιτόλη) ενσωματώνεται στο μονοπάτι της σύνθεσης φωσφολιπιδίων.

#### 4.4.2. Πιθανοί μοριακοί μηχανισμοί μέσω των οποίων η καταστολή της δράσης της *FRA10AC1* τροποποιεί την έκφραση των άλλων μορίων

Οι αλλαγές στην αφθονία των πρωτεϊνών ή των μεταγράφων που παράγονται από τα αντίστοιχα γονίδια, έπειτα από την καταστολή της έκφρασης του *FRA10AC1*, θα μπορούσαν να εξηγηθούν από την πιθανή συμμετοχή της *FRA10AC1* σε διαφόρους ρυθμιστικούς μηχανισμούς που αναλύονται παρακάτω.

##### ➤ *Πιθανός ρόλος της *FRA10AC1* στην εναλλακτική ή μη συναρμογή*

Η δράση της *FRA10AC1* στην εναλλακτική ή μη συναρμογή μπορεί να είναι άμεση εφόσον συνδέεται στο πρόδρομο μετάγραφο ή έμμεση λειτουργώντας ως σημείο στρατολόγησης άλλων πρωτεϊνών (πρωτεΐνη-προσαρμογέας).

Η άμεση δράση της *FRA10AC1* στην εναλλακτική ή μη συναρμογή μπορεί να οφείλεται στην συνεργασία της με την *DGCR14*. Οι δυο πρωτεΐνες φαίνεται να έχουν συντηρημένη λειτουργία στη διαδικασία της συναρμογής διότι, σύμφωνα με τους Dutcher et al. (2018), βρέθηκε ότι συνεισφέρουν, στο μονοκύτταρο οργανισμό *Chlamydomonas reinhardtii*, στην αναγνώριση των θέσεων συναρμογής. Έτσι, η *FRA10AC1* σε συνεργασία με την *DGCR14* μπορεί να επηρεάζει άμεσα τη συναρμογή, δεδομένου ότι η άμεση αλληλεπίδραση μεταξύ των δυο πρωτεϊνών έχει επιβεβαιωθεί μέσω διαφορετικών πειραματικών προσεγγίσεων και είναι εξελικτικά συντηρημένη. Την υπόθεση συμμετοχής της *FRA10AC1* σε αυτό το στάδιο της συναρμογής ενισχύει και η άμεση αλληλεπίδρασή της με την *U2AF1* η οποία έχει εδραιωμένο ρόλο στην αναγνώριση των θέσεων συναρμογής. Έτσι, η καταστολή της *FRA10AC1* θα

μπορούσε να επιφέρει αλλαγές στη συναρμογή των γονιδίων, σε γενικότερο πλαίσιο και όχι μόνο εντός των διαδικασιών επεξεργασίας του mRNA, προκαλώντας την αύξηση ή μείωση της έκφρασης των εναλλακτικών ή μη μεταγράφων και των αντίστοιχων πρωτεϊνών.

Ένας από τους πιθανούς μηχανισμούς έμμεσης δράσης της FRA10AC1 στη διαδικασία της συναρμογής είναι ο ρόλος της ως συνδεδετικός κρίκος παραγόντων συναρμογής. Η διαταραχή της λειτουργίας αυτών των παραγόντων, η οποία προκαλείται μετά από την καταστολή της FRA10AC1, μπορεί να έχει σοβαρές επιπτώσεις στη συναρμογή των γονιδίων. Ο ρόλος αυτός της FRA10AC1 μπορεί να υποδεικνύεται από την αλληλουχία της, η οποία μπορεί να συμβάλλει στην αλληλεπίδρασή της με άλλες πρωτεΐνες οι οποίες εμφανίζουν βασικό ρόλο στη συναρμογή (π.χ. πρωτεΐνες SR). Ειδικότερα, η αλληλουχία της περιέχει εγγενώς μη δομημένες περιοχές (Intrinsically Disordered Regions, IDR) (Korneta & Bujnicki 2012), δηλαδή, περιοχές που δεν παρουσιάζουν σταθερή δευτεροταγή και τριτοταγή δομή, είναι ιδιαίτερα εύκαμπτες και μπορούν να αυξήσουν την ταχύτητα σύνδεσης και αποσύνδεσης μεταξύ των μορίων (RNA ή πρωτεϊνών), περιέχουν συντηρημένα μοτίβα τα οποία αποτελούν θέσεις αναγνώρισης άλλων πρωτεϊνών ή μορίων RNA, και αποτελούν συχνά θέσεις μετα-μεταφραστικών τροποποιήσεων. Οι περιοχές IDR της FRA10AC1 είναι πλούσιες σε αργινίνη και σερίνη (RS-like) (Korneta & Bujnicki 2012). Οι περιοχές «RS-like» βοηθούν στη σύνδεση με παρόμοια μοτίβα (π.χ. RS domains) τα οποία διαθέτουν οι πρωτεΐνες SR και άλλα συστατικά του σωματίου συναρμογής (Calarco et al. 2009). Με δεδομένο ότι η U2AF1 που αποτελεί πρώτο «γείτονα» της FRA10AC1 και ανήκει στην οικογένεια των πρωτεϊνών SR, θα μπορούσε να υποτεθεί ότι οι περιοχές IDR μεσολαβούν σε αυτήν την αλληλεπίδραση, υποδεικνύοντας την πιθανή συμμετοχή της FRA10AC1, σε συνεργασία με άλλες πρωτεΐνες, σε βασικά βήματα της συναρμογής, όπως η αναγνώριση των εναλλακτικών ή μη θέσεων συναρμογής. Η διαταραχή της αλληλεπίδρασης αυτής μετά από την απουσία της FRA10AC1 θα μπορούσε πιθανώς να έχει επιπτώσεις και στις αλληλεπιδράσεις της U2AF1. Για παράδειγμα, η U2AF1 αλληλεπιδρά με την U2AF2 και την PUF60, και η συνεργασία της PUF60, με τις άλλες δύο πρωτεΐνες, όπως προαναφέρθηκε, συμβάλλει στην αναγνώριση της 3' θέσης συναρμογής (Hastings et al. 2007). Έτσι, σύμφωνα με αυτό το μηχανισμό, η καταστολή της έκφρασης της FRA10AC1 θα μπορούσε να προκαλέσει διαταραχή στις αλληλεπιδράσεις μεταξύ των πρωτεϊνών που αποτελούν βασικούς παράγοντες εναλλακτικής ή μη συναρμογής ώστε να μην μπορεί να διεξαχθεί η αποκοπή των ιντρονίων. Άρα, οι επιδράσεις της διαταραχής αυτής δε θα περιορίζονταν μόνο εντός των διαδικασιών επεξεργασίας του mRNA αλλά και σε γενικότερο πλαίσιο καθώς δε θα ήταν εφικτή να γίνει η συναρμογή των γονιδίων που συμμετέχουν σε ποικίλες διεργασίες του κυττάρου, επηρεάζοντας τα επίπεδα έκφρασής τους.

Ένας δεύτερος μηχανισμός έμμεσης συμμετοχής της FRA10AC1 στη συναρμογή, μπορεί να είναι η δράση της πρωτεΐνης ως συνδεδετικού κρίκου μεταξύ πρωτεϊνών που συμμετέχουν στην αναγνώριση των θέσεων συναρμογής και αυτών που λαμβάνουν μέρος στα καταλυτικά βήματα της διαδικασίας ή ακόμα μεταξύ παραγόντων εναλλακτικής ή μη συναρμογής που συμμετέχουν σε διαφορετικά στάδια της

διαδικασίας (π.χ. πρωτεΐνες hnRNP). Σχετικά με την πρώτη υπόθεση αυτού του μηχανισμού, η FRA10AC1, όπως αναφέρθηκε και παραπάνω, φαίνεται να έχει συντηρημένη λειτουργία στην αναγνώριση των θέσεων συναρμογής καθώς αυτός ο ρόλος της δείχθηκε σε μονοκύτταρο οργανισμό (Dutcher et al. 2018). Επιπλέον, έχει απομονωθεί ως συστατικό των τελευταίων υποσυμπλόκων του σωματίου συναρμογής (B act, C, P) και η ευρύτερη «γειτονιά» της στο δίκτυο αλληλεπιδράσεων του σωματίου συναρμογής είναι εμπλουτισμένη με παράγοντες που συμμετέχουν σε αυτά τα στάδια της διαδικασίας (βλ. Ενότητα 4.4.3.). Άρα, η FRA10AC1 θα μπορούσε, μέσω αλληλεπιδράσεων, να «φέρνει κοντά» τους παράγοντες που βρίσκονται στις θέσεις συναρμογής με αυτούς που δρουν κατά τη διάρκεια των καταλυτικών βημάτων της διαδικασίας, δεδομένου ότι η πρωτεΐνη εμφανίζεται σε αυτά τα στάδια. Έτσι, η καταστολή της έκφρασής της μπορεί να διαταράσσει τις αλληλεπιδράσεις μεταξύ παραγόντων που συμμετέχουν σε διαφορετικά στάδια της συναρμογής, γεγονός που μπορεί να έχει γενικότερες επιπτώσεις στη συναρμογή γονιδίων που συμμετέχουν σε πολλές διαφορετικές διεργασίες του κυττάρου. Όσον αφορά τη δεύτερη υπόθεση του μηχανισμού, έχειδειχθεί ότι πρωτεΐνες hnRNP παίζουν ρόλο, στα καρκινικά κύτταρα, στον έλεγχο της επιλογής της πορείας της γλυκόλυσης προς την αερόβια γλυκόλυση ή τον κύκλο του κιτρικού οξέος και κατ' επέκταση την οξειδωτική φωσφορυλίωση, ρυθμίζοντας την εναλλακτική συναρμογή του γονιδίου της πυροσταφυλικής κινάσης που καταλύει την αντίδραση του φωσφοενολοπυροσταφυλικού οξέος σε πυροσταφυλικό οξύ (Chen et al. 2010). Δεδομένου ότι στην ευρύτερη «γειτονιά» της FRA10AC1 στο δίκτυο του σωματίου συναρμογής υπάρχουν πρωτεΐνες hnRNP, η καταστολή της έκφρασής της θα μπορούσε να έχει επιπτώσεις και στην συναρμογή των γονιδίων, για παράδειγμα αυτών που κωδικοποιούν ένζυμα που συμμετέχουν σε μεταβολικές διαδικασίες (π.χ. οξειδωτική φωσφορυλίωση), αν ληφθεί υπόψη ότι η απουσία της πρωτεΐνης προκάλεσε διαταραχές σε διάφορα μεταβολικά μονοπάτια (βλ. Ενότητα 4.4.1.).

Ένας τελευταίος μηχανισμός έμμεσης συμμετοχής της FRA10AC1 στη διαδικασία της συναρμογής στηρίζεται στη δράση της και πάλι ως συνδετικού κρίκου μεταξύ πρωτεϊνών που δεν αποτελούν συστατικά του σωματίου και παραγόντων εναλλακτικής ή μη συναρμογής. Για παράδειγμα, έχειδειχθεί σε διάφορες μελέτες ότι η οικογένεια των κινασών σερίνης/θρεονίνης (MAPKs), μέσω της φωσφορυλίωσης πρωτεϊνών που περιλαμβάνονται στα σηματοδοτικά μονοπάτια που αυτές συμμετέχουν, μπορεί να ρυθμίσουν την συναρμογή γονιδίων καθώς οι πρωτεΐνες που φωσφορυλιώνουν αλληλεπιδρούν με παράγοντες συναρμογής (π.χ. U2AF2, SPF45) (Al-Ayoubi et al. 2012, Tisserant & König 2008). Έτσι αν ληφθεί υπόψη ότι ένας πρώτος «γείτονας» της FRA10AC1 εκτός του σωματίου συναρμογής (MAP3K12) ανήκει στην οικογένεια των MAP κινασών, η καταστολή της FRA10AC1 ίσως θα μπορούσε, δρώντας ως συνδετικός κρίκος, να προκαλέσει διαταραχές στη συναρμογή γονιδίων που συμμετέχουν σε πληθώρα βιολογικών διαδικασιών.

- *Πιθανός ρόλος της FRA10AC1, ως παράγοντας εναλλακτικής ή μη συναρμογής, στη μεταγραφή-ρύθμιση της μεταγραφής*

Πιθανοί μηχανισμοί αυτής της κατηγορίας αφορούν τη δράση της FRA10AC1 ως παράγοντα εναλλακτικής ή μη συναρμογής, ο οποίος, έχοντας αυτήν την ιδιότητα, μπορεί να ρυθμίζει την έναρξη ή επιμήκυνση της μεταγραφής. Αυτή η υπόθεση λειτουργίας έχει δειχθεί σε αρκετές ερευνητικές μελέτες. Ενδεικτικά, κατά την έναρξη της μεταγραφής, έχει δειχθεί ότι η καταστολή πρωτεϊνών (π.χ. CHD1) που συνδέουν συστατικά του συμπλόκου SF3a του U2 snRNP με τις τροποποιημένες ιστόνες (π.χ. H3K4me3) προκαλεί διαταραχές στη διαδικασία της συναρμογής και στη στρατολόγηση ρυθμιστών της μεταγραφής στις τροποποιημένες ιστόνες με αποτέλεσμα να καταστέλλεται ή να ενεργοποιείται η μεταγραφή γονιδίων (Sims et al. 2007). Έτσι, η καταστολή της FRA10AC1 θα μπορούσε να διαταράσσει τις αλληλεπιδράσεις των παραγόντων συναρμογής είτε με τις τροποποιημένες ιστόνες είτε με τις πρωτεΐνες-προσαρμογείς (π.χ. CHD1) προκαλώντας αλλαγές στην μεταγραφή των γονιδίων. Ακόμη, στο στάδιο της επιμήκυνσης, έχει δειχθεί ότι ρυθμιστές της εναλλακτικής συναρμογής (π.χ. πρωτεΐνες SR) λειτουργούν ως θέσεις στρατολόγησης παραγόντων που είναι υπεύθυνοι για το ρυθμό εξέλιξης της επιμήκυνσης της μεταγραφής ή συνεισφέρουν με έμμεσο τρόπο στην φωσφορυλίωση του καρβοξυτελικού άκρου της RNA πολυμεράσης II ρυθμίζοντας την πορεία της μεταγραφής (Braunschweig et al. 2013). Ένας εναλλακτικός τρόπος ρύθμισης της επιμήκυνσης της μεταγραφής από παράγοντες εναλλακτικής συναρμογής είναι η έμμεση συνεισφορά τους στην αποφυγή σχηματισμού υβριδίων RNA-DNA (R-loops) που είναι πιθανό να σχηματιστούν κατά τη δημιουργία του νεοσυντιθέμενου mRNA (Lin et al. 2008). Ένα τελευταίο παράδειγμα αποτελεί μια μελέτη στην οποία έχει δειχθεί ότι ένα σύμπλοκο δυο πρωτεϊνών (σύμπλοκο DBIRD) το οποίο ρυθμίζει την εναλλακτική συναρμογή εξονίων που βρίσκονται κοντά σε περιοχές που είναι πλούσιες σε αδενίνη και θυμίνη συμμετέχει και στην ρύθμιση της πορείας της εξέλιξης της επιμήκυνσης της μεταγραφής, χωρίς να είναι γνωστός ο μηχανισμός (Close et al. 2012). Σύμφωνα με τους παραπάνω μηχανισμούς, η καταστολή της FRA10AC1 θα μπορούσε να προκαλεί είτε την ενεργοποίηση είτε την καταστολή της μεταγραφής γονιδίων.

- *Πιθανός διττός ρόλος της FRA10AC1 στη συναρμογή και στη μεταγραφή*

Ένας πιθανός μηχανισμός της κατηγορίας αυτής στηρίζεται στον πιθανό διττό ρόλο της FRA10AC1 τόσο στη συναρμογή όσο και στη μεταγραφή. Ένα παράδειγμα που συνάδει με αυτήν την υπόθεση είναι η λειτουργία του πρώτου «γείτονα» της FRA10AC1, της DGCR14, η οποία έχει δειχθεί ότι συμμετέχει και στις δύο παραπάνω λειτουργίες. Συγκεκριμένα, έχει δειχθεί, στο *C. elegans*, ότι συνεισφέρει στη διασφάλιση της σωστής συναρμογής του mRNA, σε περιπτώσεις που υπάρχουν μη φυσιολογικές αλληλουχίες στην 3' θέση συναρμογής (Noma et al. 2014) και, στο μονοκύτταρο οργανισμό *Chlamydomonas reinhardtii*, ότι συνεισφέρει στην αναγνώριση των θέσεων συναρμογής (Dutcher et al. 2018). Παράλληλα, το ορθόλογο μόριό της στον ποντικό έχει προταθεί ότι λειτουργεί τόσο ως θετικός ρυθμιστής των πυρηνικών υποδοχέων RORγ, που λειτουργούν ως μεταγραφικοί παράγοντες, όσο και ως «γέφυρα» μεταξύ της συναρμογής και της μεταγραφής καθώς διαφορετικά τμήματα

της πρωτεϊνικής της ακολουθίας αλληλεπιδρούν με τα snRNAs και με τους πυρηνικούς υποδοχείς RORγ (Takada 2018). Ένα επιπρόσθετο παράδειγμα αποτελεί μια μελέτη στον ποντικό στην οποία δείχθηκε ότι δύο ρυθμιστές της μεταγραφής, οι Nacc1 και Zfp871, ρυθμίζουν την έκφραση γονιδίων που αποτελούν ρυθμιστές της εναλλακτικής ή μη συναρμογής, και επιπλέον, μπορούν να συνδεθούν και στο mRNA, χωρίς να διαθέτουν μοτίβα πρόσδεσης στο RNA, λειτουργώντας ως ρυθμιστές της εναλλακτικής ή μη συναρμογής (Han et al. 2017). Σύμφωνα με αυτό το μηχανισμό, η καταστολή της FRA10AC1, ως πιθανός παράγοντας συναρμογής και ως ρυθμιστής των παραγόντων συναρμογής, θα μπορούσε να προκαλέσει τη λανθασμένη ή εναλλακτική συναρμογή των γονιδίων, προκαλώντας την αύξηση ή μείωση της έκφρασης των εναλλακτικών ή μη μεταγράφων και των αντίστοιχων πρωτεϊνών.

➤ *Πιθανός ρόλος της FRA10AC1 στη μεταγραφή-ρύθμιση της μεταγραφής*

Ένας πιθανός μηχανισμός δράσης της FRA10AC1 είναι η άμεση συμμετοχή της στη μεταγραφή δρώντας είτε ως μεταγραφικός παράγοντας ή ως συμπαράγοντας των μεταγραφικών παραγόντων, δηλαδή ως συν-ενεργοποιητής (co-activator) ή συν-καταστολέας (co-repressor). Σύμφωνα με αυτό το μηχανισμό, η καταστολή της έκφρασης της FRA10AC1 θα μπορούσε να προκαλέσει είτε την αυξημένη είτε την μειωμένη έκφραση των μεταγράφων και των αντίστοιχων πρωτεϊνών. Σύμφωνα με τα αποτελέσματα των ομικών αναλύσεων τα μόρια που παρουσίασαν μειωμένη έκφραση ήταν περισσότερα από αυτά με αυξημένη έκφραση, γεγονός που μπορεί να υποδείξει ότι η καταστολή της FRA10AC1 εμποδίζει την έκφραση των γονιδίων σε επίπεδο μεταγραφής, δηλαδή δρα ως πιθανός συν-ενεργοποιητής της μεταγραφής. Ωστόσο, αυτή η αναλογία αλλάζει εντός του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής με το 74% των μορίων με τροποποιημένη έκφραση να εμφανίζουν αυξημένα επίπεδα μετά την καταστολή της έκφρασης του γονιδίου FRA10AC1.

➤ *Πιθανός βιολογικός ρόλος της FRA10AC1 ως πρωτεΐνη-προσαρμογέας μεταξύ της συναρμογής και των υπολοίπων μετα-μεταγραφικών διαδικασιών*

Ένα πιθανός μηχανισμός που θα μπορούσε να προταθεί για την FRA10AC1, ο οποίος εξηγεί την εμφάνιση αλλαγών μετά την καταστολή της, είναι η συνεισφορά της, ως πρωτεΐνη-προσαρμογέας η οποία μεσολαβεί ώστε να αλληλεπιδράσουν άλλες πρωτεΐνες μεταξύ τους. Αυτό θα μπορούσε να γίνει μέσω των μη δομημένων περιοχών που διαθέτει (IDRs). Έτσι, θα μπορούσε να υποτεθεί ότι η FRA10AC1, μέσω της αλληλεπίδρασής της με άλλες πρωτεΐνες μπορεί να συμβάλλει στο να αποκτήσουν οι άλλες πρωτεΐνες την κατάλληλη διαμόρφωση ώστε να μπορούν και αυτές με τη σειρά τους να αλληλεπιδράσουν. Έτσι, η έλλειψή της θα μπορούσε να οδηγήσει σε τροποποιημένη έκφραση είτε των πρωτεϊνών που φυσιολογικά θα στρατολογούσε, λόγω π.χ. μειωμένης σταθερότητάς τους ή/και των μορίων mRNA που θα αποτελούσαν στόχο του μηχανισμού στον οποίο συμμετέχει.

Μία πρώτη υπόθεση αυτού του μηχανισμού είναι η δράση της ως πρωτεΐνη προσαρμογέας μεταξύ της εναλλακτικής συναρμογής και της αποικοδόμησης εναλλακτικών ή μη μεταγράφων. Έχει δείχθει ότι αρκετά από τα γεγονότα



φυσιολογικής ή μη εναλλακτικής συναρμογής (~10-20%) είναι πιθανόν να δημιουργήσουν πρόωρα κωδικόνια λήξης τα οποία οδηγούν σε αποικοδόμηση των μεταγράφων (Nonsense-Mediated Decay, NMD) και κατ' επέκταση σε αλλαγές στα επίπεδα έκφρασης τους (Pan et al. 2006). Σε αρκετές περιπτώσεις έχει αναφερθεί ότι ο συνδυασμός μεταξύ γεγονότων εναλλακτικής συναρμογής και αποικοδόμησης, όταν υπάρχει πρόωρο κωδικόνιο λήξης, ρυθμίζει τα επίπεδα έκφρασης βασικών παραγόντων της συναρμογής και των άλλων διαδικασιών επεξεργασίας του mRNA (Ni et al. 2007, Saltzman et al. 2008), αλλά και ομάδων πρωτεϊνών που συμμετέχουν σε άλλες διεργασίες (Barash et al. 2010, Lareau et al. 2007). Στο ίδιο πλαίσιο, έχει δειχθεί ότι η hnRNP I η οποία αποτελεί ρυθμιστή της εναλλακτικής συναρμογής, καταστέλλει την συναρμογή εναλλακτικών εξονίων, σε μη νευρικά κύτταρα, μιας ομάδας τεσσάρων γονιδίων που παρουσιάζουν ιστοειδική έκφραση στους νευρώνες με αποτέλεσμα τα μετάγραφα που δεν υφίστατο εναλλακτική συναρμογή να παραμένουν στο πυρήνα και να αποικοδομούνται από το πυρηνικό εξώσωμα. Αντίθετα, σε νευρικά κύτταρα, η αναστολή της δράσης της hnRNP I μέσω miRNA οδηγεί στην συναρμογή των εναλλακτικών εξονίων επιτρέποντας την έξοδο των μεταγράφων από τον πυρήνα και την μετάφρασή τους (Yap et al. 2012). Σύμφωνα με τον παραπάνω μηχανισμό, η FRA10AC1 μπορεί να «αναγνωρίζει» τη λανθασμένη συναρμογή και να μεσολαβεί ώστε να στρατολογηθούν παράγοντες αποικοδόμησης. Έτσι, η απουσία της μπορεί να προκαλούσε την αύξηση της έκφρασης αρκετών μεταγράφων ή/και την αστάθεια των παραγόντων αποικοδόμησης.

Μια εναλλακτική υπόθεση του μηχανισμού λειτουργίας της FRA10AC1 ως πρωτεΐνη-προσαρμογέας είναι η συνεργασία της με παράγοντες που ρυθμίζουν την εναλλακτική αποκοπή και πολυαδενυλίωση του 3' άκρου των mRNA και τη σταθερότητα αυτών. Οι εναλλακτικές θέσεις αποκοπής και πολυαδενυλίωσης του 3' άκρου των mRNA, οι οποίες έχουν βρεθεί τουλάχιστον στο 50% των γονιδίων του ανθρώπου (Tian et al. 2005), οδηγούν στην παραγωγή μεταγράφων με μεγαλύτερο ή μικρότερο μήκος 3' μη μεταφραζόμενων περιοχών ως προς το φυσιολογικό, γεγονός που μπορεί να επηρεάσει τη σταθερότητα και τη μετάφραση των mRNA (Katz et al. 2010, Millevoi & Vagner 2010). Αυτό μπορεί να γίνει μέσω της ρύθμισης της πρόσδεσης των miRNA και των πρωτεϊνών που συνδέονται με το RNA (RNA-binding) στις 3' μη μεταφραζόμενες περιοχές η οποία επηρεάζει τα επίπεδα έκφρασης των μεταγράφων και των πρωτεϊνών (Mayr & Bartel 2009). Έτσι, η FRA10AC1 θα μπορούσε να αλληλεπιδρά με τους παραπάνω παράγοντες, λειτουργώντας, ίσως, ως σημείο στρατολόγησής τους πάνω στο μετάγραφο, με αποτέλεσμα η καταστολή της έκφρασής της να οδηγεί είτε στην αύξηση είτε στην μείωση των μεταγράφων και των αντίστοιχων πρωτεϊνών, αν ληφθεί υπόψη ότι δυο δεύτεροι «γείτονές» της (PCBP1, PCBP2) συμμετέχουν στην επεξεργασία του 3' άκρου.

Μία επιπρόσθετη υπόθεση του μηχανισμού λειτουργίας της FRA10AC1 ως πρωτεΐνη-προσαρμογέας είναι η δράση της ως «μεσολαβητή» μεταξύ της συναρμογής και της εξόδου του mRNA από τον πυρήνα. Δεδομένου ότι η ευρύτερη «γειτονιά» της είναι εμπλουτισμένη με παράγοντες εξόδου του mRNA από τον πυρήνα, εκ των οποίων μία (THOC1) είναι πρώτος «γείτονας» της (βλ. Ενότητα 4.4.3.), θα μπορούσε να υποτεθεί ότι η FRA10AC1 λειτουργεί ως σημείο στρατολόγησης πρωτεϊνών που

κατευθύνουν το μετάγραφο εκτός πυρήνα. Έτσι, η καταστολή της έκφρασής της μπορεί να διαταράσσει τις αλληλεπιδράσεις με αυτές τις πρωτεΐνες οι οποίες πιθανώς δεν μπορούν να οδηγήσουν το μετάγραφο εκτός πυρήνα για να μεταφραστεί με αποτέλεσμα να μειώνεται η έκφρασή του.

Μία τελευταία υπόθεση του μηχανισμού λειτουργίας της FRA10AC1 ως πρωτεΐνη-προσαρμογέας είναι η συμμετοχή της στο ρυθμιστικό μηχανισμό της μετα-μεταγραφικής γονιδιακής έκφρασης, «RNA regulon», στους ευκαρυώτες. Τα «RNA regulons» είναι πρωτεΐνες οι οποίες συνδέονται με το RNA (RNA binding) ή μη κωδικά RNAs (non-coding RNAs) και συντονίζουν-ρυθμίζουν ομάδες mRNAs που κωδικοποιούν λειτουργικά συσχετιζόμενες πρωτεΐνες ώστε να διασφαλιστεί η συντονισμένη μετάφραση αυτών των πρωτεϊνών (Halbeisen et al. 2008, Morris et al. 2010). Η FRA10AC1 είναι πιθανό να συνεργάζεται με τέτοιου είδους πρωτεΐνες, δεδομένου ότι ένα αρκετά μεγάλο ποσοστό (67%) της ευρύτερης «γειτονιάς» της (πρώτοι & δεύτεροι γείτονες) στο δίκτυο του σωματίου συναρμογής ανήκουν σε αυτήν την κατηγορία πρωτεϊνών (RNA binding), 7 εκ των οποίων είναι πρώτοι «γείτονες» της (SF3B2, PRPF3, CHERP, U2AF1, DGCR14, NKAP, MFAP1). Αυτή η κατηγορία πρωτεϊνών συμμετέχει στη ρύθμιση της σταθερότητας, της αποικοδόμησης και της μετάφρασης των mRNA που παράγονται από γονίδια που συμμετέχουν σε κοινές βιολογικές διαδικασίες. Αυτό επιτυγχάνεται είτε μέσω της σύνδεσής τους με τις 3' μη μεταφραζόμενες περιοχές των μεταγράφων είτε δρώντας συναγωνιστικά ή ανταγωνιστικά με τα μη κωδικά RNAs (π.χ. miRNAs) ή λειτουργώντας ως σημεία στρατολόγησης άλλων πρωτεϊνών που συμβάλλουν και αυτές στους μετα-μεταγραφικούς μηχανισμούς (Iadevaia & Gerber 2015, Imig et al. 2012). Η καταστολή λοιπόν της έκφρασης της FRA10AC1 θα μπορούσε να μειώσει ή να αυξήσει την έκφραση των πρωτεϊνών ενεργοποιώντας ή καταστέλλοντας, μέσω αλληλεπιδράσεων, την πρόσδεση πρωτεϊνών που συνδέονται με το mRNA εντός του πυρήνα, και κατ' επέκταση αυτή η διαταραχή της σύνδεσης των πρωτεϊνών στο mRNA εντός του πυρήνα θα μπορούσε να επηρεάσει και την επακόλουθη δράση των miRNAs στο κυτταρόπλασμα.

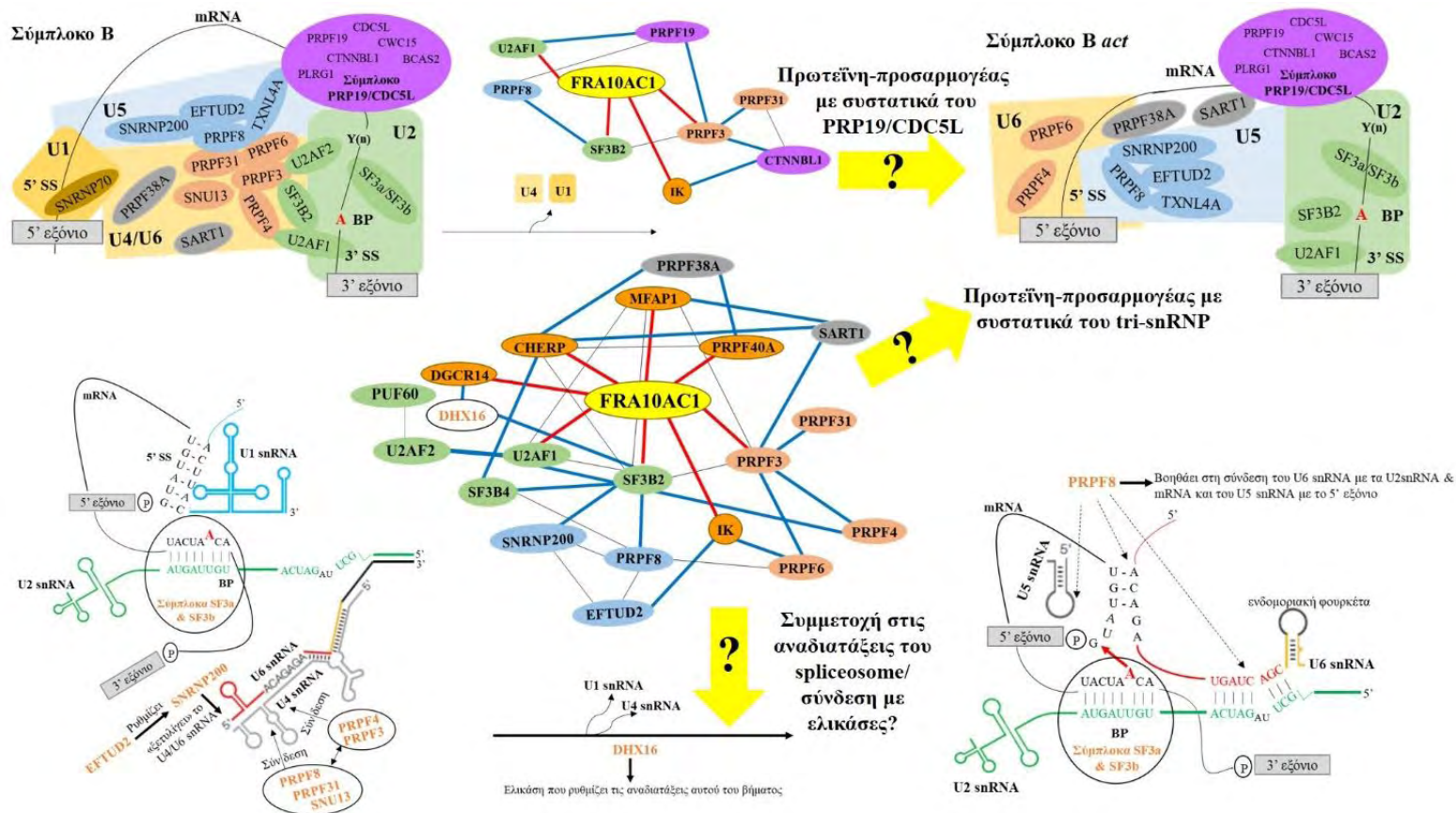
#### 4.4.3. Πιθανά μοντέλα λειτουργίας της FRA10AC1 εντός του δικτύου πρωτεϊνικών αλληλεπιδράσεων του σωματίου συναρμογής

Λαμβάνοντας υπόψη τους παραπάνω πιθανούς μηχανισμούς λειτουργίας της FRA10AC1 σε συνδυασμό με τις πρωτεϊνικές αλληλεπιδράσεις της εντός του δικτύου του σωματίου συναρμογής, ορισμένα μοντέλα λειτουργίας της αναδεικνύονται ως πιο πιθανά. Στην πλειοψηφία των μηχανισμών δεν αναμένεται να λαμβάνουν χώρα αλληλεπιδράσεις μεταξύ της FRA10AC1 και των στόχων της που εμφανίζουν αύξηση ή μείωση στην έκφρασή τους καθώς η ρύθμιση των επιπέδων έκφρασης πραγματοποιείται σε επίπεδο mRNA ενώ το παραπάνω δίκτυο περιλαμβάνει αποκλειστικά πρωτεϊνικές αλληλεπιδράσεις. Ωστόσο, ο μηχανισμός δράσης της FRA10AC1 ως πρωτεΐνη-προσαρμογέα που συνεισφέρει στο να λάβουν χώρα αλληλεπιδράσεις μεταξύ πρωτεϊνών (βλ. παραπάνω), μπορεί να αιτιολογήσει τροποποιημένα επίπεδα έκφρασης και των αλληλεπιδρώντων μορίων της. Έτσι, βάσει

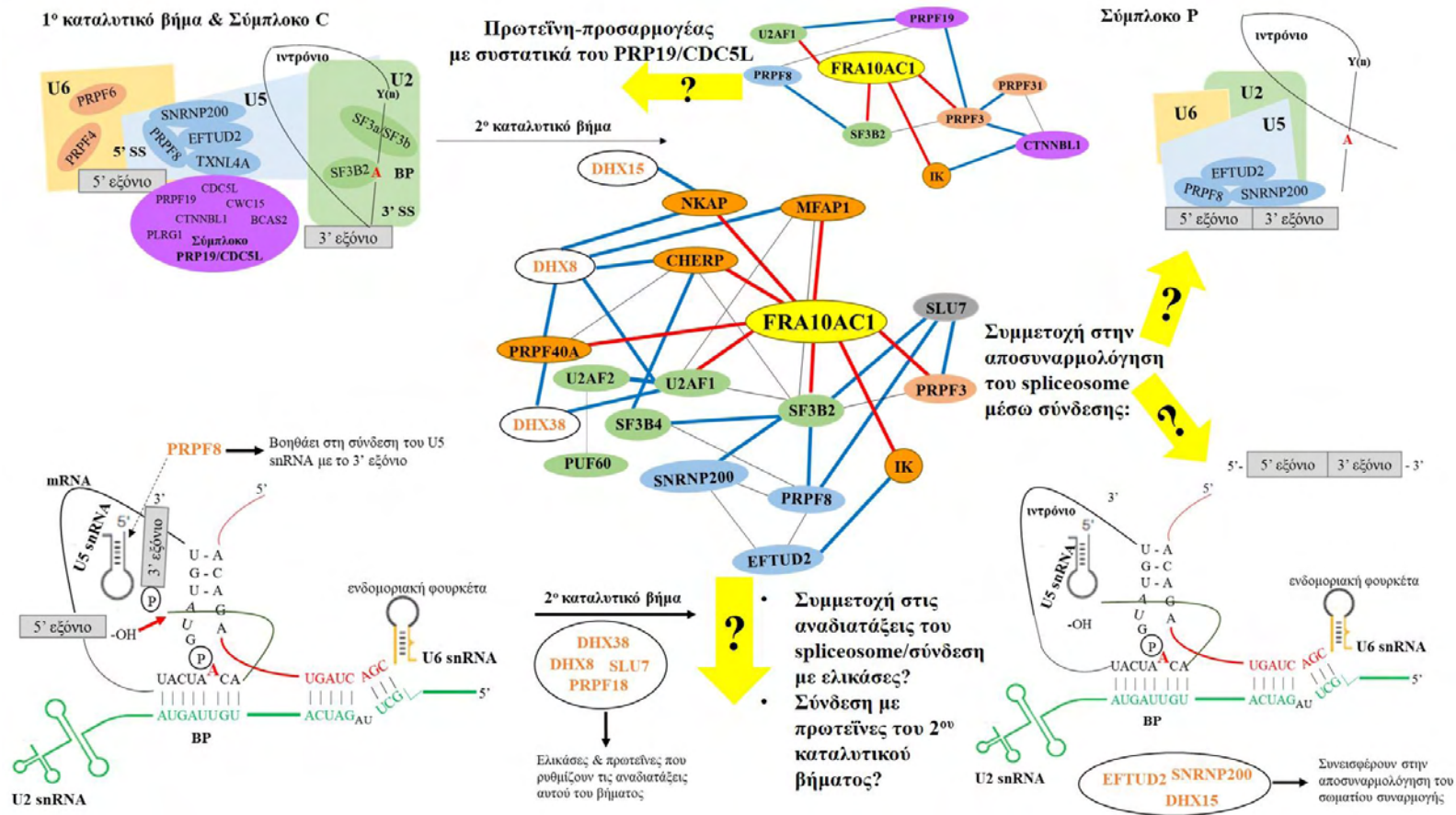
των αλληλεπιδράσεων της FRA10AC1 εντός του δικτύου του σωματίου συναρμογής, τα πιο πιθανά μοντέλα λειτουργίας της πρωτεΐνης είναι δύο.

➤ *Πιθανό μοντέλο λειτουργίας της FRA10AC1 ως πρωτεΐνη-προσαρμογέας μεταξύ παραγόντων συναρμογής*

Το πρώτο μοντέλο λειτουργίας περιλαμβάνει τη δράση της ως πρωτεΐνη-προσαρμογέα μεταξύ παραγόντων συναρμογής που συμμετέχουν στην αναγνώριση των 5' και 3' θέσεων συναρμογής και παραγόντων των καταλυτικών σταδίων της διαδικασίας. Στους πρώτους «γείτονες» της τοποθετούνται δυο βασικοί παράγοντες συναρμογής (Εικ. 83, 84), η U2AF1 η οποία αναγνωρίζει τη συντηρημένη αλληλουχία AG στην 3' θέση συναρμογής και είναι απαραίτητη τόσο για τη συναρμολόγηση των συστατικών του σωματίου συναρμογής όσο και για το δεύτερο καταλυτικό βήμα της συναρμογής (Voith von Voithenberg et al. 2016, Wu & Krainer 1999) και η SF3B2 που συμμετέχει στην αναγνώριση της θέσης διακλάδωσης στο ιντρόνιο και στη στρατολόγηση σε αυτήν πρωτεϊνών του U2 snRNP (Spadaccini et al. 2006, Golas et al. 2003, Lardelli et al. 2010). Η «γειτονιά» της (δευτεροί γείτονες) εμπλουτίζεται με παράγοντες που συμμετέχουν σε αυτά τα στάδια της διαδικασίας, όπως την U2AF2 και PUF60 οι οποίες συνεργάζονται με την U2AF1, και την SF3B4 που δρα συνεργατικά με την SF3B2 (Εικ. 83, 84). Επιπλέον, δυο πρώτοι «γείτονες» της, η PRPF3 και η MFAP1 συμμετέχουν πριν ή κατά τα καταλυτικά βήματα της διαδικασίας (Εικ. 83, 84). Η PRPF3 είναι συστατικό ενός συμπλόκου (U4/U6.U5) που διευκολύνει τις αναδιατάξεις του U4 και U6 RNA και των U1 και U4 snRNPs έτσι ώστε να ξεκινήσει το πρώτο καταλυτικό βήμα της συναρμογής (Maita et al. 2005, Liu 2006). Εξετάζοντας τους δεύτερους «γείτονες» της παρατηρήθηκε ότι οκτώ πρωτεΐνες (PRPF8, SNRNP200, EFTUD2, PRPF31, PRPF4, PRPF6, SART1, PRPF38A) έχουν απομονωθεί ως συστατικά του τριμερούς συμπλόκου U4/U6.U5 snRNP (Εικ. 83, 84) το οποίο έχει συνολικά 30 πρωτεΐνες εκ των οποίων οι 18 είναι Sm και Lsm (Nguyen et al. 2013, Wan et al. 2016). Επιπλέον, η σύνδεσή της με το σύμπλοκο U4/U6.U5 snRNP ενισχύεται με την αλληλεπίδραση των πρώτων γειτόνων της, MFAP1 και IK, με συστατικά του συμπλόκου αυτού (PRPF6, PRPF3) (Hegele et al. 2012). Ακόμη υποστηρίζεται και από τη δημοσίευση των Akerman et al. (2015) στην οποία η FRA10AC1 συμμετέχει σε ένα cluster 25 πρωτεϊνών, 4 από τις οποίες (PRPF3, PRPF38A, PRPF4, SART1) αποτελούν συστατικά του U4/U6.U5 snRNP, και δύο (SNRNP27, SART3) συνδέονται με το σύμπλοκο αυτό. Επίσης, στη «γειτονιά» της FRA10AC1 τοποθετούνται δυο πρωτεΐνες που αποτελούν συστατικά του συμπλόκου PRP19/CDC5L (Εικ. 83, 84) το οποίο περιλαμβάνει συνολικά 7 πρωτεΐνες και ρυθμίζει το σχηματισμό και την εξέλιξη των απαραίτητων αναδιατάξεων που συμβαίνουν στα δυο καταλυτικά βήματα της συναρμογής (Grote et al. 2010, Chanarat & Sträber 2013). Τέλος, στους δεύτερους «γείτονες» της υπάρχουν αρκετές ελικάσες (π.χ. DHX16, DHX38, DHX8) που συνεισφέρουν στη ρύθμιση των αναδιατάξεων στα καταλυτικά βήματα της συναρμογής (Gencheva et al. 2010; Warkocki et al. 2009) (Εικ. 83, 84). Έτσι παρατηρείται ότι η ευρύτερη «γειτονιά» της FRA10AC1 είναι εμπλουτισμένη σε παράγοντες που συμμετέχουν στην αναγνώριση των θέσεων συναρμογής και στα καταλυτικά βήματα της διαδικασίας (Εικ. 83, 84).



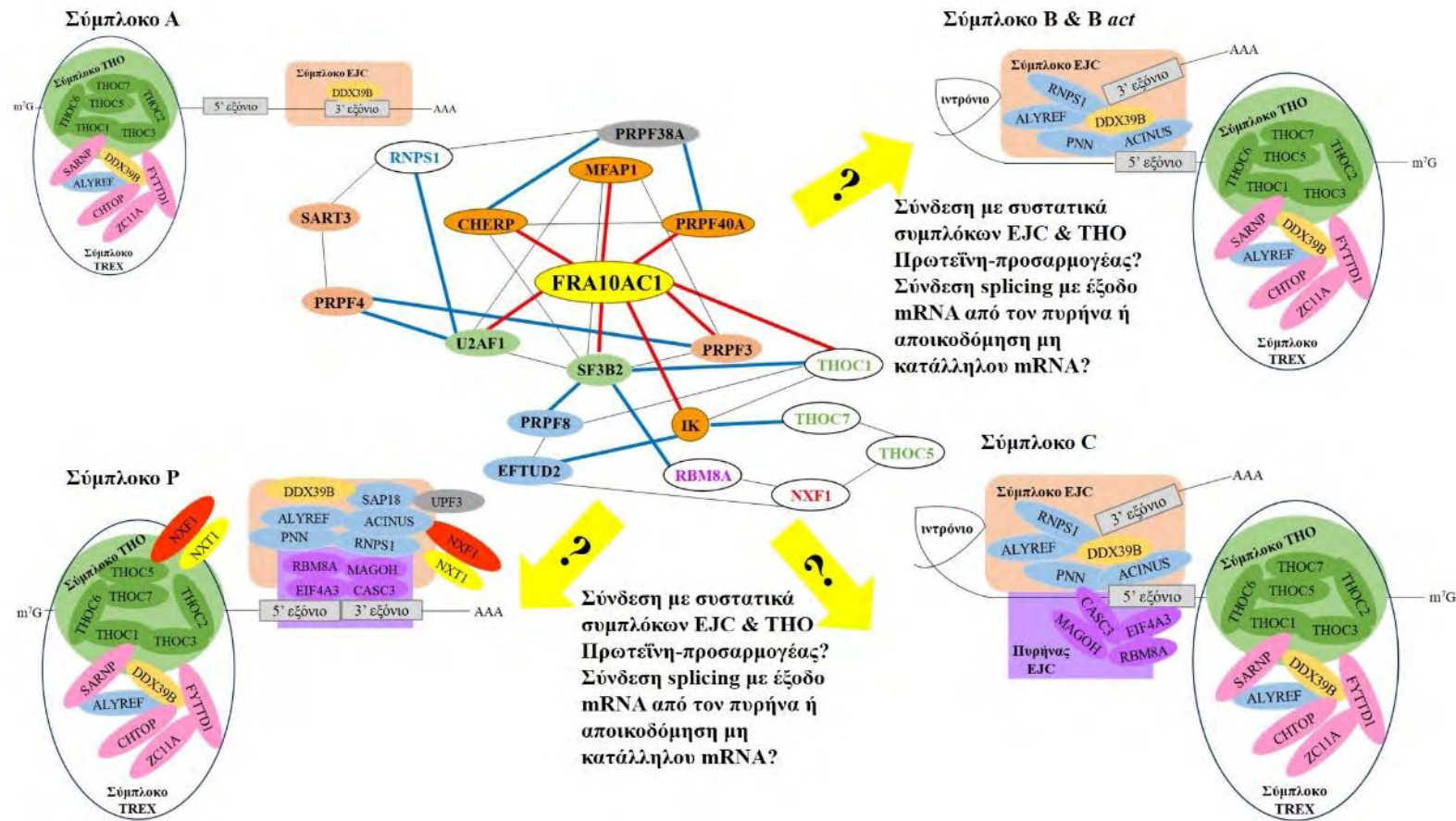
**Εικ. 83:** Πιθανός βιολογικός ρόλος της **FRA10AC1** ως πρωτεΐνη-προσαρμογέας μεταξύ των αρχικών (αναγνώριση θέσεων συναρμογής) και καταλυτικών σταδίων (υποσύμπλοκα B και B act) της συναρμογής βάσει των «γειτόνων» της στο δίκτυο αλληλεπιδράσεων του σωματίου συναρμογής. Απεικονίζονται οι πρώτοι και δεύτεροι «γείτονες» της FRA10AC1 που συμμετέχουν στο σχηματισμό των υποσυμπλόκων B και B act του σωματίου συναρμογής (πάνω μέρος), και οι αναδιατάξεις του mRNA και των snRNAs που συμβαίνουν σε αυτά τα στάδια της συναρμογής (κάτω μέρος). Με πράσινο, μπλε, κόκκινο και μοβ χρώμα φαίνονται τα συστατικά των συμπλόκων U2snRNP, U5snRNP, U4/U6snRNP, PRP19/CDC5L, αντίστοιχα, και με γκρι οι βοηθητικοί παράγοντες του συμπλόκου U4/U6snRNP. Οι κόκκινες και οι μπλε γραμμές δείχνουν τις αλληλεπιδράσεις της FRA10AC1 με τους πρώτους και δεύτερους «γείτονές» της.



Εικ. 84: Πιθανός βιολογικός ρόλος της FRA10AC1 ως πρωτεΐνη-προσαρμογέας μεταξύ των αρχικών (αναγνώριση θέσεων συναρμογής) και καταλυτικών σταδίων (υποσύμπλοκα C και P) της συναρμογής βάσει των «γειτόνων» της στο δίκτυο αλληλεπιδράσεων του σωματίου συναρμογής. Φαίνονται οι πρώτοι και δεύτεροι «γείτονες» της FRA10AC1 που συμμετέχουν στο σχηματισμό των υποσυσμπλόκων C και P του σωματίου συναρμογής (πάνω μέρος), και οι αναδιατάξεις του mRNA και των snRNAs που συμβαίνουν σε αυτά τα στάδια της συναρμογής (κάτω μέρος). Ο χρωματικός κώδικας είναι ίδιος με αυτόν της Εικόνας 83.

- *Πιθανό μοντέλο λειτουργίας της FRA10AC1 ως πρωτεΐνη-προσαρμογέας μεταξύ συναρμογής και εξόδου του mRNA από τον πυρήνα ή αποικοδόμησης μη κατάλληλων mRNA*

Ένα δεύτερο μοντέλο λειτουργίας της FRA10AC1, βάσει των πρώτων και δεύτερων «γειτόνων» της, είναι η δράση της ως πρωτεΐνη-προσαρμογέας μεταξύ της συναρμογής και της εξόδου του mRNA από τον πυρήνα ή της αποικοδόμησης μη κατάλληλων mRNA. Η πρωτεΐνη FRA10AC1 θα μπορούσε να δρα ως μεσολαβητής στις αλληλεπιδράσεις των πρωτεϊνών των συμπλόκων TREX, NXF1-P15 και EJC που σχετίζονται με την έξοδο του mRNA από τον πυρήνα και την αποικοδόμηση μη κατάλληλων mRNA. Η διευρυμένη «γειτονιά» της σε ολόκληρο το δίκτυο (Εικ. 85) είναι εμπλουτισμένη με: α) παράγοντες του συμπλόκου EJC (RNPS1, NXF1, RBM8A), β) πρωτεΐνες που καθοδηγούν το ώριμο mRNA στους πυρηνικούς πόρους (π.χ. THOC1, THOC5, THOC7), εκ των οποίων η THOC1 είναι πρώτος «γειτόνάς» της, και γ) πρωτεΐνες που καθοδηγούν το ώριμο mRNA προς την αποικοδόμησή του, εφόσον δεν είναι κατάλληλο για να μεταφραστεί (π.χ. RNPS1). Το σύμπλοκο TREX αποτελείται από 13 πρωτεΐνες, 6 από τις οποίες (THOC1, THOC2, THOC3, THOC5, THOC6, THOC7) σχηματίζουν ένα υποσύμπλοκο (σύμπλοκο THO). Στην αρχή της συναρμογής και κατά τη μεταγραφή το TREX συνδέεται με το mRNA. Κατά το τέλος της συναρμογής το ετεροδιμερές σύμπλοκο NXF1-P15 συνδέεται με συστατικά του συμπλόκου TREX (ALYREF, THOC5) και μεταφέρει το mRNA στους πυρηνικούς πόρους. Εκεί, το NXF1-P15 συνδέεται με το μοτίβο φαινυλαλανίνης/γλυκίνης (FG) των νουκλεοπορινών (π.χ. NUP214, NUP358) και βοηθάει στη μεταφορά του mRNP στο κυτταρόπλασμα (Wickramasinghe & Laskey 2015). Το σύμπλοκο EJC αποτελείται από ένα βασικό πυρήνα τεσσάρων πρωτεϊνών (EIF4A3, MAGOH, RBM8A, CASC3), άλλες έξι πρωτεΐνες (RNPS1, ALYREF, ACIN1, PNN, SAP18, UPF3) που συνδέονται με τον βασικό πυρήνα και τρεις πρωτεΐνες (DDX39B, NXF1, NXT1) που συνδέονται παροδικά με αυτό το σύμπλοκο. Η συναρμογή του EJC γίνεται σταδιακά κατά τη διάρκεια της συναρμογής και ολοκληρώνεται στο υποσύμπλοκο C. Όπως παρατηρείται, περιλαμβάνει και παράγοντες που συμμετέχουν στην έξοδο του mRNA από τον πυρήνα (ALYREF, NXF1, NXT1). Πριν την έξοδο του mRNA από τον πυρήνα, δυο συστατικά (ACIN1, PNN) του EJC απομακρύνονται. Στο κυτταρόπλασμα, το mRNA μπορεί να ακολουθήσει δυο διαδρομές. Αν το mRNA είναι φυσιολογικό, το συστατικό του EJC, EIF4A3, στρατολογεί παράγοντες της μετάφρασης και έτσι ξεκινάει η μετάφραση. Αν το mRNA έχει πρόωρο κωδικόνιο λήξης, τότε ένα άλλο συστατικό του EJC, UPF3, ενεργοποιεί το μονοπάτι αποικοδόμησης μέσω της σύνδεσης του με πρωτεΐνες που συμμετέχουν σε αυτό (π.χ. UPF2) (Le Hir et al. 2016).



**Εικ. 85:** Πιθανός βιολογικός ρόλος της *FRA10AC1* ως πρωτεΐνη-προσαρμογέας μεταξύ της συναρμογής και της εξόδου του mRNA από τον πυρήνα ή της αποικοδόμησης μη κατάλληλων mRNAs. Φαίνεται η αναδιάταξη των συστατικών των συμπλόκων THO, TREX και EJC κατά τη διάρκεια της συναρμογής (υποσύμπλοκα A, B, *B act*, C και P). Με πράσινη, μωβ, μπλε και κόκκινη γραμματοσειρά φαίνονται κατ' αντιστοιχία οι γείτονες της *FRA10AC1* που είναι συστατικά του συμπλόκου THO, του πυρήνα του EJC, του EJC και του ετεροδιμερούς NXF1-NXT1. Οι υπόλοιποι γείτονες της *FRA10AC1* και του δικτύου που απεικονίζεται ακολουθούν το χρωματικό κώδικα της Εικόνας 84.

Το γεγονός ότι οι αλλαγές που προκάλεσε η καταστολή της έκφρασης του γονιδίου *FRA10AC1* στα επίπεδα έκφρασης μορίων αφορούν τουλάχιστον 1500 μετάγραφα και 500 πρωτεΐνες υποδεικνύει ότι επιτελεί βασική κυτταρική λειτουργία η οποία όμως δεν είναι πολύ γενικευμένη καθώς οι παραπάνω αριθμοί αντιστοιχούν στο ~10% του συνολικού αριθμού των γονιδίων του ανθρώπου. Μία γενική υπόθεση είναι η συνεργατική δράση της με άλλες πρωτεΐνες στο πλαίσιο διαδικασιών επεξεργασίας του mRNA, οι οποίες παύουν να είναι αποτελεσματικές όταν υπάρχει έλλειψή της, με συνέπεια κάποιο ποσοστό μεταγράφων και αντίστοιχων πρωτεϊνών να μην παράγονται στη σωστή ποσότητα. Επιπλέον, το ποσοστό των αλλαγών στα επίπεδα έκφρασης των μορίων εντός και εκτός του σωματίου συναρμογής κυμαίνεται στα ίδια επίπεδα (~10%). Ωστόσο, τόσο εντός όσο και εκτός του σωματίου συναρμογής, η έλλειψη της *FRA10AC1* οδηγεί σε αλλαγές στα επίπεδα έκφρασης μορίων τα οποία στην πλειοψηφία τους κατηγοριοποιούνται σε συγκεκριμένες λειτουργικές ομάδες (βλ. Ενότητα 4.4.1.). Έτσι, υποδεικνύεται ότι ομάδες λειτουργικά παρόμοιων μορίων συνεπηρεάζονται μετά την καταστολή της έκφρασης του γονιδίου *FRA10AC1*. Αυτό είναι αναμενόμενο λαμβάνοντας υπ' όψιν την δράση της *FRA10AC1* σε επίπεδο επεξεργασίας mRNA και τον συντονισμό της γονιδιακής έκφρασης, δηλαδή το συντονισμό-ρύθμιση ομάδων mRNAs που κωδικοποιούν λειτουργικά συσχετιζόμενες πρωτεΐνες, σε διαφορετικά επίπεδα μετα-μεταγραφικής ρύθμισης, όπως, στη συναρμογή, στην έξοδο από τον πυρήνα, στην σταθερότητα, στο καθορισμό της τοπολογίας και στην μετάφραση των μεταγράφων (Keene 2007, Morris et al. 2010).

Η συμμετοχή της *FRA10AC1* στη διαδικασία της συναρμογής θα μπορούσε να επιβεβαιωθεί με στοχευμένα βιοχημικά πειράματα ταυτοποίησης άμεσων αλληλεπιδράσεων με παράγοντες που συμμετέχουν στα καταλυτικά βήματα της συναρμογής. Επιπλέον, θα μπορούσε να γίνει *in vitro* δοκιμή συναρμογής ενός συνθετικού mRNA χρησιμοποιώντας το εκχύλισμα των κυττάρων στα οποία η *FRA10AC1* έχει κατασταλεί ώστε να διαπιστωθούν οι αλλαγές που συμβαίνουν στην συναρμογή. Ακόμη, για να ελεγχθούν τα μοριακά μονοπάτια που εμπλέκονται στην παθοφυσιολογία της απλοανεπάρκειας του *FRA10AC1* σε επίπεδο οργανισμού, απαραίτητη είναι η ανάπτυξη μοντέλου ποντικού με απενεργοποιημένο (knockout) το *Fra10Ac1*. Ένα τέτοιο μοντέλο έχει δημιουργηθεί στο πλαίσιο της Υποδομής INFRAFRONTIER.GR (<http://www.infrafrontier.gr/>, επικεφαλής Δρ. Δ. Κοντογιάννης, Ερευνητικό Κέντρο Βιοϊατρικών Επιστημών "Αλέξανδρος Φλέμινγκ") από εμβρυικά βλαστοκύτταρα (ES) ποντικού, στελέχους C57BL6/N, τα οποία προήλθαν από το EUCOMM (European Conditional Mouse Mutagenesis Program, <http://www.mousephenotype.org>). Σε αυτήν τη φάση έχει δημιουργηθεί το πραγματικό knockout *tm1b* (συνεργασία με Δρ. Α. Κλινάκη, Ίδρυμα Ιατροβιολογικών Ερευνών, Ακαδημίας Αθηνών, ΙΒΕΑΑ). Το ζωικό μοντέλο θα αξιοποιηθεί σε αναλύσεις συμπεριφοράς ώστε να ελεγχθεί η συσχέτιση της απλοανεπάρκειας του *FRA10AC1* με τα συμπτώματα που έχουν αναφερθεί στους φορείς της εύθραυστης θέσης *FRA10A*. Επιπλέον θα αποτελέσει ένα χρήσιμο εργαλείο για την εξέταση των επιδράσεων, μέσω της χρήσης ομικών αναλύσεων, που επιφέρει η απενεργοποίηση του γονιδίου *FRA10AC1* σε επίπεδο οργανισμού αλλά και εστιασμένα σε συγκεκριμένους ιστούς με



προτεραιότητα τον εγκέφαλο. Ο συνδυασμός λοιπόν τόσο των αποτελεσμάτων της παρούσας εργασίας όσο και των αναλύσεων του ζωικού μοντέλου ποντικού που θα ακολουθήσουν θα συνεισφέρει αφενός στην αποσαφήνιση της βιολογικής λειτουργίας της *FRA10AC1* και αφετέρου στην πιθανή συσχέτιση της απλοανεπάρκειας του *FRA10AC1* με τον συγκεκριμένο φαινότυπο.

## **SUMMARY**

The *FRA10AC1* gene is mapped in the rare fragile site *FRA10A*, which is induced in the absence of folic acid in the cell culture medium, in the human chromosomal region 10q23.3. *FRA10A* is the most prevalent among the rare folate-sensitive autosomal fragile sites in the human genome, and its expression is estimated about 1 in 500 individuals. *FRA10A* carriers exhibit mental and developmental disabilities, such as dysmorphic features, short stature, hypospadias and speech difficulty. The molecular basis of its cytogenetic expression is the expansion (~200 repeat units) of trinucleotide repeats of the type (CGG)<sub>n</sub> which are located in the 5' untranslated region of the *FRA10AC1* gene. The expansion results in hypermethylation of this region results in and the transcriptional repression of the corresponding allele.

The *FRA10AC1* gene is transcriptionally active in all adult tissues, exhibiting higher levels of expression in organs with high transcriptional activity such as brain, heart, skeletal muscles and liver. The major transcript of the *FRA10AC1* gene encodes a protein of 315aa, which exhibits exclusively nuclear topology. FRA10AC1 is a conserved protein as it presents orthologs in a number of eukaryotic, multicellular or unicellular organisms, but not in prokaryotic organisms. The FRA10AC1 protein has been repeatedly identified as a component of the major spliceosome and its subcomplexes, B *act* (activated), C and P. Also, genetic experiments in *Chlamydomonas reinhardtii* indicate that FRA10AC1 contributes in splice site recognition, and, its protein-protein interactions with spliceosomal components (DGCR14, SF3B2) were identified by different biochemical methods. Based on the data mentioned above, it is indicated that FRA10AC1 involves in mRNA processing.

To understand the functional role of FRA10AC1, in a first place, its protein-protein interaction network was constructed, and, then, the protein-protein interaction network of whole spliceosome. PICKLE meta-database, DroID and Worm Interactome Database, which include protein-protein interaction data in human and in model organisms, *D. melanogaster* and *C. elegans*, respectively, were used for network reconstruction. Only direct protein-protein interactions were retrieved from these databases. A key criterion for the integration of one interaction in the protein-protein interaction network of spliceosome was the biochemical identification of proteins as components of the spliceosome and its subcomplexes. The analysis of the network topological parameters performed by computational methods, and gene ontology analysis of the FRA10AC1 broader 'neighborhood' within the spliceosome network contributed to imply possible biological function of FRA10AC1. At the same time, a *HeLa* cell model with stable suppressed expression of the *FRA10AC1* gene was developed, through the use of shRNA clones which target the coding region of the endogenous *FRA10AC1* gene. In this cell model, omic approaches were applied on three basic levels of cellular function (transcriptomics, proteomics, metabolomics) in order to investigate the effects of suppression of *FRA10AC1* gene expression in cell physiology. The study of these changes in the context of the protein-protein interactions of the spliceosome, but also beyond this context, also contributed to the construction of proposed FRA10AC1 functional models within spliceosome.

## **ΒΙΒΛΙΟΓΡΑΦΙΑ**

- Abelson, J. (2008). Is the spliceosome a ribonucleoprotein enzyme? *Nature Structural and Molecular Biology*.  
<https://doi.org/10.1038/nsmb1208-1235>
- Acencio, M. L., & Lemke, N. (2009). Towards the prediction of essential genes by integration of network topology, cellular localization and biological process information. *BMC Bioinformatics*, 10, 290. <https://doi.org/10.1186/1471-2105-10-290>
- Adikesavan, A. K., Katsonis, P., Marciano, D. C., Lua, R., Herman, C., & Lichtarge, O. (2011). Separation of Recombination and SOS Response in Escherichia coli RecA Suggests LexA Interaction Sites. *PLoS Genetics*, 7(9), e1002244.  
<https://doi.org/10.1371/journal.pgen.1002244>
- Adrian, J., Chang, J., Ballenger, C. E., Bargmann, B. O. R., Alassimone, J., Davies, K. A., ... Bergmann, D. C. (2015). Transcriptome dynamics of the stomatal lineage: Birth, amplification, and termination of a self-renewing population. *Developmental Cell*, 33(1), 107–119. <https://doi.org/10.1016/j.devcel.2015.01.025>
- Aebersold, R., & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, 422(6928), 198–207.  
<https://doi.org/10.1038/nature01511>
- Agafonov, D. E., Deckert, J., Wolf, E., Odenwalder, P., Bessonov, S., Will, C. L., ... Luhmann, R. (2011). Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Molecular and Cellular Biology*, 31(13), 2667–2682. <https://doi.org/10.1128/MCB.05266-11>
- Aggarwal, K., Choe, L. H., & Lee, K. H. (2006). Shotgun proteomics using the iTRAQ isobaric tags. *Briefings in Functional Genomics and Proteomics*. <https://doi.org/10.1093/bfpp/ell018>
- Agrawal, A. A., Salsi, E., Chatrikhi, R., Henderson, S., Jenkins, J. L., Green, M. R., ... Kielkopf, C. L. (2016). An extended U2AF65-RNA-binding domain recognizes the 3' splice site signal. *Nature Communications*, 7.  
<https://doi.org/10.1038/ncomms10950>
- Ahmed, A., & Xing, E. P. (2009). Recovering time-varying networks of dependencies in social and biological studies. *Proceedings of the National Academy of Sciences*, 106(29), 11878–11883. <https://doi.org/10.1073/pnas.0901910106>
- Ahmed, S. S. S. J., Ahameethunisa, A. R., Santosh, W., Chakravarthy, S., & Kumar, S. (2011). Systems biological approach on neurological disorders: A novel molecular connectivity to aging and psychiatric diseases. *BMC Systems Biology*, 5.  
<https://doi.org/10.1186/1752-0509-5-6>
- Akerman, M., Fregoso, O. I., Das, S., Ruse, C., Jensen, M. A., Pappin, D. J., ... Krainer, A. R. (2015). Differential connectivity of splicing activators and repressors to the human spliceosome. *Genome Biology*, 16(1). <https://doi.org/10.1186/s13059-015-0682-5>
- Al-Ayoubi, A. M., Zheng, H., Liu, Y., Bai, T., & Eblen, S. T. (2012). Mitogen-Activated Protein Kinase Phosphorylation of Splicing Factor 45 (SPF45) Regulates SPF45 Alternative Splicing Site Utilization, Proliferation, and Cell Adhesion. *Molecular and Cellular Biology*, 32(14), 2880–2893. <https://doi.org/10.1128/MCB.06327-11>
- Albert, R., Jeong, H., & Barabasi, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, 406: 378–482, 2000. *Nature*, 406(6794), 378–382. <https://doi.org/10.1038/35019019>
- Albertazzi, L., Arosio, D., Marchetti, L., Ricci, F., & Beltram, F. (2009). Quantitative FRET analysis with the E0GFP-mCherry fluorescent protein pair. *Photochemistry and Photobiology*, 85(1), 287–297. <https://doi.org/10.1111/j.1751-1097.2008.00435.x>
- Alberts, B. (1998). The cell as a collection of protein machines: Preparing the next generation of molecular biologists. *Cell*. [https://doi.org/10.1016/S0092-8674\(00\)80922-8](https://doi.org/10.1016/S0092-8674(00)80922-8)
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular Biology of the Cell, 4th edition*. Garland Science. <https://doi.org/10.3389/fimmu.2015.00171>
- Alioto, T. S. (2007). U12DB: A database of orthologous U12-type spliceosomal introns. *Nucleic Acids Research*, 35(SUPPL. 1). <https://doi.org/10.1093/nar/gkl796>
- Allemand, E., Myers, M. P., Garcia-Bernardo, J., Harel-Bellan, A., Krainer, A. R., & Muchardt, C. (2016). A Broad Set of Chromatin Factors Influences Splicing. *PLoS Genetics*, 12(9). <https://doi.org/10.1371/journal.pgen.1006318>
- Almaas, E. (2007). Biological impacts and context of network theory. *Journal of Experimental Biology*, 210(9), 1548–1558. <https://doi.org/10.1242/jeb.003731>
- Alpert, T., Herzl, L., & Neugebauer, K. M. (2017). Perfect timing: splicing and transcription rates in living cells. *Wiley Interdisciplinary Reviews: RNA*. <https://doi.org/10.1002/wrna.1401>
- Altman, B. J., & Rathmell, J. C. (2012). Metabolic stress in autophagy and cell death pathways. *Cold Spring Harbor Perspectives in Biology*. <https://doi.org/10.1101/cshperspect.a008763>
- Altman, B. J., Stine, Z. E., & Dang, C. V. (2016). From Krebs to clinic: Glutamine metabolism to cancer therapy. *Nature Reviews Cancer*. <https://doi.org/10.1038/nrc.2016.71>
- Amberger, J. S., & Hamosh, A. (2017). Searching online mendelian inheritance in man (OMIM): A knowledgebase of human genes and genetic phenotypes. *Current Protocols in Bioinformatics*, 2017, 1.2.1-1.2.12. <https://doi.org/10.1002/cpbi.27>
- Andersen, D. S., & Tapon, N. (2008). Drosophila MFAP1 is required for Pre-mRNA processing and G2/M progression. *Journal of Biological Chemistry*, 283(45), 31256–31267. <https://doi.org/10.1074/jbc.M803512200>
- Andersen, G. R., Valente, L., Pedersen, L., Kinzy, T. G., & Nyborg, J. (2001). Crystal structures of nucleotide exchange intermediates in the eEF1A-eEF1B $\alpha$  complex. *Nature Structural Biology*, 8(6), 531–534. <https://doi.org/10.1038/88598>
- Anderson, N. G., & Anderson, N. L. (1996). Twenty years of two-dimensional electrophoresis: Past, present and future. *Electrophoresis*, 17(3), 443–453. <https://doi.org/10.1002/elps.1150170303>
- Antrobus, R., & Borner, G. H. H. (2011). Improved elution conditions for native co-immunoprecipitation. *PLoS ONE*, 6(3). <https://doi.org/10.1371/journal.pone.0018218>
- Arnoldo, A., Curak, J., Kittanakom, S., Chevelev, I., Lee, V. T., Sahebol-Amri, M., ... Stagljar, I. (2008). Identification of Small Molecule Inhibitors of Pseudomonas aeruginosa Exoenzyme S Using a Yeast Phenotypic Screen. *PLoS Genetics*, 4(2), e1000005. <https://doi.org/10.1371/journal.pgen.1000005>
- Arshavsky, V. Y. (2002). Like night and day: Rods and cones have different pigment regeneration pathways. *Neuron*. [https://doi.org/10.1016/S0896-6273\(02\)00937-6](https://doi.org/10.1016/S0896-6273(02)00937-6)
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*. <https://doi.org/10.1038/75556>

- Asplund, A., Edqvist, P. H. D., Schwenk, J. M., & Pontén, F. (2012). Antibodies for profiling the human proteome-The Human Protein Atlas as a resource for cancer research. *Proteomics*, *12*(13), 2067–2077. <https://doi.org/10.1002/pmic.201100504>
- Assenov, Y., Ramirez, F., Schelhorn, S. E. S. E., Lengauer, T., & Albrecht, M. (2008). Computing topological parameters of biological networks. *Bioinformatics*, *24*(2), 282–284. <https://doi.org/10.1093/bioinformatics/btm554>
- Aubol, B. E., Wu, G., Keshwani, M. M., Movassat, M., Fattet, L., Hertel, K. J., ... Adams, J. A. (2016). Release of SR Proteins from CLK1 by SRPK1: A Symbiotic Kinase System for Phosphorylation Control of Pre-mRNA Splicing. *Molecular Cell*, *63*(2), 218–228. <https://doi.org/10.1016/j.molcel.2016.05.034>
- Bagni, C., Tassone, F., Neri, G., & Hagerman, R. (2012). Fragile X syndrome: Causes, diagnosis, mechanisms, and therapeutics. *Journal of Clinical Investigation*. <https://doi.org/10.1172/JCI63141>
- Bai, B., Hales, C. M., Chen, P.-C., Gozal, Y., Dammer, E. B., Fritz, J. J., ... Peng, J. (2013). U1 small nuclear ribonucleoprotein complex and RNA splicing alterations in Alzheimer's disease. *Proceedings of the National Academy of Sciences*, *110*(41), 16562–16567. <https://doi.org/10.1073/pnas.1310249110>
- Bai, S. W., Herrera-Abreu, M. T., Rohn, J. L., Racine, V., Tajadura, V., Suryavanshi, N., ... Ridley, A. J. (2011). Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration. *BMC Biology*, *9*. <https://doi.org/10.1186/1741-7007-9-54>
- Bakhach, J. (2009). The cryopreservation of composite tissues: Principles and recent advancement on cryopreservation of different type of tissues. *Organogenesis*. <https://doi.org/10.4161/org.5.3.9583>
- Barabási, A.-L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, *12*(1), 56–68. <https://doi.org/10.1038/nrg2918>
- Barabási, A.-L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, *5*(2), 101–113. <https://doi.org/10.1038/nrg1272>
- Barabási, A. L. (2009). Scale-free networks: A decade and beyond. *Science*. <https://doi.org/10.1126/science.1173299>
- Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512. <https://doi.org/10.1126/science.286.5439.509>
- Baranzini, S. E. (2009). The genetics of autoimmune diseases: a networked perspective. *Current Opinion in Immunology*. <https://doi.org/10.1016/j.coi.2009.09.014>
- Barash, Y., Calarco, J. A., Gao, W., Pan, Q., Wang, X., Shai, O., ... Frey, B. J. (2010). Deciphering the splicing code. *Nature*, *465*(7294), 53–59. <https://doi.org/10.1038/nature09000>
- Barbosa, I., Haque, N., Fiorini, F., Barrandon, C., Tomasetto, C., Blanchette, M., & Le Hir, H. (2012). Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly. *Nature Structural and Molecular Biology*, *19*(10), 983–991. <https://doi.org/10.1038/nsmb.2380>
- Barrell, D., Dimmer, E., Huntley, R. P., Binns, D., O'Donovan, C., & Apweiler, R. (2009). The GOA database in 2009 - An integrated Gene Ontology Annotation resource. *Nucleic Acids Research*, *37*(SUPPL. 1). <https://doi.org/10.1093/nar/gkn803>
- Barrios-Rodiles, M., Brown, K. R., Ozdamar, B., Bose, R., Liu, Z., Donovan, R. S., ... Wrana, J. L. (2005). High-throughput mapping of a dynamic signaling network in mammalian cells. *Science*, *307*(5715), 1621–1625. <https://doi.org/10.1126/science.1105776>
- Bartkova, J., Rezaei, N., Liontos, M., Karakaidos, P., Kletsas, D., Issaeva, N., ... Gorgoulis, V. G. (2006). Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature*, *444*(7119), 633–637. <https://doi.org/10.1038/nature05268>
- Bastiaens, P. I., & Squire, A. (1999). Fluorescence lifetime imaging microscopy: spatial resolution of biochemical processes in the cell. *Trends Cell Biol*, *9*(2), 48–52. [https://doi.org/S0962-8924\(98\)01410-X \[pii\]](https://doi.org/S0962-8924(98)01410-X [pii])
- Bauer-Mehren, A., Bundschuh, M., Rautschka, M., Mayer, M. A., Sanz, F., & Furlong, L. I. (2011). Gene-disease network analysis reveals functional modules in mendelian, complex and environmental diseases. *PLoS ONE*, *6*(6). <https://doi.org/10.1371/journal.pone.0020284>
- Bayley, J.-P., & Devilee, P. (2012). The Warburg effect in 2012. *Current Opinion in Oncology*, *24*(1), 62–67. <https://doi.org/10.1097/CCO.0b013e32834deb9e>
- Becerra, S., Montes, M., Hernández-Munain, C., & Suñé, C. (2015). Prp40 pre-mRNA processing factor 40 homolog B ( PRPF40B ) associates with SF1 and U2AF 65 and modulates alternative pre-mRNA splicing in vivo. *Rna*, *21*, 1–20. <https://doi.org/10.1261/ma.047258.114.3>
- Beggs, J. D. (2005). Lsm proteins and RNA processing <sup>1</sup>. *Biochemical Society Transactions*, *33*(3), 433–438. <https://doi.org/10.1042/BST0330433>
- Behzadnia, N., Golas, M. M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., ... Lüthmann, R. (2007). Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO Journal*, *26*(6), 1737–1748. <https://doi.org/10.1038/sj.emboj.7601631>
- Bellare, P., Small, E. C., Huang, X., Wohlschlegel, J. A., Staley, J. P., & Sontheimer, E. J. (2008). A role for ubiquitin in the spliceosome assembly pathway. *Nature Structural and Molecular Biology*, *15*(5), 444–451. <https://doi.org/10.1038/nsmb.1401>
- Beloribi-Djefailia, S., Vasseur, S., & Guillaume, F. (2016). Lipid metabolic reprogramming in cancer cells. *Oncogenesis*, *5*(1), e189. <https://doi.org/10.1038/oncsis.2015.49>
- Benjamin, D. I., Cravatt, B. F., & Nomura, D. K. (2012). Global profiling strategies for mapping dysregulated metabolic pathways in cancer. *Cell Metabolism*. <https://doi.org/10.1016/j.cmet.2012.09.013>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*. <https://doi.org/10.2307/2346101>
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Wheeler, D. L. (2007). GenBank. *Nucleic Acids Research*, *36*(Database), D25–D30. <https://doi.org/10.1093/nar/gkm929>
- Bentley, D. L. (2005). Rules of engagement: Co-transcriptional recruitment of pre-mRNA processing factors. *Current Opinion in Cell Biology*. <https://doi.org/10.1016/j.ceb.2005.04.006>
- Berger, B., Peng, J., & Singh, M. (2013). Computational solutions for omics data. *Nature Reviews Genetics*. <https://doi.org/10.1038/nrg3433>
- Berman, D. E., Ringe, D., Petsko, G. A., & Small, S. A. (2015). The Use of Pharmacological Retromer Chaperones in Alzheimer's Disease and other Endosomal-related Disorders. *Neurotherapeutics*. <https://doi.org/10.1007/s13311-014-0321-y>
- Bessonov, S., Anokhina, M., Krasauskas, A., Golas, M. M., Sander, B., Will, C. L., ... Lüthmann, R. (2010). Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome

- activation and first step catalysis. *RNA (New York, N.Y.)*, 16(12), 2384–2403. <https://doi.org/10.1261/rna.2456210>
- Bessonov, S., Anokhina, M., Will, C. L., Urlaub, H., & Lührmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. *Nature*, 452(7189), 846–850. <https://doi.org/10.1038/nature06842>
- Bian, Y., Ye, M., Wang, C., Cheng, K., Song, C., Dong, M., ... Zou, H. (2013). Global screening of CK2 kinase substrates by an integrated phosphoproteomics workflow. *Scientific Reports*, 3. <https://doi.org/10.1038/srep03460>
- Biesecker, L. G. (2013). Hypothesis-generating research and predictive medicine. *Genome Research*, 23(7), 1051–1053. <https://doi.org/10.1101/gr.157826.113>
- Bird, I. M. (1989). High performance liquid chromatography: principles and clinical applications. *BMJ (Clinical Research Ed.)*, 299(6702), 783–787. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2508923>
- Blakeley, P., Siepen, J. A., Lawless, C., & Hubbard, S. J. (2010). Investigating protein isoforms via proteomics: A feasibility study. *Proteomics*, 10(6), 1127–1140. <https://doi.org/10.1002/pmic.200900445>
- Bloch, K. D., & Grossmann, B. (2001). Digestion of DNA with Restriction Endonucleases. In *Current Protocols in Molecular Biology* (Vol. Chapter 3, p. Unit3.1). Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/0471142727.mb0301s31>
- Bodenreider, O. (2008). Biomedical ontologies in action: role in knowledge management, data integration and decision support. *Yearbook of Medical Informatics*, 3841, 67–79. <https://doi.org/me08010067> [pii]
- Bomszyk, K., Denisenko, O., & Ostrowski, J. (2004). hnRNP K: One protein multiple processes. *BioEssays*. <https://doi.org/10.1002/bies.20048>
- Borner, G. H. H., Hein, M. Y., Hirst, J., Edgar, J. R., Mann, M., & Robinson, M. S. (2014). Fractionation profiling: a fast and versatile approach for mapping vesicle proteomes and protein-protein interactions. *Molecular Biology of the Cell*, 25(20), 3178–3194. <https://doi.org/10.1091/mbc.E14-07-1198>
- Bortfeldt, R. H., Schuster, S., & Koch, I. (2010). Exhaustive analysis of the modular structure of the spliceosomal assembly network: A Petri net approach. *In Silico Biology*, 10(1–2), 89–123. <https://doi.org/10.3233/ISB-2010-0419>
- Boudrez, A., Beullens, M., Waelkens, E., Stalmans, W., & Bollen, M. (2002). Phosphorylation-dependent interaction between the splicing factors SAP155 and NIPP1. *Journal of Biological Chemistry*, 277(35), 31834–31841. <https://doi.org/10.1074/jbc.M204427200>
- Boyce, A. E., McGrath, J. A., Techanukul, T., Murrell, D. F., Chow, C. W., McGregor, L., & Warren, L. J. (2012). Ectodermal dysplasia-skin fragility syndrome due to a new homozygous internal deletion mutation in the PKP1 gene. *Australasian Journal of Dermatology*, 53(1), 61–65. <https://doi.org/10.1111/j.1440-0960.2011.00846.x>
- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72(1–2), 248–254. [https://doi.org/10.1016/0003-2697\(76\)90527-3](https://doi.org/10.1016/0003-2697(76)90527-3)
- Branzei, D., & Foiani, M. (2010). Maintaining genome stability at the replication fork. *Nature Reviews Molecular Cell Biology*. <https://doi.org/10.1038/nrm2852>
- Braunschweig, U., Gueroussou, S., Plocik, A. M., Graveley, B. R., & Blencowe, B. J. (2013). Dynamic Integration of Splicing within Gene Regulatory Pathways. *Cell*, 152(6), 1252–1269. <https://doi.org/10.1016/J.CELL.2013.02.034>
- Brennan, D., Hu, Y., Kljuic, A., Choi, Y. W., Joubert, S., Bashkin, M., ... Mahoney, M. G. (2004). Differential structural properties and expression patterns suggest functional significance for multiple mouse desmoglein 1 isoforms. *Differentiation*, 72(8), 434–449. <https://doi.org/10.1111/j.1432-0436.2004.07208009.x>
- Bressan, G. C., Quaresma, A. J. C., Moraes, E. C., Manfiolli, A. O., Passos, D. O., Gomes, M. D., & Kobarg, J. (2009). Functional association of human Ki-1/57 with pre-mRNA splicing events. *The FEBS Journal*, 276(14), 3770–3783. <https://doi.org/10.1111/j.1742-4658.2009.07092.x>
- Brown, G. R., Hem, V., Katz, K. S., Ovetsky, M., Wallin, C., Ermolaeva, O., ... Murphy, T. D. (2015). Gene: A gene-centered information resource at NCBI. *Nucleic Acids Research*, 43(D1), D36–D42. <https://doi.org/10.1093/nar/gku1055>
- Brückner, A., Polge, C., Lentze, N., Auerbach, D., & Schlattner, U. (2009). Yeast two-hybrid, a powerful tool for systems biology. *International Journal of Molecular Sciences*. <https://doi.org/10.3390/ijms10062763>
- Brunner, H. G., & Van Driel, M. A. (2004). From syndrome families to functional genomics. *Nature Reviews Genetics*. <https://doi.org/10.1038/nrg1383>
- Brzyżek, G., & Świeżewski, S. (2015). Mutual interdependence of splicing and transcription elongation. *Transcription*, 6(2), 37–39. <https://doi.org/10.1080/21541264.2015.1040146>
- Buck, C. A., & Horwitz, A. F. (1987). Integrin, a transmembrane glycoprotein complex mediating cell-substratum adhesion. *Journal of Cell Science. Supplement*, 8, 231–250. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/3332662>
- Bullock, N., & Oltean, S. (2017). The many faces of SRPK1. *Journal of Pathology*. <https://doi.org/10.1002/path.4846>
- Burge, C. B., Padgett, R. A., & Sharp, P. A. (1998). Evolutionary fates and origins of U12-type introns. *Molecular Cell*, 2(6), 773–785. [https://doi.org/10.1016/S1097-2765\(00\)80292-0](https://doi.org/10.1016/S1097-2765(00)80292-0)
- Burgess, D. J. (2016). A space for transcriptomics. *Nature Reviews Genetics*, 17(8), 436–437. <https://doi.org/10.1038/nrg.2016.94>
- Busch, A., & Hertel, K. J. (2012). Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdisciplinary Reviews: RNA*. <https://doi.org/10.1002/wrna.100>
- Cabral, R. M., Wan, H., Cole, C. L., Abrams, D. J., Kelsell, D. P., & South, A. P. (2010). Identification and characterization of DSP1a, a novel isoform of human desmoplakin. *Cell and Tissue Research*, 341(1), 121–129. <https://doi.org/10.1007/s00441-010-0989-1>
- Calarco, J. A., Superina, S., O'Hanlon, D., Gabut, M., Raj, B., Pan, Q., ... Blencowe, B. J. (2009). Regulation of Vertebrate Nervous System Alternative Splicing and Development by an SR-Related Protein. *Cell*, 138(5), 898–910. <https://doi.org/10.1016/j.cell.2009.06.012>
- Calderaro, A., Arcangeletti, M.-C., Rodighiero, I., Buttrini, M., Gorrini, C., Motta, F., ... De Conto, F. (2015). Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry applied to virus identification. *Scientific Reports*, 4(1), 6803. <https://doi.org/10.1038/srep06803>
- Callahan, K. P., & Butler, J. S. (2010). TRAMP complex enhances RNA degradation by the nuclear exosome component Rrp6. *Journal of Biological Chemistry*, 285(6), 3540–3547. <https://doi.org/10.1074/jbc.M109.058396>
- Cannistraci, C. V., Alanis-Lobato, G., & Ravasi, T. (2013). Minimum curvilinearity to enhance topological prediction of protein interactions by network embedding. In *Bioinformatics* (Vol. 29). <https://doi.org/10.1093/bioinformatics/btt208>
- Carlsson, N., Borde, A., Wölfel, S., Kerman, B., & Larsson, A. (2011). Quantification of protein concentration by the Bradford method in the presence of pharmaceutical polymers. *Analytical Biochemistry*, 411(1), 116–121. <https://doi.org/10.1016/j.ab.2010.12.026>

- Caro-Maldonado, A. (2011). Dying for Something to Eat: How Cells Respond to Starvation. *The Open Cell Signaling Journal*, 3(1), 42–51. <https://doi.org/10.2174/1876390101103010042>
- Carpenter, A. E., & Sabatini, D. M. (2004). Systematic genome-wide screens of gene function. *Nature Reviews Genetics*, 5(1), 11–22. <https://doi.org/10.1038/nrg1248>
- Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., ... Hentze, M. W. (2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell*, 149(6), 1393–1406. <https://doi.org/10.1016/j.cell.2012.04.031>
- Cha, R. S., & Kleckner, N. (2002). ATR homolog Mec1 promotes fork progression, thus averting breaks in replication slow zones. *Science*, 297(5581), 602–606. <https://doi.org/10.1126/science.1071398>
- Chalifa-Caspi, V., Yanai, I., Ophir, R., Rosen, N., Shmoish, M., Benjamin-Rodrig, H., ... Lancet, D. (2004). GeneAnnot: comprehensive two-way Chalifa-Caspi, V., Yanai, I., Ophir, R., Rosen, N., Shmoish, M., Benjamin-Rodrig, H., ... Lancet, D. (2004). GeneAnnot: comprehensive two-way linking between oligonucleotide array probesets and GeneCards genes. *Bioinformatics (Oxford, England)*, 20(9), 1457–1458. <https://doi.org/10.1093/bioinformatics/bth081>
- Champion-Arnaud, P., & Reed, R. (1994). The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. *Genes and Development*, 8(16), 1974–1983. <https://doi.org/10.1101/gad.8.16.1974>
- Chan, S. P., & Cheng, S. C. (2005). The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. *Journal of Biological Chemistry*, 280(35), 31190–31199. <https://doi.org/10.1074/jbc.M505060200>
- Chan, S. Y., & Loscalzo, J. (2012). The emerging paradigm of network medicine in the study of human disease. *Circulation Research*. <https://doi.org/10.1161/CIRCRESAHA.111.258541>
- Chanarat, S., & Str aber, K. (2013). Splicing and beyond: The many faces of the Prp19 complex. *Biochimica et Biophysica Acta - Molecular Cell Research*. <https://doi.org/10.1016/j.bbamcr.2013.05.023>
- Chatr-Aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N. K., ... Tyers, M. (2017). The BioGRID interaction database: 2017 update. *Nucleic Acids Research*, 45(D1), D369–D379. <https://doi.org/10.1093/nar/gkw1102>
- Chaudhry, M. A., Chodosh, L. A., McKenna, W. G., & Muschel, R. J. (2002). Gene expression profiling of HeLa cells in G1 or G2 phases. *Oncogene*, 21(12), 1934–1942. <https://doi.org/10.1038/sj/onc/1205264>
- Chavez, J. D., Eng, J. K., Schweppe, D. K., Cilia, M., Rivera, K., Zhong, X., ... Bruce, J. E. (2016). A General Method for Targeted Quantitative Cross-Linking Mass Spectrometry. *PLOS ONE*, 11(12), e0167547. <https://doi.org/10.1371/journal.pone.0167547>
- Chazal, P.-E., Daguene, E., Wendling, C., Ulryck, N., Tomasetto, C., Sargueil, B., & Le Hir, H. (2013). EJC core component MLN51 interacts with eIF3 and activates translation. *Proceedings of the National Academy of Sciences*, 110(15), 5903–5908. <https://doi.org/10.1073/pnas.1218732110>
- Chen, B., Fan, W., Liu, J., & Wu, F. X. (2014). Identifying protein complexes and functional modules-from static PPI networks to dynamic PPI networks. *Briefings in Bioinformatics*, 15(2), 177–194. <https://doi.org/10.1093/bib/bbt039>
- Chen, D., Li, Z., Yang, Q., Zhang, J., Zhai, Z., & Shu, H. B. (2003). Identification of a nuclear protein that promotes NF-kappaB activation. *Biochem Biophys Res Commun*, 310(3), 720–724. <https://doi.org/S0006291X03018758> [pii]
- Chen, H., Gao, Y., Su, Y., & Lv, Y. (2008). Photo-induced chemiluminescent method for determination of reducing sugars. *Luminescence*, 23(5), 287–291. <https://doi.org/10.1002/bio.1030>
- Chen, H., Puhl, H. L., Koushik, S. V., Vogel, S. S., & Ikeda, S. R. (2006). Measurement of FRET Efficiency and Ratio of Donor to Acceptor Concentration in Living Cells. *Biophysical Journal*, 91(5). <https://doi.org/10.1529/biophysj.106.088773>
- Chen, J., Li, Y. M., Li, J. F., Liu, L. D., Liao, Y., Na, R. X., ... Li, Q. H. (2010). Transcriptional regulation by HSV-1 induced HTRP via acetylation system. *Virologica Sinica*, 25(6), 417–424. <https://doi.org/10.1007/s12250-010-3147-8>
- Chen, M., Zhang, J., & Manley, J. L. (2010). Turning on a fuel switch of cancer: hnRNP proteins regulate alternative splicing of pyruvate kinase mRNA. *Cancer Research*, 70(22), 8977–8980. <https://doi.org/10.1158/0008-5472.CAN-10-2513>
- Chen, W., & Moore, M. J. (2014). The spliceosome: Disorder and dynamics defined. *Current Opinion in Structural Biology*. <https://doi.org/10.1016/j.sbi.2014.01.009>
- Chen, Y. I. G., Moore, R. E., Ge, H. Y., Young, M. K., Lee, T. D., & Stevens, S. W. (2007). Proteomic analysis of in vivo-assembled pre-mRNA splicing complexes expands the catalog of participating factors. *Nucleic Acids Research*, 35(12), 3928–3944. <https://doi.org/10.1093/nar/gkm347>
- Cheng, C., & Pounds, S. (2007). False discovery rate paradigms for statistical analyses of microarray gene expression data. *Bioinformatics*, 1(10), 436–446. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17597936>
- Cheng, H., Dufu, K., Lee, C. S., Hsu, J. L., Dias, A., & Reed, R. (2006). Human mRNA Export Machinery Recruited to the 5' End of mRNA. *Cell*, 127(7), 1389–1400. <https://doi.org/10.1016/j.cell.2006.10.044>
- Cheng, N. H., Zhang, W., Chen, W. Q., Jin, J., Cui, X., Butte, N. F., ... Hirschi, K. D. (2011). A mammalian monothiol glutaredoxin, Grx3, is critical for cell cycle progression during embryogenesis. *FEBS Journal*, 278(14), 2525–2539. <https://doi.org/10.1111/j.1742-4658.2011.08178.x>
- Cherbas, L., Moss, R., & Cherbas, P. (1994). Transformation techniques for Drosophila cell lines. *Methods in Cell Biology*, 44, 161–179. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7707950>
- Chernushevich, I. V., Loboda, A. V., & Thomson, B. A. (2001). An introduction to quadrupole-time-of-flight mass spectrometry. *Journal of Mass Spectrometry*, 36(8), 849–865. <https://doi.org/10.1002/jms.207>
- Chiarugi, A., Dolle, C., Felici, R., Ziegler, M., D lle, C., Felici, R., ... Ziegler, M. (2012). The NAD metabolome--a key determinant of cancer cell biology. *Nature Reviews. Cancer*, 12(11), 741–752. <https://doi.org/10.1038/nrc3340>
- Chinwalla, A. T., Cook, L. L., Delehaunty, K. D., Fewell, G. A., Fulton, L. A., Fulton, R. S., ... Zody, M. C. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915), 520–562. <https://doi.org/10.1038/nature01262>
- Cho, H., Berger, B., & Peng, J. (2016). Compact Integration of Multi-Network Topology for Functional Analysis of Genes. *Cell Systems*, 3(6), 540–548.e5. <https://doi.org/10.1016/j.cels.2016.10.017>
- Cho, I., Tsai, P. F., Lake, R. J., Basheer, A., & Fan, H. Y. (2013). ATP-Dependent Chromatin Remodeling by Cockayne Syndrome Protein B and NAPI-Like Histone Chaperones Is Required for Efficient Transcription-Coupled DNA Repair. *PLoS Genetics*, 9(4). <https://doi.org/10.1371/journal.pgen.1003407>
- Chu, C., Zhang, Q. C., Da Rocha, S. T., Flynn, R. A., Bharadwaj, M., Calabrese, J. M., ... Chang, H. Y. (2015). Systematic discovery of Xist RNA binding proteins. *Cell*, 161(2), 404–416. <https://doi.org/10.1016/j.cell.2015.03.025>
- Chua, H. N., Sung, W. K., & Wong, L. (2006). Exploiting indirect neighbours and topological weight to predict protein function

- from protein-protein interactions. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 3916 LNBI, p. 1). [https://doi.org/10.1007/11691730\\_1](https://doi.org/10.1007/11691730_1)
- Chuang, T. W., Lee, K. M., & Tarn, W. Y. (2015). Function and pathological implications of exon junction complex factor Y14. *Biomolecules*. <https://doi.org/10.3390/biom5020343>
- Clark, T. A., Sugnet, C. W., & Ares, M. (2002). Genomewide Analysis of mRNA Processing in Yeast Using Splicing-Specific Microarrays. *Science*, 296(5569), 907–910. <https://doi.org/10.1126/science.1069415>
- Claverie, J. M. (2001). Gene number. What if there are only 30,000 human genes? *Science (New York, N.Y.)*, 291(5507), 1255–1257. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11233450>
- Clegg, R. M. (1995). Fluorescence resonance energy transfer. *Current Opinion in Biotechnology*, 6(1), 103–110. [https://doi.org/10.1016/0958-1669\(95\)80016-6](https://doi.org/10.1016/0958-1669(95)80016-6)
- Clegg, R. M. (2004). The vital contributions of Perrin and Förster. *Biophotonics International*.
- Close, P., East, P., Dirac-Svejstrup, A. B., Hartmann, H., Heron, M., Maslen, S., ... Svejstrup, J. Q. (2012). DBIRD complex integrates alternative mRNA splicing with RNA polymerase II transcript elongation. *Nature*, 484(7394), 386–389. <https://doi.org/10.1038/nature10925>
- Cohen, S. N., Chang, A. C., Boyer, H. W., & Helling, R. B. (1973). Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 70(11), 3240–3244. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/4594039>
- Coltri, P., Effenberger, K., Chalkley, R. J., Burlingame, A. L., & Jurica, M. S. (2011). Breaking up the C complex spliceosome shows stable association of proteins with the lariat intron intermediate. *PLoS ONE*, 6(4). <https://doi.org/10.1371/journal.pone.0019061>
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., ... Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17(1), 13. <https://doi.org/10.1186/s13059-016-0881-8>
- Consortium, T. E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57–74. <https://doi.org/10.1038/nature11247>
- Consortium, T. U. (2008). The Universal Protein Resource. *Nucleic Acid Research*, 36(November 2007), D190–D195. <https://doi.org/10.1093/nar/gkl929>
- Conti, M., & Beavo, J. (2007). Biochemistry and Physiology of Cyclic Nucleotide Phosphodiesterases: Essential Components in Cyclic Nucleotide Signaling. *Annual Review of Biochemistry*, 76(1), 481–511. <https://doi.org/10.1146/annurev.biochem.76.060305.150444>
- Cooper, T. A., Wan, L., & Dreyfuss, G. (2009). RNA and Disease. *Cell*. <https://doi.org/10.1016/j.cell.2009.02.011>
- Corda, D., Kudo, T., Zizza, P., Jurisci, C., Kawai, E., Kato, N., ... Mariggi?, S. (2009). The developmentally regulated osteoblast phosphodiesterase GDE3 is glycerophosphoinositol-specific and modulates cell growth. *Journal of Biological Chemistry*, 284(37), 24848–24856. <https://doi.org/10.1074/jbc.M109.035444>
- Cordin, O., & Beggs, J. D. (2013). RNA helicases in splicing. *RNA Biology*. <https://doi.org/10.4161/rna.22547>
- Corrionero, A., Miñana, B., & Valcárcel, J. (2011). Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. *Genes and Development*, 25(5), 445–459. <https://doi.org/10.1101/gad.2014311>
- Crosetto, N., Bienko, M., & van Oudenaarden, A. (2015). Spatially resolved transcriptomics and beyond. *Nature Reviews Genetics*, 16(1), 57–66. <https://doi.org/10.1038/nrg3832>
- Cusick, M. E., Yu, H., Smolyar, A., Venkatesan, K., Carvunis, A. R., Simonis, N., ... Vidal, M. (2009). Literature-curated protein interaction datasets. *Nature Methods*, 6(1), 39–46. <https://doi.org/10.1038/nmeth.1284>
- Cvitkovic, I., & Jurica, M. S. (2013). Spliceosome database: A tool for tracking components of the spliceosome. *Nucleic Acids Research*, 41(D1). <https://doi.org/10.1093/nar/gks999>
- da Silva, J. P. M., Acencio, M. L., Mombach, J. C. M., Vieira, R., da Silva, J. C., Lemke, N., & Sinigaglia, M. (2008). In silico network topology-based prediction of gene essentiality. *Physica A: Statistical Mechanics and Its Applications*, 387(4), 1049–1055. <https://doi.org/10.1016/j.physa.2007.10.044>
- Davidson, L., & West, S. (2013). Splicing-coupled 3' end formation requires a terminal splice acceptor site, but not intron excision. *Nucleic Acids Research*, 41(14), 7101–7114. <https://doi.org/10.1093/nar/gkt446>
- Davis, D., Yaverolu, Ö. N., Malod-Dognin, N., Stojmirovic, A., & Pržulj, N. (2015). Topology-function conservation in protein-protein interaction networks. *Bioinformatics*, 31(10), 1632–1639. <https://doi.org/10.1093/bioinformatics/btv026>
- Day, R. N., & Davidson, M. W. (2012). Fluorescent proteins for FRET microscopy: monitoring protein interactions in living cells. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 34(5), 341–350. <https://doi.org/10.1002/bies.201100098>
- de Almeida, S. F., & Carmo-Fonseca, M. (2012). Design principles of interconnections between chromatin and pre-mRNA splicing. *Trends in Biochemical Sciences*. <https://doi.org/10.1016/j.tibs.2012.02.002>
- De Berardinis, R. J., & Chandel, N. S. (2016). Fundamentals of cancer metabolism. *Science Advances*. <https://doi.org/10.1126/sciadv.1600200>
- De Bodt, S., Proost, S., Vandepoele, K., Rouzé, P., & Van de Peer, Y. (2009). Predicting protein-protein interactions in *Arabidopsis thaliana* through integration of orthology, gene ontology and co-expression. *BMC Genomics*, 10(1), 288. <https://doi.org/10.1186/1471-2164-10-288>
- De Graaf, K., Czajkowska, H., Rottmann, S., Packman, L. C., Lilischkis, R., Lüscher, B., & Becker, W. (2006). The protein kinase DYRK1A phosphorylates the splicing factor SF3b1/SAP155 at Thr434, a novel in vivo phosphorylation site. *BMC Biochemistry*, 7. <https://doi.org/10.1186/1471-2091-7-7>
- de Las Rivas, J., & Fontanillo, C. (2010). Protein-protein interactions essentials: Key concepts to building and analyzing interactome networks. *PLoS Computational Biology*, 6(6), 1–8. <https://doi.org/10.1371/journal.pcbi.1000807>
- Debacker, K., & Frank Kooy, R. (2007). Fragile sites and human disease. *Human Molecular Genetics*, 16(R2). <https://doi.org/10.1093/hmg/ddm136>
- Debacker, K., Winnepenninckx, B., Longman, C., Colgan, J., Tolmie, J., Murray, R., ... Kooy, F. (2007). The molecular basis of the folate-sensitive fragile site FRA11A at 11q13. *Cytogenetic and Genome Research*, 119(1–2), 9–14. <https://doi.org/10.1159/000109612>
- Debatisse, M., Le Tallec, B., Letessier, A., Dutrillaux, B., & Brison, O. (2012). Common fragile sites: Mechanisms of instability revisited. *Trends in Genetics*. <https://doi.org/10.1016/j.tig.2011.10.003>
- DeBerardinis, R. J., Lum, J. J., Hatzivassiliou, G., & Thompson, C. B. (2008). The Biology of Cancer: Metabolic Reprogramming Fuels Cell Growth and Proliferation. *Cell Metabolism*, 7(1), 11–20. <https://doi.org/10.1016/j.cmet.2007.10.002>
- Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C. L., Kastner, B., ... Luhrmann, R. (2006). Protein Composition

- and Electron Microscopy Structure of Affinity-Purified Human Spliceosomal B Complexes Isolated under Physiological Conditions. *Molecular and Cellular Biology*, 26(14), 5528–5543. <https://doi.org/10.1128/MCB.00582-06>
- Deepak, S., Kottapalli, K., Rakwal, R., Oros, G., Rangappa, K., Iwahashi, H., ... Agrawal, G. (2007). Real-Time PCR: Revolutionizing Detection and Expression Analysis of Genes. *Current Genomics*, 8(4), 234–251. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18645596>
- Denli, A. M., Trops, B. B. J., Plasterk, R. H. A., Ketting, R. F., & Hannon, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature*, 432(7014), 231–235. <https://doi.org/10.1038/nature03049>
- Dettmer, K., Nürnberg, N., Kaspar, H., Gruber, M. A., Almstetter, M. F., & Oefner, P. J. (2011). Metabolite extraction from adherently growing mammalian cells for metabolomics studies: optimization of harvesting and extraction protocols. *Analytical and Bioanalytical Chemistry*, 399(3), 1127–1139. <https://doi.org/10.1007/s00216-010-4425-x>
- Di Girolamo, F., Lante, I., Muraca, M., & Putignani, L. (2013). The Role of Mass Spectrometry in the ‘Omics’ Era. *Current Organic Chemistry*, 17(23), 2891–2905. <https://doi.org/10.2174/1385272817888131118162725>
- Di Micco, R., Fumagalli, M., Cicalese, A., Piccinin, S., Gasparini, P., Luise, C., ... D’Adda Di Fagagna, F. (2006). Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature*, 444(7119), 638–642. <https://doi.org/10.1038/nature05327>
- Di Pietro, S. M., & Dell’Angelica, E. C. (2005). The cell biology of Hermansky-Pudlak syndrome: recent advances. *Traffic (Copenhagen, Denmark)*, 6(7), 525–533. <https://doi.org/10.1111/j.1600-0854.2005.00299.x>
- Dimmer, E. C., Huntley, R. P., Alam-Faruque, Y., Sawford, T., O’Donovan, C., Martin, M. J., ... Apweiler, R. (2012). The UniProt-GO Annotation database in 2011. *Nucleic Acids Research*, 40(D1). <https://doi.org/10.1093/nar/gkr1048>
- Dobkin, C., Rabe, A., Dumas, R., El Idrissi, A., Haubenstock, H., & Ted Brown, W. (2000). Fmr1 knockout mouse has a distinctive strain-specific learning impairment. *Neuroscience*, 100(2), 423–429. [https://doi.org/10.1016/S0306-4522\(00\)00292-X](https://doi.org/10.1016/S0306-4522(00)00292-X)
- Doncheva, N. T., Assenov, Y., Domingues, F. S., & Albrecht, M. (2012). Topological analysis and interactive visualization of biological networks and protein structures. *Nature Protocols*, 7(4), 670–685. <https://doi.org/10.1038/nprot.2012.004>
- Dong, J., & Horvath, S. (2007). Understanding network concepts in modules. *BMC Systems Biology*, 1. <https://doi.org/10.1186/1752-0509-1-24>
- Dong, Z., & Chen, Y. (2013). Transcriptomics: Advances and approaches. *Science China Life Sciences*, 56(10), 960–967. <https://doi.org/10.1007/s11427-013-4557-2>
- Drescher, D. G., Drescher, M. J., & Ramakrishnan, N. A. (2009). Surface plasmon resonance (SPR) analysis of binding interactions of proteins in inner-ear sensory epithelia. *Methods in Molecular Biology*, 493, 323–343. [https://doi.org/10.1007/978-1-59745-523-7\\_20](https://doi.org/10.1007/978-1-59745-523-7_20)
- Drissi, R., Dubois, M.-L., Douziech, M., & Boisvert, F.-M. (2015). Quantitative Proteomics Reveals Dynamic Interactions of the Minichromosome Maintenance Complex (MCM) in the Cellular Response to Etoposide Induced DNA Damage. *Molecular & Cellular Proteomics*, 14(7), 2002–2013. <https://doi.org/10.1074/mcp.M115.048991>
- Du, Y., Xu, N., Lu, M., & Li, T. (2011). Hubiquitome: A database of experimentally verified ubiquitination cascades in humans. *Database*, 2011. <https://doi.org/10.1093/database/bar055>
- Duan, G., & Walther, D. (2015). The Roles of Post-translational Modifications in the Context of Protein Interaction Networks. *PLoS Computational Biology*, 11(2). <https://doi.org/10.1371/journal.pcbi.1004049>
- Dudley, E., Yousef, M., Wang, Y., & Griffiths, W. J. (2010). Targeted metabolomics and mass spectrometry. In *Advances in protein chemistry and structural biology* (Vol. 80, pp. 45–83). <https://doi.org/10.1016/B978-0-12-381264-3.00002-3>
- Dunham, W. H., Mullin, M., & Gingras, A. C. (2012). Affinity-purification coupled to mass spectrometry: Basic principles and strategies. *Proteomics*. <https://doi.org/10.1002/pmic.201100523>
- Dunn, W. B., Bailey, N. J. C., & Johnson, H. E. (2005). Measuring the metabolome: current analytical technologies. *The Analyst*, 130(5), 606. <https://doi.org/10.1039/b418288j>
- Durek, P., & Walther, D. (2008). The integrated analysis of metabolic and protein interaction networks reveals novel molecular organizing principles. *BMC Systems Biology*, 2(1), 100. <https://doi.org/10.1186/1752-0509-2-100>
- Durkin, S. G., & Glover, T. W. (2007). Chromosome Fragile Sites (2007).pdf. *Annual Review of Genetics*, 41, 169–192. <https://doi.org/10.1146/annurev.genet.41.042007.165900>
- Dutcher, S. K., Lin, H., Zhang, Z., & Iomini, C. (2018). Identifying RNA splicing factors using IFT genes in Chlamydomonas reinhardtii. *Open Biology*, 8(3), 170211. <https://doi.org/10.1098/rsob.170211>
- Dybkov, O., Will, C. L., Deckert, J., Behzadnia, N., Hartmuth, K., & Lüthmann, R. (2006). U2 snRNA-protein contacts in purified human 17S U2 snRNPs and in spliceosomal A and B complexes. *Molecular and Cellular Biology*, 26(7), 2803–2816. <https://doi.org/10.1128/MCB.26.7.2803-2816.2006>
- Echeverri, C. J., & Perrimon, N. (2006). High-throughput RNAi screening in cultured cells: a user’s guide. *Nature Reviews Genetics*, 7(5), 373–384. <https://doi.org/10.1038/nrg1836>
- Einarson, M. B. (2004). Detection of protein-protein interactions using the GST fusion protein pull-down technique. *Nature Methods*, 1(3), 275–276. <https://doi.org/10.1038/nmeth1204-275>
- Eisenberg, T., Knauer, H., Schauer, A., Büttner, S., Ruckstuhl, C., Carmona-Gutierrez, D., ... Madeo, F. (2009). Induction of autophagy by spermidine promotes longevity. *Nature Cell Biology*, 11(11), 1305–1314. <https://doi.org/10.1038/ncb1975>
- Elbarbary, R. A., & Maquat, L. E. (2016). Coupling pre-mRNA splicing and 3’ end formation to mRNA export: Alternative ways to punch the nuclear export clock. *Genes and Development*, 30(5), 487–488. <https://doi.org/10.1101/gad.278937.116>
- Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., & Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, 411(6836), 494–498. <https://doi.org/10.1038/35078107>
- Ellis, J. D., Barrios-Rodiles, M., Çolak, R., Irimia, M., Kim, T. H., Calarco, J. A., ... Blencowe, B. J. (2012). Tissue-Specific Alternative Splicing Remodels Protein-Protein Interaction Networks. *Molecular Cell*, 46(6), 884–892. <https://doi.org/10.1016/j.molcel.2012.05.037>
- Emmert-Streib, F., Tripathi, S., Simoes, R. de M., Hawwa, A. F., & Dehmer, M. (2013). The human disease network. *Systems Biomedicine*, 1(1), 20–28. <https://doi.org/10.4161/sysb.22816>
- Estrada, E. (2006). Virtual identification of essential proteins within the protein interaction network of yeast. *Proteomics*, 6(1), 35–40. <https://doi.org/10.1002/pmic.200500209>
- Ewing, R. M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., ... Figeys, D. (2007). Large-scale mapping of human protein-protein interactions by mass spectrometry. *Molecular Systems Biology*, 3. <https://doi.org/10.1038/msb4100134>
- Expert-Bezancon, A., Caer, J. P. Le, & Marie, J. (2002). Heterogeneous nuclear ribonucleoprotein (hnRNP) K is a component of an intronic splicing enhancer complex that activates the splicing of the alternative exon 6A from chicken β-tropomyosin pre-mRNA. *Journal of Biological Chemistry*, 277(19), 16614–16623. <https://doi.org/10.1074/jbc.M201083200>



- Fabrizio, P., & Lüthmann, R. (2012). The spliceosome in constitutive splicing. In *Alternative pre-mRNA Splicing: Theory and Protocols* (pp. 49–64). <https://doi.org/10.1002/9783527636778.ch5>
- Fadhal, E., Gamielidien, J., & Mwambene, E. C. (2014). Protein interaction networks as metric spaces: A novel perspective on distribution of hubs. *BMC Systems Biology*, *8*(1). <https://doi.org/10.1186/1752-0509-8-6>
- Fang, W. F., Douglas, I. S., Wang, C. C., Kao, H. C., Chang, Y. T., Tseng, C. C., ... Lin, M. C. (2014). 5-lipoxygenase activating protein (FLAP) dependent leukotriene biosynthesis inhibition (MK591) attenuates lipid a endotoxin-induced inflammation. *PLoS ONE*, *9*(7). <https://doi.org/10.1371/journal.pone.0102622>
- Fazekas, J., Grunt, T. W., Jensen-Jarolim, E., & Singer, J. (2017). Long term storage in liquid nitrogen leads to only minor phenotypic and gene expression changes in the mammary carcinoma model cell line BT474. *Oncotarget*, *8*(21), 35076–35087. <https://doi.org/10.18632/oncotarget.16623>
- Ferguson, A. D., McKeever, B. M., Xu, S., Wisniewski, D., Miller, D. K., Yamin, T. T., ... Becker, J. W. (2007). Crystal structure of inhibitor-bound human 5-lipoxygenase-activating protein. *Science*, *317*(5837), 510–512. <https://doi.org/10.1126/science.1144346>
- Ferrari, F., Bortoluzzi, S., Coppe, A., Sirota, A., Safran, M., Shmoish, M., ... Bicciato, S. (2007). Novel definition files for human GeneChips based on GeneAnnot. *BMC Bioinformatics*, *8*(1), 446. <https://doi.org/10.1186/1471-2105-8-446>
- Fields, S., & Song, O. (1989). A novel genetic system to detect protein–protein interactions. *Nature*, *340*(6230), 245–246. <https://doi.org/10.1038/340245a0>
- Figey, D. (2008). Mapping the human protein interactome. *Cell Research*. <https://doi.org/10.1038/cr.2008.72>
- Fluhr, R. (2008). Regulation of splicing by protein phosphorylation. *Current Topics in Microbiology and Immunology*. [https://doi.org/10.1007/978-3-540-76776-3\\_7](https://doi.org/10.1007/978-3-540-76776-3_7)
- Fong, N., Bird, G., Vigneron, M., & Bentley, D. L. (2003). A 10 residue motif at the C-terminus of the RNA pol II CTD is required for transcription, splicing and 3' end processing. *EMBO Journal*, *22*(16), 4274–4282. <https://doi.org/10.1093/emboj/cdg396>
- Formstecher, E., Aresta, S., Collura, V., Hamburger, A., Meil, A., Trehin, A., ... Daviet, L. (2005). Protein interaction mapping: A Drosophila case study. *Genome Research*, *15*(3), 376–384. <https://doi.org/10.1101/gr.2659105>
- Förster, J., Famili, I., Fu, P., Palsson, B., & Nielsen, J. (2003). Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. *Genome Research*, *13*(2), 244–253. <https://doi.org/10.1101/gr.234503>
- Franchitto, A. (2013). Genome instability at common fragile sites: Searching for the cause of their instability. *BioMed Research International*. <https://doi.org/10.1155/2013/730714>
- Freeman, L. C., Roeder, D., & Mulholland, R. R. (1979). Centrality in social networks: ii. experimental results. *Social Networks*, *2*(2), 119–141. [https://doi.org/10.1016/0378-8733\(79\)90002-9](https://doi.org/10.1016/0378-8733(79)90002-9)
- Friedel, C. C., & Zimmer, R. (2007). Influence of degree correlations on network structure and stability in protein-protein interaction networks. *BMC Bioinformatics*, *8*. <https://doi.org/10.1186/1471-2105-8-297>
- Frilander, M. J., & Steitz, J. A. (2001). Dynamic exchanges of RNA interactions leading to catalytic core formation in the U12-dependent spliceosome. *Molecular Cell*, *7*(1), 217–226. [https://doi.org/10.1016/S1097-2765\(01\)00169-1](https://doi.org/10.1016/S1097-2765(01)00169-1)
- Froger, A., & Hall, J. E. (2007). Transformation of plasmid DNA into E. coli using the heat shock method. *Journal of Visualized Experiments : JoVE*, (6), 253. <https://doi.org/10.3791/253>
- Frontini, M. J., Nong, Z., Gros, R., Drangova, M., O'Neil, C., Rahman, M. N., ... Pickering, J. G. (2011). Fibroblast growth factor 9 delivery during angiogenesis produces durable, vasoresponsive microvessels wrapped by smooth muscle cells. *Nature Biotechnology*, *29*(5), 421–427. <https://doi.org/10.1038/nbt.1845>
- Fujimura, A., Kishimoto, H., Yanagisawa, J., & Kimura, K. (2012). Enhancer of rudimentary homolog (ERH) plays an essential role in the progression of mitosis by promoting mitotic chromosome alignment. *Biochemical and Biophysical Research Communications*, *423*(3), 588–592. <https://doi.org/10.1016/j.bbrc.2012.06.018>
- Funato, H., Sato, M., Sinton, C. M., Gautron, L., Williams, S. C., Skach, A., ... Yanagisawa, M. (2010). Loss of Goosecoid-like and DiGeorge syndrome critical region 14 in interpeduncular nucleus results in altered regulation of rapid eye movement sleep. *Proceedings of the National Academy of Sciences*, *107*(42), 18155–18160. <https://doi.org/10.1073/pnas.1012764107>
- Furukawa, T., Kuboki, Y., Tanji, E., Yoshida, S., Hatori, T., Yamamoto, M., ... Shiratori, K. (2011). Whole-exome sequencing uncovers frequent GNAS mutations in intraductal papillary mucinous neoplasms of the pancreas. *Scientific Reports*, *1*. <https://doi.org/10.1038/srep00161>
- Fuster, M. M., & Esko, J. D. (2005). The sweet and sour of cancer: Glycans as novel therapeutic targets. *Nature Reviews Cancer*. <https://doi.org/10.1038/nrc1649>
- Futschik, M. E., Tschaut, A., Chaurasia, G., & Herzel, H. (2007). Graph-theoretical comparison reveals structural divergence of human protein interaction networks. *Genome Informatics. International Conference on Genome Informatics*, *18*, 141–151. [https://doi.org/9781860949920\\_0014](https://doi.org/9781860949920_0014) [pii]
- Galbraith, S., Daniel, J. A., & Vissel, B. (2010). A Study of Clustered Data and Approaches to Its Analysis. *Journal of Neuroscience*, *30*(32), 10601–10608. <https://doi.org/10.1523/JNEUROSCI.0362-10.2010>
- Gallagher, S. R. (2006). One-Dimensional SDS Gel Electrophoresis of Proteins. In *Current Protocols in Molecular Biology* (Vol. Chapter 10, p. Unit 10.2A). Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/0471142727.mb1002as75>
- Gallo, K. A., & Johnson, G. L. (2002). Mixed-lineage kinase control of JNK and p38 MAPK pathways. *Nature Reviews Molecular Cell Biology*. <https://doi.org/10.1038/nrm906>
- Garbis, S. D., Tyritzis, S. I., Roumeliotis, T., Zerefos, P., Giannopoulou, E. G., Vlahou, A., ... Constantinides, C. A. (2008). Search for potential markers for prostate cancer diagnosis, prognosis and treatment in clinical tissue specimens using amine-specific isobaric tagging (iTRAQ) with two-dimensional liquid chromatography and tandem mass spectrometry. *J Proteome Res*, *7*(8), 3146–3158. <https://doi.org/10.1021/pr800060r>
- Gariyban, L., & Avashia, N. (2013). Polymerase Chain Reaction. *Journal of Investigative Dermatology*, *133*(3), 1–4. <https://doi.org/10.1038/jid.2013.1>
- Gatenby, R. A., & Gillies, R. J. (2004). Why do cancers have high aerobic glycolysis? *Nature Reviews Cancer*. <https://doi.org/10.1038/nrc1478>
- Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., ... Superti-Furga, G. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature*, *440*(7084), 631–636. <https://doi.org/10.1038/nature04532>
- Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., ... Superti-Furga, G. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, *415*(6868), 141–147. <https://doi.org/10.1038/415141a>

- Ge, H., Walhout, A. J. M., & Vidal, M. (2003). Integrating “omic” information: A bridge between genomics and systems biology. *Trends in Genetics*. <https://doi.org/10.1016/j.tig.2003.08.009>
- Gebhardt, A., Habjan, M., Benda, C., Meiler, A., Haas, D. A., Hein, M. Y., ... Pichlmair, A. (2015). mRNA export through an additional cap-binding complex consisting of NCBP1 and NCBP3. *Nature Communications*, 6. <https://doi.org/10.1038/ncomms9192>
- Géczy, J., Oostra, B. A., Hockey, A., Carbonell, P., Turner, G., Haan, E. A., ... Mulley, J. C. (1997). FMR2 expression in families with FRAXE mental retardation. *Human Molecular Genetics*, 6(3), 435–441. <https://doi.org/10.1093/hmg/6.3.435>
- Gelfman, S., Cohen, N., Yearim, A., & Ast, G. (2013). DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. *Genome Research*, 23(5), 789–799. <https://doi.org/10.1101/gr.143503.112>
- Gencheva, M., Kato, M., Newo, A. N. S., & Lin, R.-J. (2010). Contribution of DEAH-box protein DHX16 in human pre-mRNA splicing. *The Biochemical Journal*, 429(1), 25–32. <https://doi.org/10.1042/BJ20100266>
- Gene Ontology Consortium. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, 32(90001), 258D–261. <https://doi.org/10.1093/nar/gkh036>
- Gene Ontology Consortium. (2015). Gene Ontology Consortium: going forward. *Nucleic Acids Research*, 43(Database issue), D1049–56. <https://doi.org/10.1093/nar/gku1179>
- Gerstein, M. B., Rozowsky, J., Yan, K. K., Wang, D., Cheng, C., Brown, J. B., ... Waterston, R. (2014). Comparative analysis of the transcriptome across distant species. *Nature*, 512(7515), 445–448. <https://doi.org/10.1038/nature13424>
- Geschwind, D. H., & Konopka, G. (2009). Neuroscience in the era of functional genomics and systems biology. *Nature*, 461(7266), 908–915. <https://doi.org/10.1038/nature08537>
- Ghosh, S., & Ghosh, I. (1975). Variation of stemline karyotype in a HeLa cell line. *Zeitschrift Für Krebsforschung Und Klinische Onkologie*, 84(2), 129–133. <https://doi.org/10.1007/BF00304038>
- Giot, L. (2003). A Protein Interaction Map of Drosophila melanogaster. *Science*, 302(5651), 1727–1736. <https://doi.org/10.1126/science.1090289>
- Gioutlakis, A., Klapa, M. I., & Moschonas, N. K. (2017). PICKLE 2.0: A human protein-protein interaction meta-database employing data integration via genetic information ontology. *PLoS ONE*, 12(10). <https://doi.org/10.1371/journal.pone.0186039>
- Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12), 7821–7826. <https://doi.org/10.1073/pnas.122653799>
- Glover, T. W., Berger, C., Coyle, J., & Echo, B. (1984). DNA polymerase ?? inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. *Human Genetics*, 67(2), 136–142. <https://doi.org/10.1007/BF00272988>
- Goehler, H., Lalowski, M., Stelzl, U., Waelter, S., Stroedicke, M., Worm, U., ... Wanker, E. E. (2004). A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington’s disease. *Molecular Cell*, 15(6), 853–865. <https://doi.org/10.1016/j.molcel.2004.09.016>
- Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., & Barabási, A.-L. (2007). The human disease network. *Proceedings of the National Academy of Sciences of the United States of America*, 104(21), 8685–8690. <https://doi.org/10.1073/pnas.0701361104>
- Goh, K. I., & Choi, I. G. (2012). Exploring the human disease network: The human disease network. *Briefings in Functional Genomics*, 11(6), 533–542. <https://doi.org/10.1093/bfpg/els032>
- Golas, M. M., Sander, B., Will, C. L., Lührmann, R., & Stark, H. (2003). Molecular architecture of the multiprotein splicing factor SF3b. *Science*, 300(5621), 980–984. <https://doi.org/10.1126/science.1084155>
- Goldfarb, S., Ziyadeh, F. N., Kern, E. F. O., & Simmons, D. A. (1991). Effects of polyol-pathway inhibition and dietary myo-inositol on glomerular hemodynamic function in experimental diabetes mellitus in rats. *Diabetes*, 40(4), 465–471. <https://doi.org/10.2337/diab.40.4.465>
- Gong, W., Emanuel, B. S., Galili, N., Kim, D. H., Roe, B., Driscoll, D. a, & Budarf, M. L. (1997). Structural and mutational analysis of a conserved gene (DGSI) from the minimal DiGeorge syndrome critical region. *Human Molecular Genetics*, 6(2), 267–276. <https://doi.org/10.1093/hmg/6.2.267>
- Goñi, J., Esteban, F. J., de Mendizábal, N. V., Sepulcre, J., Ardanza-Trevijano, S., Agirrezabal, I., & Villoslada, P. (2008). A computational analysis of protein-protein interaction networks in neurodegenerative diseases. *BMC Systems Biology*, 2. <https://doi.org/10.1186/1752-0509-2-52>
- Goodacre, R., Vaidyanathan, S., Dunn, W. B., Harrigan, G. G., & Kell, D. B. (2004). Metabolomics by numbers: Acquiring and understanding global metabolite data. *Trends in Biotechnology*. <https://doi.org/10.1016/j.tibtech.2004.03.007>
- Gosetti, F., Mazzucco, E., Gennaro, M. C., & Marengo, E. (2013). Ultra high performance liquid chromatography tandem mass spectrometry determination and profiling of prohibited steroids in human biological matrices. A review. *Journal of Chromatography B*, 927, 22–36. <https://doi.org/10.1016/j.jchromb.2012.12.003>
- Gossen, M., Freundlieb, S., Bender, G., Müller, G., Hillen, W., & Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. *Science (New York, N.Y.)*, 268(5218), 1766–1769. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7792603>
- Gout, S., Brambilla, E., Boudria, A., Drissi, R., Lantuejoul, S., Gazzeri, S., & Eymin, B. (2012). Abnormal Expression of the Pre-mRNA Splicing Regulators SRSF1, SRSF2, SRPK1 and SRPK2 in Non Small Cell Lung Carcinoma. *PLoS ONE*, 7(10). <https://doi.org/10.1371/journal.pone.0046539>
- Govindarajan, R., Duraiyan, J., Kaliyappan, K., & Palanisamy, M. (2012). Microarray and its applications. *Journal of Pharmacy & Bioallied Sciences*, 4(Suppl 2), S310-2. <https://doi.org/10.4103/0975-7406.100283>
- Graham, F. L., & van der Eb, A. J. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology*, 52(2), 456–467. [https://doi.org/10.1016/0042-6822\(73\)90341-3](https://doi.org/10.1016/0042-6822(73)90341-3)
- GRAINGER, R. J. (2005). Prp8 protein: At the heart of the spliceosome. *RNA*, 11(5), 533–557. <https://doi.org/10.1261/rna.2220705>
- Greenbaum, D., Jansen, R., & Gerstein, M. (2002). Analysis of mRNA expression and protein abundance data: An approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts. *Bioinformatics*, 18(4), 585–596. <https://doi.org/10.1093/bioinformatics/18.4.585>
- Grimm, S. (2004). The art and design of genetic screens: mammalian culture cells. *Nature Reviews Genetics*, 5(3), 179–189. <https://doi.org/10.1038/nrg1291>
- Grote, M., Wolf, E., Will, C. L., Lemm, I., Agafonov, D. E., Schomburg, A., ... Luhrmann, R. (2010). Molecular Architecture of the Human Prp19/CDC5L Complex. *Molecular and Cellular Biology*, 30(9), 2105–2119. <https://doi.org/10.1128/MCB.01505-09>
- Guo, D., Hazbun, T. R., Xu, X. J., Ng, S. L., Fields, S., & Kuo, M. H. (2004). A tethered catalysis, two-hybrid system to identify

- protein-protein interactions requiring post-translational modifications. *Nature Biotechnology*, 22(7), 888–892. <https://doi.org/10.1038/nbt985>
- Guruharsha, K. G., Rual, J.-F., Zhai, B., Mintseris, J., Vaidya, P., Vaidya, N., ... Artavanis-Tsakonas, S. (2011). A protein complex network of *Drosophila melanogaster*. *Cell*, 147(3), 690–703. <https://doi.org/10.1016/j.cell.2011.08.047>
- Gustafsson, M., Nestor, C. E., Zhang, H., Barabási, A. L., Baranzini, S., Brunak, S., ... Benson, M. (2014). Modules, networks and systems medicine for understanding disease and aiding diagnosis. *Genome Medicine*. <https://doi.org/10.1186/s13073-014-0082-6>
- Guthrie, C., & Patterson, B. (1988). Spliceosomal snRNAs. *Annual Review of Genetics*, 22(2), 387–419. <https://doi.org/10.1146/annurev.ge.22.120188.002131>
- Habelhah, H., Shah, K., Huang, L., Ostareck-Lederer, A., Burlingame, A. L., Shokat, K. M., ... Ronai, Z. (2001). ERK phosphorylation drives cytoplasmic accumulation of hnRNP-K and inhibition of mRNA translation. *Nature Cell Biology*, 3(3), 325–330. <https://doi.org/10.1038/35060131>
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., ... Tuschl, T. (2010). Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell*, 141(1), 129–141. <https://doi.org/10.1016/j.cell.2010.03.009>
- Halbeisen, R. E., Galgano, A., Scherrer, T., & Gerber, A. P. (2008). Post-transcriptional gene regulation: From genome-wide studies to principles. *Cellular and Molecular Life Sciences*. <https://doi.org/10.1007/s00018-007-7447-6>
- Hall, B., Limaye, A., & Kulkarni, A. B. (2009). Overview: Generation of Gene Knockout Mice. In *Current Protocols in Cell Biology* (Vol. Chapter 19, p. Unit 19.12.1-17). Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/0471143030.cb1912s44>
- Han, H., Braunschweig, U., Gonatopoulos-Pournatzis, T., Weatheritt, R. J., Hirsch, C. L., Ha, K. C. H., ... Blencowe, B. J. (2017). Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors. *Molecular Cell*, 65(3), 539–553.e7. <https://doi.org/10.1016/j.molcel.2017.01.011>
- Han, X., Aslanian, A., Yates, J. R., & III. (2008). Mass spectrometry for proteomics. *Current Opinion in Chemical Biology*, 12(5), 483–490. <https://doi.org/10.1016/j.cbpa.2008.07.024>
- Hao, T., Peng, W., Wang, Q., Wang, B., & Sun, J. (2016). Reconstruction and application of protein–protein interaction network. *International Journal of Molecular Sciences*. <https://doi.org/10.3390/ijms17060907>
- Hao, Y., Wu, W., Li, H., Yuan, J., Luo, J., Zhao, Y., & Chen, R. (2016). NPInter v3.0: An upgraded database of noncoding RNA-associated interactions. *Database*, 2016. <https://doi.org/10.1093/database/baw057>
- Haque, N., & Oberdoerffer, S. (2014). Chromatin and splicing. *Methods in Molecular Biology*, 1126, 97–113. [https://doi.org/10.1007/978-1-62703-980-2\\_7](https://doi.org/10.1007/978-1-62703-980-2_7)
- Hardy, S., Legagneux, V., Audic, Y., & Paillard, L. (2010). Reverse genetics in eukaryotes. *Biology of the Cell*, 102(10), 561–580. <https://doi.org/10.1042/BC20100038>
- Hartmuth, K., Urlaub, H., Vornlocher, H.-P., Will, C. L., Gentzel, M., Wilm, M., & Lührmann, R. (2002). Protein composition of human pre-spliceosomes isolated by a tobramycin affinity-selection method. *Proceedings of the National Academy of Sciences of the United States of America*, 99(26), 16719–16724. <https://doi.org/10.1073/pnas.262483899>
- Hastings, M. L., Allemand, E., Duelli, D. M., Myers, M. P., & Krainer, A. R. (2007). Control of Pre-mRNA Splicing by the General Splicing Factors PUF60 and U2AF 65. *PLoS ONE*, 2(6). <https://doi.org/10.1371/journal.pone.0000538>
- Hatzfeld, M. (2007). Plakophilins: Multifunctional proteins or just regulators of desmosomal adhesion? *Biochimica et Biophysica Acta - Molecular Cell Research*. <https://doi.org/10.1016/j.bbamcr.2006.04.009>
- Haunhorst, P., Berndt, C., Eitner, S., Godoy, J. R., & Lillig, C. H. (2010). Characterization of the human monothiol glutaredoxin 3 (PICOT) as iron-sulfur protein. *Biochemical and Biophysical Research Communications*, 394(2), 372–376. <https://doi.org/10.1016/j.bbrc.2010.03.016>
- Haunhorst, P., Hanschmann, E.-M., Bräutigam, L., Stehling, O., Hoffmann, B., Mühlhoff, U., ... Lillig, C. H. (2013). Crucial function of vertebrate glutaredoxin 3 (PICOT) in iron homeostasis and hemoglobin maturation. *Molecular Biology of the Cell*, 24(12), 1895–1903. <https://doi.org/10.1091/mbc.E12-09-0648>
- Hayes, S., Malacrida, B., Kiely, M., & Kiely, P. A. (2016). Studying protein-protein interactions: progress, pitfalls and solutions. *Biochemical Society Transactions*, 44(4), 994–1004. <https://doi.org/10.1042/BST20160092>
- He, H., Liyanarachchi, S., Akagi, K., Nagy, R., Li, J., Dietrich, R. C., ... De La Chapelle, A. (2011). Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD i. *Science*, 332(6026), 238–240. <https://doi.org/10.1126/science.1200587>
- Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., ... Stelzl, U. (2012). Dynamic Protein-Protein Interaction Wiring of the Human Spliceosome. *Molecular Cell*, 45(4), 567–580. <https://doi.org/10.1016/j.molcel.2011.12.034>
- Hein, M. Y., Hubner, N. C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., ... Mann, M. (2015). A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell*, 163(3), 712–723. <https://doi.org/10.1016/j.cell.2015.09.053>
- Helmrich, A., Ballarino, M., & Tora, L. (2011). Collisions between Replication and Transcription Complexes Cause Common Fragile Site Instability at the Longest Human Genes. *Molecular Cell*, 44(6), 966–977. <https://doi.org/10.1016/j.molcel.2011.10.013>
- Herold, N., Will, C. L., Wolf, E., Kastner, B., Urlaub, H., & Lührmann, R. (2009). Conservation of the Protein Composition and Electron Microscopy Structure of *Drosophila melanogaster* and Human Spliceosomal Complexes. *Molecular and Cellular Biology*, 29(1), 281–301. <https://doi.org/10.1128/MCB.01415-08>
- Hewett, D. R., Handt, O., Hobson, L., Mangelsdorf, M., Eyre, H. J., Baker, E., ... Richards, R. I. (1998). FRA10B structure reveals common elements in repeat expansion and chromosomal fragile site genesis. *Molecular Cell*, 1(6), 773–781. [https://doi.org/10.1016/S1097-2765\(00\)80077-5](https://doi.org/10.1016/S1097-2765(00)80077-5)
- Highley, J. R., Kirby, J., Jansweijer, J. A., Webb, P. S., Hewamadduma, C. A., Heath, P. R., ... Ince, P. G. (2014). Loss of nuclear TDP-43 in amyotrophic lateral sclerosis (ALS) causes altered expression of splicing machinery and widespread dysregulation of RNA splicing in motor neurones. *Neuropathology and Applied Neurobiology*, 40(6), 670–685. <https://doi.org/10.1111/nan.12148>
- Hillenkamp, F., Karas, M., Beavis, R. C., & Chait, B. T. (1991). Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers. *Analytical Chemistry*, 63(24), 1193A–1203A. <https://doi.org/10.1021/ac00024a716>
- Hir, H. Le, Saulière, J., & Wang, Z. (2016). The exon junction complex as a node of post-transcriptional networks. *Nature Reviews Molecular Cell Biology*. <https://doi.org/10.1038/nrm.2015.7>
- Hogg, R., McGrail, J. C., & O’Keefe, R. T. (2010). The function of the NineTeen Complex (NTC) in regulating spliceosome

- conformations and fidelity during pre-mRNA splicing. *Biochemical Society Transactions*, 38(4), 1110–1115. <https://doi.org/10.1042/BST0381110>
- Hong, Y., Chan, C. B., Kwon, I.-S., Li, X., Song, M., Lee, H.-P., ... Ye, K. (2012). SRPK2 Phosphorylates Tau and Mediates the Cognitive Defects in Alzheimer's Disease. *Journal of Neuroscience*, 32(48), 17262–17272. <https://doi.org/10.1523/JNEUROSCI.3300-12.2012>
- Hood, L., Heath, J. R., Phelps, M. E., & Lin, B. (2004). Systems Biology and New Technologies Enable Predictive and Preventative Medicine. *Science*, 306(5696), 640–643. <https://doi.org/10.1126/science.1104635>
- Horrocks, C., Halse, R., Suzuki, R., & Shepherd, P. R. (2003). Human cell systems for drug discovery. *Current Opinion in Drug Discovery & Development*, 6(4), 570–575. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12951820>
- Hosios, A. M., Hecht, V. C., Danai, L. V., Johnson, M. O., Rathmell, J. C., Steinhäuser, M. L., ... Vander Heiden, M. G. (2016). Amino Acids Rather than Glucose Account for the Majority of Cell Mass in Proliferating Mammalian Cells. *Developmental Cell*, 36(5), 540–549. <https://doi.org/10.1016/j.devcel.2016.02.012>
- Houdebine, L.-M. (2007). Transgenic Animal Models in Biomedical Research. In *Target Discovery and Validation Reviews and Protocols* (Vol. 360, pp. 163–202). New Jersey: Humana Press. <https://doi.org/10.1385/1-59745-165-7:163>
- Hu, H., Yan, X., Huang, Y., Han, J., & Zhou, X. J. (2005). Mining coherent dense subgraphs across massive biological networks for functional discovery. *Bioinformatics*, 21(SUPPL. 1). <https://doi.org/10.1093/bioinformatics/bti1049>
- Huang, D. W., Sherman, B. T., Tan, Q., Collins, J. R., Alvord, W. G., Roayaei, J., ... Lempicki, R. A. (2007). The DAVID Gene Functional Classification Tool: A novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biology*, 8(9). <https://doi.org/10.1186/gb-2007-8-9-r183>
- Huang, H.-L., Hsing, H.-W., Lai, T.-C., Chen, Y.-W., Lee, T.-R., Chan, H.-T., ... Chan, H.-L. (2010). Trypsin-induced proteome alteration during cell subculture in mammalian cells. *Journal of Biomedical Science*, 17(1), 36. <https://doi.org/10.1186/1423-0127-17-36>
- Huang, H., Asimaki, A., Lo, D., McKenna, W., & Saffitz, J. E. (2008). Disparate effects of different mutations in plakoglobin on cell mechanical behavior. *Cell Motility and the Cytoskeleton*, 65(12), 964–978. <https://doi.org/10.1002/cm.20319>
- Huang, M., Ding, S., Wang, H., & Zhu, X. (2008). Mining physical protein-protein interactions from the literature. *Genome Biology*, 9(SUPPL. 2). <https://doi.org/10.1186/gb-2008-9-s2-s12>
- Huh, S.-H., & Ornitz, D. M. (2010). -catenin deficiency causes DiGeorge syndrome-like phenotypes through regulation of Tbx1. *Development*, 137(7), 1137–1147. <https://doi.org/10.1242/dev.045534>
- Humphrey-Smith, I. (2015). The 20th anniversary of proteomics and some of its origins. *PROTEOMICS*, 15(11), 1773–1776. <https://doi.org/10.1002/pmic.201400582>
- Huthmacher, C., Gille, C., & Holzhütter, H.-G. (2008). A computational analysis of protein interactions in metabolic networks reveals novel enzyme pairs potentially involved in metabolic channeling. *Journal of Theoretical Biology*, 252(3), 456–464. <https://doi.org/10.1016/j.jtbi.2007.09.042>
- Huttlin, E. L., Bruckner, R. J., Paulo, J. A., Cannon, J. R., Ting, L., Baltier, K., ... Wade Harper, J. (2017). Architecture of the human interactome defines protein communities and disease networks. *Nature*, 545(7655), 505–509. <https://doi.org/10.1038/nature22366>
- Iadevaia, V., & Gerber, A. (2015). Combinatorial Control of mRNA Fates by RNA-Binding Proteins and Non-Coding RNAs. *Biomolecules*, 5(4), 2207–2222. <https://doi.org/10.3390/biom5042207>
- Ilagan, J. O., Chalkley, R. J., Burlingame, A. L., & Jurica, M. S. (2013). Rearrangements within human spliceosomes captured after exon ligation. *Rna*, 19(3), 400–412. <https://doi.org/10.1261/rna.034223.112>
- Ilagan, J. O., Ramakrishnan, A., Hayes, B., Murphy, M. E., Zebari, A. S., Bradley, P., & Bradley, R. K. (2015). U2AF1 mutations alter splice site recognition in hematological malignancies. *Genome Research*, 25(1), 14–26. <https://doi.org/10.1101/gr.181016.114>
- Ilagan, J., Yuh, P., Chalkley, R. J., Burlingame, A. L., & Jurica, M. S. (2009). The Role of Exon Sequences in C Complex Spliceosome Structure. *Journal of Molecular Biology*, 394(2), 363–375. <https://doi.org/10.1016/j.jmb.2009.09.019>
- Imielinski, M., Berger, A. H., Hammerman, P. S., Hernandez, B., Pugh, T. J., Hodis, E., ... Meyerson, M. (2012). Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*, 150(6), 1107–1120. <https://doi.org/10.1016/j.cell.2012.08.029>
- Imig, J., Kanitz, A., & Gerber, A. P. (2012). RNA regulons and the RNA-protein interaction network. *Biomolecular Concepts*. <https://doi.org/10.1515/bmc-2012-0016>
- Ingham, R. J., Colwill, K., Howard, C., Dettwiler, S., Lim, C. S., Yu, J., ... Pawson, T. (2005). WW domains provide a platform for the assembly of multiprotein networks. *Mol Cell Biol*, 25(16), 7092–7106. <https://doi.org/10.1128/MCB.25.16.7092-7106.2005>
- Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., & Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences*, 98(8), 4569–4574. <https://doi.org/10.1073/pnas.061034498>
- Ito, T., Ota, K., Kubota, H., Yamaguchi, Y., Chiba, T., Sakuraba, K., & Yoshida, M. (2002). Roles for the Two-hybrid System in Exploration of the Yeast Protein Interactome. *Molecular & Cellular Proteomics*, 1(8), 561–566. <https://doi.org/10.1074/mcp.R200005-MCP200>
- Ivanov, I., McKenzie, B. S., Zhou, L., Tadokoro, C. E., Lepelley, A., Lafaille, J. J., ... Littman, D. R. (2006). The orphan nuclear receptor ROR $\gamma$  directs the differentiation program of proinflammatory IL-17+ T helper cells. *Cell*, 126(6), 1121–1133. [https://doi.org/S0092-8674\(06\)01105-6](https://doi.org/S0092-8674(06)01105-6) [pii] 10.1016/j.cell.2006.07.035
- Jalili, M., Salehzadeh-Yazdi, A., Gupta, S., Wolkenhauer, O., Yaghmaie, M., Resendis-Antonio, O., & Alimoghaddam, K. (2016). Evolution of centrality measurements for the detection of essential proteins in biological networks. *Frontiers in Physiology*. <https://doi.org/10.3389/fphys.2016.00375>
- Jangi, M., Fleet, C., Cullen, P., Gupta, S. V., Mekhoubad, S., Chiao, E., ... Staropoli, J. F. (2017). SMN deficiency in severe models of spinal muscular atrophy causes widespread intron retention and DNA damage. *Proceedings of the National Academy of Sciences*, 114(12), E2347–E2356. <https://doi.org/10.1073/pnas.1613181114>
- Jansen, R., Greenbaum, D., & Gerstein, M. (2002). Relating whole-genome expression data with protein-protein interactions. *Genome Research*, 12(1), 37–46. <https://doi.org/10.1101/gr.205602>
- Jeong, H., Mason, S. P., Barabási, A. L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, 411(6833), 41–42. <https://doi.org/10.1038/35075138>
- Jetten. (2007). Retinoid-related orphan receptors (RORs): critical roles in development, immunity, circadian rhythm, and cellular metabolism. *Nuclear Receptor Signaling*, 4. <https://doi.org/10.1621/nrs.07003>
- Jiang, Y., Lucas, I., Young, D. J., Davis, E. M., Karrison, T., Rest, J. S., & Le Beau, M. M. (2009). Common fragile sites are

- characterized by histone hypoacetylation. *Human Molecular Genetics*, 18(23), 4501–4512. <https://doi.org/10.1093/hmg/ddp410>
- Jin, P., & Warren, S. T. (2000). Understanding the molecular basis of fragile X syndrome. *Human Molecular Genetics*, 9(6), 901–908. <https://doi.org/10.1093/hmg/9.6.901>
- Joga, M. R., Zotti, M. J., Smagghe, G., & Christiaens, O. (2016). RNAi Efficiency, Systemic Properties, and Novel Delivery Methods for Pest Insect Control: What We Know So Far. *Frontiers in Physiology*, 7, 553. <https://doi.org/10.3389/fphys.2016.00553>
- Jones, C. E., Brown, A. L., & Baumann, U. (2007). Estimating the annotation error rate of curated GO database sequence annotations. *BMC Bioinformatics*, 8. <https://doi.org/10.1186/1471-2105-8-170>
- Jones, C., Müllenbach, R., Grossfeld, P., Auer, R., Favier, R., Chien, K., ... Cotter, F. (2000). Co-localisation of CCG repeats and chromosome deletion breakpoints in Jacobsen syndrome: evidence for a common mechanism of chromosome breakage. *Hum. Mol. Genet.*, 9, 1201–1208. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10767345>
- Jones, C., Penny, L., Mattina, T., Yu, S., Baker, E., Voullaire, L., ... Tunnacliffe, A. (1995). Association of a chromosome deletion syndrome with a fragile site within the proto-oncogene CBL2. *Nature*, 376(6536), 145–149. <https://doi.org/10.1038/376145a0>
- Jones, C., Slljepevic, P., Marsh, S., Baker, E., Langdon, W. Y., Richards, R. I., & Tunnacliffe, A. (1994). Physical linkage of the fragile site FRA11B and a Jacobsen syndrome chromosome deletion breakpoint in 11q23. 3. *Human Molecular Genetics*, 3(12), 2123–2130. <https://doi.org/10.1093/hmg/3.12.2123>
- Joyce, A. R., & Palsson, B. Ø. (2006). The model organism as a system: integrating “omics” data sets. *Nature Reviews Molecular Cell Biology*, 7(3), 198–210. <https://doi.org/10.1038/nrm1857>
- Junge, A., Refsgaard, J. C., Garde, C., Pan, X., Santos, A., Alkan, F., ... Gorodkin, J. (2017). RAIN: RNA-protein Association and Interaction Networks. *Database*, 2017(1). <https://doi.org/10.1093/database/baw167>
- Jurica, M. S., Licklider, L. J., Gygi, S. R., Grigorieff, N., & Moore, M. J. (2002). Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA (New York, N.Y.)*, 8(4), 426–439. <https://doi.org/10.1017/S1355838202021088>
- Jurica, M. S., & Moore, M. J. (2003). Pre-mRNA splicing: Awash in a sea of proteins. *Molecular Cell*. [https://doi.org/10.1016/S1097-2765\(03\)00270-3](https://doi.org/10.1016/S1097-2765(03)00270-3)
- Kähkönen, M., Tengström, C., Alitalo, T., Matilainen, R., Kaski, M., & Airaksinen, E. (1989). Population cytogenetics of folate-sensitive fragile sites - II. Autosomal rare fragile sites. *Human Genetics*, 82(1), 3–8. <https://doi.org/10.1007/BF00288261>
- Kaida, D., Berg, M. G., Younis, I., Kasim, M., Singh, L. N., Wan, L., & Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature*, 468(7324), 664–668. <https://doi.org/10.1038/nature09479>
- Kaitsuka, T., & Matsushita, M. (2015). Regulation of translation factor EEF1D gene function by alternative splicing. *International Journal of Molecular Sciences*. <https://doi.org/10.3390/ijms16023970>
- Kaitsuka, T., Tomizawa, K., & Matsushita, M. (2011). Transformation of eEF1B into heat-shock response transcription factor by alternative splicing. *EMBO Reports*, 12(7), 673–681. <https://doi.org/10.1038/embor.2011.82>
- Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1), 7–15. [https://doi.org/10.1016/0020-0190\(89\)90102-6](https://doi.org/10.1016/0020-0190(89)90102-6)
- Kamburov, A., Stelzl, U., Lehrach, H., & Herwig, R. (2013). The ConsensusPathDB interaction database: 2013 Update. *Nucleic Acids Research*, 41(D1). <https://doi.org/10.1093/nar/gks1055>
- Kamtchueng, C., Stébenne, M. É., Delannoy, A., Wilhelm, E., Léger, H., Benecke, A. G., & Bell, B. (2014). Alternative splicing of TAF6: Downstream transcriptome impacts and upstream RNA splice control elements. *PLoS ONE*, 9(7). <https://doi.org/10.1371/journal.pone.0102399>
- Kanani, H., Chrysanthopoulos, P. K., & Klapa, M. I. (2008). Standardizing GC–MS metabolomics. *Journal of Chromatography B*, 871(2), 191–201. <https://doi.org/10.1016/j.jchromb.2008.04.049>
- Kanani, H. H., & Klapa, M. I. (2007). Data correction strategy for metabolomics analysis using gas chromatography–mass spectrometry. *Metabolic Engineering*, 9(1), 39–51. <https://doi.org/10.1016/j.ymben.2006.08.001>
- Kapoor, R. V., Coyle, R., Staton, C. A., Brown, N. J., & Vaidyanathan, S. (2015). Cell line dependence of metabolite leakage in metabolome analyses of adherent normal and cancer cell lines. *Metabolomics*, 11(6), 1743–1755. <https://doi.org/10.1007/s11306-015-0833-4>
- Kapoor, R. V., Coyle, R., Staton, C. A., Brown, N. J., & Vaidyanathan, S. (2017). Influence of washing and quenching in profiling the metabolome of adherent mammalian cells: a case study with the metastatic breast cancer cell line MDA-MB-231. *The Analyst*, 142(11), 2038–2049. <https://doi.org/10.1039/C7AN00207F>
- Karni, R., De Stanchina, E., Lowe, S. W., Sinha, R., Mu, D., & Krainer, A. R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nature Structural and Molecular Biology*, 14(3), 185–193. <https://doi.org/10.1038/nsmb1209>
- Karpova, T. S., Baumann, C. T., He, L., Wu, X., Grammer, A., Lipsky, P., ... McNally, J. G. (2003). Fluorescence resonance energy transfer from cyan to yellow fluorescent protein detected by acceptor photobleaching using confocal microscopy and a single laser. *Journal of Microscopy*, 209(1), 56–70. <https://doi.org/10.1046/j.1365-2818.2003.01100.x>
- Katagiri, S., Hayashi, T., Yoshitake, K., Sergeev, Y., Akahori, M., Furuno, M., ... Iwata, T. (2015). Congenital Achromatopsia and Macular Atrophy Caused by a Novel Recessive PDE6C Mutation (p.E591K). *Ophthalmic Genetics*, 36(2), 137–144. <https://doi.org/10.3109/13816810.2014.991932>
- Katahira, J. (2015). Nuclear export of messenger RNA. *Genes*, 6(2), 163–184. <https://doi.org/10.3390/genes6020163>
- Katz, Y., Wang, E. T., Airolidi, E. M., & Burge, C. B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods*, 7(12), 1009–1015. <https://doi.org/10.1038/nmeth.1528>
- Kazdoba, T. M., Leach, P. T., Silverman, J. L., & Crawley, J. N. (2014). Modeling fragile X syndrome in the Fmr1 knockout mouse. *Intractable & Rare Diseases Research*, 3(4), 118–133. <https://doi.org/10.5582/irdr.2014.01024>
- Kebarle, P. (2000). A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *Journal of Mass Spectrometry*, 35(7), 804–817. [https://doi.org/10.1002/1096-9888\(200007\)35:7<804::AID-JMS22>3.0.CO;2-Q](https://doi.org/10.1002/1096-9888(200007)35:7<804::AID-JMS22>3.0.CO;2-Q)
- Keene, J. D. (2007). RNA regulons: coordination of post-transcriptional events. *Nature Reviews Genetics*, 8(7), 533–543. <https://doi.org/10.1038/nrg2111>
- Keene, J. D., Komisarow, J. M., & Friedersdorf, M. B. (2006). RIP-Chip: The isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nature Protocols*, 1(1), 302–307. <https://doi.org/10.1038/nprot.2006.47>
- Kell, D. B., & Oliver, S. G. (2004). Here is the evidence, now what is the hypothesis? The complementary roles of inductive and

- hypothesis-driven science in the post-genomic era. *BioEssays*, 26(1), 99–105. <https://doi.org/10.1002/bies.10385>
- Kenworthy, A. K., & Edidin, M. (1998). Distribution of a glycosylphosphatidylinositol-anchored protein at the apical surface of MDCK cells examined at a resolution of less-than-100 angstrom using imaging fluorescence resonance energy transfer. *Journal of Cell Biology*, 142(1), 69–84.
- Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., ... Hermjakob, H. (2012). The IntAct molecular interaction database in 2012. *Nucleic Acids Research*, 40(D1). <https://doi.org/10.1093/nar/gkr1088>
- Keshava Prasad, T. S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., ... Pandey, A. (2009). Human Protein Reference Database--2009 update. *Nucleic Acids Research*, 37(Database issue), D767–72. <https://doi.org/10.1093/nar/gkn892>
- Kfir, N., Lev-Maor, G., Glaich, O., Alajem, A., Datta, A., Sze, S. K., ... Ast, G. (2015). SF3B1 Association with Chromatin Determines Splicing Outcomes. *Cell Reports*, 11(4), 618–629. <https://doi.org/10.1016/j.celrep.2015.03.048>
- Khin, T. M., Uehara, T., Aoshima, K., & Oda, Y. (2009). Polar anionic metabolome analysis by nano-LC/MS with a metal chelating agent. *Analytical Chemistry*, 81(18), 7766–7772. <https://doi.org/10.1021/ac901269h>
- Khuri, S., & Wuchty, S. (2015). Essentiality and centrality in protein interaction networks revisited. *BMC Bioinformatics*, 16(1). <https://doi.org/10.1186/s12859-015-0536-x>
- Kilchert, C., Wittmann, S., & Vasiljeva, L. (2016). The regulation and functions of the nuclear RNA exosome complex. *Nature Reviews Molecular Cell Biology*. <https://doi.org/10.1038/nrm.2015.15>
- Kim, D., & Rossi, J. (2008). RNAi mechanisms and applications. *BioTechniques*, 44(5), 613–616. <https://doi.org/10.2144/000112792>
- Kim, H. J., Kim, N. C., Wang, Y. D., Scarborough, E. A., Moore, J., Diaz, Z., ... Taylor, J. P. (2013). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature*, 495(7442), 467–473. <https://doi.org/10.1038/nature11922>
- Kim, T. K., & Eberwine, J. H. (2010). Mammalian cell transfection: The present and the future. *Analytical and Bioanalytical Chemistry*, 397(8), 3173–3178. <https://doi.org/10.1007/s00216-010-3821-6>
- Kinoshita, J. H. (1990). A thirty year journey in the polyol pathway. *Experimental Eye Research*, 50(6), 567–573. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2115448>
- Klapa, M. I., Tsafou, K., Theodoridis, E., Tsakalidis, A., & Moschonas, N. K. (2013). Reconstruction of the experimentally supported human protein interactome: What can we learn? *BMC Systems Biology*, 7. <https://doi.org/10.1186/1752-0509-7-96>
- Klingström, T., & Plewczynski, D. (2011). Protein-protein interaction and pathway databases, a graphical review. *Briefings in Bioinformatics*, 12(6), 702–713. <https://doi.org/10.1093/bib/bbq064>
- Knight, S. J. L., Flannery, A. V., Hirst, M. C., Campbell, L., Christodoulou, Z., Phelps, S. R., ... Davies, K. E. (1993). Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell*, 74(1), 127–134. [https://doi.org/10.1016/0092-8674\(93\)90300-F](https://doi.org/10.1016/0092-8674(93)90300-F)
- Kobrynski, L. J., & Sullivan, K. E. (2007). Velocardiofacial syndrome, DiGeorge syndrome: the chromosome 22q11.2 deletion syndromes. *Lancet*. [https://doi.org/10.1016/S0140-6736\(07\)61601-8](https://doi.org/10.1016/S0140-6736(07)61601-8)
- Köcher, T., & Superti-Furga, G. (2007). Mass spectrometry-based functional proteomics: From molecular machines to protein networks. *Nature Methods*. <https://doi.org/10.1038/nmeth1093>
- Koh, G. C. K. W., Porras, P., Aranda, B., Hermjakob, H., & Orchard, S. E. (2012). Analyzing protein-protein interaction networks. *Journal of Proteome Research*, 11(4), 2014–2031. <https://doi.org/10.1021/pr201211w>
- Komono, Y., Huang, Y.-J., Qiu, J., Lin, L., Xu, Y., Zhou, Y., ... Zhang, D.-E. (2015). SRSF2 Is Essential for Hematopoiesis, and Its Myelodysplastic Syndrome-Related Mutations Dysregulate Alternative Pre-mRNA Splicing. *Molecular and Cellular Biology*, 35(17), 3071–3082. <https://doi.org/10.1128/MCB.00202-15>
- Kong, F., James, G., Gordon, S., Zelynski, A., & Gilbert, G. L. (2001). Species-Specific PCR for Identification of Common Contaminant Mollicutes in Cell Culture. *Applied and Environmental Microbiology*, 67(7), 3195–3200. <https://doi.org/10.1128/AEM.67.7.3195-3200.2001>
- Kong, K. Y. E., Tang, H. M. V., Pan, K., Huang, Z., Lee, T. H. J., Hinnebusch, A. G., ... Wong, C. M. (2014). Cotranscriptional recruitment of yeast TRAMP complex to intronic sequences promotes optimal pre-mRNA splicing. *Nucleic Acids Research*, 42(1), 643–660. <https://doi.org/10.1093/nar/gkt888>
- Korneta, I., & Bujnicki, J. M. (2012). Intrinsic Disorder in the Human Spliceosomal Proteome. *PLoS Computational Biology*, 8(8). <https://doi.org/10.1371/journal.pcbi.1002641>
- Kosho, T., Okamoto, N., Ohashi, H., Tsurusaki, Y., Imai, Y., Hibi-Ko, Y., ... Matsumoto, N. (2013). Clinical correlations of mutations affecting six components of the SWI/SNF complex: Detailed description of 21 patients and a review of the literature. *American Journal of Medical Genetics, Part A*, 161(6), 1221–1237. <https://doi.org/10.1002/ajmg.a.35933>
- Koskinen, P., Törönen, P., Nokso-Koivisto, J., & Holm, L. (2015). PANNZER: High-throughput functional annotation of uncharacterized proteins in an error-prone environment. *Bioinformatics*, 31(10), 1544–1552. <https://doi.org/10.1093/bioinformatics/btu851>
- Krallinger, M., Morgan, A., Smith, L., Leitner, F., Tanabe, L., Wilbur, J., ... Valencia, A. (2008). Evaluation of text-mining systems for biology: Overview of the Second BioCreative community challenge. *Genome Biology*. <https://doi.org/10.1186/gb-2008-9-s2-s1>
- Krauthammer, M., Kaufmann, C. A., Gilliam, T. C., & Rzhetsky, A. (2004). Molecular triangulation: Bridging linkage and molecular-network information for identifying candidate genes in Alzheimer's disease. *Proceedings of the National Academy of Sciences*, 101(42), 15148–15153. <https://doi.org/10.1073/pnas.0404315101>
- Kremer, E. J., Yu, S., Pritchard, M., Nagaraja, R., Heitz, D., Lynch, M., ... et al. (1991). Isolation of a human DNA sequence which spans the fragile X. *Am J Hum Genet*, 49(3), 656–661. Retrieved from [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=1882843](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1882843)
- Kruger, N. J. (1994). The Bradford method for protein quantitation. *Methods in Molecular Biology (Clifton, N.J.)*, 32, 9–15. <https://doi.org/10.1385/0-89603-268-X:9>
- Kuraoka, I., Ito, S., Wada, T., Hayashida, M., Lee, L., Saijo, M., ... Tanaka, K. (2008). Isolation of XAB2 complex involved in pre-mRNA splicing, transcription, and transcription-coupled repair. *Journal of Biological Chemistry*, 283(2), 940–950. <https://doi.org/10.1074/jbc.M706647200>
- Kyburz, A., Friedlein, A., Langen, H., & Keller, W. (2006). Direct Interactions between Subunits of CPSF and the U2 snRNP Contribute to the Coupling of Pre-mRNA 3' End Processing and Splicing. *Molecular Cell*, 23(2), 195–205. <https://doi.org/10.1016/j.molcel.2006.05.037>
- Labak, C. M., Wang, P. Y., Arora, R., Guda, M. R., Asuthkar, S., Tsung, A. J., & Velpula, K. K. (2016). Glucose transport:

- Meeting the metabolic demands of cancer, and applications in glioblastoma treatment. *American Journal of Cancer Research*.
- Lage, K., Karlberg, E. O., Størling, Z. M., Ólason, P. Í., Pedersen, A. G., Rigina, O., ... Brunak, S. (2007). A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nature Biotechnology*, 25(3), 309–316. <https://doi.org/10.1038/nbt1295>
- Lahner, B., Gong, J., Mahmoudian, M., Smith, E. L., Abid, K. B., Rogers, E. E., ... Salt, D. E. (2003). Genomic scale profiling of nutrient and trace elements in *Arabidopsis thaliana*. *Nature Biotechnology*, 21(10), 1215–1221. <https://doi.org/10.1038/nbt865>
- Laird, C. D., Hansen, R. S., Canfield, T. K., Lamb, M. M., & Gartler, S. M. (1993). Chromosomal fragile sites: Molecular test of the delayed-replication model. In *Cold Spring Harbor Symposia on Quantitative Biology* (Vol. 58, pp. 633–636).
- Landry, J. J. M., Pyl, P. T., Rausch, T., Zichner, T., Tekkedil, M. M., Stütz, A. M., ... Steinmetz, L. M. (2013). The genomic and transcriptomic landscape of a HeLa cell line. *G3 (Bethesda, Md.)*, 3(8), 1213–1224. <https://doi.org/10.1534/g3.113.005777>
- Landsverk, H. B., Kirkhus, M., Bollen, M., Küntziger, T., & Collas, P. (2005). PNUITS enhances in vitro chromosome decondensation in a PPI-dependent manner. *The Biochemical Journal*, 390(Pt 3), 709–717. <https://doi.org/10.1042/BJ20050678>
- Lane, A. N., & Fan, T. W. M. (2015). Regulation of mammalian nucleotide metabolism and biosynthesis. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkv047>
- Lapointe, C. P., Wilinski, D., Saunders, H. A. J., & Wickens, M. (2015). Protein-RNA networks revealed through covalent RNA marks. *Nature Methods*, 12(12), 1163–1170. <https://doi.org/10.1038/nmeth.3651>
- Lardelli, R. M., Thompson, J. X., Yates, J. R., & Stevens, S. W. (2010). Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA*, 16(3), 516–528. <https://doi.org/10.1261/rna.2030510>
- Lareau, L. F., Brooks, A. N., Soergel, D. A. W., Meng, Q., & Brenner, S. E. (2007). The coupling of alternative splicing and nonsense-mediated mRNA decay. *Advances in Experimental Medicine and Biology*, 623, 190–211. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18380348>
- Latosinska, A., Vougas, K., Makridakis, M., Klein, J., Mullen, W., Abbas, M., ... Jankowski, V. (2015). Comparative analysis of label-free and 8-plex iTRAQ approach for quantitative tissue proteomic analysis. *PLoS ONE*, 10(9). <https://doi.org/10.1371/journal.pone.0137048>
- Le, A., Lane, A. N., Hamaker, M., Bose, S., Gouw, A., Barbi, J., ... Dang, C. V. (2012). Glucose-Independent Glutamine Metabolism via TCA Cycling for Proliferation and Survival in B Cells. *Cell Metabolism*, 15(1), 110–121. <https://doi.org/10.1016/j.cmet.2011.12.009>
- Lee, B., Wu, C. Y., Lin, Y. W., Park, S. W., & Wei, L. N. (2016). Synergistic activation of Arg1 gene by retinoic acid and IL-4 involves chromatin remodeling for transcription initiation and elongation coupling. *Nucleic Acids Research*, 44(16), 7568–7579. <https://doi.org/10.1093/nar/gkw392>
- Lee, D.-S., Park, J., Kay, K. A., Christakis, N. A., Oltvai, Z. N., & Barabási, A.-L. (2008). The implications of human metabolic network topology for disease comorbidity. *Proceedings of the National Academy of Sciences of the United States of America*, 105(29), 9880–9885. <https://doi.org/10.1073/pnas.0802208105>
- Lee, P. Y., Costumbrado, J., Hsu, C.-Y., & Kim, Y. H. (2012). Agarose gel electrophoresis for the separation of DNA fragments. *Journal of Visualized Experiments : JoVE*, (62), 1–5. <https://doi.org/10.3791/3923>
- León, I. R., Schwämmle, V., Jensen, O. N., & Sprenger, R. R. (2013). Quantitative Assessment of In-solution Digestion Efficiency Identifies Optimal Protocols for Unbiased Protein Analysis. *Molecular & Cellular Proteomics*, 12(10), 2992–3005. <https://doi.org/10.1074/mcp.M112.025585>
- Letovsky, S., & Kasif, S. (2003). Predicting protein function from protein/protein interaction data: a probabilistic approach. *Bioinformatics*, 19(Suppl 1), i197–i204. <https://doi.org/10.1093/bioinformatics/btg1026>
- Letteboer, S. J. F., & Roepman, R. (2008). Versatile screening for binary protein-protein interactions by yeast two-hybrid mating. *Methods in Molecular Biology*, 484, 145–159. [https://doi.org/10.1007/978-1-59745-398-1\\_10](https://doi.org/10.1007/978-1-59745-398-1_10)
- Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C., & Mattaj, L. W. (1996). A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes and Development*, 10(13), 1683–1698. <https://doi.org/10.1101/gad.10.13.1683>
- Li, L., Mu, Q., Zhang, B., & Yan, B. (2010). Analytical strategies for detecting nanoparticle–protein interactions. *The Analyst*, 135(7), 1519. <https://doi.org/10.1039/c0an00075b>
- Li, M., Zhang, H., Wang, J., & Pan, Y. (2012). A new essential protein discovery method based on the integration of protein-protein interaction and gene expression data. *BMC Systems Biology*, 6(1), 15. <https://doi.org/10.1186/1752-0509-6-15>
- Li, Q. S., Parrado, A. R., Samtani, M. N., & Narayan, V. A. (2015). Variations in the FRA10AC1 fragile site and 15q21 are associated with Cerebrospinal fluid Aβ<sup>1-42</sup> level. *PLoS ONE*, 10(8). <https://doi.org/10.1371/journal.pone.0134000>
- Li, X., Foley, E. A., Molloy, K. R., Li, Y., Chait, B. T., & Kapoor, T. M. (2012). Quantitative chemical proteomics approach to identify post-translational modification-mediated protein-protein interactions. *Journal of the American Chemical Society*, 134(4), 1982–1985. <https://doi.org/10.1021/ja210528v>
- Li, X., Song, J., Liu, J., Wu, S., Wang, L., Gong, L., & Lin, X. (2014). Serine-arginine protein kinase 1 is associated with breast cancer progression and poor patient survival. *Medical Oncology (Northwood, London, England)*, 31(8), 83. <https://doi.org/10.1007/s12032-014-0083-8>
- Licata, L., & Orchard, S. (2016). The MintAct project and molecular interaction databases. In *Methods in Molecular Biology* (Vol. 1415, pp. 55–69). [https://doi.org/10.1007/978-1-4939-3572-7\\_3](https://doi.org/10.1007/978-1-4939-3572-7_3)
- Lin, C. Y., Lee, T. L., Chiu, Y. Y., Lin, Y. W., Lo, Y. S., Lin, C. T., & Yang, J. M. (2015). Module organization and variance in protein-protein interaction networks. *Scientific Reports*, 5. <https://doi.org/10.1038/srep09386>
- Lin, P. C., & Xu, R. M. (2012). Structure and assembly of the SF3a splicing factor complex of U2 snRNP. *EMBO Journal*, 31(6), 1579–1590. <https://doi.org/10.1038/emboj.2012.7>
- Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., & Fu, X.-D. (2008). The splicing factor SC35 has an active role in transcriptional elongation. *Nature Structural & Molecular Biology*, 15(8), 819–826. <https://doi.org/10.1038/nsmb.1461>
- Lindsay, E. a, Vitelli, F., Su, H., Morishima, M., Huynh, T., Pramparo, T., ... Baldini, a. (2001). Tbx1 haploinsufficiency in the DiGeorge syndrome region causes aortic arch defects in mice. *Nature*, 410(6824), 97–101. <https://doi.org/10.1038/35065105>
- Lipp, J. J., Marvin, M. C., Shokat, K. M., & Guthrie, C. (2015). SR protein kinases promote splicing of nonconsensus introns. *Nature Structural and Molecular Biology*, 22(8), 611–617. <https://doi.org/10.1038/nsmb.3057>

- Liu, M. M., & Zack, D. J. (2013). Alternative splicing and retinal degeneration. *Clinical Genetics*.  
<https://doi.org/10.1111/cge.12181>
- Liu, S. (2006). The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. *RNA*, *12*(7), 1418–1430.  
<https://doi.org/10.1261/rna.55406>
- Liu, Y. C., & Cheng, S. C. (2015). Functional roles of DEXD/H-box RNA helicases in Pre-mRNA splicing. *Journal of Biomedical Science*. <https://doi.org/10.1186/s12929-015-0161-z>
- Liu, Y., Ge, Q., Chan, B., Liu, H., Singh, S. R., Manley, J., ... Hou, S. X. (2016). Whole-animal genome-wide RNAi screen identifies networks regulating male germline stem cells in *Drosophila*. *Nature Communications*, *7*, 12149.  
<https://doi.org/10.1038/ncomms12149>
- Liu, Z.-P., Wang, Y., Zhang, X.-S., & Chen, L. (2012). Network-based analysis of complex diseases. *IET Systems Biology*, *6*(1), 22. <https://doi.org/10.1049/iet-syb.2010.0052>
- Lohman, G. J. S., Tabor, S., & Nichols, N. M. (2011). DNA Ligases. In *Current Protocols in Molecular Biology* (Vol. Chapter 3, p. Unit3.14). Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/0471142727.mb0314s94>
- Londry, F. A., & Hager, J. W. (2003). Mass selective axial ion ejection from a linear quadrupole ion trap. *Journal of the American Society for Mass Spectrometry*, *14*(10), 1130–1147. [https://doi.org/10.1016/S1044-0305\(03\)00446-X](https://doi.org/10.1016/S1044-0305(03)00446-X)
- Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., ... Moore, H. F. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*. <https://doi.org/10.1038/ng.2653>
- Lorenz, M. A., Burant, C. F., & Kennedy, R. T. (2011). Reducing time and increasing sensitivity in sample preparation for adherent mammalian cell metabolomics. *Analytical Chemistry*, *83*(9), 3406–3414. <https://doi.org/10.1021/ac103313x>
- Lorenz, M. G., & Wackernagel, W. (1994). Bacterial gene transfer by natural genetic transformation in the environment. *Microbiological Reviews*, *58*(3), 563–602.
- Lovero, K. L., Fukata, Y., Granger, A. J., Fukata, M., & Nicoll, R. A. (2015). The LGI1–ADAM22 protein complex directs synapse maturation through regulation of PSD-95 function. *Proceedings of the National Academy of Sciences*, *112*(30), E4129–E4137. <https://doi.org/10.1073/pnas.1511910112>
- Lu, Z. X., Jiang, P., & Xing, Y. (2012). Genetic variation of pre-mRNA alternative splicing in human populations. *Wiley Interdiscip Rev RNA*, *3*(4), 581–592. <https://doi.org/10.1002/wrna.120>
- Lucey, B. P., Nelson-Rees, W. A., & Hutchins, G. M. (2009). Henrietta Lacks, HeLa cells, and cell culture contamination. *Archives of Pathology and Laboratory Medicine*. <https://doi.org/10.1043/1543-2165-133.9.1463>
- Luco, R., Pan, Q., & Tominaga, K. (2010). Regulation of alternative splicing by histone modifications. *Science*, *327*(5968), 996–1000. <https://doi.org/10.1126/science.1184208.Regulation>
- Lukusa, T., & Fryns, J. P. (2008). Human chromosome fragility. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. <https://doi.org/10.1016/j.bbagr.2007.10.005>
- Lunt, S. Y., & Vander Heiden, M. G. (2011). Aerobic Glycolysis: Meeting the Metabolic Requirements of Cell Proliferation. *Annual Review of Cell and Developmental Biology*, *27*(1), 441–464. <https://doi.org/10.1146/annurev-cellbio-092910-154237>
- Lutz, S. M., Cho, M. H., Young, K., Hersh, C. P., Castaldi, P. J., McDonald, M. L., ... Silverman, E. K. (2015). A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC Genetics*, *16*(1). <https://doi.org/10.1186/s12863-015-0299-4>
- Ma, C.-Y., Phoebe Chen, Y.-P., Berger, B., & Liao, C.-S. (2017). Identification of Protein Complexes by Integrating Multiple Alignment of Protein Interaction Networks. *Bioinformatics*, btx043. <https://doi.org/10.1093/bioinformatics/btx043>
- Ma, N., Matsunaga, S., Morimoto, A., Sakashita, G., Urano, T., Uchiyama, S., & Fukui, K. (2011). The nuclear scaffold protein SAF-A is required for kinetochore-microtubule attachment and contributes to the targeting of Aurora-A to mitotic spindles. *Journal of Cell Science*, *124*(3), 394–404. <https://doi.org/10.1242/jcs.063347>
- Ma, N., Matsunaga, S., Takata, H., Ono-Maniwa, R., Uchiyama, S., & Fukui, K. (2007). Nucleolin functions in nucleolus formation and chromosome congression. *Journal of Cell Science*, *120*(12), 2091–2105. <https://doi.org/10.1242/jcs.008771>
- Ma, X. M., Yoon, S. O., Richardson, C. J., Jülich, K., & Blenis, J. (2008). SKAR Links Pre-mRNA Splicing to mTOR/S6K1-Mediated Enhanced Translation Efficiency of Spliced mRNAs. *Cell*, *133*(2), 303–313.  
<https://doi.org/10.1016/j.cell.2008.02.031>
- Macé, G., Bogliolo, M., Guervilly, J. H., Du Villard, J. A. D., & Rosselli, F. (2005). 3R coordination by Fanconi anemia proteins. In *Biochimie* (Vol. 87, pp. 647–658). <https://doi.org/10.1016/j.biochi.2005.05.003>
- Maeda, O., Usami, N., Kondo, M., Takahashi, M., Goto, H., Shimokata, K., ... Sekido, Y. (2004). Plakoglobin (gamma-catenin) has TCF/LEF family-dependent transcriptional activity in beta-catenin-deficient cell line. *Oncogene*, *23*(4), 964–972.  
<https://doi.org/10.1038/sj.onc.1207254>
- Maga-Nteve, C., & Klapa, M. I. (2016). Streamlining GC-MS metabolomic analysis using the M-IOLITE software suite. *IFAC-PapersOnLine*, *49*(26), 286–288. <https://doi.org/10.1016/j.ifacol.2016.12.140>
- Magalska, A., Schellhaus, A., Moreno-Andrés, D., Zanini, F., Schooley, A., Sachdev, R., ... Antonin, W. (2014). RuvB-like atpases function in chromatin decondensation at the end of mitosis. *Developmental Cell*, *31*(3), 305–318.  
<https://doi.org/10.1016/j.devcel.2014.09.001>
- Magger, O., Waldman, Y. Y., Ruppin, E., & Sharan, R. (2012). Enhancing the Prioritization of Disease-Causing Genes through Tissue Specific Protein Interaction Networks. *PLoS Computational Biology*, *8*(9).  
<https://doi.org/10.1371/journal.pcbi.1002690>
- Mahmood, T., & Yang, P. C. (2012). Western blot: Technique, theory, and trouble shooting. *North American Journal of Medical Sciences*, *4*(9), 429–434. <https://doi.org/10.4103/1947-2714.100998>
- Maier, T., Güell, M., & Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. *FEBS Letters*.  
<https://doi.org/10.1016/j.febslet.2009.10.036>
- Maita, H., Kitaura, H., Ariga, H., & Iguchi-Ariga, S. M. M. (2005). Association of PAP-1 and Prp3p, the products of causative genes of dominant retinitis pigmentosa, in the tri-snRNP complex. *Experimental Cell Research*, *302*(1), 61–68.  
<https://doi.org/10.1016/j.yexcr.2004.08.022>
- Majumder, A., Pahlberg, J., Muradov, H., Boyd, K. K., Sampath, A. P., & Artemyev, N. O. (2015). Exchange of Cone for Rod Phosphodiesterase 6 Catalytic Subunits in Rod Photoreceptors Mimics in Part Features of Light Adaptation. *Journal of Neuroscience*, *35*(24), 9225–9235. <https://doi.org/10.1523/JNEUROSCI.3563-14.2015>
- Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L., Wilm, M., & Lührmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science*, *298*(5601), 2205–2208.  
<https://doi.org/10.1126/science.1077783>
- Makarov, E. M., Owen, N., Bottrill, A., & Makarova, O. V. (2012). Functional mammalian spliceosomal complex e contains



- SMN complex proteins in addition to U1 and U2 snRNPs. *Nucleic Acids Research*, 40(6), 2639–2652. <https://doi.org/10.1093/nar/gkr1056>
- Makarova, O. V., Makarov, E. M., Urlaub, H., Will, C. L., Gentzel, M., Wilm, M., & Lührmann, R. (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *EMBO Journal*, 23(12), 2381–2391. <https://doi.org/10.1038/sj.emboj.7600241>
- Malcovati, L., Karimi, M., Papaemmanuil, E., Ambaglio, I., Jädersten, M., Jansson, M., ... Lindberg, E. H. (2015). SF3B1 mutation identifies a distinct subset of myelodysplastic syndrome with ring sideroblasts. *Blood*, 126(2), 233–241. <https://doi.org/10.1182/blood-2015-03-633537>
- Maniatis, T., & Reed, R. (2002). An extensive network of coupling among gene expression machines. *Nature*, 416(6880), 499–506. <https://doi.org/10.1038/416499a>
- Mann, M., Hendrickson, R. C., & Pandey, A. (2001). Analysis of Proteins and Proteomes by Mass Spectrometry. *Annual Review of Biochemistry*, 70(1), 437–473. <https://doi.org/10.1146/annurev.biochem.70.1.437>
- Marques, J. T., Kim, K., Wu, P.-H., Alleyne, T. M., Jafari, N., & Carthew, R. W. (2010). Loqs and R2D2 act sequentially in the siRNA pathway in Drosophila. *Nature Structural & Molecular Biology*, 17(1), 24–30. <https://doi.org/10.1038/nsmb.1735>
- Maslov, S., & Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science*, 296(5569), 910–913. <https://doi.org/10.1126/science.1065103>
- Matera, A. G., & Wang, Z. (2014). A day in the life of the spliceosome. *Nature Reviews Molecular Cell Biology*. <https://doi.org/10.1038/nrm3742>
- Mathew, R., Hartmuth, K., Möhlmann, S., Urlaub, H., Ficner, R., & Lührmann, R. (2008). Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. *Nature Structural and Molecular Biology*, 15(5), 435–443. <https://doi.org/10.1038/nsmb.1415>
- Mathivanan, S., Periaswamy, B., Gandhi, T. K., Kandasamy, K., Suresh, S., Mohmood, R., ... Pandey, A. (2006). An evaluation of human protein-protein interaction data in the public domain. *BMC Bioinformatics*, 7 Suppl 5(1471–2105 (Electronic)), S19. <https://doi.org/10.1186/1471-2105-7-S5-S19>
- Matsunaga, S., Takata, H., Morimoto, A., Hayashihara, K., Higashi, T., Akatsuchi, K., ... Fukui, K. (2012). RBMX: A Regulator for Maintenance and Centromeric Protection of Sister Chromatid Cohesion. *Cell Reports*, 1(4), 299–308. <https://doi.org/10.1016/j.celrep.2012.02.005>
- Matthews, L. R., Vaglio, P., Reboul, J., Ge, H., Davis, B. P., Garrels, J., ... Vidal, M. (2001). Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or “interologs.” *Genome Research*, 11(12), 2120–2126. <https://doi.org/10.1101/gr.205301>
- Mavrou, A., Syyrou, M., Tsenghi, C., & Metaxotou, C. (1991). Autosomal folate sensitive fragile sites in normal and mentally retarded individuals in Greece. In *American Journal of Medical Genetics* (Vol. 38, pp. 437–439). <https://doi.org/10.1002/ajmg.1320380259>
- Mayers, J. R., & Vander Heiden, M. G. (2015). Famine versus feast: Understanding the metabolism of tumors in vivo. *Trends in Biochemical Sciences*. <https://doi.org/10.1016/j.tibs.2015.01.004>
- Mayr, C., & Bartel, D. P. (2009). Widespread Shortening of 3'UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. *Cell*, 138(4), 673–684. <https://doi.org/10.1016/j.cell.2009.06.016>
- Mazza, C., Ohno, M., Segref, A., Mattaj, I. W., & Cusack, S. (2001). Crystal structure of the human nuclear cap binding complex. *Molecular Cell*, 8(2), 383–396. [https://doi.org/10.1016/S1097-2765\(01\)00299-4](https://doi.org/10.1016/S1097-2765(01)00299-4)
- McHugh, C. A., Russell, P., & Guttman, M. (2014). Methods for comprehensive experimental identification of RNA-protein interactions. *Genome Biology*. <https://doi.org/10.1186/gb4152>
- McKay, S. L., & Johnson, T. L. (2010). A bird's-eye view of post-translational modifications in the spliceosome and their roles in spliceosome dynamics. *Molecular BioSystems*, 6(11), 2093. <https://doi.org/10.1039/c002828b>
- Meek, S., Buehr, M., Sutherland, L., Thomson, A., Mullins, J. J., Smith, A. J., & Burdon, T. (2010). Efficient Gene Targeting by Homologous Recombination in Rat Embryonic Stem Cells. *PLoS ONE*, 5(12), e14225. <https://doi.org/10.1371/journal.pone.0014225>
- Memmott, J., Waser, N. M., & Price, M. V. (2004). Tolerance of pollination networks to species extinctions. *Proceedings of the Royal Society B: Biological Sciences*, 271(1557), 2605–2611. <https://doi.org/10.1098/rspb.2004.2909>
- Merkhofer, E. C., Hu, P., & Johnson, T. L. (2014). Introduction to cotranscriptional RNA splicing. *Methods in Molecular Biology*. [https://doi.org/10.1007/978-1-62703-980-2\\_6](https://doi.org/10.1007/978-1-62703-980-2_6)
- Metsu, S., Rooms, L., Rainger, J., Taylor, M. S., Bengani, H., Wilson, D. L., ... Kooy, R. F. (2014). FRA2A Is a CGG Repeat Expansion Associated with Silencing of AFF3. *PLoS Genetics*, 10(4). <https://doi.org/10.1371/journal.pgen.1004242>
- Miccoli, L., Denis, D. S. F., Frouin, I., Harper, F., Maga, G., & Angulo, J. F. (2003). Selective interactions of human kin17 and RPA proteins with chromatin and the nuclear matrix in a DNA damage- and cell cycle-regulated manner. *Nucleic Acids Research*, 31(14), 4162–4175. <https://doi.org/10.1093/nar/gkg459>
- Michaelis, R. C., Velagaleti, G. V. N., Jones, C., Pivnick, E. K., Phelan, M. C., Boyd, E., ... Tharapel, A. T. (1998). Most jacobson syndrome deletion breakpoints occur distal to FRA11b. *American Journal of Medical Genetics*, 76(3), 222–228. [https://doi.org/10.1002/\(SICI\)1096-8628\(19980319\)76:3<222::AID-AJMG5>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1096-8628(19980319)76:3<222::AID-AJMG5>3.0.CO;2-S)
- Michlewski, G., Sanford, J. R., & Cáceres, J. F. (2008). The Splicing Factor SF2/ASF Regulates Translation Initiation by Enhancing Phosphorylation of 4E-BP1. *Molecular Cell*, 30(2), 179–189. <https://doi.org/10.1016/j.molcel.2008.03.013>
- Millevoi, S., & Vagner, S. (2010). Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Research*, 38(9), 2757–2774. <https://doi.org/10.1093/nar/gkp1176>
- Mitra, K., Carvunis, A.-R., Ramesh, S. K., & Ideker, T. (2013). Integrative approaches for finding modular structure in biological networks. *Nature Reviews Genetics*, 14(10), 719–732. <https://doi.org/10.1038/nrg3552>
- Moehle, E. A., Braberg, H., Krogan, N. J., & Guthrie, C. (2014). Adventures in time and space: Splicing efficiency and RNA polymerase II elongation rate. *RNA Biology*. <https://doi.org/10.4161/rna.28646>
- Moir, R. D., Lee, J., & Willis, I. M. (2012). Recovery of RNA polymerase III transcription from the glycerol-repressed state: Revisiting the role of protein kinase CK2 in MAF1 phosphorylation. *Journal of Biological Chemistry*, 287(36), 30833–30841. <https://doi.org/10.1074/jbc.M112.378828>
- Montes, M., Becerra, S., Sánchez-Álvarez, M., & Suñé, C. (2012). Functional coupling of transcription and splicing. *Gene*. <https://doi.org/10.1016/j.gene.2012.04.006>
- Mooney, S., Leuendorf, J. E., Hendrickson, C., & Hellmann, H. (2009). Vitamin B6: A long known compound of surprising complexity. *Molecules*. <https://doi.org/10.3390/molecules14010329>
- Morante-Redolat, J. M., Gorostidi-Pagola, A., Piquer-Sirerol, S., Sáenz, A., Poza, J. J., Galán, J., ... López de Munain, A. (2002). Mutations in the LGI1/Epitempin gene on 10q24 cause autosomal dominant lateral temporal epilepsy. *Human*

- Molecular Genetics*, 11(9), 1119–1128. <https://doi.org/10.1093/hmg/11.9.1119>
- Mordes, D., Yuan, L., Xu, L., Kawada, M., Molday, R. S., & Wu, J. Y. (2007). Identification of photoreceptor genes affected by PRPF31 mutations associated with autosomal dominant retinitis pigmentosa. *Neurobiology of Disease*, 26(2), 291–300. <https://doi.org/10.1016/j.nbd.2006.08.026>
- Morris, A. R., Mukherjee, N., & Keene, J. D. (2010). Systematic analysis of posttranscriptional gene expression. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*. <https://doi.org/10.1002/wsbm.54>
- Moumen, A., Masterson, P., O'Connor, M. J., & Jackson, S. P. (2005). hnRNP K: An HDM2 target and transcriptional coactivator of p53 in response to DNA damage. *Cell*, 123(6), 1065–1078. <https://doi.org/10.1016/j.cell.2005.09.032>
- Mrasek, K., Schoder, C., Teichmann, A. C., Behr, K., Franze, B., Wilhelm, K., ... Weise, A. (2010). Global screening and extended nomenclature for 230 aphidicolin-inducible fragile sites, including 61 yet unreported ones. *International Journal of Oncology*, 36(4), 929–940. <https://doi.org/10.3892/ijo-00000572>
- Mruk, D. D., & Cheng, C. Y. (2011). Enhanced chemiluminescence (ECL) for routine immunoblotting. *Spermatogenesis*, 1(2), 121–122. <https://doi.org/10.4161/spmg.1.2.16606>
- Mullen, A. R., Wheaton, W. W., Jin, E. S., Chen, P. H., Sullivan, L. B., Cheng, T., ... Deberardinis, R. J. (2012). Reductive carboxylation supports growth in tumour cells with defective mitochondria. *Nature*, 481(7381), 385–388. <https://doi.org/10.1038/nature10642>
- Murali, T., Pacifico, S., Yu, J., Guest, S., Roberts, G. G., & Finley, R. L. (2011). DroID 2011: A comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila. *Nucleic Acids Research*, 39(SUPPL. 1). <https://doi.org/10.1093/nar/gkq1092>
- Murray-Stewart, T. R., Woster, P. M., & Casero, R. A. (2016). Targeting polyamine metabolism for cancer therapy and prevention. *The Biochemical Journal*, 473(19), 2937–2953. <https://doi.org/10.1042/BCJ20160383>
- Muschet, C., Mühlner, G., Prehn, C., de Angelis, M. H., Adamski, J., & Tokarz, J. (2016). Removing the bottlenecks of cell culture metabolomics: fast normalization procedure, correlation of metabolites to cell number, and impact of the cell harvesting method. *Metabolomics*, 12(10). <https://doi.org/10.1007/s11306-016-1104-8>
- Nabieva, E., Jim, K., Agarwal, A., Chazelle, B., & Singh, M. (2005). Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps. *Bioinformatics*, 21(SUPPL. 1). <https://doi.org/10.1093/bioinformatics/bti1054>
- Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., ... Mann, M. (2011). Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular Systems Biology*, 7. <https://doi.org/10.1038/msb.2011.81>
- Nakao, M., Barrero, R. A., Mukai, Y., Motono, C., Suwai, M., & Nakai, K. (2005). Large-scale analysis of human alternative protein isoforms: Pattern classification and correlation with subcellular localization signals. *Nucleic Acids Research*, 33(8), 2355–2363. <https://doi.org/10.1093/nar/gki520>
- Nancarrow, J. K., Kremer, E., Holman, K., Eyre, H., Doggett, N. A., Le Paslier, D., ... Richards, R. I. (1994). Implications of FRA16A structure for the mechanism of chromosomal fragile site genesis. *Science*, 264(5167), 1938–1941. <https://doi.org/10.1126/science.8009225>
- Nepusz, T., Yu, H., & Paccanaro, A. (2012). Detecting overlapping protein complexes in protein-protein interaction networks. *Nature Methods*, 9(5), 471–472. <https://doi.org/10.1038/nmeth.1938>
- Nesvizhskii, A. I. (2012). Computational and informatics strategies for identification of specific protein interaction partners in affinity purification mass spectrometry experiments. *Proteomics*. <https://doi.org/10.1002/pmic.201100537>
- Neumann, B., Walter, T., Hériché, J. K., Bulkescher, J., Erfle, H., Conrad, C., ... Ellenberg, J. (2010). Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature*, 464(7289), 721–727. <https://doi.org/10.1038/nature08869>
- Newman, M. E. J. (2002). Assortative Mixing in Networks. *Physical Review Letters*, 89(20). <https://doi.org/10.1103/PhysRevLett.89.208701>
- Ngounou Wetie, A. G., Sokolowska, I., Woods, A. G., Roy, U., Deinhardt, K., & Darie, C. C. (2014). Protein-protein interactions: Switch from classical methods to proteomics and bioinformatics-based approaches. *Cellular and Molecular Life Sciences*. <https://doi.org/10.1007/s00018-013-1333-1>
- Nguyen, P. V., Srihari, S., & Leong, H. W. (2013). Identifying conserved protein complexes between species by constructing interolog networks. *BMC Bioinformatics*, 14(SUPPL16). <https://doi.org/10.1186/1471-2105-14-S16-S8>
- Ni, J. Z., Grate, L., Donohue, J. P., Preston, C., Nobida, N., O'Brien, G., ... Ares, M. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes & Development*, 21(6), 708–718. <https://doi.org/10.1101/gad.1525507>
- Nicholson, J. K., & Wilson, I. D. (2003). Understanding “Global” Systems Biology: Metabonomics and the Continuum of Metabolism. *Nature Reviews Drug Discovery*, 2(8), 668–676. <https://doi.org/10.1038/nrd1157>
- Nielsen, J., & Oliver, S. (2005). The next wave in metabolome analysis. *Trends in Biotechnology*. <https://doi.org/10.1016/j.tibtech.2005.08.005>
- Nilse, L., Sigloch, F. C., Biniossek, M. L., & Schilling, O. (2015). Toward improved peptide feature detection in quantitative proteomics using stable isotope labeling. *PROTEOMICS - Clinical Applications*, 9(7–8), 706–714. <https://doi.org/10.1002/prca.201400173>
- Niwa, M., Rose, S. D., & Berget, S. M. (1990). In vitro polyadenylation is stimulated by the presence of an upstream intron. *Genes and Development*, 4(9), 1552–1559. <https://doi.org/10.1101/gad.4.9.1552>
- Noma, K., Goncharov, A., & Jin, Y. (2014). Systematic analyses of rpm-1 suppressors reveal roles for ESS-2 in mRNA splicing in *Caenorhabditis elegans*. *Genetics*, 198(3), 1101–1115. <https://doi.org/10.1534/genetics.114.167841>
- Nussbacher, J. K., Batra, R., Lagier-Tourenne, C., & Yeo, G. W. (2015). RNA-binding proteins in neurodegeneration: Seq and you shall receive. *Trends in Neurosciences*. <https://doi.org/10.1016/j.tins.2015.02.003>
- Obdlik, P., El-Bakkoury, M., Hamacher, T., Cappellaro, C., Vilarino, C., Fleischer, C., ... Frommer, W. B. (2004). K+ channel interactions detected by a genetic system optimized for systematic studies of membrane protein interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 101(33), 12242–12247. <https://doi.org/10.1073/pnas.0404467101>
- Ohr, T., Merkle, D., Birkenfeld, K., Echeverri, C. J., & Schulle, P. (2006). In situ fluorescence analysis demonstrates active siRNA exclusion from the nucleus by Exportin 5. *Nucleic Acids Research*, 34(5), 1369–1380. <https://doi.org/10.1093/nar/gkl001>
- Oka, S., Kato, J., & Moss, J. (2006). Identification and characterization of a mammalian 39-kDa poly(ADP-ribose) glycohydrolase. *Journal of Biological Chemistry*, 281(2), 705–713. <https://doi.org/10.1074/jbc.M510290200>
- Orchard, S., Amari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., ... Hermjakob, H. (2014). The MIntAct project - IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Research*, 42(D1).

- <https://doi.org/10.1093/nar/gkt1115>
- Orchard, S., Kerrien, S., Abbani, S., Aranda, B., Bhate, J., Bidwell, S., ... Hermjakob, H. (2012). Protein interaction data curation: The International Molecular Exchange (IMEx) consortium. *Nature Methods*. <https://doi.org/10.1038/nmeth.1931>
- Orrantia, E., & Chang, P. L. (1990). Intracellular distribution of DNA internalized through calcium phosphate precipitation. *Experimental Cell Research*, *190*(2), 170–174. [https://doi.org/10.1016/0014-4827\(90\)90181-9](https://doi.org/10.1016/0014-4827(90)90181-9)
- Oti, M., Snel, B., Huynen, M. A., & Brunner, H. G. (2006). Predicting disease genes using protein-protein interactions. *Journal of Medical Genetics*, *43*(8), 691–698. <https://doi.org/10.1136/jmg.2006.041376>
- Ou-Yang, L., Dai, D.-Q., Li, X.-L., Wu, M., Zhang, X.-F., & Yang, P. (2014a). Detecting temporal protein complexes from dynamic protein-protein interaction networks. *BMC Bioinformatics*, *15*(1), 335. <https://doi.org/10.1186/1471-2105-15-335>
- Ou-Yang, L., Dai, D. Q., Li, X. L., Wu, M., Zhang, X. F., & Yang, P. (2014b). Detecting temporal protein complexes from dynamic protein-protein interaction networks. *BMC Bioinformatics*, *15*(1). <https://doi.org/10.1186/1471-2105-15-335>
- Owen, O. E., Kalthan, S. C., & Hanson, R. W. (2002). The key role of anaplerosis and cataplerosis for citric acid cycle function. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.R200006200>
- Pabis, M., Neufeld, N., Steiner, M. C., Bojic, T., Shav-Tal, Y., & Neugebauer, K. M. (2013). The nuclear cap-binding complex interacts with the U4/U6-U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. *RNA (New York, N.Y.)*, *19*(8), 1054–1063. <https://doi.org/10.1261/rna.037069.112>
- Pacheco, T. R., Moita, L. F., Gomes, A. Q., Hacoheh, N., & Carmo-Fonseca, M. (2006). RNA interference knockdown of hU2AF35 impairs cell cycle progression and modulates alternative splicing of Cdc25 transcripts. *Molecular Biology of the Cell*, *17*(10), 4187–4199. <https://doi.org/10.1091/mbc.E06-01-0036>
- Padi, M., & Quackenbush, J. (2015). Integrating transcriptional and protein interaction networks to prioritize condition-specific master regulators. *BMC Systems Biology*, *9*(1), 80. <https://doi.org/10.1186/s12918-015-0228-1>
- Pagani, I., Liolios, K., Jansson, J., Chen, I.-M. A., Smirnova, T., Nosrat, B., ... Kyrpidis, N. C. (2012). The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Research*, *40*(Database issue), D571-9. <https://doi.org/10.1093/nar/gkr1100>
- Pan, K., Lee, J. T. H., Huang, Z., & Wong, C. M. (2015). Coupling and coordination in gene expression processes with pre-mRNA splicing. *International Journal of Molecular Sciences*. <https://doi.org/10.3390/ijms16035682>
- Pan, Q., Saltzman, A. L., Kim, Y. K., Misquitta, C., Shai, O., Maquat, L. E., ... Blencowe, B. J. (2006). Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. *Genes & Development*, *20*(2), 153–158. <https://doi.org/10.1101/gad.1382806>
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, *40*(12), 1413–1415. <https://doi.org/10.1038/ng.259>
- Papasaiakas, P., Tejedor, J. R., Vigevani, L., & Valcárcel, J. (2015). Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. *Molecular Cell*, *57*(1), 7–22. <https://doi.org/10.1016/j.molcel.2014.10.030>
- Papasaiakas, P., & Valcárcel, J. (2015). The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends in Biochemical Sciences*, *41*(1), 33–45. <https://doi.org/10.1016/j.tibs.2015.11.003>
- Park, J. M., Park, J. H., Mun, D. G., Bae, J., Jung, J. H., Back, S., ... Lee, S. W. (2015). Integrated analysis of global proteome, phosphoproteome, and glycoproteome enables complementary interpretation of disease-related protein networks. *Scientific Reports*, *5*. <https://doi.org/10.1038/srep18189>
- Parrish, J. E., Oostra, B. A., Verkerk, A. J. M. H., Richards, C. S., Reynolds, J., Spikes, A. S., ... Nelson, D. L. (1994). Isolation of a GCC repeat showing expansion in FRAXF, a fragile site distal to FRAXA and FRAXE. *Nature Genetics*, *8*(3), 229–235. <https://doi.org/10.1038/ng1194-229>
- Parrish, J. R., Gulyas, K. D., & Finley, R. L. (2006). Yeast two-hybrid contributions to interactome mapping. *Current Opinion in Biotechnology*. <https://doi.org/10.1016/j.copbio.2006.06.006>
- Pascarella, A., Ferrandino, G., Credendino, S. C., Moccia, C., D'Angelo, F., Miranda, B., ... Amendola, E. (2018). DNAJC17 is localized in nuclear speckles and interacts with splicing machinery components. *Scientific Reports*, *8*(1), 7794. <https://doi.org/10.1038/s41598-018-26093-1>
- Patel, D. M., & Green, K. J. (2014). Desmosomes in the heart: A review of clinical and mechanistic analyses. *Cell Communication and Adhesion*. <https://doi.org/10.3109/15419061.2014.906533>
- Patel, S., & Ahmed, S. (2015). Emerging field of metabolomics: Big promise for cancer biomarker identification and drug discovery. *Journal of Pharmaceutical and Biomedical Analysis*, *107*, 63–74. <https://doi.org/10.1016/j.jpba.2014.12.020>
- Patterson, S. D., & Aebersold, R. H. (2003). Proteomics: the first decade and beyond. *Nature Genetics*, *33*(3s), 311–323. <https://doi.org/10.1038/ng1106>
- Paushkin, S., Gubitza, A. K., Massenet, S., & Dreyfuss, G. (2002). The SMN complex, an assemblysome of ribonucleoproteins. *Current Opinion in Cell Biology*. [https://doi.org/10.1016/S0955-0674\(02\)00332-0](https://doi.org/10.1016/S0955-0674(02)00332-0)
- Pavlopoulos, G. A., Secrier, M., Moschopoulos, C. N., Soldatos, T. G., Kossida, S., Aerts, J., ... Bagos, P. G. (2011). Using graph theory to analyze biological networks. *BioData Mining*. <https://doi.org/10.1186/1756-0381-4-10>
- Pearson, C. E., & Sinden, R. R. (1996). Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry*, *35*(15), 5041–5053. <https://doi.org/10.1021/bi9601013>
- Pellegrini, M., Baglioni, M., & Geraci, F. (2016). Protein complex prediction for large protein-protein interaction networks with the Core&Peel method. *BMC Bioinformatics*, *17*. <https://doi.org/10.1186/s12859-016-1191-6>
- Pelliccia, F., Bosco, N., Curatolo, A., & Rocchi, A. (2008). Replication timing of two human common fragile sites: FRA1H and FRA2G. *Cytogenetic and Genome Research*, *121*(3–4), 196–200. <https://doi.org/10.1159/000138885>
- Penny, L., Dell'Aquila, M., Jones, M. C., Bergoffen, J., Cunniff, C., Fryns, J.-P., ... Evans, G. (1995). Clinical and molecular characterization of patients with distal 11q deletions. *American Journal of Human Genetics*, *56*, 676–683.
- Peters, L. L., Robledo, R. F., Bult, C. J., Churchill, G. A., Paigen, B. J., & Svenson, K. L. (2007). The mouse as a model for human biology: a resource guide for complex trait analysis. *Nature Reviews Genetics*, *8*(1), 58–69. <https://doi.org/10.1038/nrg2025>
- Petit, P., Fryns, J. P., Berghe, H. van den, & Hecht, F. (1986). Population cytogenetics of autosomal fragile sites. *Clinical Genetics*, *29*(2), 96–100. <https://doi.org/10.1111/j.1399-0004.1986.tb01229.x>
- Pham, K., Pal, R., Qu, Y., Liu, X., Yu, H., Shiao, S. L., ... Cheng, N. (2015). Nuclear glutaredoxin 3 is critical for protection against oxidative stress-induced cell death. *Free Radical Biology and Medicine*, *85*, 197–206. <https://doi.org/10.1016/j.freeradbiomed.2015.05.003>

- Phan, L. M., Yeung, S.-C. J., & Lee, M.-H. (2014). Cancer metabolic reprogramming: importance, main features, and potentials for precise targeted anti-cancer therapies. *Cancer Biology & Medicine*, *11*(1), 1–19. <https://doi.org/10.7497/j.issn.2095-3941.2014.01.001>
- Phang, J. M., Liu, W., Hancock, C. N., & Fischer, J. W. (2015). Proline metabolism and cancer. *Current Opinion in Clinical Nutrition and Metabolic Care*, *18*(1), 71–77. <https://doi.org/10.1097/MCO.0000000000000121>
- Phelan, M. C., & Lawler, G. (2001). Cell Counting. In *Current Protocols in Cytometry* (Vol. Appendix 3, p. Appendix 3A). Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/0471142956.cya03as00>
- Pichiorri, F., Ishii, H., Okumura, H., Trapasso, F., Wang, Y., & Huebner, K. (2008). Molecular parameters of genome instability: roles of fragile genes at common fragile sites. *Journal of Cellular Biochemistry*, *104*(5), 1525–1533. <https://doi.org/10.1002/jcb.21560>
- Picotti, P., Bodenmiller, B., & Aebersold, R. (2013). Proteomics meets the scientific method. *Nature Methods*, *10*(1), 24–27. <https://doi.org/10.1038/nmeth.2291>
- Pieretti, M., Zhang, F., Fu, Y. H., Warren, S. T., Oostra, B. A., Caskey, C. T., & Nelson, D. L. (1991). Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell*, *66*(4), 817–822. [https://doi.org/10.1016/0092-8674\(91\)90125-I](https://doi.org/10.1016/0092-8674(91)90125-I)
- Pires, M. M., Cantor, M., Guimarães, P. R., De Aguiar, M. A. M., Dos Reis, S. F., & Coltri, P. P. (2015). The network organization of protein interactions in the spliceosome is reproduced by the simple rules of food-web models. *Scientific Reports*, *5*. <https://doi.org/10.1038/srep14865>
- Piston, D. W., & Kremers, G.-J. (2007). Fluorescent protein FRET: the good, the bad and the ugly. *Trends in Biochemical Sciences*, *32*(9), 407–414. <https://doi.org/10.1016/j.tibs.2007.08.003>
- Pitt, J. J. (2009). Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *The Clinical Biochemist Reviews*, *30*(1), 19–34. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/19224008>
- Pontén, F. K., Schwenk, J. M., Asplund, A., & Edqvist, P. H. D. (2011). The Human Protein Atlas as a proteomic resource for biomarker discovery. *Journal of Internal Medicine*. <https://doi.org/10.1111/j.1365-2796.2011.02427.x>
- Proudfoot, N. J., Furger, A., & Dye, M. J. (2002). Integrating mRNA processing with transcription. *Cell*. [https://doi.org/10.1016/S0092-8674\(02\)00617-7](https://doi.org/10.1016/S0092-8674(02)00617-7)
- Pržulj, N., Wigle, D. A., & Jurisica, I. (2004). Functional topology in a network of protein interactions. *Bioinformatics*, *20*(3), 340–348. <https://doi.org/10.1093/bioinformatics/btg415>
- Pujana, M. A., Han, J. D. J., Starita, L. M., Stevens, K. N., Tewari, M., Ahn, J. S., ... Vidal, M. (2007). Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nature Genetics*, *39*(11), 1338–1349. <https://doi.org/10.1038/ng.2007.2>
- Pyle, A. M. (2008). Translocation and Unwinding Mechanisms of RNA and DNA Helicases. *Annual Review of Biophysics*, *37*(1), 317–336. <https://doi.org/10.1146/annurev.biophys.37.032807.125908>
- Qi, T., Tang, W., Wang, L., Zhai, L., Guo, L., & Zeng, X. (2011). G-actin participates in RNA polymerase II-dependent transcription elongation by recruiting positive transcription elongation factor b (P-TEFb). *Journal of Biological Chemistry*, *286*(17), 15171–15181. <https://doi.org/10.1074/jbc.M110.184374>
- Qu, Y., Wang, J., Ray, P. S., Guo, H., Huang, J., Shin-Sim, M., ... Cui, X. (2011). Thioredoxin-like 2 regulates human cancer cell growth and metastasis via redox homeostasis and NF-κB signaling. *The Journal of Clinical Investigation*, *121*(1), 212–225. <https://doi.org/10.1172/JCI43144>
- Quesada, V., Conde, L., Villamor, N., Ordonez, G. R., Jares, P., Bassaganyas, L., ... Lopez-Otin, C. (2012). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat Genet*, *44*(1), 47–52. <https://doi.org/10.1038/ng.1032>
- Ragland, R. L., Glynn, M. W., Arlt, M. F., & Glover, T. W. (2008). Stably transfected common fragile site sequences exhibit instability at ectopic sites. *Genes Chromosomes and Cancer*, *47*(10), 860–872. <https://doi.org/10.1002/gcc.20591>
- Rain, J. C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., ... Legrain, P. (2001). The protein-protein interaction map of *Helicobacter pylori*. *Nature*, *409*(6817), 211–215. <https://doi.org/10.1038/35051615>
- Raj, B., & Blencowe, B. J. (2015). Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. *Neuron*. <https://doi.org/10.1016/j.neuron.2015.05.004>
- Rakesh, R., Joseph, A. P., Bhaskara, R. M., & Srinivasan, N. (2016). Structural and mechanistic insights into human splicing factor SF3b complex derived using an integrated approach guided by the cryo-EM density maps. *RNA Biology*, *13*(10), 1025–1040. <https://doi.org/10.1080/15476286.2016.1218590>
- Ramírez-Peinado, S., León-Annicchiarico, C. L., Galindo-Moreno, J., Iurlaro, R., Caro-Maldonado, A., Prehn, J. H. M., ... Muñoz-Pinedo, C. (2013). Glucose-starved cells do not engage in prosurvival autophagy. *Journal of Biological Chemistry*, *288*(42), 30387–30398. <https://doi.org/10.1074/jbc.M113.490581>
- Rao, V. S., Srinivas, K., Sujini, G. N., & Kumar, G. N. S. (2014). Protein-Protein Interaction Detection: Methods and Analysis. *International Journal of Proteomics*, *2014*, 1–12. <https://doi.org/10.1155/2014/147648>
- Rappsilber, J., Ryder, U., Lamond, A. I., & Mann, M. (2002). Large-Scale Proteomic Analysis of the Human Spliceosome Large-Scale Proteomic Analysis of the Human Spliceosome, 1231–1245. <https://doi.org/10.1101/gr.473902>
- Rath, A., Cunningham, F., & Deber, C. M. (2013). Acrylamide concentration determines the direction and magnitude of helical membrane protein gel shifts. *Proceedings of the National Academy of Sciences*, *110*(39), 15668–15673. <https://doi.org/10.1073/pnas.1311305110>
- Raulien, N., Friedrich, K., Strobel, S., Rubner, S., Baumann, S., von Bergen, M., ... Wagner, U. (2017). Fatty acid oxidation compensates for lipopolysaccharide-induced Warburg effect in glucose-deprived monocytes. *Frontiers in Immunology*, *8*(MAY). <https://doi.org/10.3389/fimmu.2017.00609>
- Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A. L. (2002). Hierarchical Organization of Modularity in Metabolic Networks. *Science*, *297*(5586), 1551–1555. <https://doi.org/10.1126/science.1073374>
- Reed, R. (2003). Coupling transcription, splicing and mRNA export. *Current Opinion in Cell Biology*. [https://doi.org/10.1016/S0955-0674\(03\)00048-6](https://doi.org/10.1016/S0955-0674(03)00048-6)
- Reguly, T., Breitkreutz, A., Boucher, L., Breitkreutz, B. J., Hon, G. C., Myers, C. L., ... Tyers, M. (2006). Comprehensive curation and analysis of global interaction networks in *Saccharomyces cerevisiae*. *Journal of Biology*, *5*. <https://doi.org/10.1186/jbiol36>
- Reshmi, G., Chandra, S. S., Babu, V. J., Babu, P. S., Santhi, W. S., Ramachandran, S., ... Pillai, M. R. (2011). Identification and analysis of novel microRNAs from fragile sites of human cervical cancer: computational and experimental approach. *Genomics*, *97*(6), 333–340. <https://doi.org/10.1016/j.ygeno.2011.02.010>
- Rill, R. L., Beheshhti, A., & Van Winkle, D. H. (2002). DNA electrophoresis in agarose gels: Effects of field and gel concentration on the exponential dependence of reciprocal mobility on DNA length. *Electrophoresis*, *23*(16), 2710–2719.

- [https://doi.org/10.1002/1522-2683\(200208\)23:16<2710::AID-ELPS2710>3.0.CO;2-0](https://doi.org/10.1002/1522-2683(200208)23:16<2710::AID-ELPS2710>3.0.CO;2-0)
- Rio, D. C., Ares, M., Hannon, G. J., & Nilsen, T. W. (2010). Purification of RNA Using TRIzol (TRI Reagent). *Cold Spring Harbor Protocols*, 2010(6), pdb.prot5439-prot5439. <https://doi.org/10.1101/pdb.prot5439>
- Roberts, G. G., Parrish, J. R., Mangiola, B. A., & Finley, R. L. (2012). High-throughput yeast two-hybrid screening. *Methods in Molecular Biology*, 812, 39–61. [https://doi.org/10.1007/978-1-61779-455-1\\_3](https://doi.org/10.1007/978-1-61779-455-1_3)
- Roberts, J. D., Herkert, J. C., Rutberg, J., Nikkel, S. M., Wiesfeld, A. C. P., Dooijes, D., ... Gollob, M. H. (2013). Detection of genomic deletions of PKP2 in arrhythmogenic right ventricular cardiomyopathy. *Clinical Genetics*, 83(5), 452–456. <https://doi.org/10.1111/j.1399-0004.2012.01950.x>
- Rodgers, J. T., Lerin, C., Gerhart-Hines, Z., & Puigserver, P. (2008). Metabolic adaptations through the PGC-1 alpha and SIRT1 pathways. *FEBS Letters*, 582(1), 46–53. <https://doi.org/10.1016/j.febslet.2007.11.034>
- Rolland, T., Taşan, M., Charlotheaux, B., Pevzner, S. J., Zhong, Q., Sahni, N., ... Vidal, M. (2014). A proteome-scale map of the human interactome network. *Cell*, 159(5), 1212–1226. <https://doi.org/10.1016/j.cell.2014.10.050>
- Rosner, B., & Grove, D. (1999). Use of the Mann-Whitney U-test for clustered data. *Statistics in Medicine*, 18(11), 1387–1400. [https://doi.org/10.1002/\(SICI\)1097-0258\(19990615\)18:11<1387::AID-SIM126>3.0.CO;2-V](https://doi.org/10.1002/(SICI)1097-0258(19990615)18:11<1387::AID-SIM126>3.0.CO;2-V)
- Rossi, S., Seignani, C., Nnadi, S. C., Siracusa, L. D., & Calin, G. A. (2008). Cancer-associated genomic regions (CAGRs) and noncoding RNAs: Bioinformatics and therapeutic implications. *Mammalian Genome*. <https://doi.org/10.1007/s00335-008-9119-8>
- Rual, J. F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., ... Vidal, M. (2005). Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, 437(7062), 1173–1178. <https://doi.org/10.1038/nature04209>
- Ruiz-Herrera, A., & Robinson, T. J. (2007). Chromosomal instability in Afrotheria: Fragile sites, evolutionary breakpoints and phylogenetic inference from genome sequence assemblies. *BMC Evolutionary Biology*, 7. <https://doi.org/10.1186/1471-2148-7-199>
- Safari-Alighiarloo, N., Taghizadeh, M., Rezaei-Tavirani, M., Goliaei, B., & Peyvandi, A. A. (2014). Protein-protein interaction networks (PP) and complex diseases. *Gastroenterology and Hepatology from Bed to Bench*.
- Saha, S., Chatterjee, P., Basu, S., Kundu, M., & Nasipuri, M. (2014). FunPred-1: Protein function prediction from a protein interaction network using neighborhood analysis. *Cellular and Molecular Biology Letters*, 19(4), 675–691. <https://doi.org/10.2478/s11658-014-0221-5>
- Sahni, N., Yi, S., Taipale, M., Fuxman Bass, J. L., Coulombe-Huntington, J., Yang, F., ... Vidal, M. (2015). Widespread macromolecular interaction perturbations in human genetic disorders. *Cell*, 161(3), 647–660. <https://doi.org/10.1016/j.cell.2015.04.013>
- Saldi, T., Cortazar, M. A., Sheridan, R. M., & Bentley, D. L. (2016). Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *Journal of Molecular Biology*. <https://doi.org/10.1016/j.jmb.2016.04.017>
- Saltzman, A. L., Kim, Y. K., Pan, Q., Fagnani, M. M., Maquat, L. E., & Blencowe, B. J. (2008). Regulation of multiple core spliceosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay. *Molecular and Cellular Biology*, 28(13), 4320–4330. <https://doi.org/10.1128/MCB.00361-08>
- Salwinski, L. (2004). The Database of Interacting Proteins: 2004 update. *Nucleic Acids Research*, 32(90001), 449D–451. <https://doi.org/10.1093/nar/gkh086>
- Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor laboratory press. New York. [https://doi.org/574.873224\\_1/1989](https://doi.org/574.873224_1/1989)
- Sandberg, A. S., Branca, R. M. M., Lehtiö, J., & Forshed, J. (2014). Quantitative accuracy in mass spectrometry based proteomics of complex samples: The impact of labeling and precursor interference. *Journal of Proteomics*, 96, 133–144. <https://doi.org/10.1016/j.jprot.2013.10.035>
- Santoni, D., Swiercz, A., Żmieńko, A., Kasprzak, M., Blazewicz, M., Bertolazzi, P., & Felici, G. (2014). An Integrated Approach (CLuster Analysis Integration Method) to Combine Expression Data and Protein-Protein Interaction Networks in Agrigenomics: Application on *Arabidopsis thaliana*. *OMICS: A Journal of Integrative Biology*, 18(2), 155–165. <https://doi.org/10.1089/omi.2013.0050>
- Santoyo-Ramos, P., Cristina, M., & Robles-Flores, M. (2012). The Role of O-Linked  $\beta$ -N-Acetylglucosamine (GlcNAc) Modification in Cell Signaling. In *Glycosylation*. InTech. <https://doi.org/10.5772/47874>
- Sarafidou, T., Kahl, C., Martinez-Garay, I., Mangelsdorf, M., Gesk, S., Baker, E., ... Gécz, J. (2004). Folate-sensitive fragile site FRA10A is due to an expansion of a CGG repeat in a novel gene, FRA10AC1, encoding a nuclear protein. *Genomics*, 84(1), 69–81. <https://doi.org/10.1016/j.ygeno.2003.12.017>
- Sasaki-Osugi, K., Imoto, C., Takahara, T., Shibata, H., & Maki, M. (2013). Nuclear ALG-2 protein interacts with Ca<sup>2+</sup> homeostasis endoplasmic reticulum protein (CHERP) Ca<sup>2+</sup>-dependently and participates in regulation of alternative splicing of inositol trisphosphate receptor type 1 (IP3R1) Pre-mRNA. *Journal of Biological Chemistry*, 288(46), 33361–33375. <https://doi.org/10.1074/jbc.M113.497479>
- Sasikumar, A. N., Perez, W. B., & Kinzy, T. G. (2012). The many roles of the eukaryotic elongation factor 1 complex. *Wiley Interdisciplinary Reviews: RNA*. <https://doi.org/10.1002/wrna.1118>
- Schaberg, K. E., Shirure, V. S., Worley, E. A., George, S. C., & Naegle, K. M. (2017). Ensemble clustering of phosphoproteomic data identifies differences in protein interactions and cell-cell junction integrity of HER2-overexpressing cells. *Integrative Biology: Quantitative Biosciences from Nano to Macro*, 9(6), 539–547. <https://doi.org/10.1039/c7ib00054e>
- Schmidt, A., Langbein, L., Rode, M., Prätzel, S., Zimbelmann, R., & Franke, W. W. (1997). Plakophilins 1a and 1b: Widespread nuclear proteins recruited in specific epithelial cells as desmosomal plaque components. *Cell and Tissue Research*, 290(3), 481–499. <https://doi.org/10.1007/s004410050956>
- Schmidt, C., Granborg, M., & Deckert, J. (2014). Mass spectrometry – based relative quantification of proteins in precatalytic and catalytically active spliceosomes by metabolic count Mass spectrometry – based relative quantification of proteins in precatalytic and catalytically active spliceosomes by m. *Rna*, 20(6), 406–420. <https://doi.org/10.1261/rna.041244.113>
- Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*. <https://doi.org/10.1038/nmeth.2089>
- Schneider, C., Will, C. L., Makarova, O. V., Makarov, E. M., & Luhrmann, R. (2002). Human U4/U6.U5 and U4atac/U6atac.U5 tri-snRNPs exhibit similar protein compositions. *Mol Cell Biol*, 22(10), 3219–3229. <https://doi.org/10.1128/MCB.22.10.3219>
- Schoes, A. M., Ream, D. C., Thorman, A. W., Babbitt, P. C., & Friedberg, I. (2013). Biases in the Experimental Annotations of Protein Function and Their Effect on Our Understanding of Protein Function Space. *PLoS Computational Biology*, 9(5). <https://doi.org/10.1371/journal.pcbi.1003063>

- Schulte, E. C., Ellwanger, D. C., Dihanich, S., Manzoni, C., Stangl, K., Schormair, B., ... Winkelmann, J. (2014). Rare variants in LRRK1 and Parkinson's disease. *Neurogenetics*, *15*(1), 49–57. <https://doi.org/10.1007/s10048-013-0383-8>
- Schwartz, J. C., & Jardine, I. (1996). Quadrupole ion trap mass spectrometry. *Methods in Enzymology*, *270*, 552–586. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8803984>
- Schwikowski, B., Uetz, P., & Fields, S. (2000). A network of protein-protein interactions in yeast. *Nature Biotechnology*, *18*(12), 1257–1261. <https://doi.org/10.1038/82360>
- Scott, M. S., & Barton, G. J. (2007). Probabilistic prediction and ranking of human protein-protein interactions. *BMC Bioinformatics*, *8*. <https://doi.org/10.1186/1471-2105-8-239>
- Scotti, M. M., & Swanson, M. S. (2016). RNA mis-splicing in disease. *Nature Reviews Genetics*. <https://doi.org/10.1038/nrg.2015.3>
- Selvaraj, S., & Natarajan, J. (2011). Microarray data analysis and mining tools. *Bioinformatics*, *6*(3), 95–99. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/21584183>
- Sen, S., Langiewicz, M., Jumaa, H., & Webster, N. J. G. (2015). Deletion of serine/arginine-rich splicing factor 3 in hepatocytes predisposes to hepatocellular carcinoma in mice. *Hepatology*, *61*(1), 171–183. <https://doi.org/10.1002/hep.27380>
- Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N. G., Palmer, A. E., & Tsien, R. Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from *Drosophila* sp. red fluorescent protein. *Nature Biotechnology*, *22*(12), 1567–1572. <https://doi.org/10.1038/nbt1037>
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., ... Ideker, T. (2003). Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504. <https://doi.org/10.1101/gr.1239303>
- Shao, W., Kim, H.-S., Cao, Y., Xu, Y.-Z., & Query, C. C. (2012). A U1-U2 snRNP Interaction Network during Intron Definition. *Molecular and Cellular Biology*, *32*(2), 470–478. <https://doi.org/10.1128/MCB.06234-11>
- Sharan, R., Ulitsky, I., & Shamir, R. (2007). Network-based prediction of protein function. *Molecular Systems Biology*. <https://doi.org/10.1038/msb4100129>
- Sharma, K., D'Souza, R. C. J., Tyanova, S., Schaab, C., Wiśniewski, J. R., Cox, J., & Mann, M. (2014). Ultra-deep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling. *Cell Reports*, *8*(5), 1583–1594. <https://doi.org/10.1016/j.celrep.2014.07.036>
- Sharp, P. A., & Burge, C. B. (1997). Classification of introns: U2-type or U12-type. *Cell*. [https://doi.org/10.1016/S0092-8674\(00\)80479-1](https://doi.org/10.1016/S0092-8674(00)80479-1)
- Shaul, Y. D., Yuan, B., Thiru, P., Nutter-Upham, A., McCallum, S., Lanzkron, C., ... Sabatini, D. M. (2016). MERAV: A tool for comparing gene expression across human tissues and cell types. *Nucleic Acids Research*, *44*(D1), D560–D566. <https://doi.org/10.1093/nar/gkv1337>
- Shaw, M. A., Chiurazzi, P., Romain, D. R., Neri, G., & Géczy, J. (2002). A novel gene, FAM11A, associated with the FRAXF CpG island is transcriptionally silent in FRAXF full mutation. *European Journal of Human Genetics*, *10*(11), 767–772. <https://doi.org/10.1038/sj.ejhg.5200881>
- Shen, S., Warzecha, C. C., Carstens, R. P., & Xing, Y. (2010). MADS+: discovery of differential splicing events from Affymetrix exon junction array data. *Bioinformatics*, *26*(2), 268–269. <https://doi.org/10.1093/bioinformatics/btp643>
- Shen, X., Yi, L., Jiang, X., Zhao, Y., Hu, X., He, T., & Yang, J. (2016). Neighbor affinity based algorithm for discovering temporal protein complex from dynamic PPI network. *Methods*, *110*, 90–96. <https://doi.org/10.1016/j.ymeth.2016.06.010>
- Shirai, C. L., Ley, J. N., White, B. S., Kim, S., Tibbitts, J., Shao, J., ... Walter, M. J. (2015). Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. *Cancer Cell*, *27*(5), 631–643. <https://doi.org/10.1016/j.ccell.2015.04.008>
- Sigurðsson, T., Stark, K. L., Karayiorgou, M., Gogos, J. A., & Gordon, J. A. (2010). Impaired hippocampal-prefrontal synchrony in a genetic mouse model of schizophrenia. *Nature*, *464*(7289), 763–767. <https://doi.org/10.1038/nature08855>
- Simonis, N., Rual, J. F., Carvunis, A. R., Tasan, M., Lemmens, I., Hirozane-Kishikawa, T., ... Vidal, M. (2009). Empirically controlled mapping of the *Caenorhabditis elegans* protein-protein interactome network. *Nature Methods*, *6*(1), 47–54. <https://doi.org/10.1038/nmeth.1279>
- Simpson, K. J., Davis, G. M., & Boag, P. R. (2012). Comparative high-throughput RNAi screening methodologies in *C. elegans* and mammalian cells. *New Biotechnology*, *29*(4), 459–470. <https://doi.org/10.1016/J.NBT.2012.01.003>
- Sims, R. J., Millhouse, S., Chen, C.-F., Lewis, B. A., Erdjument-Bromage, H., Tempst, P., ... Reinberg, D. (2007). Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Postinitiation Factors and Pre-mRNA Splicing. *Molecular Cell*, *28*(4), 665–676. <https://doi.org/10.1016/J.MOLCEL.2007.11.010>
- Sinden, R. R. (1999). Biological Implications of the DNA Structures Associated with Disease-Causing Triplet Repeats. *The American Journal of Human Genetics*, *64*(2), 346–353. <https://doi.org/http://dx.doi.org/10.1086/302271>
- Singh, J., & Padgett, R. A. (2009). Rates of in situ transcription and splicing in large human genes. *Nature Structural and Molecular Biology*, *16*(11), 1128–1133. <https://doi.org/10.1038/nsmb.1666>
- Singh, R. K., & Cooper, T. A. (2012). Pre-mRNA splicing in disease and therapeutics. *Trends in Molecular Medicine*. <https://doi.org/10.1016/j.molmed.2012.06.006>
- Sinha, A., & Rae, R. (2016). Genome-Wide RNAi Screens in *C. elegans* to Identify Genes Influencing Lifespan and Innate Immunity. In *Methods in molecular biology (Clifton, N.J.)* (Vol. 1470, pp. 171–182). [https://doi.org/10.1007/978-1-4939-6337-9\\_14](https://doi.org/10.1007/978-1-4939-6337-9_14)
- Sjöblom, T., Jones, S., Wood, L. D., Parsons, D. W., Lin, J., Barber, T. D., ... Velculescu, V. E. (2006). The consensus coding sequences of human breast and colorectal cancers. *Science (New York, N.Y.)*, *314*(5797), 268–274. <https://doi.org/10.1126/science.1133427>
- Škunca, N., Altenhoff, A., & Dessimoz, C. (2012). Quality of computationally inferred gene ontology annotations. *PLoS Computational Biology*, *8*(5). <https://doi.org/10.1371/journal.pcbi.1002533>
- Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., ... Lewis, S. (2007). The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*. <https://doi.org/10.1038/nbt1346>
- Smith, B., & Brochhausen, M. (2010). Putting biomedical ontologies to work. *Methods of Information in Medicine*, *49*(2), 135–140. <https://doi.org/10.3414/ME9302>
- Smith, B. J. (1984). SDS Polyacrylamide Gel Electrophoresis of Proteins. *Methods in Molecular Biology (Clifton, N.J.)*, *1*, 41–55. <https://doi.org/10.1385/0-89603-062-8:41>
- Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Analytical Chemistry*, *78*(3), 779–787. <https://doi.org/10.1021/ac051437y>

- Smith, D. I., Zhu, Y., McAvoy, S., & Kuhn, R. (2006). Common fragile sites, extremely large genes, neural development and cancer. In *Cancer Letters* (Vol. 232, pp. 48–57). <https://doi.org/10.1016/j.canlet.2005.06.049>
- Smith, G. R., & Sternberg, M. J. E. (2002). Prediction of protein-protein interactions by docking methods. *Current Opinion in Structural Biology*. [https://doi.org/10.1016/S0959-440X\(02\)00285-3](https://doi.org/10.1016/S0959-440X(02)00285-3)
- Smith, R. (2009). Two-Dimensional Electrophoresis: An Overview. In *Methods in molecular biology (Clifton, N.J.)* (Vol. 519, pp. 2–17). [https://doi.org/10.1007/978-1-59745-281-6\\_1](https://doi.org/10.1007/978-1-59745-281-6_1)
- Snider, J., Kotlyar, M., Saraon, P., Yao, Z., Jurisica, I., & Stagljar, I. (2015). Fundamentals of protein interaction network mapping. *Molecular Systems Biology*, *11*(12), 848. <https://doi.org/10.15252/msb.20156351>
- Sobolik-Delmaire, T., Reddy, R., Pashaj, A., Roberts, B. J., & Wahl, J. K. (2010). Plakophilin-1 localizes to the nucleus and interacts with single-stranded DNA. *Journal of Investigative Dermatology*, *130*(11), 2638–2646. <https://doi.org/10.1038/jid.2010.191>
- Sofola, O. A., Jin, P., Botas, J., & Nelson, D. L. (2007). Argonaute-2-dependent rescue of a Drosophila model of FXTAS by FRAXE premutation repeat. *Human Molecular Genetics*, *16*(19), 2326–2332. <https://doi.org/10.1093/hmg/ddm186>
- Soltis, R. D., Hasz, D., Morris, M. J., & Wilson, I. D. (1979). The effect of heat inactivation of serum on aggregation of immunoglobulins. *Immunology*, *36*(1), 37–45. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1457381&tool=pmcentrez&rendertype=abstract>
- Sossey-Alaoui, K., Li, X., & Cowell, J. K. (2007). c-Abl-mediated phosphorylation of WAVE3 is required for lamellipodia formation and cell migration. *Journal of Biological Chemistry*, *282*(36), 26257–26265. <https://doi.org/10.1074/jbc.M701484200>
- SPADACCINI, R. (2006). Biochemical and NMR analyses of an SF3b155-p14-U2AF-RNA interaction network involved in branch point definition during pre-mRNA splicing. *RNA*, *12*(3), 410–425. <https://doi.org/10.1261/rna.2271406>
- Spartz, A. K., Herman, R. K., & Shaw, J. E. (2004). SMU-2 and SMU-1, Caenorhabditis elegans homologs of mammalian spliceosome-associated proteins RED and tSAP57, work together to affect splice site choice. *Mol Cell Biol*, *24*(15), 6811–6823. <https://doi.org/10.1128/MCB.24.15.6811-6823.2004> [pii]
- Spirin, V., & Mirny, L. A. (2003). Protein complexes and functional modules in molecular networks. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(21), 12123–12128. <https://doi.org/10.1073/pnas.2032324100>
- Spradling, A. C., & Rubin, G. M. (1983). The effect of chromosomal position on the expression of the Drosophila xanthine dehydrogenase gene. *Cell*, *34*(1), 47–57. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/6309411>
- Sprinzak, E., & Margalit, H. (2001). Correlated sequence-signatures as markers of protein-protein interaction. *Journal of Molecular Biology*, *311*(4), 681–692. <https://doi.org/10.1006/jmbi.2001.4920>
- Srinivasan, K., Friedman, B. A., Larson, J. L., Lauffer, B. E., Goldstein, L. D., Appling, L. L., ... Hansen, D. V. (2016). Untangling the brain's neuroinflammatory and neurodegenerative transcriptional responses. *Nature Communications*, *7*. <https://doi.org/10.1038/ncomms11295>
- Srisawat, C., & Engelke, D. R. (2001). Streptavidin aptamers: Affinity tags for the study of RNAs and ribonucleoproteins. *RNA*, *7*(4), 632–641. <https://doi.org/10.1017/S135583820100245X>
- Stark, C. (2006). BioGRID: a general repository for interaction datasets. *Nucleic Acids Research*, *34*(90001), D535–D539. <https://doi.org/10.1093/nar/gkj109>
- Starr, D. A., & Fridolfsson, H. N. (2010). Interactions Between Nuclei and the Cytoskeleton Are Mediated by SUN-KASH Nuclear-Envelope Bridges. *Annual Review of Cell and Developmental Biology*, *26*(1), 421–444. <https://doi.org/10.1146/annurev-cellbio-100109-104037>
- Steele, J. G., Johnson, G., & Underwood, P. A. (1992). Role of serum vitronectin and fibronectin in adhesion of fibroblasts following seeding onto tissue culture polystyrene. *Journal of Biomedical Materials Research*, *26*(7), 861–884. <https://doi.org/10.1002/jbm.820260704>
- Stelling, J., Sauer, U., Szallasi, Z., Doyle, F. J., & Doyle, J. (2004). Robustness of cellular functions. *Cell*. <https://doi.org/10.1016/j.cell.2004.09.008>
- Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler, H., ... Wanker, E. E. (2005). A human protein-protein interaction network: A resource for annotating the proteome. *Cell*, *122*(6), 957–968. <https://doi.org/10.1016/j.cell.2005.08.029>
- STEWART, S. A. (2003). Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA*, *9*(4), 493–501. <https://doi.org/10.1261/rna.2192803>
- Stuart, L. M., Boulais, J., Charriere, G. M., Hennessy, E. J., Brunet, S., Jutras, I., ... Ezekowitz, R. A. B. (2007). A systems biology analysis of the Drosophila phagosome. *Nature*, *445*(7123), 95–101. <https://doi.org/10.1038/nature05380>
- Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., ... Hogenesch, J. B. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences*, *101*(16), 6062–6067. <https://doi.org/10.1073/pnas.0400782101>
- Su, N., Xu, X., Li, C., He, Q., Zhao, L., Li, C., ... Chen, L. (2010). Generation of Fgfr3 conditional knockout mice. *International Journal of Biological Sciences*, *6*(4), 327–332. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/20582225>
- Sullivan, L. B., Gui, D. Y., Hosios, A. M., Bush, L. N., Freinkman, E., & Vander Heiden, M. G. (2015). Supporting Aspartate Biosynthesis Is an Essential Function of Respiration in Proliferating Cells. *Cell*, *162*(3), 552–563. <https://doi.org/10.1016/j.cell.2015.07.017>
- Sun, J., Sun, J., Zhao, Z., & Zhao, Z. (2010). A comparative study of cancer proteins in the human protein-protein interaction network. *BMC Genomics*, *11 Suppl 3*, S5. <https://doi.org/10.1186/1471-2164-11-S3-S5> [pii]n10.1186/1471-2164-11-S3-S5 [doi]
- Sun, Y., Hays, N. M., Periasamy, A., Davidson, M. W., & Day, R. N. (2012). Monitoring protein interactions in living cells with fluorescence lifetime imaging microscopy. *Methods in Enzymology*, *504*, 371–391. <https://doi.org/10.1016/B978-0-12-391857-4.00019-7>
- Sundaramoorthy, S., Vazquez-Novelle, M. D., Lekomtsev, S., Howell, M., & Petronczki, M. (2014). Functional genomics identifies a requirement of pre-mRNA splicing factors for sister chromatid cohesion. *Embo J*, *33*(22), 2623–2642. <https://doi.org/10.15252/embj.201488244>
- Supper, V., Hartl, I., Boulègue, C., Ohradanova-Repic, A., & Stockinger, H. (2017). Dynamic Interaction- and Phospho-Proteomics Reveal Lck as a Major Signaling Hub of CD147 in T Cells. *The Journal of Immunology*, *198*(6), 2468–2478. <https://doi.org/10.4049/jimmunol.1600355>
- Sussman, M. R., Huttlin, E. L., & Wohlbach, D. J. (2009). Democratization and Integration of Genomic Profiling Tools. In *Methods in molecular biology (Clifton, N.J.)* (Vol. 553, pp. 373–393). [https://doi.org/10.1007/978-1-60327-563-7\\_20](https://doi.org/10.1007/978-1-60327-563-7_20)
- Sutandy, F. X. R., Qian, J., Chen, C.-S., & Zhu, H. (2013). Overview of Protein Microarrays. *Current Protocols in Protein*

- Science*, 72(1), 27.1.1-27.1.16. <https://doi.org/10.1002/0471140864.ps2701s72>
- Sutherland, G. R. (1982). Heritable fragile sites on human chromosomes. VIII. Preliminary population cytogenetic data on the folic-acid-sensitive fragile sites. *Am J Hum Genet*, 34(3), 452–458.
- SUTHERLAND, G. R. (1985). Heritable fragile sites on human chromosomes XII. Population Cytogenetics. *Annals of Human Genetics*, 49(2), 153–161. <https://doi.org/10.1111/j.1469-1809.1985.tb01687.x>
- Sutherland, G. R., Baker, E., & Richards, R. I. (1998). Fragile sites still breaking. *Trends in Genetics*. [https://doi.org/10.1016/S0168-9525\(98\)01628-X](https://doi.org/10.1016/S0168-9525(98)01628-X)
- Sutherland, G. R., & Richards, R. I. (1995). The molecular basis of fragile sites in human chromosomes. *Current Opinion in Genetics & Development*, 5(3), 323–327. [https://doi.org/10.1016/0959-437X\(95\)80046-8](https://doi.org/10.1016/0959-437X(95)80046-8)
- Suthram, S., Dudley, J. T., Chiang, A. P., Chen, R., Hastie, T. J., & Butte, A. J. (2010). Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. *PLoS Computational Biology*, 6(2). <https://doi.org/10.1371/journal.pcbi.1000662>
- Swarbrick, J. D., Shaw, D. J., Chhabra, S., Ghai, R., Valkov, E., Norwood, S. J., ... Collins, B. M. (2011). VPS29 is not an active metallo-phosphatase but is a rigid scaffold required for retromer interaction with accessory proteins. *PLoS ONE*, 6(5). <https://doi.org/10.1371/journal.pone.0020420>
- Taipale, M., Tucker, G., Peng, J., Krykbaeva, I., Lin, Z. Y., Larsen, B., ... Lindquist, S. (2014). A quantitative chaperone interaction network reveals the architecture of cellular protein homeostasis pathways. *Cell*, 158(2), 434–448. <https://doi.org/10.1016/j.cell.2014.05.039>
- Takada, I. (2015). DGCR14 Induces *Il17a* Gene Expression through the ROR $\gamma$ /BAZ1B/RSKS2 Complex. *Molecular and Cellular Biology*, 35(2), 344–355. <https://doi.org/10.1128/MCB.00926-14>
- Takahashi, E. ichi, Hori, T. aki, & Murata, M. (1988). Population cytogenetics of rare fragile sites in Japan. *Human Genetics*, 78(2), 121–126. <https://doi.org/10.1007/BF00278179>
- Tanackovic, G., Ransijn, A., Thibault, P., Elela, S. A., Klinck, R., Berson, E. L., ... Rivolta, C. (2011). PRPF mutations are associated with generalized defects in spliceosome formation and pre-mRNA splicing in patients with retinitis pigmentosa. *Human Molecular Genetics*, 20(11), 2116–2130. <https://doi.org/10.1093/hmg/ddr094>
- Tang, X., Wang, J., Liu, B., Li, M., Chen, G., & Pan, Y. (2011). A comparison of the functional modules identified from time course and static PPI network data. *BMC Bioinformatics*, 12. <https://doi.org/10.1186/1471-2105-12-339>
- Taylor, R. G., Walker, D. C., & McInnes, R. R. (1993). E. coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Research*, 21(7), 1677–1678. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8479929>
- Tedeschi, A., & Bradke, F. (2013). The DLK signalling pathway—a double-edged sword in neural development and regeneration. *EMBO Reports*, 14(7), 605–614. <https://doi.org/10.1038/embor.2013.64>
- Telesford, Q. K., Joyce, K. E., Hayasaka, S., Burdette, J. H., & Laurienti, P. J. (2011). The Ubiquity of Small-World Networks. *Brain Connectivity*, 1(5), 367–375. <https://doi.org/10.1089/brain.2011.0038>
- Ten Have, S., Boulon, S., Ahmad, Y., & Lamond, A. I. (2011). Mass spectrometry-based immuno-precipitation proteomics - The user's guide. *Proteomics*, 11(6), 1153–1159. <https://doi.org/10.1002/pmic.201000548>
- Tenenbaum, S. A., & Aguirre-Ghiso, J. (2005). Dephosphorylation shows SR proteins the way out. *Molecular Cell*. <https://doi.org/10.1016/j.molcel.2005.11.005>
- Teng, Q., Huang, W., Collette, T. W., Ekman, D. R., & Tan, C. (2009). A direct cell quenching method for cell-culture based metabolomics. *Metabolomics*, 5(2), 199–208. <https://doi.org/10.1007/s11306-008-0137-z>
- Teng, Y., Yu, Y., Ferreira, J. A., & Waters, R. (2005). Histone acetylation, chromatin remodelling, transcription and nucleotide excision repair in *S. cerevisiae*: Studies with two model genes. *DNA Repair*. <https://doi.org/10.1016/j.dnarep.2005.04.006>
- Terada, Y., & Yasuda, Y. (2006). Human Immunodeficiency Virus Type 1 Vpr Induces G2 Checkpoint Activation by Interacting with the Splicing Factor SAP145. *Molecular and Cellular Biology*, 26(21), 8149–8158. <https://doi.org/10.1128/MCB.01170-06>
- Tew, K. L., Li, X.-L., & Tan, S.-H. (2007). Functional centrality: detecting lethality of proteins in protein interaction networks. In *Genome informatics. International Conference on Genome Informatics* (Vol. 19, pp. 166–177). <https://doi.org/10.1093/nar/gkq973>
- The UniProt Consortium. (2011). Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Research*, 39(Database issue), D214–9. <https://doi.org/10.1093/nar/gkq1020>
- Thomas, T., & Thomas, T. J. (2003). Polyamine metabolism and cancer. *J. Cell. Mol. Med.*, 7(2), 113–126. <https://doi.org/10.1111/j.1582-4934.2003.tb00210.x>
- Tian, B., Hu, J., Zhang, H., & Lutz, C. S. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Research*, 33(1), 201–212. <https://doi.org/10.1093/nar/gki158>
- Tisserant, A., & König, H. (2008). Signal-Regulated Pre-mRNA Occupancy by the General Splicing Factor U2AF. *PLoS ONE*, 3(1), e1418. <https://doi.org/10.1371/journal.pone.0001418>
- Tochitani, S., & Hayashizaki, Y. (2007). Functional screening revisited in the postgenomic era. *Molecular BioSystems*, 3(3), 195. <https://doi.org/10.1039/b614882b>
- Tong, C., Li, P., Wu, N. L., Yan, Y., & Ying, Q.-L. (2010). Production of p53 gene knockout rats by homologous recombination in embryonic stem cells. *Nature*, 467(7312), 211–213. <https://doi.org/10.1038/nature09368>
- Towbin, H., Staehelin, T., & Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the National Academy of Sciences of the United States of America*, 76(9), 4350–4354. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/388439>
- Trembley, J. H., Tatsumi, S., Sakashita, E., Loyer, P., Slaughter, C. A., Suzuki, H., ... Mayeda, A. (2005). Activation of pre-mRNA splicing by human RNPS1 is regulated by CK2 phosphorylation. *Molecular and Cellular Biology*, 25(4), 1446–1457. <https://doi.org/10.1128/MCB.25.4.1446-1457.2005>
- Trousdale, C., & Kim, K. (2015). Retromer: Structure, function, and roles in mammalian disease. *European Journal of Cell Biology*, 94(11), 513–521. <https://doi.org/10.1016/j.ejcb.2015.07.002>
- Tsantoulis, P. K., Kotsinas, A., Sfrikakis, P. P., Evangelou, K., Sideridou, M., Levy, B., ... Gorgoulis, V. G. (2008). Oncogene-induced replication stress preferentially targets common fragile sites in preneoplastic lesions. A genome-wide study. *Oncogene*, 27(23), 3256–3264. <https://doi.org/10.1038/sj.onc.1210989>
- Tseng, C.-K., Liu, H.-L., & Cheng, S.-C. (2011). DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. *RNA*, 17(1), 145–154. <https://doi.org/10.1261/ma.2459611>
- Tseng, C. K., & Cheng, S. C. (2008). Both catalytic steps of nuclear pre-mRNA splicing are reversible. *Science*, 320(5884), 1782–1784. <https://doi.org/10.1126/science.1158993>



- Tseng, Y. T., Li, W., Chen, C. H., Zhang, S., Chen, J. J. W., Zhou, X. J., & Liu, C. C. (2015). IIDB: A database for isoform-isoform interactions and isoform network modules. *BMC Genomics*, *16*. <https://doi.org/10.1186/1471-2164-16-S2-S10>
- Tsien, R. Y. (1998). THE GREEN FLUORESCENT PROTEIN. *Annual Review of Biochemistry*, *67*(1), 509–544. <https://doi.org/10.1146/annurev.biochem.67.1.509>
- Turinsky, A. L., Razick, S., Turner, B., Donaldson, I. M., & Wodak, S. J. (2010). Literature curation of protein interactions: measuring agreement across major public databases. *Database: The Journal of Biological Databases and Curation*, *2010*, baq026. <https://doi.org/10.1093/database/baq026>
- Turriziani, B., Garcia-Munoz, A., Pilkington, R., Raso, C., Kolch, W., & von Kriegsheim, A. (2014). On-Beads Digestion in Conjunction with Data-Dependent Mass Spectrometry: A Shortcut to Quantitative and Dynamic Interaction Proteomics. *Biology*, *3*(2), 320–332. <https://doi.org/10.3390/biology3020320>
- Turunen, J. J., Niemelä, E. H., Verma, B., & Frilander, M. J. (2013). The significant other: Splicing by the minor spliceosome. *Wiley Interdisciplinary Reviews: RNA*. <https://doi.org/10.1002/wrna.1141>
- Tusher, V. G., Tibshirani, R., & Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences*, *98*(9), 5116–5121. <https://doi.org/10.1073/pnas.091062498>
- Uetz, P., Giot, L., Cagney, G., Mansfield, T. a, Judson, R. S., Knight, J. R., ... Rothberg, J. M. (2000). A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature*, *403*(6770), 623–627. <https://doi.org/10.1038/35001009>
- Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., ... Ponten, F. (2010). Towards a knowledge-based Human Protein Atlas. *Nature Biotechnology*. <https://doi.org/10.1038/nbt1210-1248>
- Ulrich, A., & Wahl, M. C. (2014). Structure and evolution of the spliceosomal peptidyl-prolyl cis-trans isomerase Cwc27. *Acta Crystallographica Section D: Biological Crystallography*, *70*(12), 3110–3123. <https://doi.org/10.1107/S1399004714021695>
- Urh, M., Simpson, D., & Zhao, K. (2009). Affinity chromatography: general methods. *Methods in Enzymology*, *463*(9), 417–438. [https://doi.org/10.1016/S0076-6879\(09\)63026-3](https://doi.org/10.1016/S0076-6879(09)63026-3)
- Urlaub, H., Raker, V. A., Kostka, S., & Lührmann, R. (2001). Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *EMBO Journal*, *20*(1–2), 187–196. <https://doi.org/10.1093/emboj/20.1.187>
- Valente, A. X. C. N., & Cusick, M. E. (2006). Yeast protein interactome topology provides framework for coordinated-functionality. *Nucleic Acids Research*, *34*(9), 2812–2819. <https://doi.org/10.1093/nar/gk1325>
- Valones, M. A. A., Guimarães, R. L., Brandão, L. A. C., De Souza, P. R. E., De Albuquerque Tavares Carvalho, A., & Crovela, S. (2009). Principles and applications of polymerase chain reaction in medical diagnostic fields: A review. *Brazilian Journal of Microbiology*, *40*(1), 1–11. <https://doi.org/10.1590/S1517-83822009000100001>
- Van Cauwenberghe, C., Van Broeckhoven, C., & Sleegers, K. (2016). The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *18*(5), 421–430. <https://doi.org/10.1038/gim.2015.117>
- van der Lelij, P., Stocsits, R. R., Ladurner, R., Petzold, G., Kreidl, E., Koch, B., ... Peters, J.-M. (2014). SNW1 enables sister chromatid cohesion by mediating the splicing of sororin and APC2 pre-mRNAs. *The EMBO Journal*, *33*(22), 2643–2658. <https://doi.org/10.15252/embj.201488202>
- Vandormael-Pournin, S., Fro, M., Nacerddine, K., Babinet, C., Cohen-Tannoudji, M., Artus, J., & Frodin, M. (2005). Impaired mitotic progression and preimplantation lethality in mice lacking OMCG1, a new evolutionarily conserved nuclear protein. *Mol Cell Biol*, *25*(14), 6289–6302. <https://doi.org/10.1128/MCB.25.14.6289-6302.2005>
- Vanunu, O., Magger, O., Ruppin, E., Shlomi, T., & Sharan, R. (2010). Associating genes and protein complexes with disease via network propagation. *PLoS Computational Biology*, *6*(1). <https://doi.org/10.1371/journal.pcbi.1000641>
- Venancio, T. M., Balaji, S., Iyer, L. M., & Aravind, L. (2009). Reconstructing the ubiquitin network - Cross-talk with other systems and identification of novel functions. *Genome Biology*, *10*(3). <https://doi.org/10.1186/gb-2009-10-3-r33>
- Venkatesh, S., & Workman, J. L. (2015). Histone exchange, chromatin structure and the regulation of transcription. *Nature Reviews Molecular Cell Biology*. <https://doi.org/10.1038/nrm3941>
- Vera, M., Pani, B., Griffiths, L. A., Muchardt, C., Abbott, C. M., Singer, R. H., & Nudler, E. (2014). The translation elongation factor eEF1A1 couples transcription to translation during heat shock response. *eLife*, *3*, e03164. <https://doi.org/10.7554/eLife.03164>
- Verkerk, A. J. M. H., Pieretti, M., Sutcliffe, J. S., Fu, Y. H., Kuhl, D. P. A., Pizzuti, A., ... Warran, S. T. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, *65*(5), 905–914. [https://doi.org/10.1016/0092-8674\(91\)90397-H](https://doi.org/10.1016/0092-8674(91)90397-H)
- Vestal, M. L., & Campbell, J. M. (2005). Tandem Time-of-Flight Mass Spectrometry. In *Methods in enzymology* (Vol. 402, pp. 79–108). [https://doi.org/10.1016/S0076-6879\(05\)02003-3](https://doi.org/10.1016/S0076-6879(05)02003-3)
- Vidal, M. (2000). Protein interaction mapping in *C. elegans* Using proteins involved in vulval development. *Science*, *287*(5450), 116–122. <https://doi.org/10.1126/science.287.5450.116>
- Vidal, M. (2009). A unifying view of 21st century systems biology. *FEBS Letters*, *583*(24), 3891–3894. <https://doi.org/10.1016/j.febslet.2009.11.024>
- Viera, A., Alsheimer, M., Gómez, R., Berenguer, I., Ortega, S., Symonds, C. E., ... Suja, J. A. (2015). CDK2 regulates nuclear envelope protein dynamics and telomere attachment in mouse meiotic prophase. *Journal of Cell Science*, *128*(1), 88–99. <https://doi.org/10.1242/jcs.154922>
- Villas-Bôas, S. G., Højer-Pedersen, J., Åkesson, M., Smedsgaard, J., & Nielsen, J. (2005). Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast*, *22*(14), 1155–1169. <https://doi.org/10.1002/yea.1308>
- Voith von Voithenberg, L., Sánchez-Rico, C., Kang, H.-S., Madl, T., Zanier, K., Barth, A., ... Lamb, D. C. (2016). Recognition of the 3' splice site RNA by the U2AF heterodimer involves a dynamic population shift. *Proceedings of the National Academy of Sciences*, *113*(46), E7169–E7175. <https://doi.org/10.1073/pnas.1605873113>
- von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., & Bork, P. (2002). Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, *417*(6887), 399–403. <https://doi.org/10.1038/nature750>
- Vuong, C. K., Black, D. L., & Zheng, S. (2016). The neurogenetics of alternative splicing. *Nature Reviews Neuroscience*. <https://doi.org/10.1038/nrn.2016.27>
- Wahl, M. C., Will, C. L., & Lührmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*. <https://doi.org/10.1016/j.cell.2009.02.009>
- Walhout, A., Sordella, R., Thierry-Mieg, N., Brasch, M., Temple, G., Hartley, J., ... Vidal, M. (1999). The *C. elegans* protein interaction mapping project: a test-case using proteins involved in vulval development. *Science*, *287*(January), 116–123.
- Wan, R., Yan, C., Bai, R., Wang, L., Huang, M., Wong, C. C. L., & Shi, Y. (2016). The 3.8 Å structure of the U4/U6.U5 tri-

- snRNP: Insights into spliceosome assembly and catalysis. *Science*, 351(6272), 466–475. <https://doi.org/10.1126/science.aad6466>
- Wanders, R. J. A., Waterham, H. R., & Ferdinandusse, S. (2016). Metabolic Interplay between Peroxisomes and Other Subcellular Organelles Including Mitochondria and the Endoplasmic Reticulum. *Frontiers in Cell and Developmental Biology*, 3, 83. <https://doi.org/10.3389/fcell.2015.00083>
- Wang, E. T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., ... Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456(7221), 470–476. <https://doi.org/10.1038/nature07509>
- Wang, G. S., & Cooper, T. A. (2007). Splicing in disease: Disruption of the splicing code and the decoding machinery. *Nature Reviews Genetics*. <https://doi.org/10.1038/nrg2164>
- Wang, H., Duan, S., Du, J., Li, X., Xu, Y., Zhang, Z., ... He, L. (2006). Transmission disequilibrium test provides evidence of association between promoter polymorphisms in 22q11 gene DGCR14 and schizophrenia. *Journal of Neural Transmission*, 113(10), 1551–1561. <https://doi.org/10.1007/s00702-005-0420-3>
- Wang, J., Peng, X., Li, M., & Pan, Y. (2013). Construction and application of dynamic protein interaction network based on time course gene expression data. *PROTEOMICS*, 13(2), 301–312. <https://doi.org/10.1002/pmic.201200277>
- Wang, J., Peng, X., Peng, W., & Wu, F.-X. (2014). Dynamic protein interaction network construction and applications. *PROTEOMICS*, 14(4–5), 338–352. <https://doi.org/10.1002/pmic.201300257>
- Wang, P., Zhou, Z., Hu, A., Pontede Albuquerque, C., Zhou, Y., Hong, L., ... Fu, X. D. (2014). Both decreased and increased SRPK1 levels promote cancer by interfering with PHLPP-mediated dephosphorylation of Akt. *Molecular Cell*, 54(3), 378–391. <https://doi.org/10.1016/j.molcel.2014.03.007>
- Wang, Q., He, J., Lynn, B., & Rymond, B. C. (2005). Interactions of the yeast SF3b splicing factor. *Molecular and Cellular Biology*, 25(24), 10745–10754. <https://doi.org/10.1128/MCB.25.24.10745-10754.2005>
- Wang, X., Gulbahce, N., & Yu, H. (2011). Network-based methods for human disease gene prediction. *Briefings in Functional Genomics*, 10(5), 280–293. <https://doi.org/10.1093/bfpg/elfr024>
- Wang, Y. H., & Griffith, J. (1996). Methylation of expanded CCG triplet repeat DNA from fragile X syndrome patients enhances nucleosome exclusion. *Journal of Biological Chemistry*, 271(38), 22937–22940. <https://doi.org/10.1074/jbc.271.38.22937>
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 57–63. <https://doi.org/10.1038/nrg2484>
- Wanka, F., Cairns, T., Boecker, S., Berens, C., Happel, A., Zheng, X., ... Meyer, V. (2016). Tet-on, or Tet-off, that is the question: Advanced conditional gene expression in *Aspergillus*. *Fungal Genetics and Biology*, 89, 72–83. <https://doi.org/10.1016/j.fgb.2015.11.003>
- Warburg, O. (1956). On the Origin of Cancer Cells On the Origin of Cance. *Source: Science, New Series*, 123(123), 309–314. <https://doi.org/10.1126/science.123.3191.309>
- Ward, A. J., & Cooper, T. A. (2010). The pathobiology of splicing. *Journal of Pathology*. <https://doi.org/10.1002/path.2649>
- Warkocki, Z., Odenwalder, P., Schmitzova, J., Platzmann, F., Stark, H., Urlaub, H., ... Luhrmann, R. (2009). Reconstitution of both steps of *Saccharomyces cerevisiae* splicing with purified spliceosomal components. *Nature Structural and Molecular Biology*, 16(12), 1237–1243. <https://doi.org/10.1038/nsmb.1729>
- Watrin, E., Demidova, M., Watrin, T., Hu, Z., & Prigent, C. (2014). Sororin pre-mRNA splicing is required for proper sister chromatid cohesion in human cells. *EMBO Reports*, 15(9), 948–955. <https://doi.org/10.15252/embr.201438640>
- Werner, R. G., & Noe, W. (1993). Mammalian cell cultures. Part I: Characterization, morphology and metabolism. *Arzneimittel-Forschung*.
- Western, L. M., & Rose, S. J. (1991). A novel DNA joining activity catalyzed by T4 DNA ligase. *Nucleic Acids Research*, 19(4), 809–813. <https://doi.org/10.1093/nar/19.4.809>
- Whitfield, M. L. (2002). Identification of Genes Periodically Expressed in the Human Cell Cycle and Their Expression in Tumors. *Molecular Biology of the Cell*, 13(6), 1977–2000. <https://doi.org/10.1091/mbc.02-02-0030>
- Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., ... Kuster, B. (2014). Mass-spectrometry-based draft of the human proteome. *Nature*, 509(7502), 582–587. <https://doi.org/10.1038/nature13319>
- Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F., & Williams, K. L. (1996). Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology & Genetic Engineering Reviews*, 13, 19–50. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8948108>
- Will, C. L., & Luhrmann, R. (2011). Spliceosome structure and function. *Cold Spring Harbor Perspectives in Biology*, 3(7), 1–2. <https://doi.org/10.1101/cshperspect.a003707>
- Wilm, M. (2011). Principles of Electrospray Ionization. *Molecular & Cellular Proteomics*, 10(7), M111.009407. <https://doi.org/10.1074/mcp.M111.009407>
- Winkler, C., Eggert, C., Gradl, D., Meister, G., Giegerich, M., Wedlich, D., ... Fischer, U. (2005). Reduced U snRNP assembly causes motor axon degeneration in an animal model for spinal muscular atrophy. *Genes and Development*, 19(19), 2320–2330. <https://doi.org/10.1101/gad.342005>
- Winnepenninckx, B., Debacker, K., Ramsay, J., Smeets, D., Smits, A., FitzPatrick, D. R., & Kooy, R. F. (2007). CGG-repeat expansion in the DIP2B gene is associated with the fragile site FRA12A on chromosome 12q13.1. *American Journal of Human Genetics*, 80(2), 221–231. <https://doi.org/10.1086/510800>
- Wise, D. R., Ward, P. S., Shay, J. E. S., Cross, J. R., Gruber, J. J., Sachdeva, U. M., ... Thompson, C. B. (2011). Hypoxia promotes isocitrate dehydrogenase-dependent carboxylation of  $\alpha$ -ketoglutarate to citrate to support cell growth and viability. *Proceedings of the National Academy of Sciences*, 108(49), 19611–19616. <https://doi.org/10.1073/pnas.1117773108>
- Wouters, F. S., Bastiaens, P. I. H., Wirtz, K. W. A., & Jovin, T. M. (1998). FRET microscopy demonstrates molecular association of non-specific lipid transfer protein (nsL-TP) with fatty acid oxidation enzymes in peroxisomes. *EMBO Journal*, 17(24), 7179–7189. <https://doi.org/10.1093/emboj/17.24.7179>
- Wu, C. H., Apweiler, R., Bairoch, A., Natale, D. A., Barker, W. C., Boeckmann, B., ... Suzek, B. (2006). The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Research*, 34(90001), D187–D191. <https://doi.org/10.1093/nar/gkj161>
- Wu, Q., & Krainer, a R. (1999). AT-AC pre-mRNA splicing mechanisms and conservation of minor introns in voltage-gated ion channel genes. *Molecular and Cellular Biology*, 19(5), 3225–3236. <https://doi.org/10.1128/MCB.19.5.3225>
- Wu, X., Liu, Q., & Jiang, R. (2009). Align human interactome with phenome to identify causative genes and networks underlying disease families. *Bioinformatics*, 25(1), 98–104. <https://doi.org/10.1093/bioinformatics/btn593>
- Wysoczanski, P., Schneider, C., Xiang, S., Munari, F., Trowitzsch, S., Wahl, M. C., ... Zweckstetter, M. (2014). Cooperative

- structure of the heterotrimeric pre-mRNA retention and splicing complex. *Nature Structural and Molecular Biology*, 21(10), 911–918. <https://doi.org/10.1038/nsmb.2889>
- Wysoczanski, P., & Zweckstetter, M. (2016). Retention and splicing complex (RES) – the importance of cooperativity. *RNA Biology*, 13(2), 128–133. <https://doi.org/10.1080/15476286.2015.1096484>
- Xia, H., Li, B., Zhang, Z., Wang, Q., Qiao, T., & Li, K. (2015). Human glutaredoxin 3 can bind and effectively transfer [4Fe-4S] cluster to apo-iron regulatory protein 1. *Biochemical and Biophysical Research Communications*, 465(3), 620–624. <https://doi.org/10.1016/j.bbrc.2015.08.073>
- Xia, X. G., Zhou, H., Ding, H., Affar, E. B., Shi, Y., & Xu, Z. (2003). An enhanced U6 promoter for synthesis of short hairpin RNA. *Nucleic Acids Research*, 31(17), e100. <https://doi.org/10.1093/nar/gng098>
- Xiao, R., Tang, P., Yang, B., Huang, J., Zhou, Y., Shao, C., ... Fu, X. D. (2012). Nuclear Matrix Factor hnRNP U/SAF-A Exerts a Global Control of Alternative Splicing by Regulating U2 snRNP Maturation. *Molecular Cell*, 45(5), 656–668. <https://doi.org/10.1016/j.molcel.2012.01.009>
- Xu, X., Song, Y., Li, Y., Chang, J., Zhang, H., & An, L. (2010). The tandem affinity purification method: An efficient system for protein complex purification and protein interaction identification. *Protein Expression and Purification*, 72(2), 149–156. <https://doi.org/10.1016/j.pep.2010.04.009>
- Yang, P., Li, X., Wu, M., Kwok, C. K., & Ng, S. K. (2011). Inferring Gene-Phenotype associations via global protein complex network propagation. *PLoS ONE*, 6(7). <https://doi.org/10.1371/journal.pone.0021502>
- Yang, X., Coulombe-Huntington, J., Kang, S., Sheynkman, G. M., Hao, T., Richardson, A., ... Vidal, M. (2016). Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell*, 164(4), 805–817. <https://doi.org/10.1016/j.cell.2016.01.029>
- Yang, Y., Hadjikyriacou, A., Xia, Z., Gayatri, S., Kim, D., Zurita-Lopez, C., ... Bedford, M. T. (2015). PRMT9 is a Type II methyltransferase that methylates the splicing factor SAP145. *Nature Communications*, 6. <https://doi.org/10.1038/ncomms7428>
- Yao, Z., Petschnigg, J., Ketteler, R., & Stagljar, I. (2015). Application guide for omics approaches to cell signaling. *Nature Chemical Biology*, 11(6), 387–397. <https://doi.org/10.1038/nchembio.1809>
- Yap, K., Lim, Z. Q., Khandelia, P., Friedman, B., & Makeyev, E. V. (2012). Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. *Genes & Development*, 26(11), 1209–1223. <https://doi.org/10.1101/gad.188037.112>
- Yeh, P. C., Yeh, C. C., Chen, Y. C., & Juang, Y. L. (2012). RED, a spindle pole-associated protein, is required for kinetochore localization of MAD1, mitotic progression, and activation of the spindle assembly checkpoint. *Journal of Biological Chemistry*, 287(15), 11704–11716. <https://doi.org/10.1074/jbc.M111.299131>
- Yokoyama, W. M., Thompson, M. L., & Ehrhardt, R. O. (2012). Cryopreservation and Thawing of Cells. In *Current Protocols in Immunology* (Vol. Appendix 3, p. 3G). Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/0471142735.ima03gs99>
- Yoon, J., Blumer, A., & Lee, K. (2006). An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality. *Bioinformatics*, 22(24), 3106–3108. <https://doi.org/10.1093/bioinformatics/btl533>
- Yoshida, K., & Ogawa, S. (2014). Splicing factor mutations and cancer. *Wiley Interdisciplinary Reviews: RNA*, 5(4), 445–459. <https://doi.org/10.1002/wrna.1222>
- Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., ... Ogawa, S. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*, 478(7367), 64–69. <https://doi.org/10.1038/nature10496>
- Young, L., Sung, J., Stacey, G., & Masters, J. R. (2010). Detection of Mycoplasma in cell cultures. *Nature Protocols*, 5(5), 929–934. <https://doi.org/10.1038/nprot.2010.43>
- Younis, I., Dittmar, K., Wang, W., Foley, S. W., Berg, M. G., Hu, K. Y., ... Dreyfuss, G. (2013). Minor introns are embedded molecular switches regulated by highly unstable U6atac snRNA. *eLife*, 2013(2). <https://doi.org/10.7554/eLife.00780>
- Yu, D., Kim, M., Xiao, G., & Hwang, T. H. (2013). Review of Biological Network Data and Its Applications. *Genomics & Informatics*, 11(4), 200. <https://doi.org/10.5808/GI.2013.11.4.200>
- Yu, G., Lu, C., & Wang, J. (2017). NoGOA: Predicting noisy GO annotations using evidences and sparse representation. *BMC Bioinformatics*, 18(1). <https://doi.org/10.1186/s12859-017-1764-z>
- Yu, J., & Finley, R. L. (2009). Combining multiple positive training sets to generate confidence scores for protein-protein interactions. *Bioinformatics*, 25(1), 105–111. <https://doi.org/10.1093/bioinformatics/btn597>
- Yu, J., Pacifico, S., Liu, G., & Finley, R. L. (2008). DroID: The Drosophila Interactions Database, a comprehensive resource for annotated gene and protein interactions. *BMC Genomics*, 9. <https://doi.org/10.1186/1471-2164-9-461>
- Yu, S., Mangelsdorf, M., Hewett, D., Hobson, L., Baker, E., Eyre, H. J., ... Richards, R. I. (1997). Human chromosomal fragile site FRA16B is an amplified AT-rich minisatellite repeat. *Cell*, 88(3), 367–374. [https://doi.org/10.1016/S0092-8674\(00\)81875-9](https://doi.org/10.1016/S0092-8674(00)81875-9)
- Zahiri, J., Bozorgmehr, J. H., & Masoudi-Nejad, A. (2013). Computational Prediction of Protein-Protein Interaction Networks: Algorithms and Resources. *Current Genomics*, 14(6), 397–414. <https://doi.org/10.2174/1389202911314060004>
- Zanzoni, A., & Brun, C. (2016). Integration of quantitative proteomics data and interaction networks: Identification of dysregulated cellular functions during cancer progression. *Methods*, 93, 103–109. <https://doi.org/10.1016/j.ymeth.2015.09.014>
- Zaretsky, J. Z., & Wreschner, D. H. (2008). Protein multifunctionality: Principles and mechanisms. *Translational Oncogenomics*.
- Zeng, S. X., Li, Y., Jin, Y., Zhang, Q., Keller, D. M., McQuaw, C. M., ... Lu, H. (2010). Structure-Specific Recognition Protein 1 Facilitates Microtubule Growth and Bundling Required for Mitosis. *Molecular and Cellular Biology*, 30(4), 935–947. <https://doi.org/10.1128/MCB.01379-09>
- Zeng, X., Han, L., Singh, S. R., Liu, H., Neumüller, R. A., Yan, D., ... Hou, S. X. (2015). Genome-wide RNAi Screen Identifies Networks Involved in Intestinal Stem Cell Regulation in Drosophila. *Cell Reports*, 10(7), 1226–1238. <https://doi.org/10.1016/j.celrep.2015.01.051>
- ZERNIKE, F. (1955). How I discovered phase contrast. *Science (New York, N.Y.)*, 121(3141), 345–349. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/13237991>
- Zhang, A., Sun, H., Xu, H., Qiu, S., & Wang, X. (2013). Cell metabolomics. *Omics: A Journal of Integrative Biology*, 17(10), 495–501. <https://doi.org/10.1089/omi.2012.0090>
- Zhang, G., Gurtu, V., & Kain, S. R. (1996). An Enhanced Green Fluorescent Protein Allows Sensitive Detection of Gene Transfer in Mammalian Cells. *Biochemical and Biophysical Research Communications*, 227(3), 707–711. <https://doi.org/10.1006/bbrc.1996.1573>
- Zhang, J., Lieu, Y. K., Ali, A. M., Penson, A., Reggio, K. S., Rabadan, R., ... Manley, J. L. (2015). Disease-associated mutation

- in *SRSF2* misregulates splicing by altering RNA-binding affinities. *Proceedings of the National Academy of Sciences*, 112(34), E4726–E4734. <https://doi.org/10.1073/pnas.1514105112>
- Zhang, X., Zhang, R., Jiang, Y., Sun, P., Tang, G., Wang, X., ... Li, X. (2011). The expanded human disease network combining protein-protein interaction information. *European Journal of Human Genetics*, 19(7), 783–788. <https://doi.org/10.1038/ejhg.2011.30>
- Zhang, Y., Bottinelli, D., Lisacek, F., Luban, J., Strambio-De-Castillia, C., Varesio, E., & Hopfgartner, G. (2015). Optimization of human dendritic cell sample preparation for mass spectrometry-based proteomic studies. *Analytical Biochemistry*, 484, 40–50. <https://doi.org/10.1016/j.ab.2015.05.007>
- Zhang, Y., Du, N., Li, K., Feng, J., Jia, K., & Zhang, A. (2014). MsiDBN: A method of identifying critical proteins in dynamic PPI networks. *BioMed Research International*, 2014. <https://doi.org/10.1155/2014/138410>
- Zheng, S., & Black, D. L. (2013). Alternative pre-mRNA splicing in neurons: Growing up and extending its reach. *Trends in Genetics*. <https://doi.org/10.1016/j.tig.2013.04.003>
- Zhou, D., & He, Y. (2008). Extracting interactions between proteins from the literature. *Journal of Biomedical Informatics*. <https://doi.org/10.1016/j.jbi.2007.11.008>
- Zhou, Y., Liu, S., Ozturk, A., & Hicks, G. G. (2014). FUS-regulated RNA metabolism and DNA damage repair: Implications for amyotrophic lateral sclerosis and frontotemporal dementia pathogenesis. *Rare Dis*, 2, e29515. <https://doi.org/10.4161/rdis.29515>
- Zhou, Z., & Fu, X. D. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma*. <https://doi.org/10.1007/s00412-013-0407-z>
- Zhou, Z., Licklider, L. J., Gygi, S. P., & Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature*, 419(6903), 182–185. <https://doi.org/10.1038/nature01031>
- Zhu, T., Xu, L., Xu, X., Wang, Z., Zhu, J., Xie, Q., ... Li, Y. (2015). Analysis of breast cancer-associated glycosphingolipids using electrospray ionization-linear ion trap quadrupole mass spectrometry. *Carbohydrate Research*, 402, 189–199. <https://doi.org/10.1016/j.carres.2014.10.006>
- Zhu, X., Gerstein, M., & Snyder, M. (2007). Getting connected: Analysis and principles of biological networks. *Genes and Development*. <https://doi.org/10.1101/gad.1528707>
- Zieske, L. R. (2006). A perspective on the use of iTRAQ™ reagent technology for protein complex and profiling studies. *Journal of Experimental Botany*, 57(7), 1501–1508. <https://doi.org/10.1093/jxb/erj168>
- Zlotorynski, E., Rahat, A., Skaug, J., Ben-Porat, N., Ozeri, E., Hershberg, R., ... Kerem, B. (2003). Molecular basis for expression of common and rare fragile sites. *Molecular and Cellular Biology*, 23(20), 7143–7151. <https://doi.org/10.1128/mcb.23.20.7143-7151.2003>
- Zois, C. E., & Harris, A. L. (2016). Glycogen metabolism has a key role in the cancer microenvironment and provides new targets for cancer therapy. *Journal of Molecular Medicine*. <https://doi.org/10.1007/s00109-015-1377-9>