

Πανεπιστήμιο Θεσσαλίας  
Σχολή Επιστημών Υγείας  
Τμήμα Βιοχημείας και Βιοτεχνολογίας

Εργαστήριο Μοριακής Βιολογίας και Γονιδιωματικής

Πτυχιακή Εργασία

***"Κατασκευή μιας βιοπληροφορικής μεθοδολογίας για τον υπολογισμό μεταγραφομικών συσχετίσεων in cis σε σταδιοποιημένες βιοψίες ασθενών με καρκίνο του προστάτη."***

***"Development of a Bioinformatics Methodology for in Transcriptomic cis Correlation Analysis in Staged Biopsies of Prostate Cancer Patients."***

Ονοματεπώνυμο: Φωτιάδης Δημήτριος

Πατρώνυμο: Αθανάσιος

Επιβλέποντες: Γιακουντής Αντώνιος, Φώτης Ψωμόπουλος

Λάρισα, 2024

## Τριμελής Συμβουλευτική Επιτροπή

Επιβλέπων:

**Φώτης Ψωμόπουλος**, Ερευνητής Β' Ινστιτούτου Εφαρμοσμένων Βιοεπιστημών (INEB) στο Εθνικό Κέντρο Έρευνας και Ανάπτυξης (ΕΚΕΤΑ)

Συνεπιβλέπων:

**Γιακουντής Αντώνιος**, Επίκουρος Καθηγητής Μοριακής Βιολογίας και Γονιδιωματικής του Τμήματος Βιοχημείας και Βιοτεχνολογίας της Σχολής Επιστημών Υγείας του Πανεπιστημίου Θεσσαλίας

Μέλη:

**Κολοβός Πέτρος**, Επίκουρος Καθηγητής Μοριακής Βιολογίας και Βιολογίας Συστημάτων του Τμήματος Μοριακής της Σχολής Επιστημών Υγείας του Δημοκρίτειου Πανεπιστημίου

# Πίνακας περιεχομένων

|                                                                                      |    |
|--------------------------------------------------------------------------------------|----|
| Τριμελής Συμβουλευτική Επιτροπή                                                      | 2  |
| Πίνακας περιεχομένων                                                                 | 3  |
| Περίληψη                                                                             | 4  |
| Abstract                                                                             | 5  |
| 1. Εισαγωγή                                                                          | 6  |
| 1.1 Καρκίνος του Προστάτη                                                            | 6  |
| 1.1.1 Επιδημιολογία                                                                  | 6  |
| 1.1.2 Μοριακή ταξινόμηση                                                             | 8  |
| 1.1.3 Διάγνωση και θεραπεία                                                          | 11 |
| 1.2 Topologically Associated Domains(TADs) και ρυθμιστικά δίκτυα γονιδίων(GRNs)      | 13 |
| 1.2.1 Ρυθμιστικά στοιχεία του γονιδιώματος                                           | 13 |
| 1.2.3 Απορρύθμιση των TADs στον καρκίνο                                              | 16 |
| 2. Υλικά και Μεθοδολογία                                                             | 17 |
| 2.1 Υλικά                                                                            | 17 |
| 3.2 Μέθοδοι                                                                          | 17 |
| 2.2.1 Φιλτράρισμα δειγμάτων με χαμηλές τιμές έκφρασης                                | 17 |
| 2.2.2 Σχηματισμός μητρών συσχέτισης                                                  | 18 |
| 2.2.3 Εύρεση γειτονικών γονιδίων                                                     | 19 |
| 2.2.4 Συνδυασμός δεδομένων                                                           | 19 |
| 2.2.5 Απεικόνιση των δεδομένων                                                       | 20 |
| 2.2.6 Δημιουργία θερμικών χαρτών και γονιδιακών δικτύων                              | 21 |
| 2.2.7 Αξιοποίηση των δεδομένων των δικτύων                                           | 22 |
| 3. Αποτελέσματα                                                                      | 23 |
| 3.1 Προετοιμασία δεδομένων                                                           | 23 |
| 3.2 Απεικόνιση δεδομένων                                                             | 24 |
| 3.2.1 Σχηματισμός ιστογράμματος κατανομής απόστασης                                  | 24 |
| 3.2.2 Σχηματισμός διαγραμμάτων κατανομής των τιμών R ανά στάδιο                      | 25 |
| 3.2.3 Ανάλυση συσχετίσεων πριν και μετά το φίλτρο στατιστικής σημαντικότητας         | 26 |
| 3.2.4: Απεικόνιση της σχέσης μεταξύ της απόστασης και των θετικών συσχετίσεων        | 29 |
| 3.2.5 Ανάλυση οντολογίας στα δεδομένα διαφορικής έκφρασης και γονιδιακών συσχετίσεων | 40 |
| 3.2.6 Κατηγοριοποίηση των βιότυπων των γονιδίων                                      | 42 |
| 4. Συζήτηση                                                                          | 45 |
| 5. Αποθετήριο δεδομένων και κώδικα                                                   | 50 |
| 6. Βιβλιογραφία                                                                      | 51 |

# Περίληψη

Ο καρκίνος του προστάτη αποτελεί μία από τις πιο συχνές κακοήθειες στους άνδρες, παρουσιάζοντας σημαντική ετερογένεια στη μοριακή του βάση. Πρόσφατες μελέτες έχουν αναδείξει τον ρόλο των Τοπολογικά Συσχετιζόμενων Περιοχών (TADs) και των *cis*-αλληλεπιδράσεων στη ρύθμιση της γονιδιακής έκφρασης, και συνεπώς και στην εξέλιξη του καρκίνου. Σκοπός ανάλυσης είναι η ανάπτυξη μιας βιοπληροφορικής μεθοδολογίας για την ανάλυση μεταγραφομικών συσχετίσεων *in cis* σε σταδιοποιημένες βιοψίες ασθενών με καρκίνο του προστάτη.

Για τον σκοπό αυτό, χρησιμοποιήθηκαν RNA-seq από τις βάσεις TCGA και GTEX, περιλαμβάνοντας τόσο φυσιολογικό όσο και καρκινικό ιστό σε διάφορα στάδια της νόσου (I-IV). Η μεθοδολογία περιλαμβάνει μια σειρά από υπολογιστικά βήματα, όπως προεπεξεργασία δεδομένων, δημιουργία μητρών συσχέτισης, αναγνώριση ζευγών γονιδίων, στατιστική φιλτράρισή τους και οπτικοποίηση των δικτύων γονιδιακής συσχέτισης. Έτσι γίνεται ταυτοποίηση συν-ρυθμιζόμενων γονιδιακών ομάδων και τον πιθανό ρόλο τους στην εξέλιξη του καρκίνου του προστάτη.

Τα αποτελέσματα αποκαλύπτουν μια μεταβολή στα πρότυπα συσχέτισης γονιδίων μεταξύ των σταδίων της νόσου, με σημαντικές ανακατατάξεις στα γονιδιακά δίκτυα. Οι θερμικοί χάρτες και τα γονιδιακά δίκτυα δείχνουν ότι η δομή των μεταγραφομικών σχέσεων τροποποιείται ανάλογα με το στάδιο του καρκίνου. Αυτό δείχνει ότι πιθανώς επηρεάζονται διαφορετικοί ρυθμιστικοί μηχανισμοί κατά την εξέλιξη του όγκου. Επιπλέον, η ανάλυση εμπλουτισμού (enrichment analysis) ανέδειξε ότι τα γονίδια με αλλοιωμένες συσχετίσεις σχετίζονται με κρίσιμες ογκογενετικές οδούς, όπως οδοί που σχετίζονται με τις ορμόνες του θυρεοειδούς, η αναδιαμόρφωση της χρωματίνης και ο μεταβολικός επαναπρογραμματισμός.

Η μελέτη περιλαμβάνει μια σειρά από διακριτά βήματα για τη διερεύνηση των μεταγραφομικών συσχετίσεων *in cis* στον καρκίνο, συμβάλλοντας στην καλύτερη κατανόηση των ρυθμιστικών αλλαγών κατά την εξέλιξη της νόσου. Τα ευρήματα αυτά μπορούν να αξιοποιηθούν για την ανακάλυψη νέων βιοδεικτών και θεραπευτικών στόχων στον καρκίνο του προστάτη.

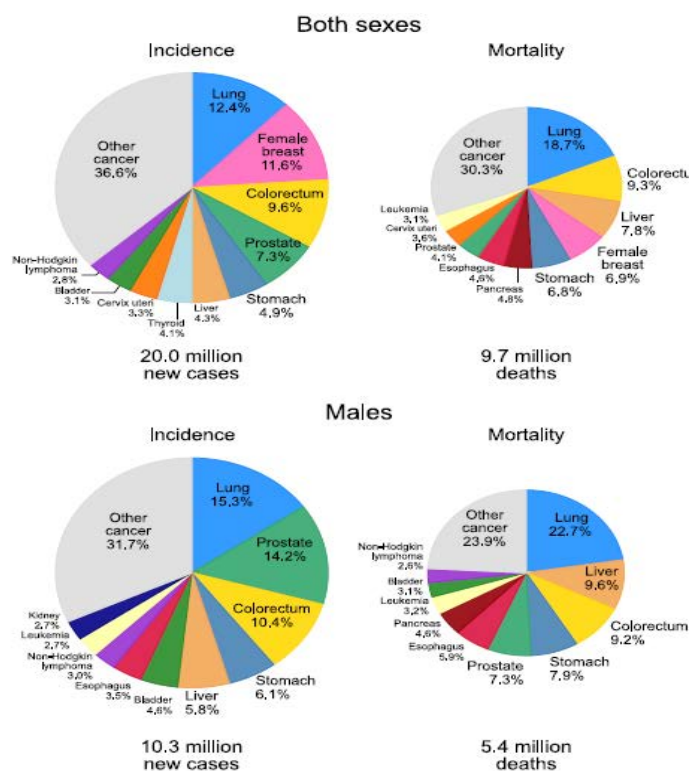
# Abstract

# 1. Εισαγωγή

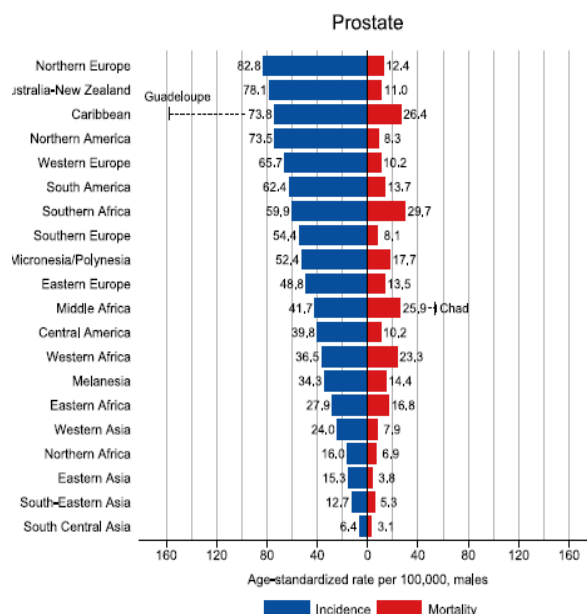
## 1.1 Καρκίνος του Προστάτη

### 1.1.1 Επιδημιολογία

Ο καρκίνος του προστάτη αποτελεί μια από τις πιο συχνές μορφές της νόσου στους άντρες. Παρόλο που θεραπεύεται εύκολα αν ανιχνευθεί στα αρχικά στάδια, είναι μία από τις κυριότερες αιτίες θανάτου που οφείλονται με τον καρκίνο [1]. Με βάση τα παγκόσμια δεδομένα για το 2022, κατατάσσεται δεύτερος, όσον αφορά την συχνότητα στους άντρες, και πέμπτος στην θνησιμότητα. που σε αριθμό περιστατικών μεταφράζονται σε 1.5 εκατομμύρια περιστατικά και περίπου 400 χιλιάδες θανάτους [2] (Εικόνα 1.1).

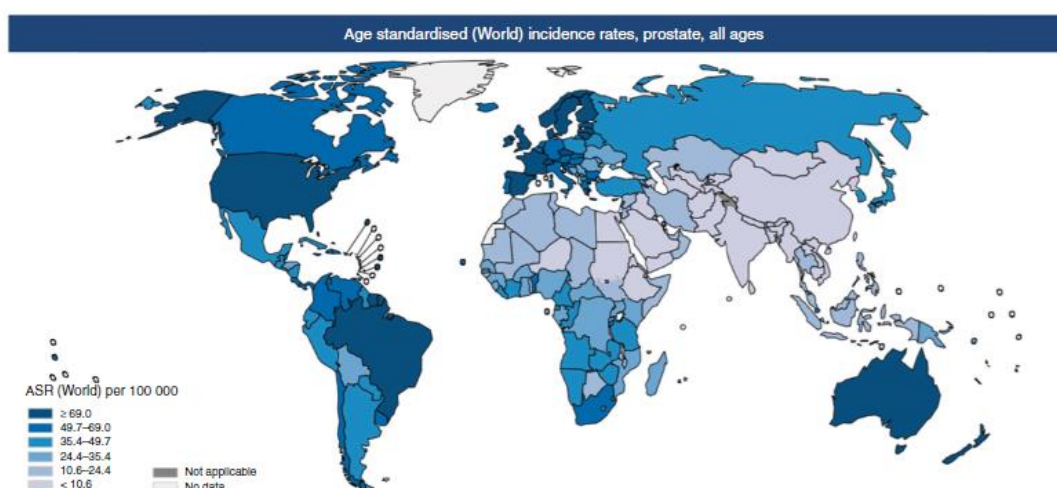


**Εικόνα 1.1:** Κυκλικά διαγράμματα απεικόνισης της συχνότητας και της θνησιμότητας της νόσου. Στην πρώτη εικόνα αναγράφονται τα στατιστικά και στα δύο φύλα, και στην δεύτερη για τους άντρες [4].

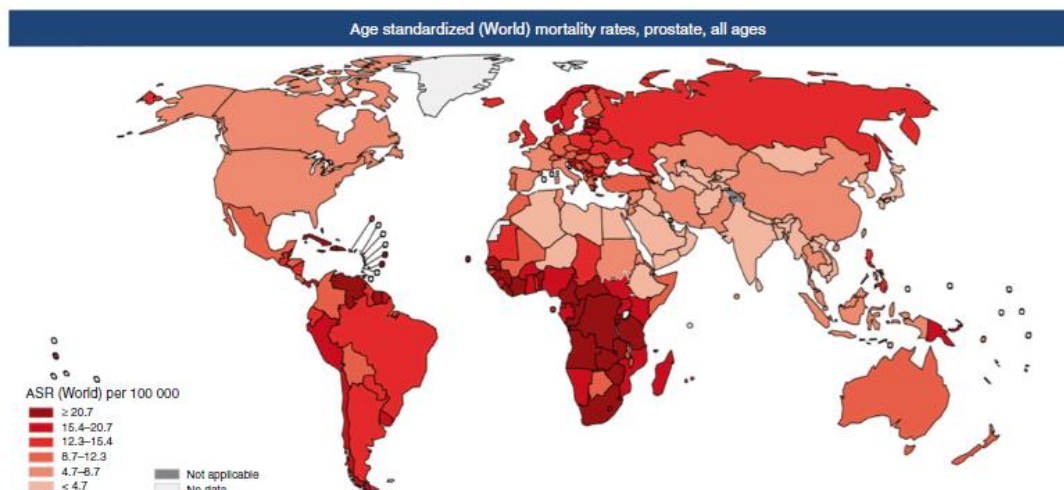


**Εικόνα 1.2:** Ιστόγραμμα συχνότητας εμφάνισης και θνησιμότητας καρκίνου του προστάτη ανά τον κόσμο [4]

Επιπλέον, διαφορετικές περιοχές του κόσμου εμφανίζουν διαφορές ως προς την εμφάνιση και την επικινδυνότητα της νόσου (**Εικόνα 1.2**). Οι δυτικές χώρες έχουν υψηλή συχνότητα, αλλά πολύ χαμηλά επίπεδα θνησιμότητας. Τον μεγαλύτερο κίνδυνο διατρέχουν οι άντρες της υποσαχάριας Αφρικής και της Καραϊβικής με Αφρικανική καταγωγή, οι οποίοι μάλιστα είναι πολύ πιθανό να εκδηλώσουν εξαιρετικά επιθετική μορφή της ασθένειας, ακόμα και σε μικρή ηλικία [2,3] (**Εικόνες 1.3/1.4**). Αρκετά υψηλά ποσοστά εμφάνισης, ανάλογης με την θνησιμότητα, παρατηρούνται μάλιστα και στην κεντρική και νότια Αμερική [2] (**Εικόνες 1.3/1.4**). Να σημειωθεί ωστόσο, ότι η οξεία διαφορά στην θνησιμότητα ανά τον κόσμο οφείλεται τόσο σε γενετικές αιτίες, όσο στην έλλειψη διαθέσιμων πόρων ώστε να γίνουν εύκολα διαγνωστικές εξετάσεις. Αυτό έχει ως αποτέλεσμα η διάγνωση να γίνεται σε τελικά και πολύ επιθετικά στάδια, αυξάνοντας έτσι τον αριθμό των θανάτων ανά αριθμό περιστατικών [2,3].



**Εικόνα 1.3:** Αριθμός περιστατικών ανά χώρα [3].



Εικόνα 1.4: Θνησιμότητα ανά χώρα [3].

Ο καρκίνος του προστάτη είναι μια πολυπαραγοντική νόσος, στην οποία διαδραματίζουν σημαντικό ρόλο τόσο περιβαλλοντικοί όσο και γενετικοί παράγοντες [4]. Ήδη αναφέρθηκε η καταγωγή ως σημαντικός παράγοντας, πράγμα που οφείλεται στην γενετική προδιάθεση που εμφανίζουν ορισμένοι πληθυσμοί [3,4]. Ένας ακόμα παράγοντας είναι η ηλικία. Το μεγαλύτερο τμήμα των διαγνώσεων αφορά άτομα άνω των 50, στατιστικό που δείχνει την αναλογική σχέση της αύξησης της ηλικίας με την πιθανότητα εκδήλωσης. Ως προς την κληρονομικότητα, άτομα με κοντινούς συγγενείς που πάσχουν, όπως αδέρφια ή ξαδέρφια πρώτου βαθμού, έχουν διπλάσια πιθανότητα να εμφανίσουν καρκίνο του προστάτη [1,4]. Ο κίνδυνος μάλιστα πολλαπλασιάζεται αν υπάρχει ιστορικό με μεγάλο αριθμό συγγενών που πάσχουν ή έπασχαν. Οι κληρονομήσιμες μορφές του καρκίνου αφορούν το 10% των περιπτώσεων και αποδίδονται σε κληρονομούμενες γονιδιακές μεταλλάξεις. Η γενετική προδιάθεση που αναφέρθηκε λοιπόν, βασίζεται σε μεταλλάξεις που επηρεάζουν γονίδια επιδιόρθωσης DNA, καταστολείς όγκων και δημιουργίας ογκογονιδίων. Τέτοια γονίδια είναι τα BRCA1/2, HOXB13, RNASEL και τα MSH2, MSH6, MLH1, και PMS2 (Σύνδρομο Lynch) [4,5]. Τέλος, εξωτερικοί παράγοντες όπως το κάπνισμα, η παχυσαρκία και η έκθεση σε ορισμένες χημικές ουσίες αυξάνουν σε έναν βαθμό την προδιάθεση για καρκίνο του προστάτη [1,4].

### 1.1.2 Μοριακή ταξινόμηση

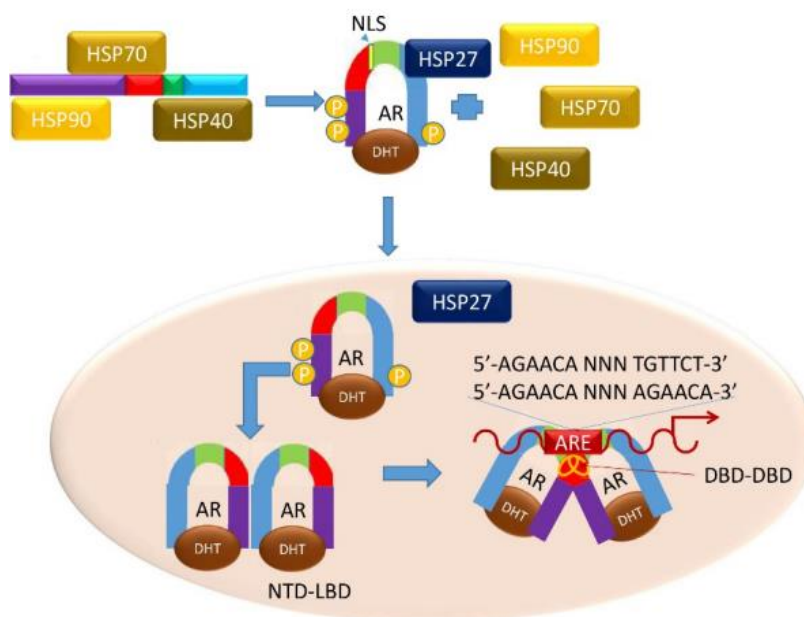
Ο καρκίνος του προστάτη εμφανίζει μεγάλη ετερογένεια ανάμεσα σε διαφορετικά περιστατικά, δημιουργώντας προβλήματα στην κλασική ιστοπαθολογική υπο-κατηγοριοποίηση του. Η κατηγοριοποίηση βασίζεται στην βαθμολογία Gleason, το σύστημα TNM και την αναγνώριση των επιπέδων του προστατικού αντιγόνου (PSA). Αν και αυτές οι μέθοδοι είναι σε μεγάλο βαθμό αξιόπιστες, αδυνατούν πολλές φορές να κάνουν μια σχετικά ακριβή εκτίμηση για την πορεία της νόσου. Ενδεικτικά, πολλοί ασθενείς με την ίδια βαθμολογία Gleason, εμφανίζουν τελείως διαφορετική κλινική εικόνα. Έτσι λοιπόν αναπτύχθηκαν αρκετοί τρόποι ώστε η ταξινόμηση να γίνει με μοριακά κριτήρια [6]. Από τους πολλούς μοριακούς τρόπους ταξινόμησης, ιδιαίτερα σημαντική είναι η συνεισφορά της ερευνητικής κοινοπραξίας The Cancer Genome Atlas (TCGA), σύμφωνα με την οποία βρέθηκαν 7 διαφορετικοί υπότυποι, που



αφορούν το 74% των περιστατικών. Αυτοί οι υπότυποι διαχωρίζονται με βάση συμβάντα γονιδιακής συγχώνευσης και με μεταλλάξεις σε συγκεκριμένα γονίδια. Έτσι οι 7 κατηγορίες είναι : (1)ERG, (2)ETV1, (3)ETV4, (4)FLI1, (5)SPOP, (6)FOXA1 και (7)DH1. Τα πρώτα 4 γονίδια αφορούν γονιδιακές συγχωνεύσεις και ανήκουν στην οικογένεια ETS, ενώ τα υπόλοιπα αποτελούν ειδικές μεταλλάξεις των εκάστοτε γονιδίων [6,7].

Πέρα από αυτό, η κάθε κατηγορία εμφανίζει διαφορές ως και προς τις επιγενετικές τροποποιήσεις και συγκεκριμένα τα επίπεδα των υπερμεθυλιώσεων σε CpG περιοχές. Οι υπερμεθυλιώσεις αυτές μειώνουν την έκφραση κάποιων σχετιζόμενων γονιδίων, οι γενετικοί τύποι των οποίων γειτνιάζουν με τις μεθυλιωμένες νησίδες. Με βάση αυτό, έχουν βρεθεί 4 ομάδες-clusters. Το cluster 1 αφορά όγκους με το υψηλότερο επίπεδο υπερμεθυλίωσης, όπου εκτεταμένες περιοχές του γονιδιώματος εμφανίζουν υψηλό ποσοστό μεθυλιωμένων CpG. Το cluster 2 έχει όγκους με μέτρια επίπεδα υπερμεθυλίωσης, όπου η μεθυλίωση καταλαμβάνει μικρότερη έκταση του γονιδιώματος σε σχέση με το cluster 1, αλλά είναι ομοιογενώς κατανομημένη σε συγκεκριμένες γονιδιακές περιοχές. Στο cluster 3, τα επίπεδα υπερμεθυλίωσης είναι επίσης μέτρια, αλλά η κατανομή είναι λιγότερο εκτεταμένη και πιο ετερογενής από το cluster 2, επηρεάζοντας λιγότερες CpG περιοχές. Στο cluster 4, οι όγκοι έχουν ετερογενές προφίλ μεθυλίωσης, με συνολικά χαμηλότερα επίπεδα υπερμεθυλίωσης και περιορισμένη επίδραση στη γονιδιακή έκφραση [7].

Παράλληλα, κάθε υπότυπος παρουσιάζει και διαφορές ως προς την ενεργότητα του υποδοχέα ανδρογόνων (AR). Ο υποδοχέας αυτός έχει ιδιαίτερη σημασία, καθώς φυσιολογικά η ενεργοποίηση του σηματοδοτεί την μεταγραφή γονιδίων που αφορούν την διαφοροποίηση, την επιβίωση και την ανάπτυξη των κυττάρων του προστάτη. Ο καρκίνος εκμεταλλεύεται αυτόν τον μηχανισμό, και στα αρχικά του στάδια είναι ανδρογόνο-εξαρτώμενος, χρησιμοποιώντας την AR σηματοδότηση για να εξελιχθεί [7,8] (**Εικόνα 1.5**).



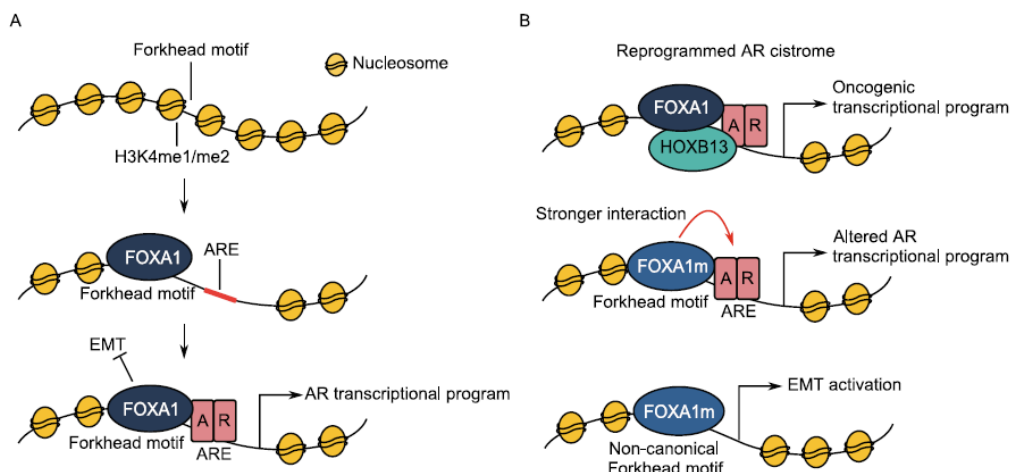
**Εικόνα 1.5:** Απεικόνιση του μονοπατιού σηματοδότησης του AR. Όταν συνδεθεί το ανδρογόνο στον υποδοχέα, οι πρωτεΐνες θερμικού σοκ (HSPs) αποσυνδέονται. Ο AR μετατρέπεται σε κυκλικά μονομερή, μέσω της σύνδεσης του αμινοτελικού και του καρβοξυτελικού άκρου του. Τα μονομερή αυτά μετακινούνται στον πυρήνα. Σχηματίζουν ομοδιμερή τα οποία είναι DNA-εξαρτώμενα και προκαλούν την μεταγραφή γονιδίων [8].

Οι γονιδιακές συντήξεις της οικογένειας ETS αφορούν το μεγαλύτερο ποσοστό γονιδιακών συμβάντων με βάση τα οποία προέκυψαν οι υποκατηγορίες, απαρτίζοντας το 53% των περιπτώσεων. Οι συντήξεις γίνονται με γονίδια που ρυθμίζονται από το ανδρογόνο όπως τα TMPRSS2, SLC45A3 και το NDRG1. Η πιο συχνή σύντηξη είναι η ERG- TMPRSS2 (46%), με τις συντήξεις των υπόλοιπων μελών της οικογένειας να καταλαμβάνουν πολύ μικρότερα ποσοστά. Μάλιστα, παρατηρήθηκε ότι συνήθως οι μεταλλάξεις αυτές δεν συνυπάρχουν στον ίδιο όγκο, δείχνοντας ότι υπάρχει διαφορετική κλωνική εξέλιξη και άρα ετερογένεια εντός του όγκου. Πέρα από αυτό οι διαφορετικές συντήξεις συνοδεύονται και από εντελώς διαφορετικά μοριακά χαρακτηριστικά εντός του όγκου. Για παράδειγμα, σε όγκους με συντήξεις θετικές για ERG, υπάρχουν με πολύ μεγάλη συχνότητα μεταλλάξεις διαγραφής του PTEN. Σε αυτούς τους όγκους συναντάται κυρίως το cluster 1 υπερμεθυλίωσης, αλλά υψηλή συχνότητα έχει και το cluster 3. Τα υπόλοιπα μέλη της οικογένειας εμφανίζουν κυρίως μεθυλίωσεις του cluster 3. Αυτή η ετερογένεια εντός της οικογένειας, φαίνεται και από την διαφορά στην ενεργότητα του AR ανάλογα με τον τύπο σύντηξης. Κοινό γνώρισμα αποτελεί πάντως ότι σε όλα τα μέλη παρατηρείται μια χαρακτηριστική σίγηση του γονιδίου STAT6 [6,7].

Όσον αφορά την υποκατηγορία με μεταλλάξεις για το SPOP, αυτές εμφανίζονται στο 11% των ασθενών. Το γονίδιο SPOP κωδικοποιεί μια E3 λιγάση της ουβικουτίνης, και μεταλλάξεις σε αυτό μειώνουν ή αφαιρούν πλήρως την ικανότητα της να οδηγεί της υποψήφιες πρωτείνες-στόχους προς αποικοδόμηση. Αυτό οδηγεί σε αυξημένη ενεργότητα του AR, επειδή δεν αποικοδομείται τόσο το ίδιο το ανδρογόνο όσο και οι συνενεργοποιητές του υποδοχέα [9]. Παράλληλα, οι όγκοι αυτοί εμφανίζουν προφίλ τύπου cluster 2. Επιπλέον γονίδια των οποίων οι τροποποιήσεις συνδυάζονται σε όγκους με μεταλλάξεις του SPOP, είναι το SPINK1 mRNA που υπερεκφράζεται, και το HEXA που εμφανίζεται μειορρυθμισμένο στο 86% των δειγμάτων [7].

Όγκοι με μεταλλάξεις στο γονίδιο FOXA1, παρατηρήθηκαν στο 3% των περιπτώσεων. Το FOXA1 αποτελεί έναν από τους λεγόμενους πρωτοπόρους μεταγραφικούς παράγοντες. Φυσιολογικά, λειτουργεί σε συνέργεια με τον AR, κάνοντας προσβάσιμη την χρωματίνη. Μεταλλάξεις στον FOXA1, μπορούν να αλλάξουν τις περιοχές στόχευσης αλλά και την ενεργότητα του, επιτρέποντας στον AR να μεταγράψει διαφορετικά γονίδια σε μεγαλύτερο βαθμό [10] (**Εικόνα 1.6**). Το προφίλ μεθυλίωσης είναι κοινό με τους SPOP μεταλλαγμένους όγκους (cluster 2), αναδεικνύοντας κάποιες κοινές επιγενετικές λειτουργίες. Από τις αναλύσεις που έχουν γίνει, φάνηκε μάλιστα ότι πολλές φορές οι δύο μεταλλάξεις μπορούν και να συνυπάρχουν στον ίδιο όγκο [7].

Τέλος, μεταλλάξεις στο IDH1 εμφανίζονται μόνο στο 1% των όγκων και διαφέρουν πιο πολύ από κάθε κατηγορία. Κυρίως προκαλούν αύξηση στην μεθυλίωση στόχων. Για αυτό τον λόγο, στον IDH1 υπότυπο, παρατηρείται το πιο έντονο προφίλ υπερμεθυλίωσης και έτσι προκαλείται σίγηση πολλών παραπάνω γονιδίων. Τα γονίδια αυτά είναι συνήθως ογκοκατασταλτικά ή συμμετέχουν στην ανάπτυξη των κυττάρων. Επιπλέον, σε αντίθεση με τις άλλες υποκατηγορίες παρατηρούνται πολύ λιγότερες αλλαγές στην αλλαγή της αλληλουχίας του DNA (γονιδιακές συντήξεις, απώλειες γονιδίων κτλ.) [7].



**Εικόνα 1.6: Δράση του FOXA1 συνεργειακά με τον AR σε φυσιολογικό προστάτη (A) και της δράσης του μεταλλαγμένου FOXA1 στον καρκίνο του προστάτη (B) [10].**

- (A) Στον φυσιολογικό προστάτη, το FOXA1 βοηθά το DNA να ανοίγει, ώστε ο υποδοχέας ανδρικών ορμονών (AR) να ενεργοποιεί γονίδια. Επίσης, αποτρέπει τη μετατροπή κυττάρων (EMT), ανεξάρτητα από τον AR.
- (B) Στον καρκίνο του προστάτη, το FOXA1 αλλάζει τη λειτουργία του AR και ενεργοποιεί διαφορετικά γονίδια. Οι μεταλλάξεις στο FOXA1 να κάνουν τον AR πιο ισχυρό και να προκαλέσουν τη μετάσταση μέσω του EMT.

Πέρα από αυτές τις βασικές υποκατηγορίες, μεγάλη μερίδα όγκων (19%) εμφανίζει μεταλλάξεις σε γονίδια αντιγραφής του DNA. Μερικά από αυτά είναι τα BRCA1, BRCA2, CDK12, ATM, FANCD2 και RAD51C. Η παρουσία τέτοιων μεταλλάξεων είναι θετική για την εξέλιξη της νόσου, καθώς με την παροχή θεραπείας (αναστολέας PARP), βελτιώνεται η κλινική εικόνα ακόμα και ασθενών με μεταστατικό καρκίνο του προστάτη σε προχωρημένο στάδιο. Επίσης πέρα από διαγραφές στο PTEN που ήδη αναφέρθηκαν, σε μεγάλο αριθμό όγκων παρατηρήθηκαν μεταλλάξεις στην PI3K και σε πρωτεΐνες της οδού MAPK [7].

### 1.1.3 Διάγνωση και θεραπεία

Μέχρι και σήμερα, δεν υπάρχει κάποιο συγκεκριμένο τεστ μέσω του οποίου μπορούμε να πούμε με 100% βεβαιότητα πως υπάρχει καρκίνος του προστάτη, χωρίς να απαιτείται καθόλου βιοψία. Η διάγνωση του καρκίνου του προστάτη ξεκινά με μη επεμβατικές μεθόδους, όπως εξετάσεις αίματος, ούρων και μοριακές αναλύσεις [5].

Το PSA (ειδικό προστατικό αντιγόνο) είναι η βασική διαγνωστική εξέταση, καθώς αποτελεί μια πρωτεΐνη που παράγεται από τον προστάτη και ανιχνεύεται στο αίμα. Όταν τα επίπεδα PSA ξεπερνούν τα 4 ng/mL, η πιθανότητα ύπαρξης καρκίνου αυξάνεται, ενώ σε τιμές άνω των 10 ng/mL, η πιθανότητα ξεπερνά το 50%. Ωστόσο, τα PSA δεν εμφανίζει υψηλά επίπεδα μόνο στην περίπτωση του καρκίνου, αφού αυξημένα επίπεδα μπορεί να εμφανιστούν και σε περιπτώσεις καλοήθους υπερπλασίας ή φλεγμονής του προστάτη. Νέες μέθοδοι, όπως το Prostate Health Index (PHI), συνδυάζουν το PSA με άλλους δείκτες για μεγαλύτερη ακρίβεια [5].

Σημαντικό ρόλο παίζουν και οι εξετάσεις στα ούρα, όπως το τεστ PCA3, που ανιχνεύει έναν μη κωδικό RNA δείκτη που υπερεκφράζεται σε σχεδόν όλες τις

περιπτώσεις καρκίνου του προστάτη. Αυτή η εξέταση μειώνει την ανάγκη για άσκοπες βιοψίες και έχει μεγαλύτερη ακρίβεια σε σχέση με άλλες μεθόδους. Συγκεκριμένα μέσω της PCR υπολογίζεται ένα "PCA3 score" που μειώνει σημαντικά τις άσκοπες βιοψίες. Ένα σκορ PCA3 άνω του 25 συνδέεται με υψηλή πιθανότητα καρκίνου [11,12].

Η μοριακή ανάλυση γονιδιακών συντήξεων, όπως το TMPRSS2-ERG που ήδη αναφέραμε, είναι εξίσου σημαντική. Λόγω του υψηλού ποσοστού εμφάνισης της συγκεκριμένης σύντηξης (50% των περιστατικών) αλλά και επειδή να εντοπιστεί σε ούρα ή ιστό, είναι ένας εύκολος τρόπος να φανεί μια πρώτη εικόνα για την κατάσταση του ασθενή. Συχνά σχετίζεται με επιθετικές μορφές καρκίνου, παρέχοντας πληροφορίες για τη θεραπεία και την πρόγνωση [12].

Η σύγχρονη απεικονιστική τεχνολογία, όπως η πολυπαραμετρική μαγνητική τομογραφία (mpMRI), έχει αναβαθμίσει σημαντικά τη διάγνωση. Η mpMRI εντοπίζει ύποπτες περιοχές του προστάτη με ακρίβεια, διευκολύνοντας τη σταδιοποίηση και την καθοδήγηση επεμβατικών εξετάσεων. Επιπλέον, η PSMA-PET, που χρησιμοποιεί ραδιοφάρμακα στοχεύοντας συγκεκριμένους δείκτες, είναι ιδιαίτερα αποτελεσματική στην ανίχνευση μεταστάσεων, ακόμα και σε αρχικά στάδια της νόσου [5,12].

Αν οι παραπάνω μέθοδοι δεν επαρκούν, η διάγνωση προχωρά σε επεμβατικές τεχνικές, όπως η βιοψία του προστάτη. Η βιοψία μπορεί να γίνει είτε δια ορθού είτε μέσω δια περινεϊκής προσέγγισης. Αν η mpMRI έχει εντοπίσει ύποπτες περιοχές, αυτές είναι που στοχεύονται για την λήψη των δειγμάτων, αυξάνοντας την ακρίβεια. Τα δείγματα, στην συνέχεια, εξετάζονται μικροσκοπικά και η επιθετικότητα του καρκίνου αξιολογείται μέσω του Gleason Score. Τα κύτταρα που μελετώνται χαρακτηρίζονται ως αρνητικά για καρκίνο, θετικά για καρκίνο ή ύποπτα για καρκίνο [5,12].

Όσον αφορά τη θεραπεία, τα τελευταία χρόνια έχει σημειωθεί σημαντική πρόοδος, προσφέροντας πλέον περισσότερες επιλογές, οι οποίες προσαρμόζονται στις ανάγκες κάθε ασθενούς. Η θεραπεία επιλέγεται με βάση το στάδιο της νόσου, την επιθετικότητα του όγκου, την ηλικία, τη γενική κατάσταση της υγείας του ασθενούς αλλά και τις προσωπικές του προτιμήσεις. Υπάρχουν διαφορετικές θεραπείες για τοπικό καρκίνο, προχωρημένο ή μεταστατικό καρκίνο, καθώς και υποστηρικτικές θεραπείες για τη βελτίωση της ποιότητας ζωής [11].

Για τον τοπικό καρκίνο του προστάτη, η ενεργός επιτήρηση εφαρμόζεται σε ασθενείς με χαμηλού κινδύνου νόσο. Περιλαμβάνει τακτική παρακολούθηση με εξετάσεις PSA, φυσικές εξετάσεις και βιοψίες, χωρίς όμως την πραγματοποίηση κάποιας επέμβασης, αν δεν υπάρχει πρόοδος στην νόσο. Δεδομένου ότι η πλειοψηφία των άλλων θεραπευτικών προσεγγίσεων συνοδεύεται από επιπλοκές, ο ασθενής μπορεί να διατηρήσει παράλληλα την ποιότητα ζωής τους [11]. Σε περίπτωση που ο καρκίνος είναι πιο επιθετικός, εφαρμόζονται χειρουργικές ή ακτινοθεραπευτικές μέθοδοι. Η ριζική προστατεκτομή, συχνά με ρομποτική υποβοήθηση, είναι πολύ πιο ακριβής και ο ασθενής μπορεί να αναρρώσει πιο γρήγορα. Παράλληλα, η ακτινοθεραπεία, με τεχνικές όπως το IMRT και η υποκλασματική ακτινοθεραπεία, μειώνουν την ακτινοβόληση των υγιών ιστών [12].

Σε περιπτώσεις μεταστατικού καρκίνου του προστάτη, η ανδρογονική αποστέρωση (ADT) αποτελεί τον βασικό τρόπο αντιμετώπισης της νόσου. Η θεραπεία στοχεύει στη μείωση των ανδρογόνων, τα οποία αξιοποιούνται για να τροφοδοτηθεί ο καρκίνος. Η ADT αποστέρωση, μπορεί να συνδυαστεί με χημειοθεραπεία, πράγμα που οδηγεί σε μεγαλύτερη πιθανότητα επιβίωσης [13]. Σε πιο ανθεκτικές περιπτώσεις, η ορμονοθεραπεία δεύτερης γενιάς με φάρμακα όπως το abiraterone και το enzalutamide έχει αποδειχθεί αποτελεσματική, εμποδίζοντας την παραγωγή ή τη λειτουργία των ανδρογόνων [12,13]. Επιπλέον, οι στοχευμένες θεραπείες, όπως οι αναστολείς PARP για ασθενείς με μεταλλάξεις στα γονίδια BRCA και η PSMA-ραδιοθεραπεία για επιθετικούς όγκους, προσφέρουν νέες δυνατότητες [12,13].

## 1.2 Topologically Associated Domains(TADs) και ρυθμιστικά δίκτυα γονιδίων(GRNs)

### 1.2.1 Ρυθμιστικά στοιχεία του γονιδιώματος

Οι υποκινητές και οι ενισχυτές αποτελούν τα 2 σημαντικότερα *cis* ρυθμιστικά στοιχεία για τον έλεγχο της γονιδιακής έκφρασης. Οι ρυθμιστικές αλληλεπιδράσεις μεταξύ τους, καθορίζουν το εάν ένα γονίδιο θα εκφραστεί, χωρικά και χρονικά [14].

Οι υποκινητές είναι περιοχές του DNA που βρίσκονται ανοδικά από την θέση έναρξης της μεταγραφής (Transcription Start Sites - TSS) και παίζουν κρίσιμο ρόλο στην εκκίνηση της.. Χαρακτηριστικά, οι ενεργοί υποκινητές φέρουν το επιγενετικό σήμα της τριμεθυλίωσης στη λυσίνη 4 της ιστόνης H3 (H3K4me3), το οποίο διευκολύνει την έλευση μεταγραφικών παραγόντων και της RNA πολυμεράσης II. Αντίθετα, σε περιπτώσεις όπου τα γονίδια καταστέλλονται, παρατηρείται τριμεθυλίωση στη λυσίνη 27 της ιστόνης H3 (H3K27me3), υποδηλώνοντας ότι ο υποκινητής είναι ανενεργός. Επιπλέον, οι υποκινητές συχνά περιέχουν νησίδες CpG, οι οποίες, όταν δεν είναι μεθυλιωμένες, επιτρέπουν τη μεταγραφή. Αντιθέτως οι μεθυλιωμένες νησίδες τροποποιούν την δομή της χρωματίνης, παρεμποδίζοντας την σύνδεση μεταγραφικών παραγόντων και την έναρξη της μεταγραφής. [15].

Οι ενισχυτές αφορούν *trans* ρυθμιστικά στοιχεία που επηρεάζουν τη δραστηριότητα των υποκινητών από απομακρυσμένες θέσεις. Οι κυριότερες επιγενετικές τροποποιήσεις τους περιλαμβάνουν μεταξύ άλλων τη μονομεθυλίωση της λυσίνης 4 της ιστόνης H3 (H3K4me1) και την ακετυλίωση της λυσίνης 27 της ιστόνης H3 (H3K27ac). Το τελευταίο αποτελεί χαρακτηριστικό σημάδι ενεργών ενισχυτών, καθώς διευκολύνει την πρόσβαση στη χρωματίνη και τη στρατολόγηση μεταγραφικών συνενεργοποιητών. Επιπλέον, οι ενεργοί ενισχυτές συχνά παράγουν μη κωδικοποιημένα RNA, γνωστά ως RNA ενισχυτών (eRNAs), τα οποία ενισχύουν περαιτέρω τον ρυθμιστικό τους ρόλο [15].

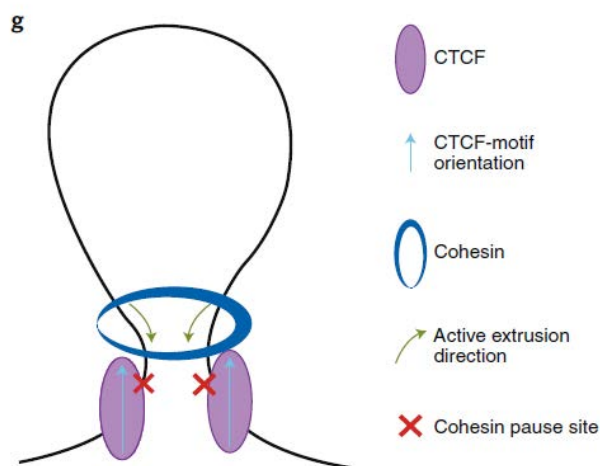
### 1.3.2 Σχηματισμός των TADs

Οι αλληλεπιδράσεις μεταξύ ενισχυτών και υποκινητών, δεν γίνονται τυχαία, αλλά εντός καλά οργανωμένων δομών που συχνά εμφανίζουν ιστοειδικότητα. Οι

δομές αυτές ονομάζονται TADs (topologically associated domains) και αφορούν τμήματα του γονιδιώματος που χαρακτηρίζονται από υψηλή συχνότητα χρωματινικών αλληλεπιδράσεων μεταξύ των ρυθμιστικών στοιχείων που περιλαμβάνονται σε αυτές. Το μέγεθος τους κυμαίνεται από 100 kb έως 1 Mb [16]. Στις δομές αυτές, γίνεται ρύθμιση της γονιδιακής έκφρασης, μέσω της συγκέντρωσης ενισχυτών και υποκινητών εντός των ορίων τους. Αυτό διασφαλίζει ότι οι ενισχυτές αλληλεπιδρούν μόνο με τους υποκινητές εντός του ίδιου TAD και επηρεάζουν μεταγραφικά μόνο τα αντίστοιχα γονίδια, αποτρέποντας τη λεγόμενη «κατάχρηση ενισχυτών» (enhancer hijacking), όπου ένας ενισχυτής θα μπορούσε να ρυθμίσει λανθασμένα γονίδια εκτός του TAD [17]. Το γονίδιο Shh αποτελεί ένα χαρακτηριστικό παράδειγμα ρύθμισης εντός των TADs. Ο ενισχυτής ZRS, παρά την απόστασή του, ρυθμίζει την έκφραση του Shh μέσω ειδικών αλληλεπιδράσεων, που αλλάζουν την δομή της χρωματίνης μόνο στις περιοχές αυτές [18].

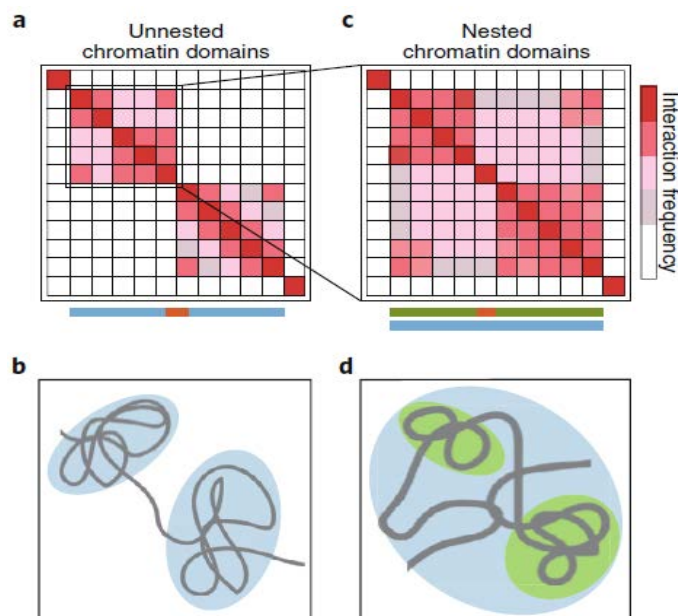
Εντός των TADs τα γονίδια οργανώνονται σε δίκτυα ρύθμισης. Αυτό αποτελεί έναν από τους κεντρικούς μηχανισμούς με τους οποίους τα κύτταρα επιτυγχάνουν τη συντονισμένη έκφραση των γονιδίων τους. Μάλιστα, σε περιπτώσεις που τα γονίδια ανήκουν στην ίδια λειτουργική κατηγορία η διαδικασία ενισχύεται ακόμα περισσότερο. Με άλλα λόγια, παρατηρείται συνμεταγραφή γονιδίων, λόγω της ταυτόχρονης ρύθμισης από πλέον εγγύς υποκινητές και ενισχυτές εντός των ορίων του TAD [19].

Ο σχηματισμός των TADs στηρίζεται στον μηχανισμό του loop extrusion. Σε αυτόν τον μηχανισμό, το σύμπλεγμα των κοχυσινών προσδένεται στο DNA και μετακινείται κατά μήκος του, σχηματίζοντας λούπες. Αυτή η διαδικασία συνεχίζεται έως ότου οι κοχυσίνες συναντήσουν επικράτειες δέσμευσης της πρωτεΐνης CTCF (CCCTC-binding factor). Τα μοτίβα αναγνώρισης και πρόσδεσης του παράγοντα CTCF είναι στρατηγικά τοποθετημένα στις θέσεις αυτές και με συγκεκριμένο προσανατολισμό (convergent orientation), ώστε να σταματούν την κίνηση του συμπλέγματος κοχυσινών με αποτέλεσμα την σταθεροποίηση του χρωματινικού βρόγχου [16]. Οι δύο περιοχές που είναι προσδεμένες οι πρωτεΐνες CTCF οριοθετούν τα TADs. Οι πρωτεΐνες αυτές δρουν ως «φράγμα» για την αποφυγή αλληλεπιδράσεων μεταξύ γειτονικών TADs. Το σύμπλεγμα των κοχυσινών παίζει επίσης ρόλο στη διατήρηση της δομής των βρόγχων μέσα στο TAD. Αυτή η συνδυαστική δράση των κοχυσινών και της CTCF οδηγεί στη σταθεροποίηση των TADs. Τα όρια αυτά είναι επίσης σημαντικά για την απομόνωση τους από τις διπλανές περιοχές του γονιδιώματος [16,17] (**Εικόνα 1.13**) .



**Εικόνα 1.13:** Γραφική απεικόνιση του μηχανισμού δημιουργίας και της μορφής των TADs [35]

Εντός των TADs εντοπίζονται μια σειρά από διαφορετικές μικρότερες λειτουργικές υπομονάδες, οι οποίες ονομάζονται nested domains. Πολλές από αυτές βρίσκονται ακόμα από διερεύνηση. Τα πιο καλά διερευνώμενα είναι τα λεγόμενα sub-TADs, που αποτελούν οργανωτικές υπομονάδες των TADs (**Εικόνα 1.14**). Σε αντίθεση με τα κύρια TADs, τα sub-TADs έχουν ασθενέστερη μονωτική δύναμη, με αποτέλεσμα οι αλληλεπιδράσεις να είναι πιο δυναμικές και να μεταβάλλονται συχνότερα. Επίσης, διαθέτουν κυτταρική εξειδίκευση. Τα sub-TADs εμφανίζονται ή εξαφανίζονται ανάλογα με την κατάσταση διαφοροποίησης ή τις επιγενετικές τροποποιήσεις ενός κυττάρου [17,20].



**Εικόνα 1.14:** Απεικόνιση σε heatmaps των TADs (a) και sub-TADs (b), και οι αντίστοιχες γραφικές απεικονίσεις τους [35]

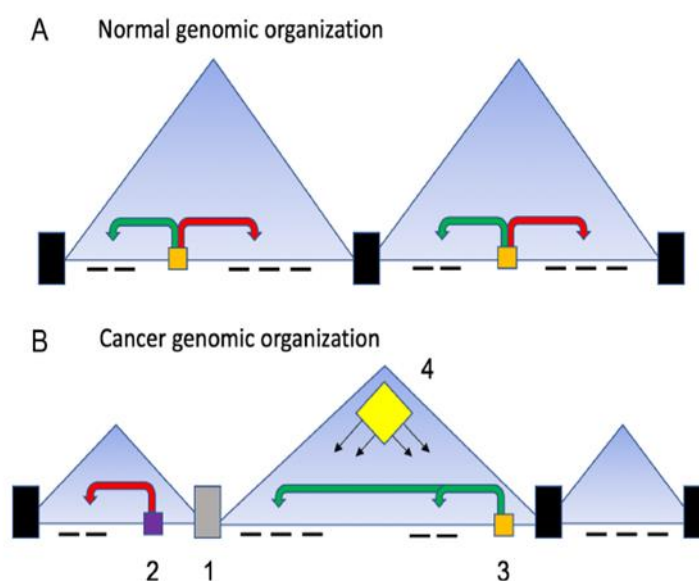
Σε ορισμένες περιπτώσεις, όπως στον καρκίνο, τα όρια των TADs μπορεί να υποστούν μεταλλάξεις ή να επηρεαστούν από επιγενετικές τροποποιήσεις, όπως η μεθυλίωση του DNA. Αυτό έχει ως αποτέλεσμα την απώλεια της μόνωσης και τη ρύθμιση γονιδίων από ενισχυτές που κανονικά θα ήταν απομονωμένοι. Παρόμοιες



διαταραχές έχουν παρατηρηθεί σε διάφορους τύπους καρκίνου, όπως και ο καρκίνος του προστάτη [17].

### 1.2.3 Απορρύθμιση των TADs στον καρκίνο

Όπως είναι λογικό, το σύνολο των γενετικών και επιγενετικών αλλαγών στον καρκίνο, επηρεάζει και τον σχηματισμό των TADs. Οι μεταλλάξεις ή οι διαγραφές στις CTCF και τις κοχεσίνες προκαλούν απώλεια μόνωσης στα σύνορα. Αυτό έχει ως αποτέλεσμα τη συγχώνευση TADs, αλλά πιο συχνά τη δημιουργία νέων, μικρότερων περιοχών. Με τον τρόπο, πολλοί ενισχυτές μπορούν να αλληλεπιδρούν με γονίδια εκτός της φυσιολογικής τους περιοχής, που κανονικά θα ήταν μονωμένα [17,21] (Εικόνα 1.15).



Εικόνα 1.15: Φυσιολογική οργάνωση των TADs (A) και η αλλαγές που παρατηρούνται στον καρκίνο (B) [36]

Στον καρκίνο του προστάτη, παρατηρούνται αντίστοιχα φαινόμενα. Για παράδειγμα, η μετατόπιση συνόρων TADs μπορεί να φέρει ενισχυτές σε επαφή με ογκογονίδια, όπως το TMPRSS2-ERG, προκαλώντας υπερέκφραση τους, προάγοντας την ογκογένεση. Επιπλέον, οι αλλαγές στις χρωματίνες A/B, που χαρακτηρίζονται ως ενεργές (A) ή ανενεργές (B), επηρεάζουν την προσβασιμότητα της χρωματίνης και τη ρύθμιση γονιδίων. Οι περιοχές A σχετίζονται με TADs πλούσια σε αλληλεπιδράσεις υποκινητών και ενισχυτών. Σε αυτές, παρατηρούνται αυξημένα ποσοστά αναδιατάξεων (structural variations, SVs) και ενισχυμένη γονιδιακή έκφραση σε γονίδια-κλειδιά όπως το AR και το WNT5A. Αυτές οι αλλαγές αυξάνουν τις πιθανότητες μετάστασης και αντοχής στη θεραπεία [22, 23]. Γενικότερα, στους ορμονοεξαρτώμενους καρκίνους, όπως ο καρκίνος του προστάτη, οι αλλαγές στα TADs επηρεάζουν τη δράση των ενισχυτών που ρυθμίζονται από ορμόνες. Πιο ειδικά, οι ανδρογονικά επαγόμενοι ενισχυτές μπορούν να ενεργοποιήσουν γονίδια που ενισχύουν τον κυτταρικό πολλαπλασιασμό, μηχανισμός που αξιοποιείται εκτενώς στα αρχικά στάδια [17].



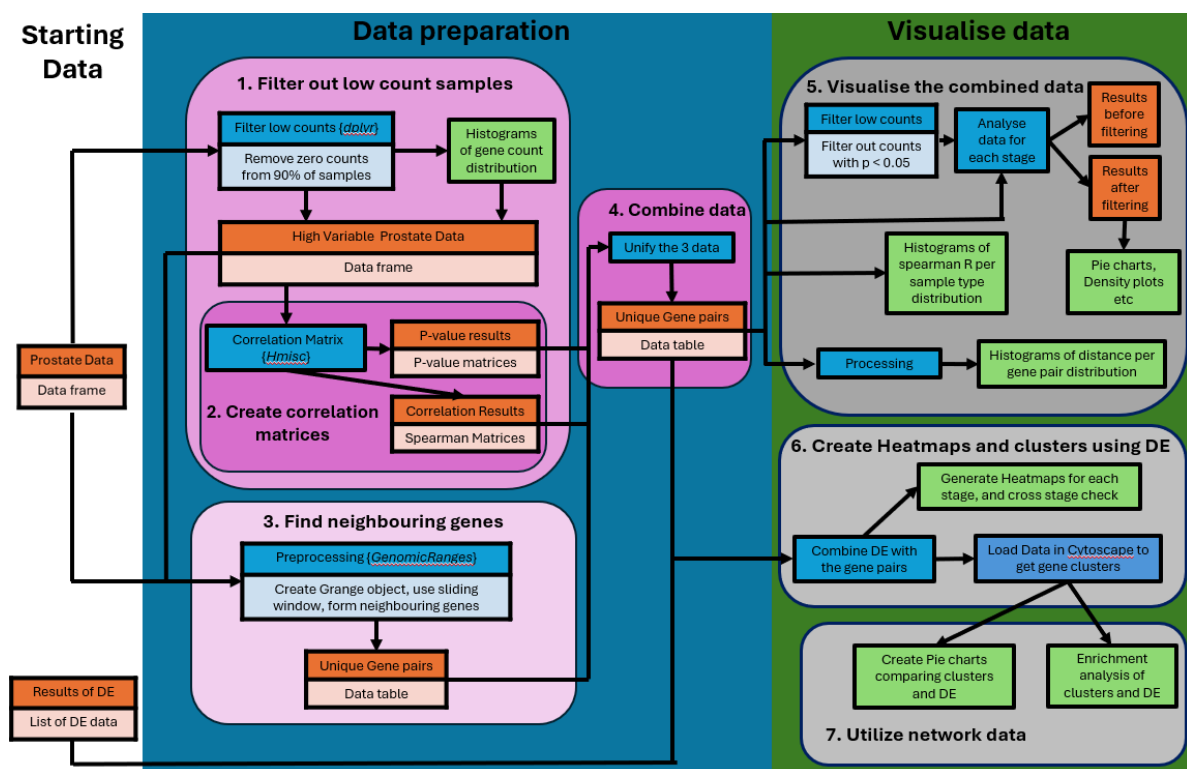
## 2. Υλικά και Μεθοδολογία

### 2.1 Υλικά

Αρχικά, χρησιμοποιήθηκαν δημόσια διαθέσιμα RNA-seq δεδομένα από τις κοινοπραξίες TCGA και GTEX, που περιγράφουν την κανονικοποιημένη έκφραση γονιδίων, σε δείγματα προστάτη. Τα δείγματα αυτά είναι κατηγοριοποιημένα σε 5 κατηγορίες ως φυσιολογικός ιστός, παρακαρκινική βιοψία, και όγκοι σταδίου I-IV. Συγκεκριμένα υπήρχαν 119 δείγματα από φυσιολογικό ιστό προστάτη, 52 από παρακαρκινικό, και 23, 167, 159 και 146 από τα στάδια I-IV για που αφορούσαν δεδομένα έκφρασης για συνολικά 56.403 γονίδια. Κάθε σύνολο δεδομένων περιλαμβάνει στοιχεία γονιδιακής έκφρασης, annotated σύμφωνα με το hg38. Παράλληλα, τα αποτελέσματα της ανάλυσης διαφορικής έκφρασης με βάση τον αλγόριθμο edgeR [24], για τα ίδια δείγματα, χρησιμοποιήθηκαν σε μεταγενέστερα βήματα της ανάλυσης.

### 3.2 Μέθοδοι

Το pipeline της πτυχιακής αποτελείται από 7 διακριτά βήματα τα οποία συνοψίζονται στην εξής εικόνα:



Εικόνα 2.1: Pipeline της συνολικής μεθοδολογίας.

#### 2.2.1 Φιλτράρισμα δειγμάτων με χαμηλές τιμές έκφρασης

Σε πρώτο στάδιο αφαιρούνται τα γονίδια με μηδενικές τιμές έκφρασης. Τα RNA-seq δεδομένα, με ονομασία Prostate Data, εισάγονται μέσω του πακέτου *openxlsx*

[v4.2.5.2][25]. Οι στήλες μεταδεδομένων αφαιρούνται, ώστε να διατηρηθούν μόνο οι τιμές έκφρασης των γονιδίων για την ανάλυση. Επιπλέον, αφαιρούνται γονίδια που εμφανίζονται παραπάνω από μια φορές, ώστε κάθε γονίδιο να αναπαρίσταται μοναδικά. Συγκεκριμένα, από την στήλη που αναγράφει τα ονόματα των γονιδίων (*gene\_name*), μέσω του function **!duplicated**, αφαιρούνται όλες οι επαναλήψεις ενός γονιδίου πέρα από την πρώτη. Τα γονίδια αυτά είχαν ίδιο όνομα αλλά διαφορετικό ENSEMBL ID (στήλη ENSEMBL ID). Οι τιμές έκφρασης που, παραμένουν ως στήλες στα δεδομένα, φιλτράρονται για την αφαίρεση γονιδίων με χαμηλή έκφραση. Ως προς αυτό το φίλτρο, ένα γονίδιο διατηρείται μόνο εάν τουλάχιστον στο 10% των δειγμάτων έχει μη μηδενικές τιμές, ώστε να αποτραπεί η αφαίρεση ισχυρά διαφορεικά εκφρασμένων γονιδίων μεταξύ φυσιολογικών και καρκινικών δειγμάτων. Με τον τρόπο αυτό διασφαλίζονται τα βιολογικά σημαντικά γονίδια. Το φιλτράρισμα αυτό εφαρμόζεται ξεχωριστά για τα δείγματα Normal και για τα δείγματα που αφορούν όγκο, δημιουργώντας δύο διακριτά φιλτραρισμένα dataset. Τα dataset αυτά χρησιμοποιούνται για την παραγωγή διαγραμμάτων, τα οποία απεικονίζουν την επίδραση του φιλτραρίσματος στις κατανομές έκφρασης των γονιδίων στις 2 συνθήκες. Τα διαγράμματα αναδιατάσσονται μέσω του πακέτου *reshape2* [v1.4.4][26], για να είναι πιο εύκολη η σύγκριση των κατανομών. Η γενικότερη διαχείριση των δεδομένων και η παραγωγή των διαγραμμάτων γίνεται μέσω του πακέτου *tidyverse* [v2.0.0][27]. Στο τέλος αυτού του βήματος, προκύπτει ένα αρχείο που περιέχει 34.906 γονίδια, με φιλτραρισμένες τις μηδενικές τιμές έκφρασης (*HighVariable.xlsx*).

## 2.2.2 Σχηματισμός μητρώων συσχέτισης

Στο δεύτερο στάδιο, σχηματίζονται πίνακες συσχέτισης αλλά και p-value, ανά ζεύγη γονιδίων σε κάθε χρωμόσωμα και για κάθε δείγμα. Αρχικά, γίνεται η φόρτωση των δεδομένων γονιδιακής έκφρασης από το αρχείο *HighVariable.xlsx*, μέσω του πακέτου *openxlsx* [v4.2.5.2]. Τα δεδομένα ταξινομούνται σε έξι στάδια (Normal, Paracancerous, Stage I, Stage II, Stage III και Stage IV), με βάση τη στήλη *Sample\_Status*. Για κάθε στάδιο, τα δεδομένα οργανώνονται ώστε να περιλαμβάνουν πληροφορίες σχετικά με τα γονίδια, όπως το όνομα του γονιδίου, το χρωμόσωμα, τη θέση έναρξης και λήξης, την περιεκτικότητα σε GC, την κατεύθυνση (*strand*) και τον βιότυπο που υποδηλώνει το είδος του μεταγράφου. Στη συνέχεια, τα δεδομένα κάθε σταδίου χωρίζονται ανά χρωμόσωμα, ώστε να διευκολυνθεί υπολογιστικά η εύρεση των συσχετίσεων. Για κάθε χρωμόσωμα, οι ήδη λογαριθμημένες τιμές έκφρασης, μετατρέπονται σε numeric και οργανώνονται σε πίνακες, όπου οι σειρές αντιστοιχούν στα ονόματα των γονιδίων. Οι συσχετίσεις υπολογίζονται χρησιμοποιώντας το function **rcorr** του πακέτου *Hmisc* [v5.1-3][28]. Έτσι, προκύπτουν πίνακες συσχέτισης (*correlation matrices*) και αντίστοιχοι πίνακες με τις p-value τιμές. Τα αποτελέσματα αποθηκεύονται σε δύο λίστες: τα *correlation\_results* και τα *p\_value\_results*, αντίστοιχα. Τα αρχεία σώζονται ως *correlation\_resultsHmisc.RData* και *p\_value\_resultsHmisc.RData*. Η γενικότερη διαχείριση των δεδομένων γίνεται μέσω του πακέτου *tidyverse* [v2.0.0]

### 2.2.3 Εύρεση γειτονικών γονιδίων

Στην συνέχεια, η ανάλυση επικεντρώνεται στον εντοπισμό γειτονικών γονιδίων και στη δημιουργία ζευγών γονιδίων με βάση τη χρωμοσωμική τους θέση (γειτονικά). Όπως και προηγουμένως φορτώνονται τα δεδομένα γονιδιακής έκφρασης από το αρχείο *HighVariable.xlsx* μέσω του πακέτου *openxlsx* [v4.2.5.2]. Από αυτό το αρχείο, δημιουργείται ένα νέο data set (*Prostate\_data*), όπου διατηρούνται μόνο οι στήλες των μεταδεδομένων, δηλαδή το όνομα του γονιδίου, το χρωμόσωμα, η κατεύθυνση, ο βιότυπος, η θέση έναρξης/λήξης και η περιεκτικότητα σε GC. Στην συνέχεια, γίνεται κυρίως χρήση του πακέτου *GenomicRanges* [v1.56.2][29]. Αρχικά, δημιουργείται ένα αντικείμενο *GRanges*, το οποίο περιγράφει τις περιοχές του ανθρώπινου γονιδιώματος σύμφωνα με την έκδοση hg38, το οποίο ονομάζεται *Hg38\_gr*. Ταυτόχρονα, τα δεδομένα του *Prostate\_data*, μετατρέπονται και αυτά σε *Granges* αντικείμενο, με ονομασία *ProstateData\_gr*. Έπειτα, πάνω στο *hg38\_gr*, χρησιμοποιείται το function *slidingWindows*, για να χωριστεί κάθε χρωμόσωμα σε επικαλυπτόμενα παράθυρα μεγέθους 500kb, με βήμα 250kb. Τα παράθυρα αυτά αντιπροσωπεύουν τμήματα του γονιδιώματος και επιτρέπουν την εντόπιση περιοχών τον μέγεθος των οποίων είναι σε αρμονία με το τυπικό μήκος των TADs. Τα γονίδια που βρίσκονται εντός των παραθύρων εντοπίζονται, με την χρήση του function ***findOverlaps***. Τα γονίδια που ανήκουν στο ίδιο παράθυρο ομαδοποιούνται. Επιπλέον, τα γονίδια που βρίσκονται εντός κάθε παραθύρου συνδυάζονται σε ζεύγη χρησιμοποιώντας το function ***combn***. Τέλος, αφαιρούνται τα επαναλαμβανόμενα ζεύγη και οργανώνονται τα δεδομένα σε πίνακες, όπου κάθε ζεύγος περιλαμβάνει: (1) το όνομα και τη χρωμοσωμική θέση (θέση έναρξης και λήξης) κάθε γονιδίου, (2) την κατεύθυνση (*strand*) και τον βιότυπο των γονιδίων, (3) την απόσταση μεταξύ των γονιδίων, η οποία υπολογίζεται μέσω του custom function ***calculate\_distance***. Τα τελικά αποτελέσματα αποθηκεύονται ως *NeighbouringGenes.xlsx*, το οποίο περιλαμβάνει όλα τα μοναδικά ζεύγη γονιδίων με τις αντίστοιχες πληροφορίες. Η διαχείριση των δεδομένων γίνεται όπως και πριν.

### 2.2.4 Συνδυασμός δεδομένων

Στο τέταρτο στάδιο, συνδυάζονται τα δεδομένα συσχέτισεων μεταξύ γονιδίων με τις χρωμοσωμικές τους θέσεις και άλλες βιολογικές πληροφορίες, με τελικό σκοπό τη δημιουργία ενός συνολικού dataset. Πιο συγκεκριμένα, γίνεται φόρτωση των δεδομένων από τα αρχεία *NeighbouringGenes.xlsx* (ως *Unique\_GenePairs*), *correlation\_resultsHmisc.RData* και *p\_value\_resultsHmisc.RData*. Το custom function ***process\_correlation\_results***, αναλαμβάνει την επεξεργασία των δεδομένων συσχέτισης για κάθε στάδιο (*Normal*, *Paracancerous*, *Stage I-IV*). Αρχικά, το function εξάγει τους πίνακες συσχέτισης και p-values για κάθε χρωμόσωμα του συγκεκριμένου σταδίου. Οι πίνακες αυτοί περιέχουν τις τιμές συσχέτισης *Spearman* και τις αντίστοιχες τιμές σημαντικότητας (p-values) για όλα τα ζεύγη γονιδίων. Στη συνέχεια, οι πίνακες συσχέτισης και p-values μετατρέπονται σε μακρά μορφή (*long format*), όπου κάθε γραμμή του πίνακα αντιπροσωπεύει ένα ζεύγος γονιδίων μαζί με τη συσχέτισή τους και το p-value. Οι πληροφορίες αυτές συνδυάζονται σε έναν ενιαίο πίνακα για κάθε χρωμόσωμα και στάδιο. Έπειτα, τα δεδομένα συσχέτισης ενώνονται με τα δεδομένα γειτονικών γονιδίων (*unique\_GenePairs*). Στο dataset που προκύπτει, για κάθε ζεύγος

γονιδίων προστίθενται οι τιμές συσχέτισης Spearman και τα αντίστοιχα p-values για το εκάστοτε στάδιο. Οι στήλες αυτές ονομάζονται R\_Stage (τιμές συσχέτισης) και P\_Stage (p-values). Το τελικό αποτέλεσμα αποθηκεύεται ως NeighbouringGenesDetails.xlsx, το οποίο περιλαμβάνει πληροφορίες όπως τα ονόματα των γονιδίων, τις χρωμοσωμικές τους θέσεις, τις βιολογικές τους ιδιότητες, καθώς και τις τιμές συσχέτισης και τα p-values για κάθε στάδιο.

### 2.2.5 Απεικόνιση των δεδομένων

Σε αυτήν την φάση γίνεται διερεύνηση στατιστικών στοιχείων και οπτικοποίηση των δεδομένων, που προκύπτουν από την ένωση που έγινε στο προηγούμενο βήμα. Η διαδικασία ξεκινά με τη φόρτωση του ολοκληρωμένου αρχείου δεδομένων NeighbouringGenesDetails.xlsx, ενώ τα δεδομένα οργανώνονται ξανά ανά στάδιο. Αρχικά, εξετάζεται η εγγύτητα των ζευγών γονιδίων. Δημιουργείται λοιπόν, ένα ιστόγραμμα που παρουσιάζει την κατανομή των αποστάσεων για όλα τα ζεύγη γονιδίων. Ταυτόχρονα, γίνεται ανάλυση για τα γονίδια που αλληλεπικαλύπτονται (με απόσταση 0). Για αυτά τα ζεύγη, λαμβάνεται υπόψιν η φορά μεταγραφής των γονιδίων, παρέχοντας πληροφορίες για το αν μεταγράφονται από την ίδια ή την αντίθετη αλυσίδα του DNA. Στη συνέχεια, υπολογίζεται η συχνότητα των θετικών και αρνητικών συσχετίσεων για κάθε στάδιο. Για να φανεί αυτό, δημιουργούνται πίνακες και πίτες που παρουσιάζουν το ποσοστό θετικών και αρνητικών συσχετίσεων ανά στάδιο. Παράλληλα, οι τιμές συσχέτισης (R) οργανώνονται σε γραφήματα κατανομής, τα οποία προσδιορίζουν το εύρος και τη συμπεριφορά των δεδομένων συσχέτισης σε κάθε στάδιο. Με την προσθήκη μάλιστα, ενώ κατωφλιού  $>0.3$  και  $<-0.3$  πάνω στις κατανομές αυτές, γίνεται ανάλυση της διαφοράς στα ποσοστά 'ίσχυρών' συσχετίσεων ανά στάδιο.

Με την ολοκλήρωση της αρχικής ανάλυσης, τα δεδομένα φιλτράρονται, ώστε να παραμείνουν οι στατιστικά σημαντικές συσχετίσεις μεταξύ γονιδίων ( $p\text{-value} < 0.05$ ). Όσες συσχετίσεις παραμείνουν μετά το φιλτράρισμα, οργανώνονται εκ νέου και αναλύονται για τον υπολογισμό των ποσοστών θετικών και αρνητικών συσχετίσεων. Η απεικόνιση της αναλογίας των συσχετίσεων αυτών γίνεται με πίτες συχνότητας. Για τον έλεγχο της σταθερότητας των αναλογιών που προέκυψαν, πραγματοποιείτε bootstrapping. Συγκεκριμένα, μέσω της μεθόδου αυτής, λαμβάνεται το 80% του συνολικού αριθμού ζευγών γονιδίων για το χρωμόσωμα 22. Αυτή η τυχαία δειγματοληψία γίνεται 1000 φορές, τόσο πριν όσο και μετά το φιλτράρισμα, και κάθε επανάληψη είναι διαφορετική από την άλλη (**replace=TRUE**). Έπειτα, από κάθε επανάληψη, λαμβάνεται ένα ποσοστό θετικών και αρνητικών συσχετίσεων, ως προς το σύνολο των ζευγών γονιδίων. Στην συνέχεια, υπολογίζεται ο μέσος όρος θετικών συσχετίσεων από το σύνολο των επαναλήψεων. Η επιλογή του χρωμοσωματος 22 είναι τυχαία. Σκοπός είναι να διαπιστωθεί αν τα ποσοστά συσχετίσεων στο σύνολο των δεδομένων είναι κοντά με αυτά που προκύπτουν από το bootstrap. Παράλληλα, δημιουργούνται ridge plots που παρουσιάζουν την συσχέτιση της απόστασης και των θετικών/αρνητικών συσχετίσεων ανά στάδιο. Για την κατασκευή τους χρησιμοποιήθηκαν τα πακέτα *ggribes* [v0.5.6][30] και *viridis* [v0.6.5][31]

## 2.2.6 Δημιουργία θερμικών χαρτών και γονιδιακών δικτύων

Στο έκτο στάδιο της ανάλυσης, χρησιμοποιούνται διαφορετικά μέσα για την απεικόνιση αλληλεπιδράσεων μεταξύ γονιδίων. Φορτώνονται ξανά τα αρχεία *NeighbouringGenesDetails.xlsx* και το αρχείο *correlation\_resultsHmisc.RData* με τους πίνακες συσχετίσεων για όλα τα στάδια και χρωμοσώματα. Παράλληλα, οι λίστες διαφορικά εκφρασμένων γονιδίων (DEGs) εισάγονται από το αρχείο *PRAD\_DEGsFilt.xlsx*. Αφού φορτωθούν τα δεδομένα, οργανώνονται τόσο ανά στάδιο (Stage I-IV) όσο και ανά και χρωμόσωμα. Σε πρώτη φάση, διατηρούνται μόνο τα ζεύγη γονιδίων που είναι στατιστικώς σημαντικά σε κάποια συνθήκη, και όχι στον φυσιολογικό ιστό ( $P\text{-value Normal} > 0.05$  &  $P\text{-value Tumor} < 0.05$ ). Στη συνέχεια, ελέγχεται αν τα γονίδια των ζευγών περιλαμβάνονται στις λίστες των διαφορικά εκφρασμένων γονιδίων (DEGs) του αντίστοιχου σταδίου. Οι πίνακες συσχετίσεων οργανώνονται ώστε να περιλαμβάνουν μόνο τα σημαντικά γονίδια, τα οποία ταξινομούνται με βάση τις χρωμοσωμικές τους θέσεις, τη σωστή σειρά που έχουν πάνω στο γονιδίωμα. Τα αρχεία αποθηκεύονται ως *Tumor\_pairs*, ανάλογα με το στάδιο. Η δημιουργία των θερμικών χαρτών πραγματοποιείται μέσω του πακέτου *rheatmap* [v1.0.12][32], το οποίο απεικονίζει τις τιμές συσχέτισης για κάθε ζεύγος γονιδίων. Οι θερμικοί χάρτες δημιουργούνται ξεχωριστά για κάθε στάδιο και χρωμόσωμα, ενώ ο κώδικας παρακάμπτει τη δημιουργία τους εάν δεν υπάρχουν επαρκή δεδομένα. Αυτό συμβαίνει όταν υπάρχουν λιγότερα από δύο γονίδια στο *correlation matrix* (σφάλμα σε προηγούμενο βήμα) ή όταν ο υποπίνακας που σχηματίζεται έχει λιγότερες από 2 γραμμές/ στήλες (ελάχιστες συσχετίσεις σε κάποιο χρωμόσωμα). Οι θερμικοί χάρτες αναλύονται και παρατηρούνται οι διαφορές ανά στάδιο. Δεδομένου ότι το φίλτρο  $P\text{-value Normal} > 0.05$  &  $P\text{-value Tumor} < 0.05$ , διαφέρει ανά στάδιο, και μάλιστα σε μεγάλο βαθμό, οι θερμικοί χάρτες εμφανίζουν μεγάλες διαφορές. Έτσι απεικονίζονται διαφορετικά γονίδια σε κάθε θερμικό χάρτη, για το ίδιο χρωμόσωμα ανά στάδιο. Για τον λόγο αυτό χρησιμοποιείται ένα δεύτερο φίλτρο συμπληρωματικά, όπου διατηρούνται μόνο τα ζεύγη γονιδίων που είναι στατιστικώς σημαντικά τουλάχιστον σε ένα από τα στάδια καρκίνου και όχι σε φυσιολογικό ιστό. Με τον τρόπο αυτό, δημιουργείται ένα υποσύνολο γονιδίων, που μέρος τους αξιοποιείται για να σχηματιστούν θερμικοί χάρτες που απεικονίζουν τμήματα των πρώτων θερμικών χαρτών που παρουσίαζαν ενδιαφέρον. Παράλληλα, δεδομένου ότι η στατιστική σημαντικότητα διαφέρει ανά στάδιο, με βάση τα γονίδια που επιλέχθηκαν σχηματίζονται και θερμικοί χάρτες που απεικονίζουν το  $p\text{-value}$  για τα ζεύγη αυτά. Συμπληρωματικά, προστίθενται και διαγράμματα που παρουσιάζουν την έκφραση των γονιδίων που επιλέχθηκαν, τροποποιημένη σε τιμές απόκλισης  $z$  ανά στάδιο της νόσου, ώστε να αντιπαραβάλλεται η διαφορική έκφραση με την συνμεταβολή της έκφρασης κάθε γονιδιακού ζεύγους. Η δημιουργία των χαρτών πραγματοποιήθηκε με μια σειρά από *custom functions*, που ονομάζονται **filter\_matrices**, **filter\_pvalue\_matrices**, **filter\_zscore**, **generate\_heatmaps**, **generate\_pvalue\_heatmaps** και **generate\_zcore\_heatmaps**. Για την ρύθμιση του χρώματος στον χάρτη των  $p\text{-values}$ , χρησιμοποιήθηκε το πακέτο *RColorBrewer* [v.1.1.3][33], με σκοπό τον πιο εύκολο διαχωρισμό των στατιστικά σημαντικών αλληλεπιδράσεων.

Σε συνδυασμό λοιπόν με τον σχηματισμό των θερμικών χαρτών, κάθε ένα από τα Tumor\_pairs αρχεία, ενσωματώνεται στο Cytoscape. Έτσι για κάθε διαφορετικό στάδιο καρκίνου του προστάτη, σχηματίζονται δίκτυα γονιδίων, καθένα από τα οποία είναι στατιστικά σημαντικό και τα συστατικά του είναι διαφορετικά εκφρασμένα. Μάλιστα με την προσθήκη ενός φίλτρου, τα δεδομένα περιλαμβάνουν μόνο ζεύγη με τιμές R άνω του 0.5, ώστε να παραμείνουν μόνο οι ισχυρές συσχετίσεις, και άρα πιο υψηλής ποιότητας δίκτυα.

### 2.2.7 Αξιοποίηση των δεδομένων των δικτύων

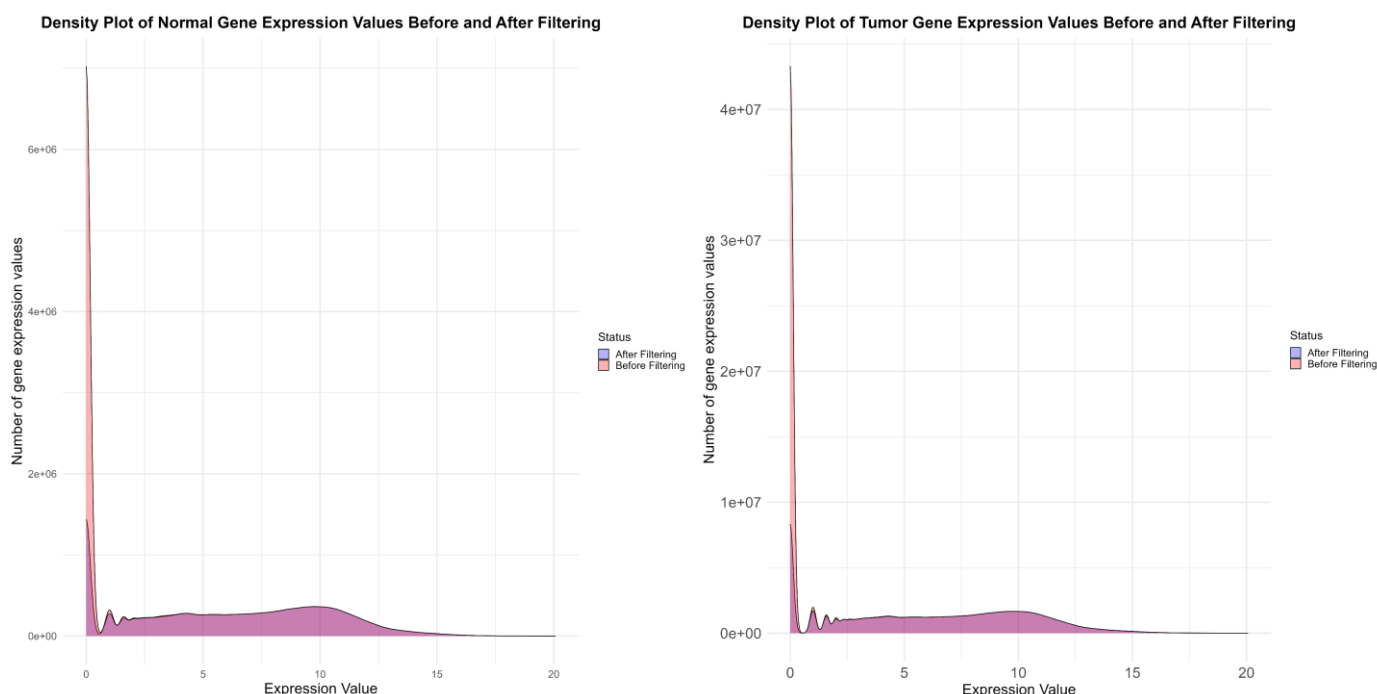
Τα δεδομένα που έχουν προκύψει από τους πίνακες του Cytoscape μπορούν να αξιοποιηθούν με μια πληθώρα διαφορετικών τρόπων. Αρχικά απομονώνονται τα μοναδικά γονίδια τόσο πριν όσο και μετά την εφαρμογή του φίλτρου για  $R < 0.5$ . Ταυτόχρονα απομονώνονται τα μοναδικά γονίδια από τις λίστες διαφορικής έκφρασης για κάθε στάδιο. Τα δεδομένα αυτά εισάγονται στο Enrichr-KG, για να πραγματοποιηθεί ανάλυση οντολογίας στο καθένα από αυτά, με βιβλιοθήκες τις WikiPathway, GO, Reactome και KEGG. Τα διαγράμματα που προκύπτουν από την εκάστοτε ανάλυση, αποθηκεύονται και προστίθενται στα αποτελέσματα της εργασίας.

Επιπλέον, γίνεται ανάλυση των δεδομένων για την δημιουργία γραφημάτων που απεικονίζουν την κατανομή των διαφορετικών βιοτύπων γονιδίων (gene biotypes) σε διαφορετικά στάδια της νόσου. Τα αρχεία από το Cytoscape τροποποιούνται με διάφορα functions ώστε να είναι πιο εύκολη η ανάλυση τους. Στη συνέχεια, οι βιότυποι των γονιδίων κατηγοριοποιούνται σε πέντε βασικές κατηγορίες: Protein, LincRNA, Antisense, Pseudogene και Other. Τα δεδομένα οργάνωνται περαιτέρω ώστε να υπολογιστούν τα ποσοστά των βιοτύπων. Για κάθε στάδιο, δημιουργούνται κυκλικά γραφήματα (pie charts) που απεικονίζουν τη κατανομή των βιοτύπων. Αυτό γίνεται τόσο για τα ζεύγη γονιδίων πριν αλλά και μετά το φίλτρο  $R < 0.5$ . Παράλληλα, γίνεται επεξεργασία δεδομένων των διαφορικά εκφρασμένων γονιδίων (DEGs) για κάθε στάδιο, τα οποία φορτώνονται από τις λίστες. Τα DEGs κατηγοριοποιούνται με αντίστοιχο τρόπο για τους διαφορετικούς βιοτύπους τους και δημιουργούνται αντίστοιχα κυκλικά γραφήματα. Ο σχηματισμός των κυκλικών γραφημάτων γίνεται μαζικά, από το custom function **create\_pie\_chart\_DE**, αρχικά για τα δεδομένα του Cytoscape και μετά για DEGs. Ο μαζικός χειρισμός των δεδομένων των DEGs γίνεται μέσω του custom function **process\_stage\_data**.

## 3. Αποτελέσματα

### 3.1 Προετοιμασία δεδομένων

Κατά την προετοιμασία των δεδομένων, φάνηκε ότι το όνομα ορισμένων γονιδίων, αναγράφεται παραπάνω από μια φορά στην στήλη `gene_names`. Τα γονίδια αυτά έχουν το ίδιο όνομα αλλά διαφορετικό ENSEMBL ID. Μέσω της R υπολογίζεται ότι υπάρχουν συνολικά 272 γονίδια με παραπάνω από μια επαναλήψεις. Τα γονίδια μπορούν να ταξινομηθούν ως γονίδια που κωδικοποιούν πρωτεΐνες (63), μακριά μη κωδικοποιητικά RNA (31), μικρά ρυθμιστικά RNA (144) και ψευδογονίδια (15). Τα υπόλοιπα 19 γονίδια δεν ταιριάζουν σε κάποια από αυτές τις κατηγορίες, καθώς 18 από αυτά είναι *miscRNAs* και το 1 είναι γονίδιο IGV.



**Εικόνα 3.1:** Μεταβολή της έκφρασης γονιδίων πριν και μετά το φιλτράρισμα, σε φυσιολογικά δείγματα προστάτη(αριστερά) και σε δείγματα προστάτη με όγκο (δεξιά).

Όπως ήδη αναφέρθηκε, τα αρχικά δεδομένα φιλτραρίστηκαν με σκοπό την αφαίρεση των γονιδίων με μηδενική έκφραση σε τουλάχιστον 90% των δειγμάτων. Με αυτόν τον τρόπο, υπολογίζεται πως ο αριθμός των γονιδίων μεταβάλλεται από 56.404 (μαζί με τις επαναλήψεις), σε 34.906 (οι επαναλήψεις έχουν ήδη φιλτραριστεί). Στα διαγράμματα πυκνότητας, οι αλλαγές στην κατανομή της έκφρασης των γονιδίων φαίνονται με ποιοτικό τρόπο. Ο διαχωρισμός γίνεται έτσι, ώστε να απεικονίζονται σε 2 διαγράμματα τα δείγματα με φυσιολογικό ιστό και αυτά με καρκινικό. Σε κάθε διάγραμμα, γίνεται σύγκριση της εκτιμώμενης κατανομής πριν και μετά την εφαρμογή

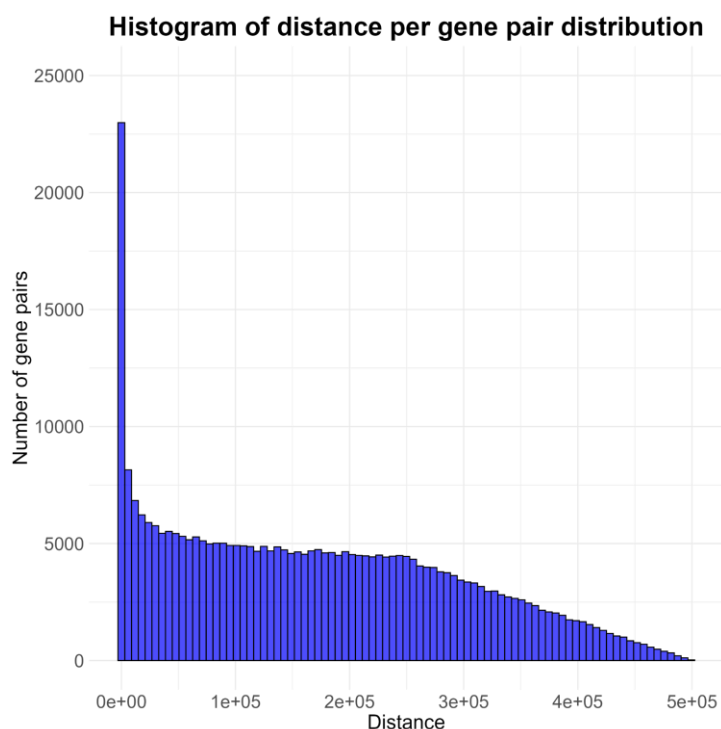
του φίλτρου, και επισημαίνεται έτσι ο μεγάλος αριθμός γονιδίων που αφαιρέθηκαν. (Εικόνες 3.1).

Στην συνέχεια, δημιουργούνται τόσο οι μήτρες συσχέτισης (Spearman R) και στατιστικής σημαντικότητας (p-value), αλλά και ο πίνακας ζευγών γειτονικών γονιδίων για κάθε χρωμόσωμα. Τα δεδομένα αυτά συνδυάζονται, ώστε να παρατεθεί ο βαθμός της αλληλεπίδρασης και η στατιστική σημαντικότητα κάθε ζεύγους, για κάθε στάδιο. Έτσι στο τέλος προκύπτουν περίπου 320.000 ζεύγη γονιδίων σε κάθε στάδιο. Ο τελικός πίνακας που δημιουργείται και συνδυάζει όλες τις πληροφορίες ονομάζεται NeighbouringGenesDetails.

## 3.2 Απεικόνιση δεδομένων

### 3.2.1 Σχηματισμός ιστογράμματος κατανομής απόστασης

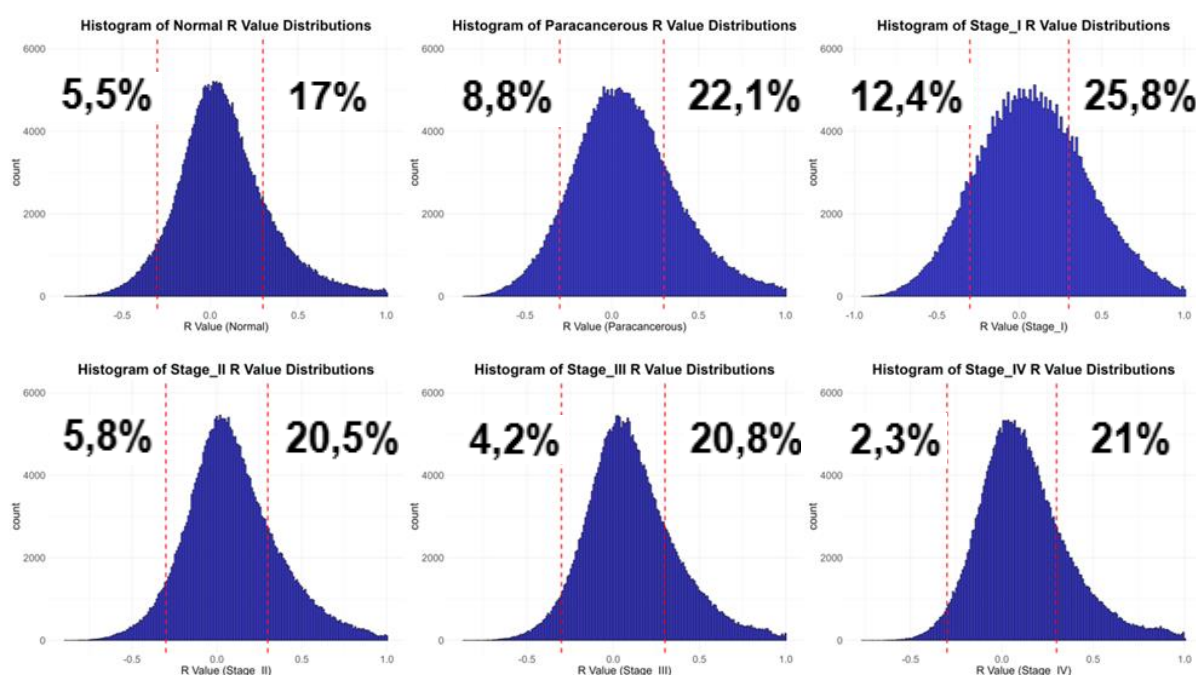
Επόμενο βήμα είναι η γραφική απεικόνιση των πληροφοριών που παρατίθενται στο NeighbouringGenesDetails. Αρχικά, αξιοποιώντας την στήλη Distance, δημιουργείται μια κατανομή των αποστάσεων μεταξύ των ζευγών γονιδίων. Δεδομένου ότι με βάση τον κώδικα, η μέγιστη απόσταση μεταξύ 2 γονιδίων είναι 500 kbs, στο ιστόγραμμα που προκύπτει, η μέγιστη τιμή του άξονα x (απόσταση) είναι 500.000. Με βάση το διάγραμμα λοιπόν, μεγάλο μέρος των ζευγών γονιδίων απέχει πολύ μικρή απόσταση μεταξύ τους. Ενδεικτικά, από το σύνολο των 320.000 αλληλεπιδράσεων, περίπου 75 χιλιάδες από αυτές κατανέμονται σε απόσταση μικρότερη των 50 kbs (Εικόνα 3.2). Μάλιστα, τα 15.580 γονίδια αλληλοεπικαλύπτονται με τα 7.895 να έχουν αντίθετο προσανατολισμό (+/-) και τα 7685 έχουν κοινό προσανατολισμό (2.554 +/- και 3.008 +/+).





### 3.2.2 Σχηματισμός διαγραμμάτων κατανομής των τιμών R ανά στάδιο

Στην συνέχεια, γίνεται απεικόνιση της κατανομής των τιμών R ανά στάδιο. Αυτό πραγματοποιείται μέσω των στήλων  $R_{<stage>}$ , και δημιουργούνται 6 διαφορετικά ιστογράμματα. Στα ιστογράμματα αυτά, αναγράφονται και όρια στους άξονες x (Spearman R), τα οποία οπτικοποιούν το πλήθος των ζευγών που ξεπερνούν τις τιμές  $R = 0.3$  και  $R = -0.3$  (Εικόνα 3.4).



Εικόνα 3.4: Κατανομή των τιμών R ανά στάδιο

Με βάση αυτό, σε όλες τις περιπτώσεις, ένα μικρό κλάσμα τους εμφανίζει συσχετίσεις υψηλότερες από αυτές τις τιμές. Ενδιαφέρον παρουσιάζει ότι οι αρνητικές τιμές συσχέτισης που ξεπερνούν το  $R = -0.3$ , είναι πολύ λιγότερες από τις θετικές, σε όλα τα στάδια. Επιπλέον, η αναλογία θετικών προς αρνητικών συσχετίσεων αυξάνεται αρκετά σε πιο προχωρημένα στάδια καρκίνου. Ενδεικτικά, στα δείγματα που βρίσκονται στο 1<sup>ο</sup> στάδιο καρκίνου, οι συνολικές υψηλές συσχετίσεις είναι 120.767, και η αναλογία θετικών προς αρνητικών συσχετίσεων είναι 67% προς 33% (81.118 προς 39.049 αντίστοιχα). Στα δείγματα που βρίσκονται σε 4<sup>ο</sup> στάδιο καρκίνου, οι συνολικές συσχετίσεις είναι 74.456 ενώ η αναλογία είναι 92% (61.117) θετικές συσχετίσεις προς 8% (7.279 αρνητικές). Τέλος, τα δείγματα που αφορούν φυσιολογικό ιστό και τα παρακαρκινικά δείγματα, εμφανίζουν περίπου ίδια ποσοστά θετικών συσχετίσεων ως μέρος των υψηλών συσχετίσεων, 76% και 71,5% (70.090 σε 97.948) αντίστοιχα. Οι αναλυτικοί αριθμοί για κάθε στάδιο φαίνονται στον **πίνακα 3.1**.

## Αριθμός ζευγών γονιδίων μετά την εφαρμογή του κατωφλίου

| Sample Type | R>0.3          | R< -0.3       | Total number of gene pairs | Number of samples |
|-------------|----------------|---------------|----------------------------|-------------------|
| Normal      | 17,3%( 55.182) | 5,5%(17.523)  | 320.828                    | 119               |
| Paracanc.   | 22,1%(70.090)  | 8,8%(27.858)  | 316.633                    | 52                |
| Stage I     | 25,8%(81.118)  | 12,4%(39.049) | 314.935                    | 23                |
| Stage II    | 20,5%(65.553)  | 5.8%(18.765)  | 320.498                    | 167               |
| Stage III   | 20,8%(66.561)  | 4.2%(13.555)  | 320.620                    | 159               |
| Stage IV    | 21%(67.177)    | 2,3%(7.279)   | 320.456                    | 146               |

Πίνακας 3.1: Απεικόνιση αριθμού θετικών και αρνητικών συσχετίσεων ανά τύπο δείγματος.

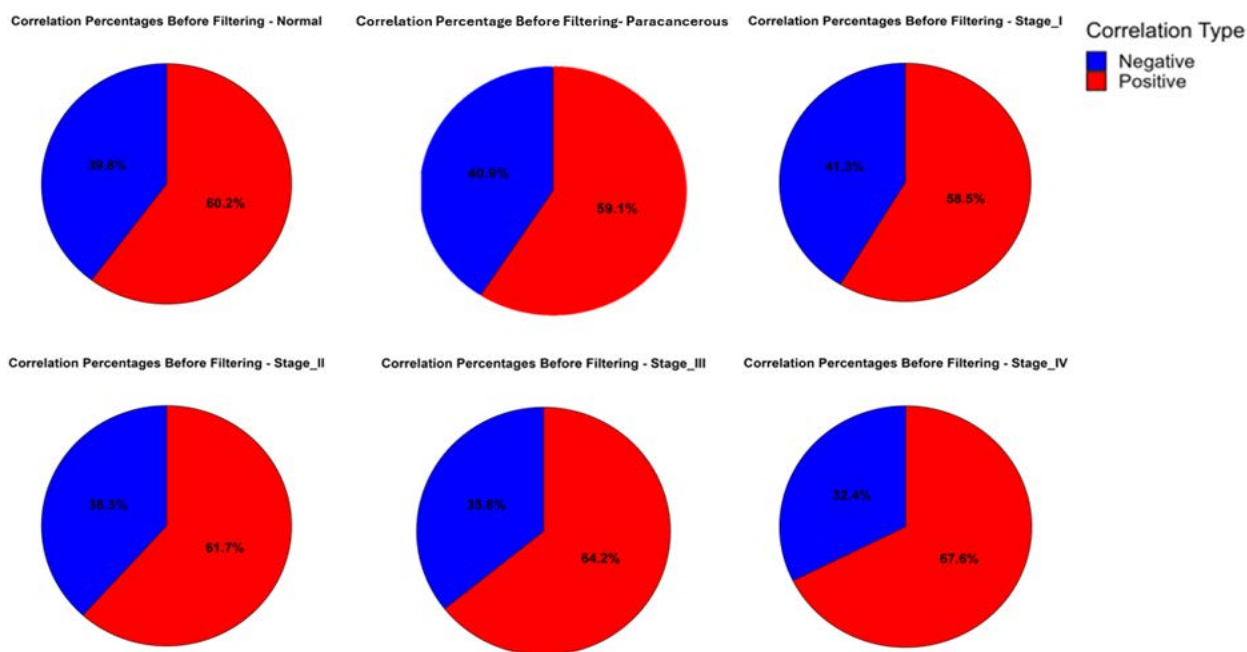
### 3.2.3 Ανάλυση συσχετίσεων πριν και μετά το φίλτρο στατιστικής σημαντικότητας

Η αναλογία θετικών και αρνητικών συσχετίσεων αναλύεται στην συνέχεια, συγκρίνοντας την ποσοστιαία, στο σύνολο των δεδομένων πριν και μετά την επιβολή ενός φίλτρου στατιστικής σημαντικότητας. Συγκεκριμένα, εφαρμόζεται ένα όριο για την τιμή  $p$ , ώστε να διατηρηθούν μόνο τα ζεύγη γονιδίων των οποίων οι συσχετίσεις έχουν  $p\text{-value} < 0.05$ . Τόσο για τις θετικές όσο και για τις αρνητικές συσχετίσεις, λαμβάνονται στοιχεία για τον συνολικό αριθμό των συσχετίσεων, ποιο κλάσμα αυτού αφορά θετικές και ποιο αρνητικές συσχετίσεις και γίνεται μετατροπή τους σε ποσοστά. Η απεικόνιση των αποτελεσμάτων γίνεται με την μορφή πίτας συχνοτήτων.

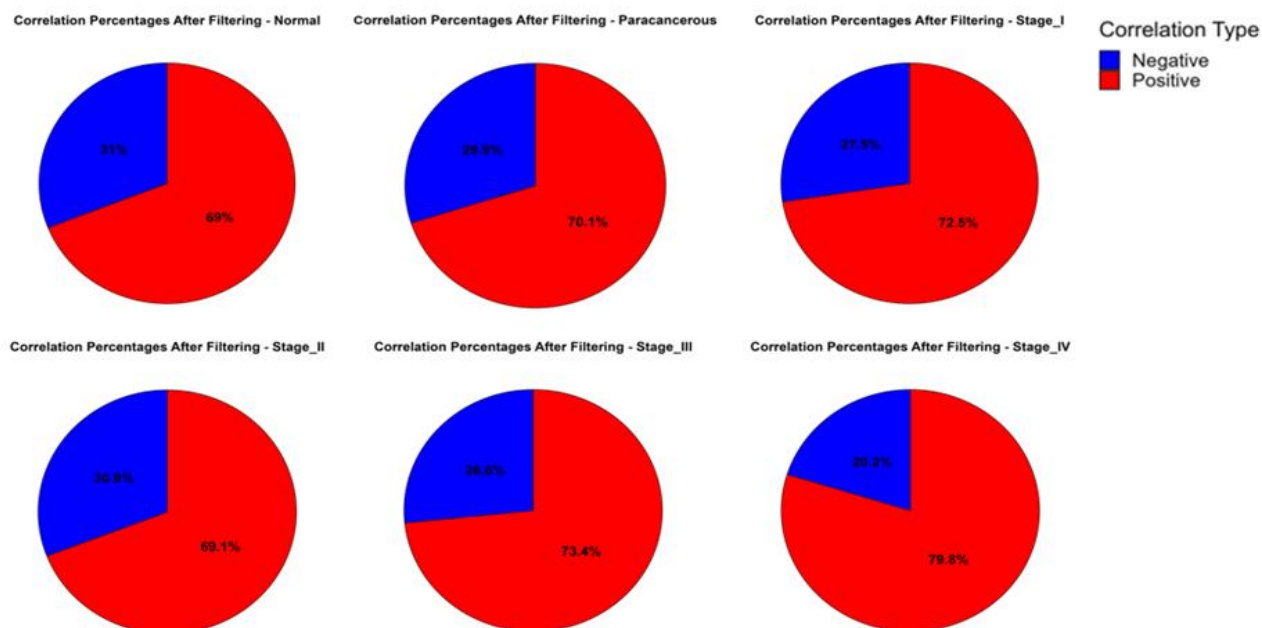
Όσον αφορά τα αποτελέσματα πριν την εφαρμογή του φίλτρου, ο συνολικός αριθμός των γονιδιακών συσχετίσεων είναι περίπου 320.000 σε κάθε τύπο δείγματος. Η αναλογία θετικών και αρνητικών συσχετίσεων παραμένει σταθερή στα φυσιολογικά και παρακαρκινικά δείγματα αλλά και στα δείγματα σταδίου 1 και 2. Ωστόσο, παρατηρείται μια σταδιακή αύξηση των θετικά σχετιζόμενων γονιδιακών ζευγών στα στάδια 3 και 4 (**Εικόνα 3.5**).

Μετά την εφαρμογή του φίλτρου, και την διατήρηση μόνο των στατιστικά σημαντικών αλληλεπιδράσεων, το προφίλ των αποτελεσμάτων αλλάζει. Αρχικά ο συνολικός αριθμός ζευγών, μειώνεται σημαντικά σε κάθε στάδιο και πλέον είναι και αρκετά διαφορετικός. Συγκεκριμένα, σε δείγματα με φυσιολογικό ιστό είναι 137.479 (42,8% των συνολικών συσχετίσεων), στα παρακαρκινικά 111.166 (35,1%), ενώ στα δείγματα από στάδιο 1 έως 4, είναι 71.4989 (22,7%), 171.814 (53,6%), 165.383 (51%) και 154.366 (48,2%) αντίστοιχα. Μεγάλη διαφορά, αποτελεί επίσης, η δραματική αύξηση της αναλογίας των θετικών συσχετίσεων προς τις αρνητικές. Στα περισσότερα

δείγματα παρατηρείται για μια αύξηση των θετικών συσχετίσεων που πλησιάζει το 10%. Στην περίπτωση των δειγμάτων του σταδίου 4, η αύξηση είναι ακόμα μεγαλύτερη, πλησιάζοντας το 12%. Μάλιστα, με εξαίρεση το στάδιο 2, η αύξηση αυτή είναι προοδευτική. Σε δείγματα φυσιολογικού ιστού, το ποσοστό ξεκινάει από 69% και φτάνει στο 80% στα δείγματα σταδίου 4 (**Εικόνα 3.6**).

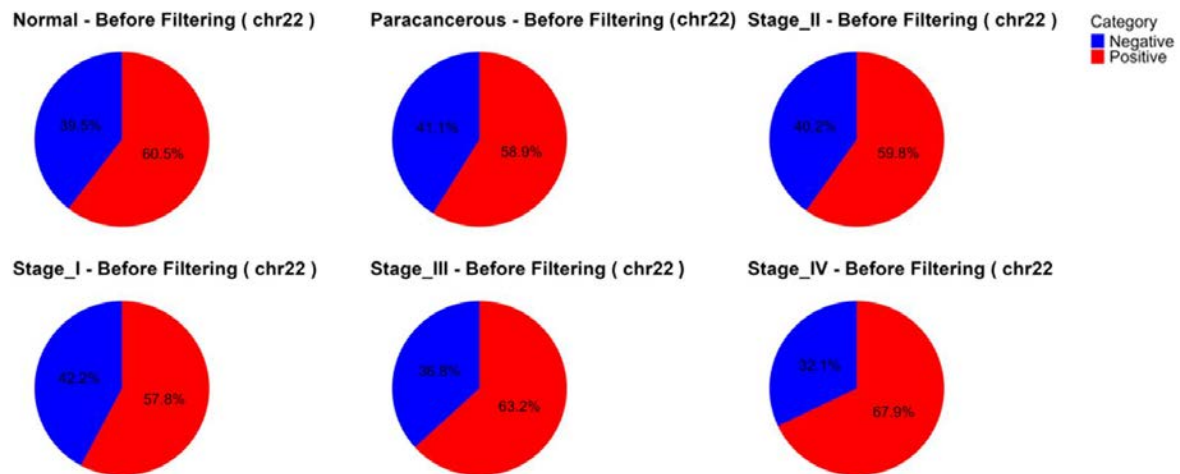


**Εικόνα 3.5:** Οπτικοποίηση ποσοστών θετικών και αρνητικών συσχετίσεων σε κάθε τύπο δείγματος, με την μορφή πίτων συχνοτήτων, πριν το φιλτράρισμα.

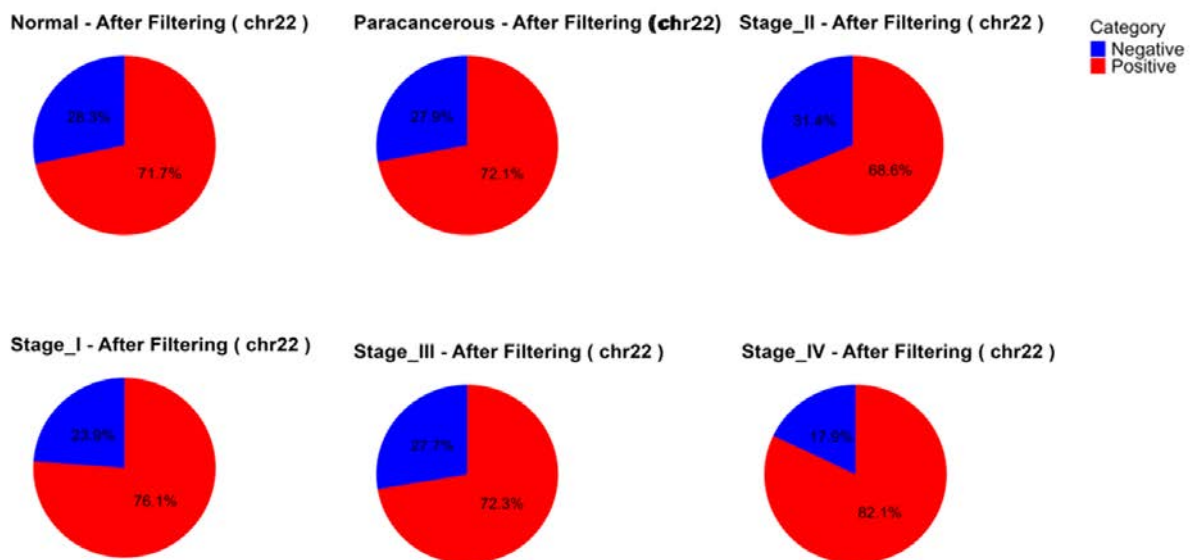


**Εικόνα 3.6:** Οπτικοποίηση ποσοστών θετικών και αρνητικών συσχετίσεων σε κάθε τύπο δείγματος, με την μορφή πίτων συχνοτήτων, μετά το φιλτράρισμα.

Για την επιβεβαίωση της στατιστικής σημαντικότητας και της μη τυχαιότητας των αποτελεσμάτων πραγματοποιείται bootstrap στο χρωμόσωμα 22. Οι μέσες τιμές των θετικών και αρνητικών συσχετίσεων από το σύνολο των τυχαίων δειγματοληψιών που γίνεται κατά το bootstrap, απεικονίζονται ξανά μέσω πιτών συχνότητας. Το μοτίβο της αύξησης της αναλογίας των θετικών και αρνητικών συσχετίσεων παραμένει το ίδιο και για τα bootstrap δείγματα, πριν και μετά το φιλτράρισμα. Μάλιστα, η αύξηση της αναλογίας μοιάζει αρκετά με τα πραγματικά δεδομένα, ακόμα και σε επίπεδο ποσοστών (**Εικόνες 3.7 και 3.8**).



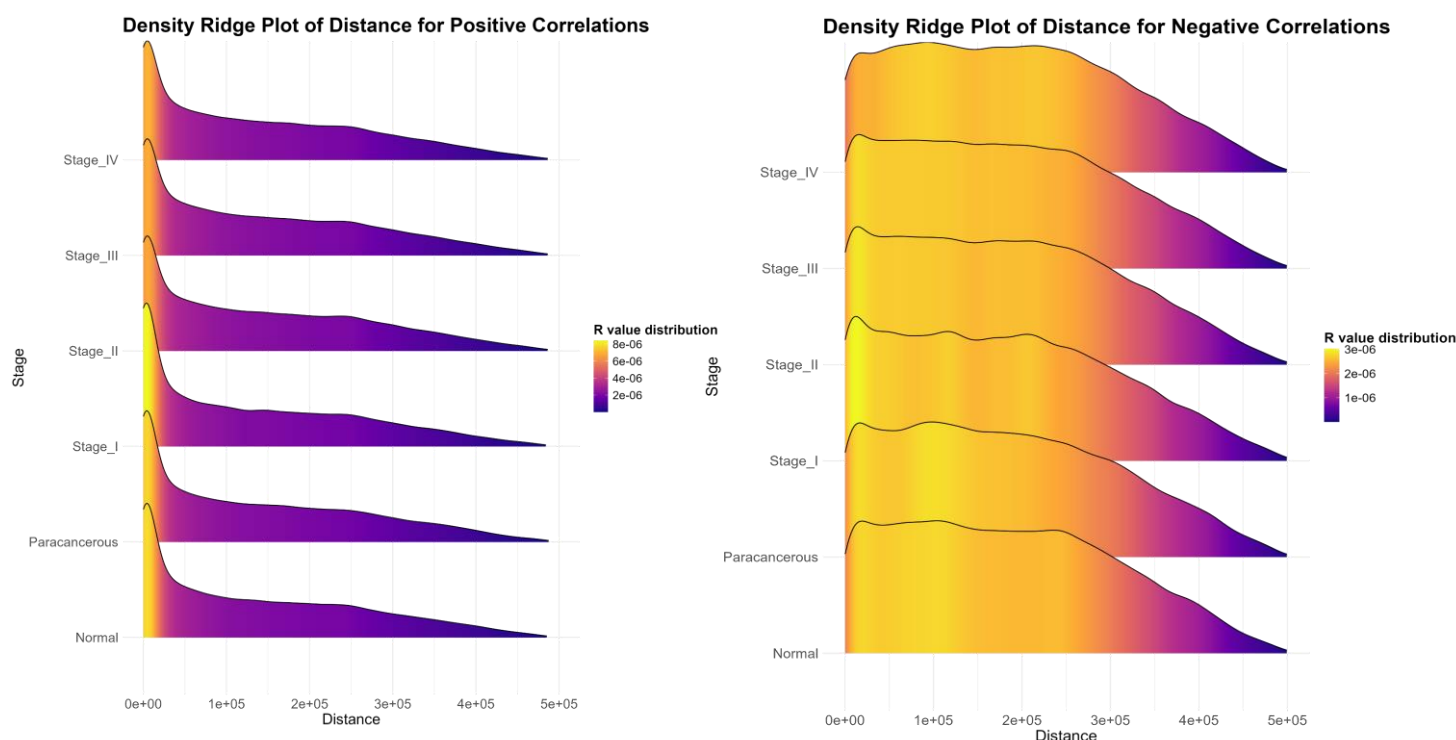
**Εικόνα 3.7:** Οπτικοποίηση των αποτελεσμάτων θετικών και αρνητικών συσχετίσεων, με τυχαία δειγματοληψία μέσω bootstrap στο χρωμόσωμα 22, πριν το φιλτράρισμα, με την μορφή πιτών συχνότητων.



**Εικόνα 3.8:** Οπτικοποίηση των αποτελεσμάτων θετικών και αρνητικών συσχετίσεων, με τυχαία δειγματοληψία μέσω bootstrap στο χρωμόσωμα 22, μετά το φιλτράρισμα, με την μορφή πιτών συχνότητων.

### 3.2.4: Απεικόνιση της σχέσης μεταξύ της απόστασης και των θετικών συσχετίσεων

Η τάση της αναλογίας των θετικών συσχετίσεων να αυξάνεται τόσο πολύ, γέννησε το ερώτημα από που μπορεί να πηγάζει αυτή η μεταβολή. Για τον λόγο αυτό, έγινε διερεύνηση της σχέσης που μπορεί να έχουν οι θετικές συσχετίσεις με την απόσταση μεταξύ των ζευγών, στα δεδομένα μετά το φιλτράρισμα. Η απεικόνιση της πιθανής αυτής σχέσης γίνεται με ridge plots (Εικόνα 3.9).



**Εικόνα 3.9:** Ridge plots που παρουσιάζουν τον αριθμό των θετικών συσχετίσεων (αριστερά) και των αρνητικών συσχετίσεων (δεξιά) ως συνάρτηση της απόστασης ανά στάδιο καρκίνου

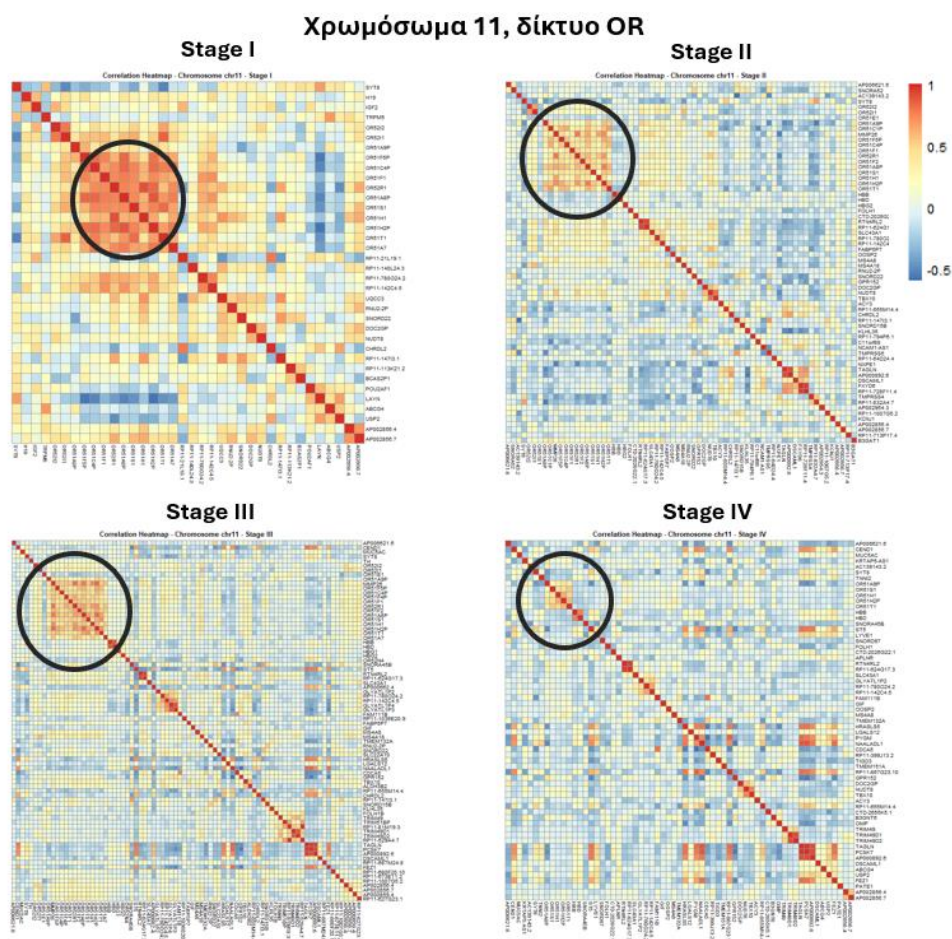
Η εικόνα της κατανομής στα διαγράμματα με τις θετικές και στα διαγράμματα με τις αρνητικές συσχετίσεις, διαφέρει αρκετά. Όσον αφορά, την πρώτη περίπτωση, η πλειοψηφία των τιμών φαίνεται να κατανέμεται στις μικρές αποστάσεις μεταξύ των γονιδίων. Αυτό ισχύει για κάθε στάδιο, και μάλιστα οι κατανομές των σταδίων μοιάζουν τόσο πολύ, που δεν φαίνεται να υπάρχει ουσιαστική διαφορά στην σχέση θετικών συσχετίσεων και απόστασης στα φυσιολογικά και καρκινικά δείγματα. Η εικόνα των κατανομών στα διαγράμματα των αρνητικών συσχετίσεων, είναι επίσης πολύ όμοια για τα όλα τα στάδια, αλλά είναι τελείως διαφορετική από αυτή των θετικών. Αρχικά, όπως ήταν και αναμενόμενο με βάση τα ποσοστά των πιτών, είναι πολύ λιγότερες σε αριθμό. Επιπλέον, η πλειοψηφία των τιμών κατανέμεται μεταξύ των αποστάσεων 0kbs έως 300 kbs. Φαίνεται λοιπόν, πως οι θετικές συσχετίσεις έχουν μια μεγαλύτερη εξάρτηση από την απόσταση των 2 γονιδίων, σε σχέση με τις αρνητικές.



### 3.2.6 Απεικόνιση δικτύων γονιδίων μέσω θερμικών χαρτών και του cytoscape

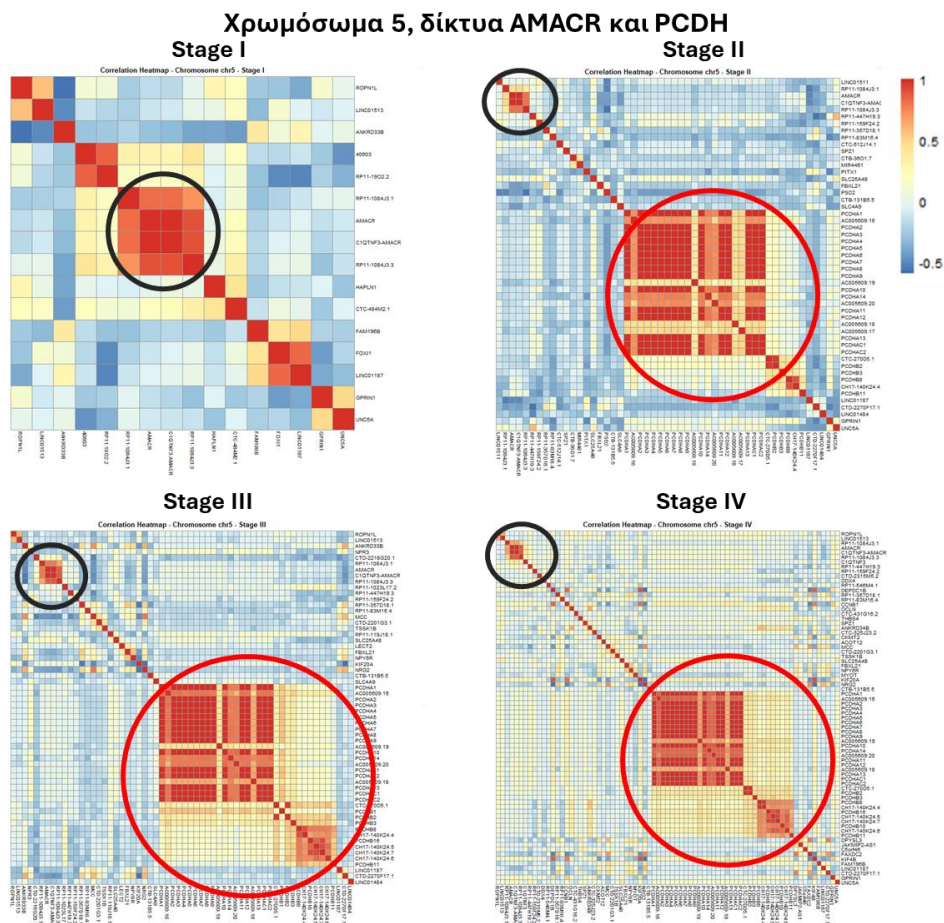
Τα δεδομένα που χρησιμοποιούνται σε αυτήν την φάση της ανάλυσης, αποτελούν ζεύγη γονιδίων, καθένα από τα οποία i) είναι διαφορετικά εκφρασμένο σε καρκινικές βιοψίες συγκριτικά με τον φυσιολογικό ιστό, και ii) οι συσχετίσεις στην έκφραση τους είναι στατιστικώς σημαντικές στο αντίστοιχο στάδιο καρκίνου που μελετάται αλλά δεν είναι στατιστικώς σημαντικές σε Normal. Δηλαδή, για κάθε στάδιο καρκίνου υπάρχει μια εξειδίκευση έτσι ώστε να μελετώνται μόνο τα γονίδια που έχουν τιμές στατιστικής σημαντικότητας ( $p$ ) για την συσχέτιση της έκφρασης χαμηλότερες του 0.05 στο στάδιο αυτό, και μεγαλύτερες από 0.05 στον φυσιολογικό ιστό.

Με βάση αυτό, σχηματίζονται θερμικοί χάρτες που απεικονίζουν τις συσχετίσεις των γονιδίων αυτών, για κάθε στάδιο, ανά χρωμόσωμα. Ο διαχωρισμός τους ανα χρωμόσωμα διευκολύνει την σύγκριση των αποτελεσμάτων μεταξύ κάθε καρκινικού σταδίου, συμπεριλαμβανομένων και των παρακαρκινικών δειγμάτων.



**Εικόνα 3.10:** Απεικόνιση μέσω θερμικών χαρτών του δικτύου OR ανά στάδιο καρκίνου. Το δίκτυο εξαλείφεται μετά το πέρας το πρώτου σταδίου.

Με την ανάγνωση των θερμικών χαρτών για το ίδιο χρωμόσωμα για κάθε στάδιο, είναι εύκολη η εύρεση πιθανών δικτύων γονιδίων. Μάλιστα μπορεί να παρατηρηθεί ακόμα και η εξέλιξη κάποιου δικτύου κατά την πρόοδο της νόσου. Για παράδειγμα, το πολυπληθές δίκτυο OR, εντοπίζεται με υψηλές τιμές R στο χρωμόσωμα 11 στο στάδιο 1. Ωστόσο, από το στάδιο 2 και μετά, οι τιμές R μειώνονται και το δίκτυο εξαφανίζεται (**Εικόνα 3.10**). Από την άλλη, μεγάλο ενδιαφέρον παρουσιάζει το δίκτυο AMACR στο χρωμόσωμα 5, το οποίο είναι παρόν σε κάθε στάδιο της νόσου (**Εικόνα 3.11**). Το γονίδιο AMACR, κωδικοποιεί μια ρακαμάση η οποία είναι απαραίτητη στην αποικοδόμηση ορισμένων ενώσεων στα μιτοχόνδρια και τα περοξεισωμάτια. Τα υπόλοιπα μέλη του δικτύου είναι το κωδικοποιητικό γονίδιο SLC45A2 (πρωτεΐνη μεταφορέας), τα lncRNA C1QTNF3-AMACR, RP11-1084J3.3, και RP11-1084J3.1. Στο ίδιο χρωμόσωμα, παρατηρείται και το μεγαλύτερο δίκτυο που προέκυψε από την ανάλυση, το δίκτυο PCDH, ένα πολύ μεγάλο cluster γονιδίων πρωτοκαδερίνης. Το συγκεκριμένο δίκτυο δεν εμφανίζεται καθόλου στο στάδιο 1 και τον φυσιολογικό ιστό αλλά μετά το στάδιο 2 η παρουσία του είναι μόνιμη (**Εικόνα 3.11**). Παράλληλα, σχηματίζονται πολλά ακόμα δίκτυα, με μικρότερο αριθμό γονιδίων, τα οποία είναι σταδιο-ειδικά, όπως το OR δίκτυο. Στο χρωμόσωμα 17 και συγκεκριμένα στο στάδιο 1, εντοπίζεται μια τριπλέτα HOX γονιδίων, τα HOXB6, HOXB2, HOXB8, και του lncRNA CTD-2377D24.6.

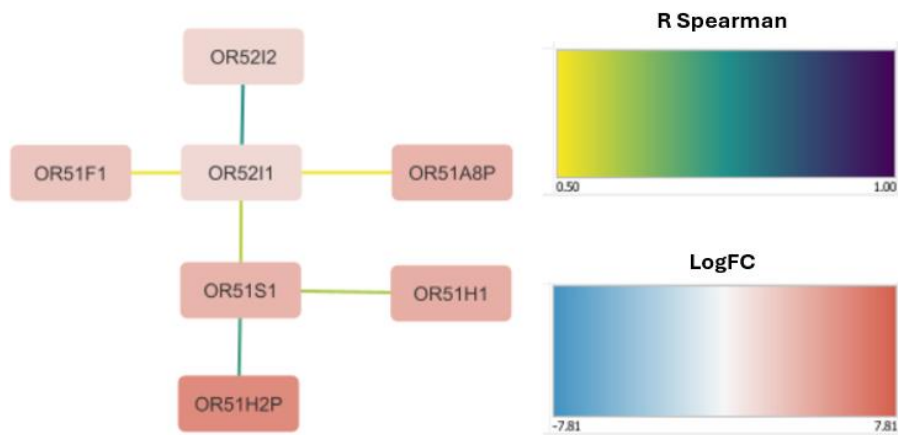


Τα δεδομένα με τις συσχετίσεις στην έκφραση των γονιδίων για κάθε στάδιο, φορτώθηκαν παράλληλα και στο Cytoscape, με σκοπό την εναλλακτική απεικόνιση των δικτύων. Μάλιστα σε αυτό το στάδιο εφαρμόζεται ένα επιπλέον φίλτρο, διατηρώντας μόνο τα ζεύγη γονιδίων των οποίων οι συσχετίσεις έχουν τιμές άνω του 0.5. Σχηματίζεται λοιπόν, ένας διαφορετικός αριθμός δικτύων, ο οποίος διαφέρει ως προς την σύσταση του ανάλογα με το στάδιο. Σε μεγάλο βαθμό τα δίκτυα που υπολογίζονται και σχηματίζονται στο Cytoscape, ταυτίζονται με αυτά που φαίνονται στους θερμικούς χάρτες. Είναι πολύ ενδιαφέρον επίσης, πως στις περισσότερες περιπτώσεις ακολουθείται και το ίδιο προφίλ εμφάνισης των δικτύων ανά στάδιο. Πιο συγκεκριμένα, το δίκτυο OR, εντοπίζεται και πάλι μόνο στο στάδιο 1 (**Εικόνα 3.11**), και το δίκτυο PCDH μόνο μετά το στάδιο 2 (**Εικόνα 3.13**).

Στο Cytoscape μάλιστα, γίνεται η ακριβής παρατήρηση των ήδη δημιουργούμενων δικτύων στα heatmaps. Ενώ στα heatmaps φαίνεται μια ευρεία περιοχή με αλληλεπιδράσεις πάνω στο χρωμόσωμα, μέσω του Cytoscape απεικονίζονται ακριβώς ποιά γονίδια από αυτές τις περιοχές συμμετέχουν για την δημιουργία των δικτύων. Το δίκτυο OR για παράδειγμα, φαίνεται να καλύπτει μια μεγάλη περιοχή στο χρωμόσωμα με πολλά γονίδια. Στην πραγματικότητα ωστόσο μόνο 7 από αυτά αλληλεπιδρούν μεταξύ τους (**Εικόνα 3.11**). Αντίστοιχα στην περίπτωση του δικτύου PCDH, στους θερμικούς χάρτες φαίνεται μια ακόμα πιο εκτεταμένη περιοχή από αυτήν του δικτύου OR. Οι πλειοψηφία των γονιδίων σε αυτήν την έκταση εμφανίζει υψηλούς συντελεστές συσχέτισης, που πολλές φορές μπορεί να ισοδυναμεί με 1. Αυτό, ενώ δίνει την εντύπωση πως στα στάδια 2 έως 4, υπάρχει αλληλεπίδραση μεταξύ του μεγαλύτερου μέρους των γονιδίων. Σύμφωνα με το Cytoscape ωστόσο, σε κάθε στάδιο τα περισσότερα γονίδια αλληλεπιδρούν μια σειρά από 4 lncRNA: AC005609.16, AC005609.17, AC005609.18, AC005609.19 και AC005609.20. Ανάλογα με το στάδιο, καθένα από αυτά τα lncRNA έχει πιο πρωταγωνιστικό ρόλο στην διαμόρφωση του δικτύου. Το πιο καθοριστικό φαίνεται να είναι το AC005609.18 που συμμετέχει σε όλα τα στάδια πέρα το στάδιο 1, και ακόμα και στα παρακαρκινικά (**Εικόνα 3.12**). Τέλος, ακόμα και σε δίκτυα που είναι λιγότερο πολυπληθή και είναι ξεκάθαρο ποιά είναι τα μέλη τους, το Cytoscape προσφέρει μια πολύ καλή εικόνα για ποιές ακριβώς είναι οι αλληλεπιδράσεις μεταξύ των γονιδίων. Το δίκτυο AMACR ενώ στα περισσότερα στάδια αποτελείται από 5 γονίδια (στα στάδια 2 και 4 το γονίδιο RP11-1084J3.3 δεν εμφανίζεται στο Cytoscape), πράγμα που είναι εμφανές και στους θερμικούς χάρτες. Ωστόσο, οι αλληλεπιδράσεις των γονιδίων συμβαίνουν μόνο με το γονίδιο RP11-1084J3.1, και όχι μεταξύ τους. Το ίδιο ισχύει και με το δίκτυο APOE, που αποτελείται από 4 γονίδια. Το γονίδιο APOE κωδικοποιεί την πρωτεΐνη E, ενώ τα υπόλοιπα μέλη είναι το APOC2 (απολιποπρωτεΐνη C2), το APOC1P1 (ψευδογονίδιο του γονιδίου APOC1) και το μη κωδικοποιητικό γονίδιο APOC4-APOC2. Το συγκεκριμένο δίκτυο εμφανίζεται στα στάδια 2 έως 4. Τα τρία πρώτα γονίδια αλληλεπιδρούν όλα μεταξύ τους, ενώ το APOC4-APOC2 μόνο με το APOC1P1.

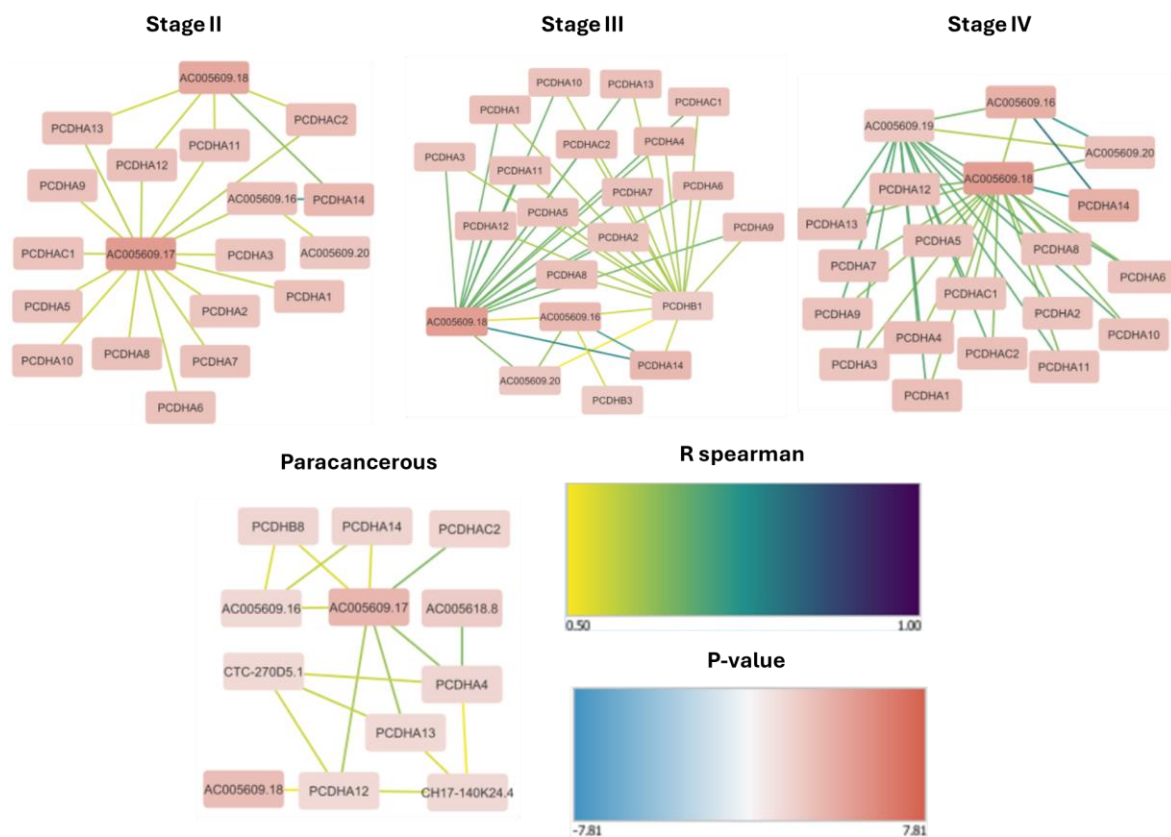


### Δίκτυο OR, Stage I



**Εικόνα 3.12:** Το δίκτυο OR, δημιουργημένο εντός του Cytoscape. Σχηματίζεται μόνο στο στάδιο 1. Για Spearman  $R > 0.5$ . Οι γραμμές δείχνουν τις αλληλεπιδράσεις μεταξύ των γονιδίων και την ένταση τους, ενώ το χρώμα των γονιδίων αν είναι αυξορρυθμιζόμενα ή μειορρυθμιζόμενα.

### PCDH network



**Εικόνα 3.13:** Το δίκτυο PCDH, δημιουργημένο εντός του Cytoscape. Σχηματίζεται μετά το στάδιο 2. Για Spearman  $R > 0.5$ . Οι γραμμές δείχνουν τις αλληλεπιδράσεις μεταξύ των γονιδίων και την ένταση τους, ενώ το χρώμα των γονιδίων αν είναι αυξορρυθμιζόμενα ή μειορρυθμιζόμενα.

Ωστόσο, όσον αφορά τους θερμικούς χάρτες, δεδομένου ότι σε κάθε στάδιο, το σύνολο των στατιστικά σημαντικών γονιδίων αλλάζει, οι συγκρίσεις γίνονται μεταξύ ανόμοιων αντικειμένων. Όπως φαίνεται για παράδειγμα στην περίπτωση του χρωμοσώματος 5, στο στάδιο 1 φαίνεται μόνο το δίκτυο AMACR και ο αριθμός των γονιδίων είναι πολύ μικρότερος από τα άλλα στάδια. Μάλιστα, διαφορές ως προς το πλήθος και την ταυτότητα των γονιδίων, υπάρχουν ακόμα και στα στάδια που το δίκτυο σχηματίζεται (**Εικόνα 3.11**).

Για τον λόγο αυτό, εφαρμόζεται ένα δεύτερο φίλτρο, για να παραχθούν θερμικοί χάρτες που αντλούν τις πληροφορίες τους από τους πρώτους. Αναλυτικότερα, αυτή την φορά διατηρούνται γονίδια που είναι στατιστικά σημαντικά σε έστω ένα στάδιο καρκίνου, και όχι σε φυσιολογικό ιστό. Με τον τρόπο αυτό, για την δημιουργία κάθε θερμικού χάρτη που απεικονίζει συσχετίσεις, θα χρησιμοποιείται ένα κοινό σύνολο γονιδίων, πράγμα που ευνοεί τις συγκρίσεις. Επιπλέον, παράγονται και θερμικοί χάρτες που δείχνουν το p-value για κάθε αλληλεπίδραση για κάθε στάδιο, αλλά και θερμικοί χάρτες που δείχνουν την μεταβολή της έκφρασης αυτών των γονιδίων όσο προοδεύει η νόσος, μέσω του Z-score.

Δημιουργούνται ξανά λοιπόν, θερμικοί χάρτες για τα δίκτυα που αναφέρθηκαν και πριν. Όπως φαίνεται, τα δίκτυα που απεικονίζονται είναι αντίστοιχα με τους πρώτους θερμικούς χάρτες, αλλά το μεγαλύτερο πλήθος γονιδίων επιτρέπει την τοποθέτηση τους με μεγαλύτερη ακρίβεια πάνω στο γονιδίωμα καθώς και την σύγκριση τους μεταξύ των σταδίων της νόσου. Ενδεικτικά απεικονίζονται τα δίκτυα OR και PCDH, ώστε να αναδειχθεί με μεγαλύτερη λεπτομέρεια η εξέλιξη τους, σε φυσιολογικά, παρακαρκινικά και καρκινικά δείγματα.

Στην περίπτωση του δικτύου PCDH, οι θερμικοί χάρτες στατιστικής σημαντικότητας, συνδυαστικά με τους θερμικούς χάρτες συσχέτισης, υπογραμμίζουν ακόμα περισσότερα την τεράστια σημασία των μη κωδικών γονιδίων AC005609. Στους φυσιολογικούς ιστούς και σε αυτούς στο στάδιο 1, κανένα από τα γονίδια AC005609.17 και AC005609.18 δεν εμφανίζουν στατιστικά σημαντικές αλληλεπιδράσεις με τα γονίδια που αναγράφονται στο Cytoscape. Το ίδιο ισχύει και για τα υπόλοιπα μη κωδικά γονίδια, ωστόσο αυτά φαίνεται να αποκτούν πιο σημαντικό ρόλο σε επόμενα στάδια της νόσου (**Εικόνα 3.15**). Όσον αφορά τα επόμενα στάδια, ο λόγος της εναλλαγής του ρυθμιστικού ρόλου μεταξύ των μη κωδικών γονιδίων γίνεται ακόμα πιο ξεκάθαρος. Στο στάδιο 2, τα περισσότερα γονίδια, αλληλεπιδρά είτε με το AC005609.17 είτε με το AC005609.18, ενώ το AC005609.16 αλληλεπιδρά με τα AC005609.17 και AC005609.20. Από τον p-value χάρτη, είναι εμφανές ότι τα γονίδια αυτά αναπτύσσουν στατιστικώς σημαντικά ζεύγη με τα υπόλοιπα του δικτύου. Παράλληλα, έχουν υψηλό συντελεστή συσχέτισης και στους χάρτες συσχέτισης με τα γονίδια PCDH, ωστόσο τα AC005609.20 και AC005609.16 φαίνεται να έχουν υψηλότερο, πράγμα που αποτελεί μια διαφοροποίηση συγκριτικά με το Cytoscape. Αντίστοιχα, το ίδιο συμβαίνει και με το στάδιο 3, που ενώ το AC005609.18 όντως φαίνεται να έχει το ίδιο σημαντικό ρόλο και στα 2 μέσα, το AC005609.16 έχει λιγότερες αλληλεπιδράσεις στο Cytoscape παρά τις πολλές και υψηλές συσχετίσεις στο αντίστοιχο heatmap. Στο στάδιο 4 ωστόσο, τα AC005609.18 και AC005609.19 είναι αυτά με τις περισσότερες συσχετίσεις και στα 2 μέσα, και το AC005609.16 αποκτά αλληλεπιδράσεις που ταυτίζονται. Το AC005609.19 εμφανίζεται πρώτη φορά σαν

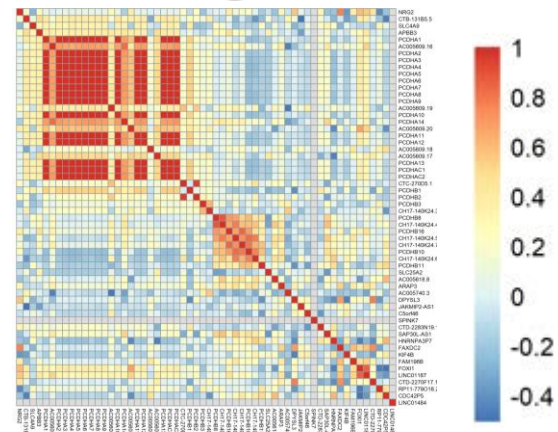
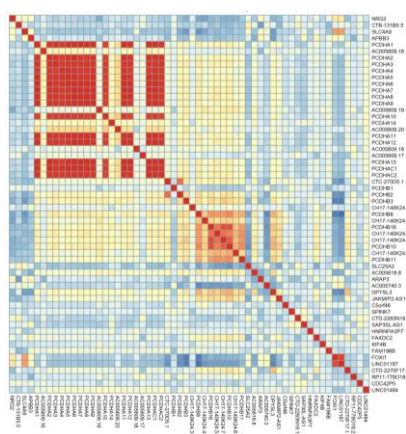
ρυθμιστής στο στάδιο 4, τόσο στο Cytoscape όσο και στους θερμικούς χάρτες συσχέτισης. Τέλος, στο παρακαρκινικό ιστό τα AC005609 που δεν συμμετέχουν έχουν χαμηλή στατιστική σημαντικότητα, και όσα εμφανίζονται στο δίκτυο έχουν πιο υψηλές συσχετίσεις σε σχέση με τα άλλα.

## Θερμικοί χάρτες για δίκτυο PCDH

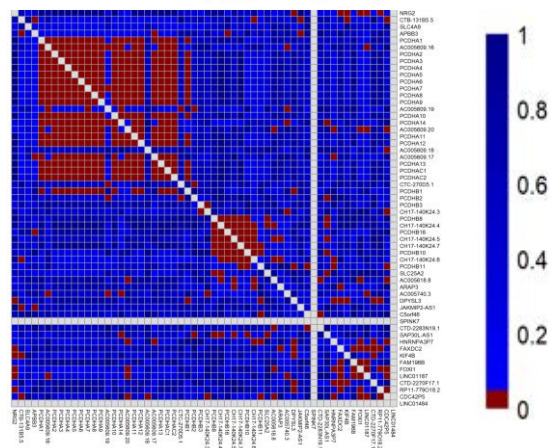
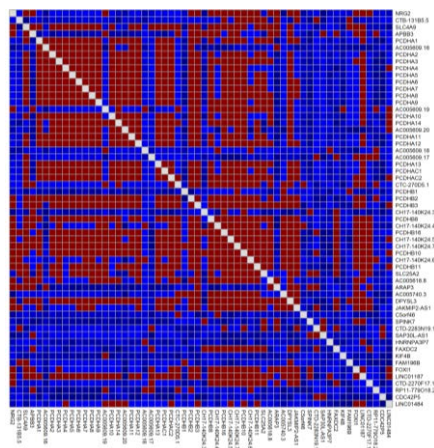
Normal

Stage I

Πίνακες  
συσχετίσεων,  
Spearman R



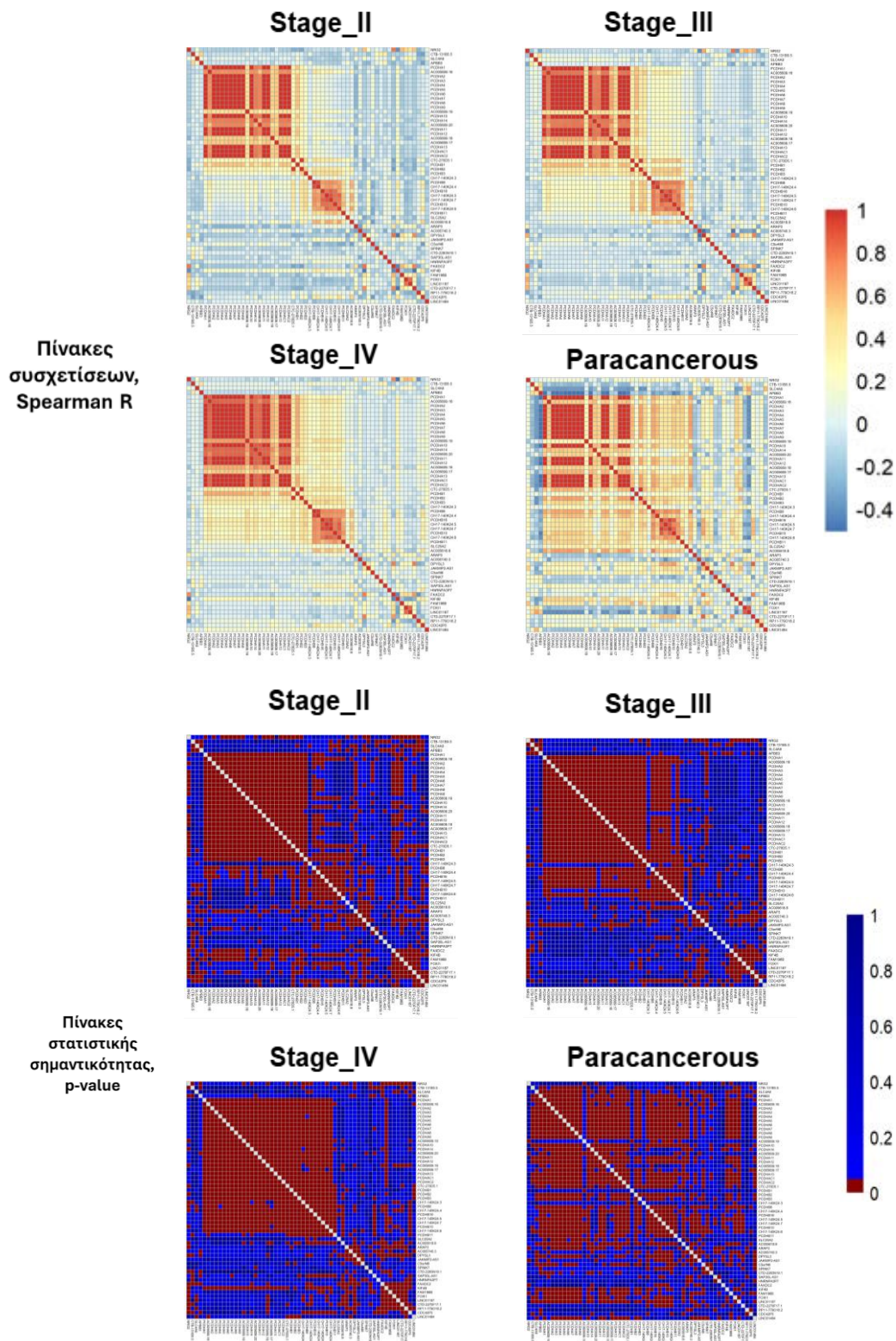
Πίνακες  
στατιστικής  
σημαντικότητας,  
p-value



**Εικόνα 3.15:** Θερμικοί χάρτες συσχέτισης και στατιστικής σημαντικότητας, για το δίκτυο PCDH, σε φυσιολογικό ιστό και σε ιστό σε στάδιο 1. Η έλλειψη στατιστικής σημαντικότητας μεταξύ των αλληλεπιδράσεων των AC005609 με τα υπόλοιπα γονίδια του cluster δικαιολογεί τα αποτελέσματα του Cytoscap



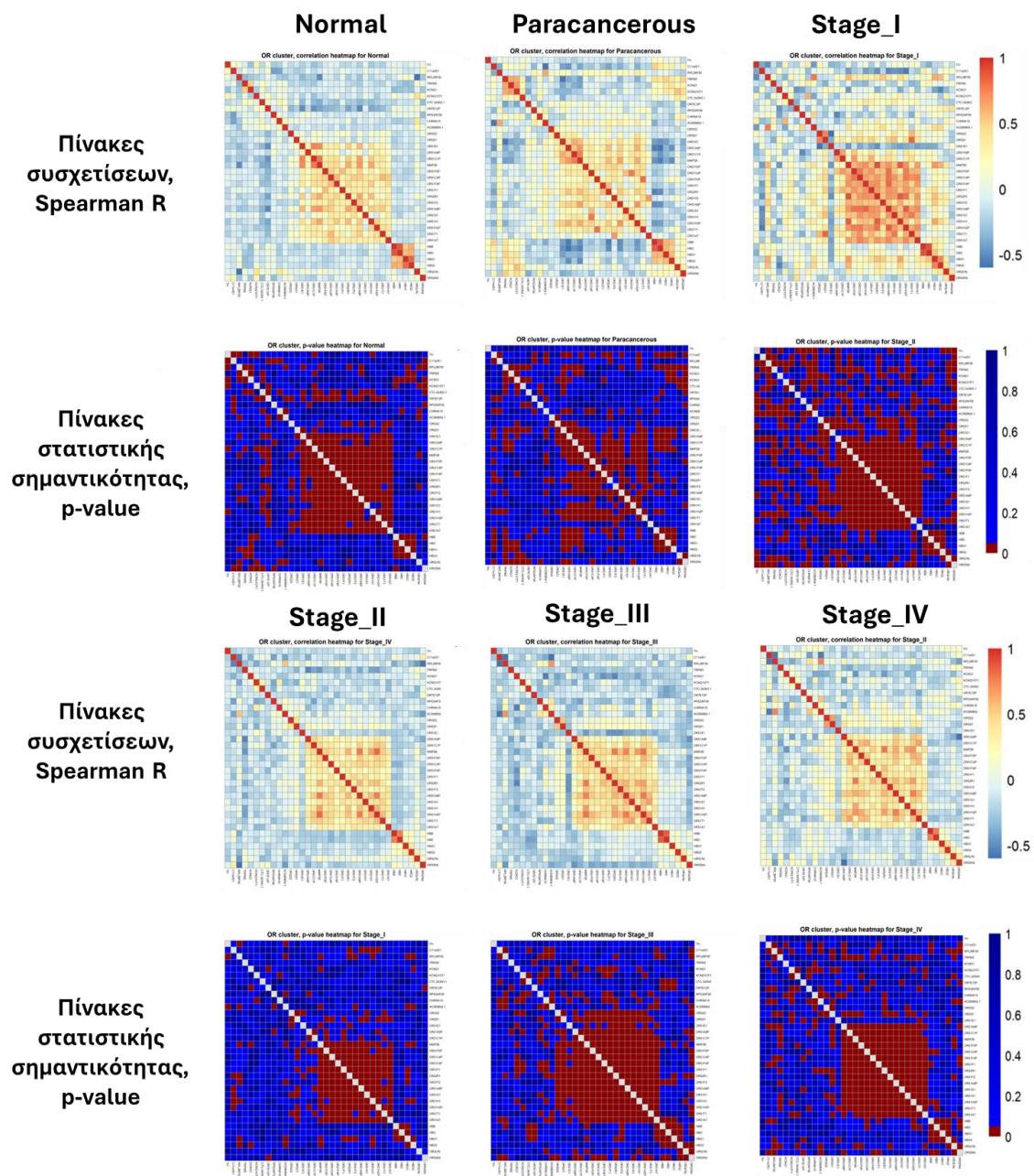
## Θερμικοί χάρτες για δίκτυο PCDH



**Εικόνα 3.16:** Θερμικοί χάρτες συσχέτισης και στατιστικής σημαντικότητας, για το δίκτυο PCDH, σε παρακαρκινικό ιστό και σε ιστούς σταδίου 2 έως 4. Οι διαφορές στην στατιστική σημαντικότητα της αλληλεπίδρασης και στην ένταση της συσχέτισης, καθορίζει ποιά AC005609 έχει πρωταγωνιστικό ρόλο ανά στάδιο.

Όσον αφορά το δίκτυο OR, οι κύριες αλληλεπιδράσεις που αναπαρίστανται στο Cytoscape δεν είναι στατιστικά σημαντικές στο θερμικούς χάρτες p-value για τα φυσιολογικά δείγματα, και για αυτόν μόνο τον λόγο δεν σχηματίζεται. Το ίδιο ακριβώς ισχύει και για τα περισσότερα από τα γονίδια στα παρακαρκινικά δείγματα. Αντίθετα οι σημαντικές αλληλεπιδράσεις είναι στατιστικά σημαντικές στο στάδιο 1 και εμφανίζουν συντελεστή συσχέτισης μεγαλύτερο 0.5. Στα υπόλοιπα στάδια όμως, ενώ οι αλληλεπιδράσεις κρίνονται ως στατιστικά σημαντικές στο θερμικούς χάρτες p-value, στους θερμικούς χάρτες συσχέτισης φαίνεται πως έχουν μικρό συντελεστή συσχέτισης ( $>0.5$ ). Αυτό συμβαδίζει και με τον μη σχηματισμό των δικτύων στο Cytoscape (**Εικόνα 3.14**).

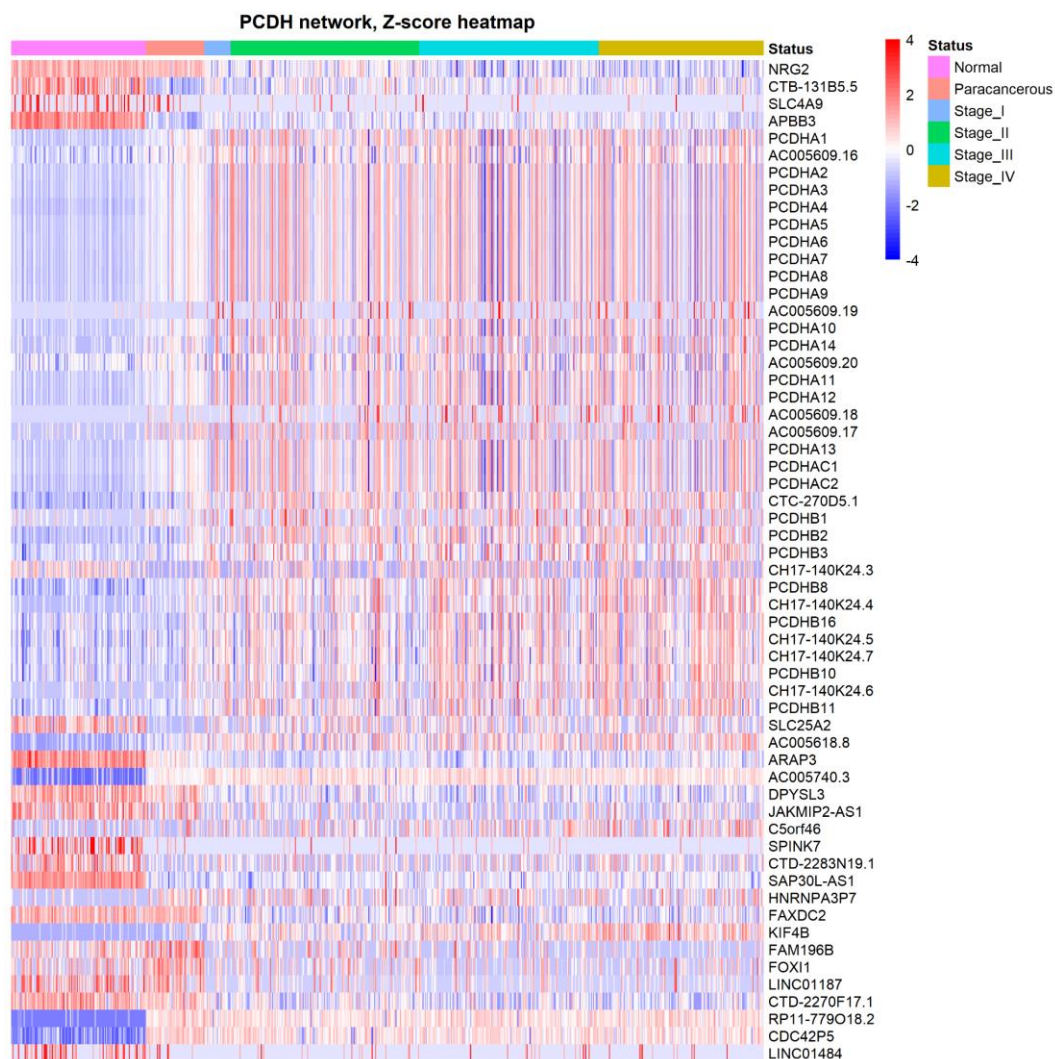
### Θερμικοί χάρτες για δίκτυο OR



**Εικόνα 3.14:** Αντιστοίχιση θερμικών χαρτών συσχετίσεων και στατιστικής σημαντικότητας, μετά το νέο φίλτρο, για όλους τους τύπους δειγμάτων. Οι χάρτες απεικονίζουν την μεταβολή στο δίκτυο OR

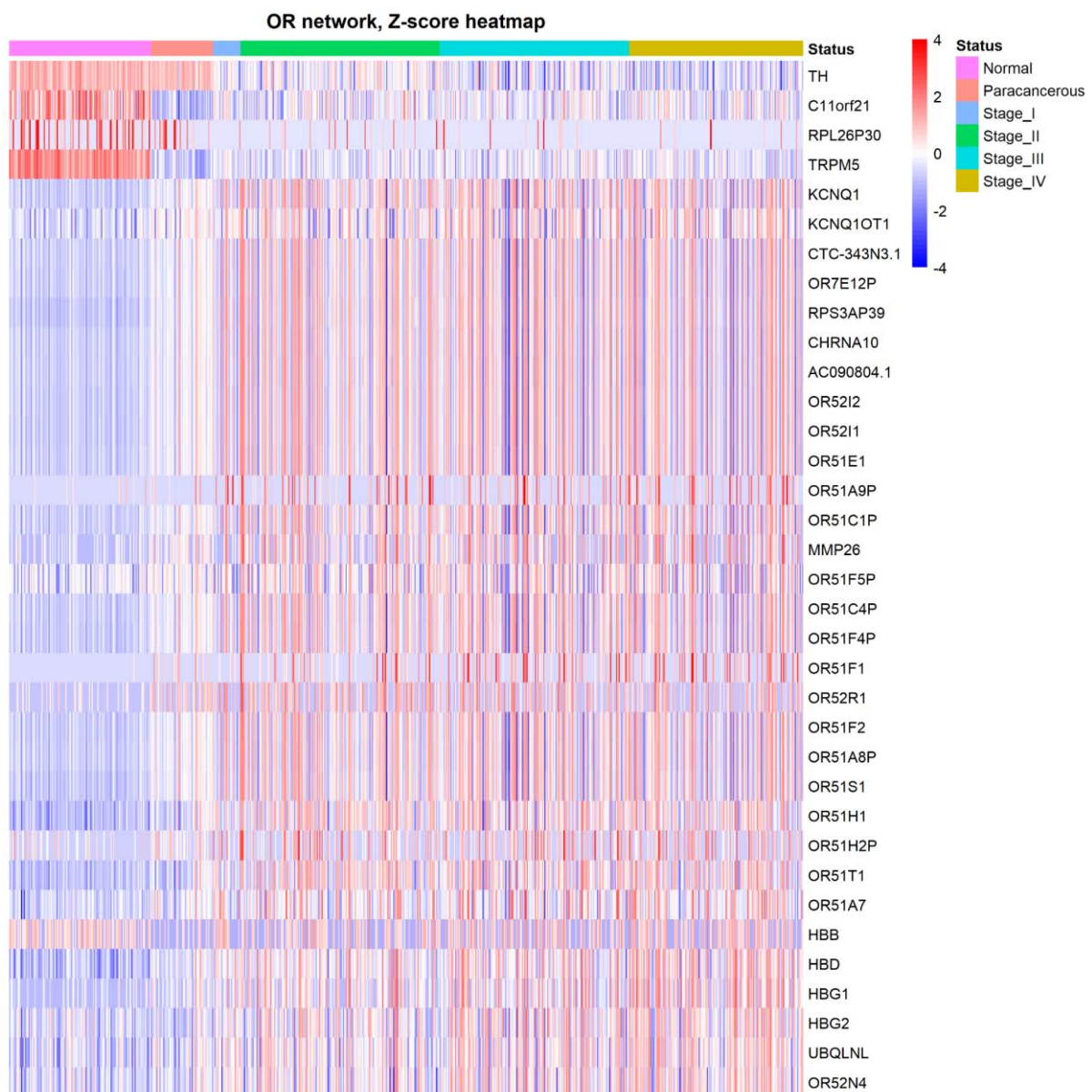


Ταυτόχρονα, μέσω των θερμικών χαρτών για τα z-score φαίνεται η μεταβολή της έκφρασης των γονιδίων για το δίκτυο PCDH. Όπως φαίνεται, η πλειοψηφία των γονιδίων ενδιαφέροντος (PCDHA και ορισμένα PCDHB), είναι αυξορρευσιζόμενα στον καρκίνο. Γενικότερα φαίνεται η τάση τα γονίδια που κωδικοποιούν πρωτεΐνες, να έχουν πιο μεγάλη διαφορά ως προς την έκφραση τους, σε σύγκριση με τον φυσιολογικό ιστό. Δεν ισχύει το ίδιο για τα μη κωδικά γονίδια AC005609, που είναι πιο δύσκολο να γίνει σύγκριση. Αρχικά, το AC005609.17, φαίνεται να έχει πιο αυξημένη έκφραση στα παρακαρκινικά δείγματα, και σε αυτά του σταδίου 2 και 3. Το συγκεκριμένο γονίδιο εμφανίζεται μόνο στα δίκτυα των παρακαρκινικών δείγματα και του σταδίου 2. Επιπλέον, τα AC005609.16, AC005609.18 και AC005609.20, εμφανίζουν αυξημένη έκφραση σε όλα τα στάδια πέρα από το στάδιο 1 και τα παρακαρκινικά δείγματα, πράγμα που επαληθεύει σε ένα βαθμό την παρουσία τους σε όλα τα δίκτυα (τα AC005609.20 και AC005609.18 δεν συμμετέχουν στο δίκτυο των παρακαρκινικών ιστών). Μάλιστα το AC005609.18 εμφανίζει πολύ πιο αυξημένη έκφραση σε κάποια δείγματα, σε σχέση με τα άλλα γονίδια. Τέλος, το AC005609.19 επίσης εμφανίζει πολύ αυξημένη έκφραση σε ορισμένα δείγματα στα στάδια 2 έως 4, ωστόσο τα δείγματα αυτά φαίνεται να είναι περισσότερα στο στάδιο 4. Αυτό ταυτίζεται με την παρουσία μόνο στο δίκτυο του σταδίου αυτού (**Εικόνα 3.15**).



**Εικόνα 3.15:** Θερμικός χάρτης του Z-score για τα γονίδια του δικτύου PCDH

Από την άλλη, λόγω του μικρότερου πλήθους των γονιδίων που απαρτίζουν το OR δίκτυο, η μελέτη της έκφρασης τους είναι πολύ πιο εύκολη. Το σύνολο τους φαίνεται να είναι αυξορρυθμιζόμενο στο σύνολο των σταδίων. Ενδιαφέρον παρουσιάζει ότι αυξημένη έκφραση δεν έχει κανένα από αυτά τα γονίδια ούτε στα παρακαρκινικά δείγματα. Το δίκτυο αυτό, εμφανίζεται μόνο στο στάδιο 1, άρα πιθανότατα τα επίπεδα έκφρασης δεν αποτελούν το αίτιο αυτής της εξειδίκευσης. Παρόλα αυτά, ενώ είναι γεγονός ότι ο αριθμός των δειγμάτων του σταδίου 1 είναι μικρός, η έκφραση των ORs του δικτύου είναι πιο ομοιογενής σε σχέση με τα άλλα στάδια. Με εξαίρεση τα OR1F1 και OR1H1, όλα τα υπόλοιπα γονίδια έχουν υψηλές τιμές z-score σε σχεδόν όλα τα δείγματα (**Εικόνα 3.16**).



**Εικόνα 3.16:** Θερμικός χάρτης του Z-score για τα γονίδια του δικτύου OR



### 3.2.5 Ανάλυση οντολογίας στα δεδομένα διαφορικής έκφρασης και γονιδιακών συσχετίσεων

Από τα δεδομένα διαφορικής έκφρασης, απομονώνονται οι λίστες διαφορικά εκφρασμένων γονιδίων, για κάθε στάδιο. Η κάθε λίστα ενσωματώνεται στο EnrichR, ώστε να πραγματοποιηθεί ανάλυση οντολογίας. Από τα αποτελέσματα, τα ειδικά για κάθε στάδιο γονίδια εμφανίζουν διαφορετική συσχέτιση με βιολογικά μονοπάτια. Στα στάδια 1 και 2, παρατηρείται συμμετοχή των γονιδίων σε μονοπάτια που σχετίζονται με τη σύσπαση των μυών, την ολίσθηση των νηματίων ακτίνης-μυοσίνης και τη συναρμολόγηση μυοϊνιδίων. Στα στάδια 2 και 3 παρατηρούμε ότι μεγάλη σημασία έχουν μονοπάτια εμβρυϊκής μορφογένεσης, κάτι που είναι αναμενόμενο, λόγω των μηχανισμών που αξιοποιεί ο καρκίνος. Επιπλέον πολλά γονίδια αφορούν την ρύθμιση πρωτεασών σερίνης και την ανάπτυξη των κερατινοκεράτων στο στάδιο 2, ενώ τα γονίδια του σταδίου 3 συμμετέχουν σε μονοπάτια κυτταρικής ομοιόστασης και ρύθμισης του μετασυναπτικού δυναμικού της μεμβράνης (**Εικόνα 3.17**).

#### Enrichment Analysis for DE genes (Normal vs Tumor)



**Εικόνα 3.17:** Διαγράμματα των πιο σημαντικών βιολογικών δραστηριοτήτων, των διαφορικά εκφρασμένων γονιδίων, ανά στάδιο καρκίνου.

Ταυτόχρονα, από τους πίνακες δεδομένων του Cytoscape, απομονώνονται τα γονίδια που συμμετέχουν στο σχηματισμό δικτύων. Τα γονίδια αυτά είναι φιλτραρισμένα, διατηρώντας μόνο αυτά που εμφανίζουν συντελεστή συσχέτισης μεγαλύτερο από 0.5. Ενώ και αυτά είναι διαφορικά εκφρασμένα, εμπλέκονται σε τελείως διαφορετικές βιολογικές λειτουργίες. Στο στάδιο 1, εμφανίζονται πολλές ορμονικές διεργασίες, όπως η βιοσύνθεση της θυροξίνης. Επιπλέον παρατηρείται ότι πολλά από αυτά τα γονίδια σχετίζονται με την οξείδωση των λιπιδίων και την βιοσύνθεση υπεροξειδίου του υδρογόνου. Στα στάδια 2 και 3, μεγάλη σημαντικότητα παρουσιάζει η συμμετοχή γονιδίων που σχετίζονται με την ανάπτυξη του νευρικού.

Επιπλέον, πολλά γονίδια συμμετέχουν στο μεταβολισμό των λιπιδίων, και πιο συγκεκριμένα την οργάνωση και τροποποίηση των χυλομικρών. Στο στάδιο 4 παρατηρείται πιο προχωρημένη αναδιαμόρφωση των λιπιδίων, με συμμετοχή της HDL και των πλασματικών λιποπρωτεϊνών. Η νευρική ανάπτυξη, αποτελεί επίσης σημαντική λειτουργία αλλά και η διαμεμβρανική μεταφορά της γλυκόζης (**Εικόνα 3.18**).

### Enrichment Analysis for all network genes after filtering for R>0.5



**Εικόνα 3.18:** Διαγράμματα των πιο σημαντικών βιολογικών δραστηριοτήτων, των υψηλά συσχετιζόμενων γονιδίων για τον σχηματισμό δικτύων, ανά στάδιο καρκίνου

Πολλά από τα δίκτυα που ήδη αναφέρθηκαν, σχετίζονται σε μεγάλο βαθμό με αυτές τις λειτουργίες. Για παράδειγμα, η μεταφορά της λίστας των γονιδίων που απαρτίζουν το PCD δίκτυο στο EnrichR, δίνει ως αποτέλεσμα την συμμετοχή του στην ανάπτυξη του νευρικού συστήματος. Ιδιαίτερα στα στάδια 3 και 4, η συγκεκριμένη βιολογική δραστηριότητα, είναι η πιο σημαντική για τα γονίδια της συγκεκριμένης γονιδιακής συστοιχίας. Ακολουθούν επίσης λειτουργίες όπως η μεταγωγή σήματος και η διακυτταρική σύνδεση (**Εικόνα 3.19**).

Αντίστοιχα το δίκτυο AMACR, συνδέεται έμμεσα με τον μεταβολισμό των λιπιδίων. Η β-οξειδωση των λιπαρών οξέων που συμμετέχει, σχετίζεται με την οξειδωση των λιπιδίων στο στάδιο 1, ενώ η βιοσύνθεση και ο μεταβολισμός των χολικών οξέων συνδέονται με την αναδιαμόρφωση των λιποπρωτεϊνών στάδια 2 με 4 (**Εικόνα 3.19**). Το ίδιο ισχύει και για το δίκτυο APOE, και μάλιστα με ακόμα πιο άμεσο τρόπο. Όλες οι βασικές λειτουργίες στις οποίες εμπλέκονται τα γονίδια του, αφορούν άμεσα τον μεταβολισμό των λιπιδίων, καθώς σχετίζονται με την οργάνωση και τροποποίηση των χυλομικρών και της VLDL (**Εικόνα 3.20**).

## Enrichment Analysis for PCD network for Stage III and Stage IV



**Εικόνα 3.19:** Διαγράμματα των πιο σημαντικών βιολογικών διεργασιών για το δίκτυο PCDH, μεταξύ των σταδίων 3 και 4

## Enrichment Analysis for important clusters



**Εικόνα 3.20:** Διαγράμματα των πιο σημαντικών βιολογικών διεργασιών για τα δίκτυα AMACR(πάνω) και APOE(κάτω)

### 3.2.6 Κατηγοριοποίηση των βιότυπων των γονιδίων

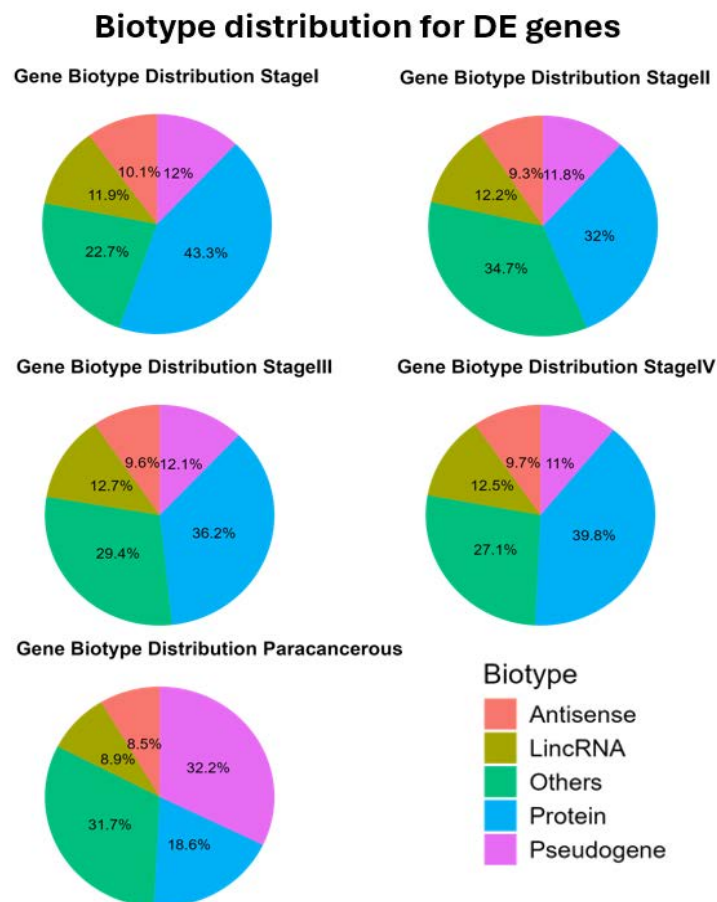
Για την καλύτερη σύγκριση των γονιδίων που συμμετέχουν στο σχηματισμό δικτύων και των διαφορεικά εκφρασμένων, πραγματοποιείται κατηγοριοποίηση των βιότυπων τους. Οι βιότυποι ομαδοποιούνται σε 5 μεγάλες κατηγορίες: protein coding, lincRNAs, antisense, pseudogenes και όλα τα υπόλοιπα. Για την καλύτερη απεικόνιση των αποτελεσμάτων, τα ποσοστά που καταλαμβάνει κάθε κατηγορία παρουσιάζονται σε πίτες συχνότητας ανά στάδιο, και για τις 2 κατηγορίες δεδομένων.

Όσον αφορά τα διαφορεικά εκφρασμένα γονίδια, αν και τα ποσοστά διαφέρουν ανά στάδιο της νόσου, οι κατανομές των βιότυπων ακολουθούν το ίδιο μοτίβο στα στάδια 1 έως 4. Τα γονίδια που κωδικοποιούν πρωτεΐνες είναι η πολυπληθέστερη κατηγορία (32%-43.3%), και ακολουθεί η κατηγορία με γονίδια που ομαδοποιήθηκαν στην κατηγορία

Others (22.7%-34.7%). Τα αντινοσηματικά, τα lincRNA και τα ψευδογονίδια έχουν σχεδόν ίδιες κατανομές (~10%-12%) και καταλαμβάνουν τις 3 τελευταίες θέσεις. Πλήρως αντίθετο σε αυτό το μοτίβο εικόνα, παρατηρείται στα παρακαρκινικά δείγματα. Ενώ πάλι το πλήθος των γονιδίων που δεν ομαδοποιήθηκαν σε μια συγκεκριμένη κατηγορία βρίσκεται στην δεύτερη θέση (31.7%), τα ψευδογονίδια καταλαμβάνουν πολύ μεγαλύτερο μέρος του ποσοστού σε σχέση με υπόλοιπα στάδια και ανέρχονται στην πρώτη θέση (32.2%). Τα κωδικοποιητικά γονίδια είναι λίγο λιγότερα από το ένα πέμπτο του συνόλου (18.6%), ενώ τα lincRNAs και τα αντινοσηματικά γονίδια έχουν αντίστοιχους πληθυσμούς με τα προηγούμενα στάδια, με 8.9% και 8.5% αντίστοιχα (**Εικόνα 3.21**).

**Εικόνα 3.21:** Κατανομή των βιότυπων των διαφορεικά εκφρασμένων γονιδίων σε antisense, LincRNAs, Proteins, Pseudogenes και Others.

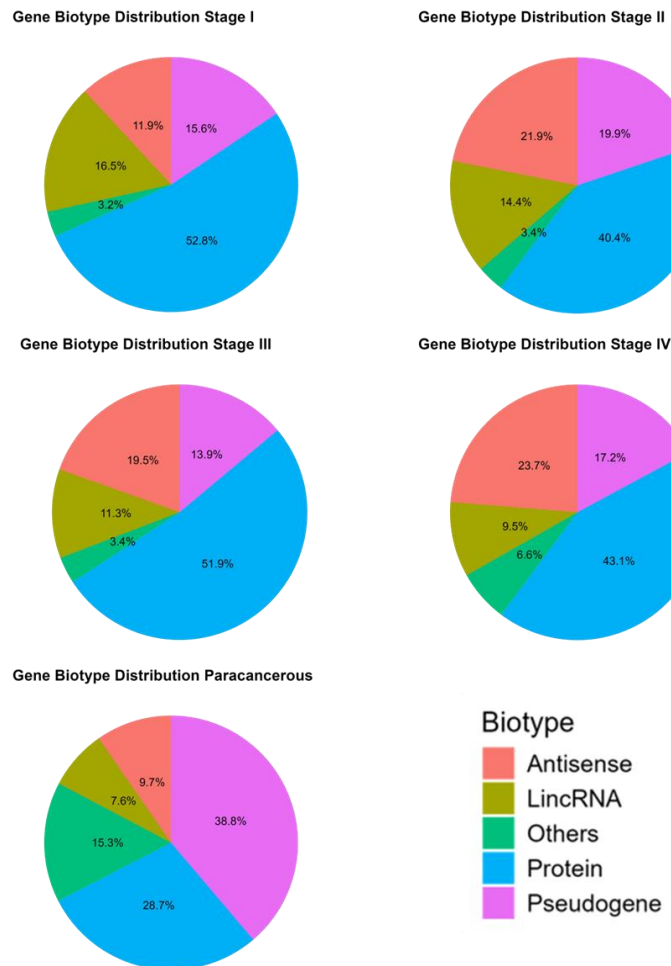
Τα γονίδια που συμμετέχουν στο σχηματισμό δικτύων από την άλλη, εμφανίζουν διαφορετικό προφίλ κατανομής (**Εικόνα 3.22**). Η κατανομή των γονιδιακών βιότυπων



παρουσιάζει παρόμοιο μοτίβο στα στάδια 1 έως 4, παρά τις διακυμάνσεις στα ποσοστά τους. Τα πρωτεϊνικά κωδικοποιητικά γονίδια αποτελούν την κυρίαρχη κατηγορία σε όλα τα στάδια, με ποσοστά που κυμαίνονται από 40.4% έως 52.8%, ενώ ακολουθούν τα γονίδια που ταξινομήθηκαν στην ομάδα "Others", με ποσοστά από 13.9% έως 24.3%. Οι υπόλοιπες κατηγορίες, δηλαδή τα αντινοσηματικά γονίδια (Antisense), τα lincRNAs και τα ψευδογονίδια (Pseudogenes), εμφανίζουν παρόμοια ποσοστά μεταξύ τους, κυμαινόμενα από 9.5% έως 23.7% σε κάθε στάδιο. Αντίθετα, στα παρακαρκινικά δείγματα, η κατανομή

των βιότυπων διαφοροποιείται σημαντικά. Τα ψευδογονίδια καταλαμβάνουν τη μεγαλύτερη αναλογία (38.8%), σε αντίθεση με τα καρκινικά στάδια, όπου το ποσοστό τους είναι σημαντικά χαμηλότερο. Τα πρωτεϊνικά γονίδια μειώνονται στο 28.7%, ενώ τα lincRNAs και τα αντινοσηματικά γονίδια παραμένουν σε χαμηλά επίπεδα (7.6%-9.7%).

### Biotype distribution of genes involved in network creation



**Εικόνα 3.22:** Κατανομή των βιότυπων των διαφορεικά εκφρασμένων γονιδίων σε antisense, LincRNAs, Proteins, Pseudogenes και Others.

## 4. Συζήτηση

Οι αλληλεπιδράσεις μεταξύ των γονιδίων αλλάζουν μεταξύ φυσιολογικών και καρκινικών ιστών. Οι αλλαγές αφορούν τόσο την ταυτότητα των αλληλεπιδράσεων, δηλαδή μεταξύ ποιών γονιδίων συμβαίνουν, όσο και την ισχύ τους. Η ισχύς περιγράφεται μέσω του συντελεστή συσχέτισης Spearman (R), ο οποίος παίρνει τόσο θετικές όσο και αρνητικές τιμές, με κλίμακα τιμών από -1 έως 1. Με βάση τα αποτελέσματα, οι θετικές συσχετίσεις είναι περισσότερες σε αναλογία από τις αρνητικές, σε όλους τους τύπους δειγμάτων και προοδευτικά αυξάνεται όσο εξελίσσεται η νόσος. Μάλιστα, με την διατήρηση των στατιστικά σημαντικών συσχετίσεων, οι θετικές συσχετίσεις καταλαμβάνουν ακόμα μεγαλύτερο μέρος του ποσοστού σε σχέση με πριν. Η ερμηνεία των αποτελεσμάτων πρέπει να βασιστεί στην βιολογική ερμηνεία των θετικών και αρνητικών συσχετίσεων στο πλαίσιο του καρκίνου.

Εντός του κυττάρου, οι θετικές και αρνητικές αλληλεπιδράσεις μεταφράζονται σε θετική ή αρνητική ρύθμιση της έκφρασης. Είναι δεδομένο ότι η ρύθμιση της γονιδιακής έκφρασης μεταβάλλεται στον καρκίνο, λόγω της αύξησης της γενωμικής αστάθειας. Αυτές οι μεταβολές μπορεί να προκύπτουν από διάφορους μηχανισμούς ρύθμισης, όπως η δράση μεταγραφικών παραγόντων, οι επιγενετικές τροποποιήσεις, τα μη κωδικά RNAs και η αναδιαμόρφωση της χρωματίνης, οι οποίοι μπορούν να λειτουργούν τόσο προωθώντας (θετική ρύθμιση) όσο και καταστέλλοντας (αρνητική ρύθμιση) την έκφραση των γονιδίων. Η κοινή ρύθμιση πολλών γονιδίων μαζί, ανεξάρτητα αν είναι θετική ή αρνητική, παρατηρείται εντός των TADs, με βάση όλους αυτούς τους μηχανισμούς.

Οι αρνητικές συσχετίσεις που παρατηρούνται μπορούν να ερμηνευτούν πιο απλά ως η διαδικασία κατά την οποία η έκφραση ενός γονιδίου συμβάλλει στην καταστολή της έκφρασης ενός άλλου. Αυτό συμβαίνει πολλές φορές όταν ο έλεγχος της μεταγραφής ή της μετάφρασης γίνεται από μη κωδικά RNA. Για παράδειγμα το lncRNA **HOTAIR** συνδέει το PRC2 με το LSD1, ενισχύοντας τη μεθυλίωση της H3K27 και την αποακετυλίωση της H3K4 αντίστοιχα. Έτσι προάγεται η σίγηση των γονιδίων στις περιοχές με αυτές τις τροποποιήσεις [34]. Επίσης, πολλά μη κωδικά γονίδια, υβριδίζουν το mRNA γονιδίου με συμπληρωματική αλληλουχία, και το καταστρέφουν, αποτρέποντας την έκφραση της πρωτεΐνης [35, 36]. Η αρνητική ρύθμιση μπορεί να γίνει επιπλέον και μέσω ανταγωνισμού γονιδίων για μεταγραφικούς παράγοντες ή ενισχυτές εντός των TADs.

Παράλληλα, οι θετικές συσχετίσεις περιγράφουν την ακριβώς αντίθετη διαδικασία, κατά την οποία η έκφραση ενός γονιδίου προάγει την έκφραση κάποιου άλλου. Οι μηχανισμοί με τους οποίους συμβαίνει αυτό, είναι αντίστοιχοι με αυτούς των αρνητικών συσχετίσεων. Η ρυθμιστική ικανότητα των μη κωδικών RNA είναι και εδώ παρούσα. Το ψευδογονίδιο PTENP1 έχει ομόλογη περιοχή 3'-UTR με το γονίδιο PTEN. Στην συγκεκριμένη περιοχή συνδέονται miRNA, με σκοπό την αποικοδόμηση του. Έτσι Αυτή η ομολογία προστατεύει το PTEN, καθώς τα miRNAs δεσμεύονται στο μετάγραφο του ψευδογονιδίου. Το αποτέλεσμα είναι η αυξημένη σταθερότητα και μετάφραση του PTEN [35, 36]. Εντός των TADs, η κοινή ρύθμιση γονιδίων από αλληλεπιδρώντες ενισχυτές και υποκινητές, έχει σαν αποτέλεσμα την συνμεταγραφή πολλών γονιδίων, από τους ίδιους μεταγραφικούς παράγοντες.



Η υπόθεση που θα μπορούσε να γίνει, είναι ότι με την αύξηση της γενωμικής αστάθειας όσο εξελίσσεται ο καρκίνος, μηχανισμοί όπως αυτοί που περιγράφηκαν αφορούν κυρίως θετική ρύθμιση. Στην πλειοψηφία τους, τόσο οι μηχανισμοί αυτοί έχουν κοινή βάση, επομένως δεν θα έπρεπε να ευνοείται κάποιος περισσότερο. Ωστόσο, στον καρκίνο παρατηρείται έντονα το πολύ ενδιαφέρον φαινόμενο που ονομάζεται enhancer hijacking εντός των TADs [37]. Ο μηχανισμός αυτός συμβαίνει μέσω διαφόρων δομικών μεταλλάξεων, όπως οι μεταθέσεις (translocations), οι αντιστροφές (inversions) και οι διαγραφές (deletions). Οι μεταθέσεις φέρνουν έναν ισχυρό ενισχυτή κοντά σε ένα ογκογονίδιο, προκαλώντας τη συνεχή ενεργοποίησή του. Οι αντιστροφές αλλάζουν την αλληλουχία της χρωματίνης, αναγκάζοντας ενισχυτές να αλληλεπιδράσουν με ακατάλληλα γονίδια. Οι διαγραφές αφαιρούν ρυθμιστικά στοιχεία που φυσιολογικά περιορίζουν την πρόσβαση του ενισχυτή στο ογκογονίδιο, επιτρέποντας τη σταθερή του υπερέκφραση. Επομένως, φαίνεται πως κατά την δημιουργία αυτού του φαινομένου που ευνοεί τις θετικές αλληλεπιδράσεις, προκαλούνται μεταλλάξεις που αφαιρούν την αρνητική ρύθμιση. Αν θεωρηθεί ότι το φαινόμενο εμφανίζεται περισσότερο σε πιο προχωρημένα στάδια, μπορεί να εξηγήσει την αύξηση της αναλογίας των θετικών αλληλεπιδράσεων.

Επιπλέον, η προσπάθεια που έγινε με βάση τα δεδομένα να γίνει σύνδεση των θετικών συσχετίσεων με την απόσταση, δεν απέδωσε. Ενώ είναι χαρακτηριστικό ότι σε όλους τους τύπους δειγμάτων η πλειοψηφία των θετικών συσχετίσεων εμφανίζεται όταν τα γονίδια είναι κοντά, δεν φάνηκε καμία ουσιαστική διαφορά στις κατανομές τους. Στις αρνητικές συσχετίσεις, το ενδιαφέρον ήταν ότι μεγάλος αριθμός τους ήταν ισοκατανεμημένος σε μεγαλύτερο εύρος αποστάσεων από ότι στις θετικές. Ωστόσο, και σε αυτήν την περίπτωση οι διαφορές μεταξύ των κατανομών για κάθε στάδιο δεν ήταν εμφανείς.

Παράλληλα, σε ένα λιγότερο μοριακό επίπεδο, σε πολλές από τις μελέτες που αφορούν τον σχηματισμό ρυθμιστικών δικτύων γονιδίων, οι αλληλεπιδράσεις που περιγράφονται είναι στην πλειοψηφία τους θετικές. Για παράδειγμα, η μελέτη ενός δικτύου που περιλαμβάνει τα διαφορικά εκφρασμένα γονίδια MAF, SOX2, WNT3A, STAT6 και FOXO1, έδειξε περίπου 10 διαφορετικές αλληλεπιδράσεις μεταξύ των γονιδίων σε φυσιολογικά δείγματα και σε καρκίνο του προστάτη. Σχεδόν όλες αυτές οι αλληλεπιδράσεις είναι θετικές (Spearman  $R > 0$ ) [38]. Σε μια άλλη έρευνα σε προχωρημένα στάδια του καρκίνου του προστάτη, αναλύθηκαν οι συσχετίσεις μεταξύ 118 γονιδίων, από τις οποίες 81 ήταν θετικές και οι 37 ήταν αρνητικές [39]. Αν και οι 2 μελέτες αφορούν μικρότερο αριθμό γονιδίων, δεδομένου ότι σε αυτή την ανάλυση μελετώνται οι αλληλεπιδράσεις στο σύνολο του γονιδιώματος, φαίνεται να μην είναι τόσο σπάνιο να είναι αυξημένες οι θετικές συσχετίσεις.

Το σύνολο αυτών των αλληλεπιδράσεων, αναλύεται και χρησιμοποιείται για τον σχηματισμό ρυθμιστικών δικτύων. Μάλιστα τα γονίδια που συμμετέχουν σε αυτήν την διαδικασία είναι διαφορικά εκφρασμένα στον καρκίνο του προστάτη. Ανάλογα λοιπόν με το στάδιο του καρκίνου, παρατηρείται ότι τα γονίδια που σχηματίζουν τα ρυθμιστικά δίκτυα διαφέρουν και συμμετέχουν σε διαφορετικά βιολογικά μονοπάτια. Ορισμένα δίκτυα είναι σταδιοειδικά ενώ άλλα σχηματίζονται σε παραπάνω από ένα στάδια.

Με βάση τα αποτελέσματα της ανάλυσης οντολογίας, στο στάδιο 1 παρατηρούνται βιολογικά μονοπάτια που αφορούν τις ορμόνες του θυρεοειδή και το οξειδωτικό στρες. Όσον αφορά, τα ορμονικά μονοπάτια που περιγράφονται (παραγωγή θυρεοειδών ορμονών και βιοσύνθεση θυροξίνης), είναι γνωστό ότι συμμετέχουν στον καρκίνο. Η τριιωδοθυρονίνη (T3), έχει αποδειχθεί πως συμβάλλει στον πολλαπλασιασμό των εξαρτώμενων από ανδρογόνο καρκινικών κυττάρων του προστάτη [40]. Αυτό το επιτυγχάνει μέσω της σύνδεσης με υποδοχείς των ορμονών του θυρεοειδή, οι οποίοι αλληλεπιδρούν με τον AR υποδοχέα, και αυτός με την σειρά του συμμετέχει στον πολλαπλασιασμό των κυττάρων του όγκου [41]. Παράλληλα, η T4 (θυροξίνη), επάγει σηματοδοτικά μονοπάτια που σχετίζονται με την επιβίωση και τον πολλαπλασιασμό, όπως το MAPK/ERK. Η T4 επίσης, φαίνεται να συμμετέχει και στην γλυκόλυση, ένα από τα βασικά χαρακτηριστικά του καρκίνου. Ο καρκίνος του προστάτη είναι εξαρτώμενος από σηματοδότηση του υποδοχέα των ανδρικών ορμονών στα αρχικά στάδια, πράγμα που συμβαδίζει με την εμφάνιση του μονοπατιού στο στάδιο 1[41]. Παράλληλα, τα μονοπάτια που αφορούν την βιοσύνθεση του υπεροξειδίου του υδρογόνου και την οξείδωση των λιπιδίων σχετίζονται επίσης με την έναρξη του καρκίνου. Το πιο ενδιαφέρον είναι ότι το δίκτυο AMACR, που εμφανίζεται σε όλα τα στάδια του καρκίνου, συνδέεται και με τα δυο μονοπάτια. Το γονίδιο AMACR, είναι απαραίτητο στην οξείδωση της β-αλυσίδας των λιπαρών οξέων και είναι υπερεκφρασμένο στον καρκίνο [42]. Η διαδικασία αυτή παράγει έμμεσα H<sub>2</sub>O<sub>2</sub>, το οποίο αυξάνει το οξειδωτικό στρες, βλάβες στο DNA και τελικά καρκινογένεση[43]. Προφανώς, στο στάδιο 1 υπάρχουν και άλλα γονίδια που είναι πιο άμεσα συνδεδεμένα με τα 2 μονοπάτια, ωστόσο, η διεργασία που περιγράφηκε αφορά πολύ πρώιμα στάδια της νόσου. Θα μπορούσε να ισχύει επίσης πως η παρουσία του δικτύου AMACR σε όλα τα στάδια να σχετίζεται με την συνεχόμενη δημιουργία οξειδωτικού στρες και άρα γενωμικής αστάθειας. Τέλος, το σταδιοεδικό δίκτυο OR, δεν φαίνεται να σχετίζεται με κανένα από τα βιολογικά μονοπάτια που περιγράφονται. Μάλιστα, δεν υπάρχει κάποια σχετική δημοσίευση για τα συγκεκριμένα γονίδια ως σχετιζόμενα με τον καρκίνο. Ωστόσο, η έκτοπη έκφραση υποδοχέων του οσφρητικού έχει περιγραφεί στον καρκίνο του προστάτη. Το γονίδιο OR51E2, για παράδειγμα, έχει αποδειχθεί πως είναι υπερεκφρασμένο σε όγκους, και προωθεί τον πολλαπλασιασμό των κυττάρων τους. Συνδέεται μάλιστα με αρχικά στάδια της νόσου [44].

Στα στάδια 2 έως 4, ένα μεγάλο μέρος γονιδίων, συμμετέχουν στην ανάπτυξη του νευρικού συστήματος. Αν και εκ πρώτης όψεως μοιάζει να μην συνδέεται με τον καρκίνο του προστάτη, όμως μεγάλο ενδιαφέρον παρουσιάζει το πως μεταβάλλεται η νόσος στα τελευταία στάδια της. Αυτό που παρατηρείται είναι η νευροενδοκρινική διαφοροποίηση του καρκίνου του προστάτη (NEPC). Στην περίπτωση αυτή τα επιθηλιακά κύτταρα του προστάτη αποκτούν χαρακτηριστικά νευροενδοκρινικών κυττάρων, όπως έκφραση νευροπεπτιδίων και διαφοροποιημένων παραγόντων ανάπτυξης. Αυτό γίνεται μέσω της μετατροπής του επιθηλιακού σε μεσεγχυματικό ιστό (EMT), μιας διαδικασίας πολύ χαρακτηριστικής στον καρκίνο. Σε αυτήν την φάση, ο καρκίνος του προστάτη εμφανίζει απώλεια της AR σηματοδότησης [45]. Με βάση τα αποτελέσματα, στα στάδια 2 έως 4, σχηματίζεται ένα πολυπληθές δίκτυο πρωτοκαδερινών. Οι πρωτοκαδερίνες αποτελούν σημαντικά γονίδια που σχετίζονται με την ανάπτυξη του νευρικού, την νευρική διαφοροποίηση και την κυτταρική προσκόλληση [46]. Στα δίκτυα εντοπίζονται κυρίως πρωτοκαδερίνες της συστάδας A,

και σε πολύ μικρότερο βαθμό και μόνο στο στάδιο 3, πρωτοκαδερίνες της συστάδας B. Μάλιστα από τα γονίδια του δικτύου, το PCDHB1, είναι το μόνο που έχει μελετηθεί συμμετέχει στον καρκίνο του προστάτη. Το δίκτυο αλλάζει όσο η νόσος εξελίσσεται, και μάλιστα το στάδιο 2 έχει παραπάνω ομοιότητες με τα παρακαρκινικά δείγματα από ότι με τα πιο προχωρημένα στάδια. Ο διαχωρισμός αυτός θα μπορούσε να σημαίνει ότι η απορρύθμιση του δικτύου ξεκινάει στο στάδιο 2. Παρά το γεγονός ότι οι συγκεκριμένες πρωτοκαδερίνες δεν έχουν αποδεδειγμένα σχέση με το EMT, δεδομένου ότι είναι αυξορρυθμιζόμενες και σχετίζονται την νευρική διαφοροποίηση, δεν θα ήταν παράλογο. Μάλιστα άλλες πρωτοκαδερίνες που δεν σχετίζονται καν με διεργασίες του, έχουν αποδεδειγμένα συμμετοχή στον NEPC. Συγκεκριμένα, πολλές μελέτες δείχνουν ότι η PCDH-PC, είναι αυξορρυθμιζόμενη στον NEPC, και συμμετέχει στο σχηματισμό του, ελέγχοντας το μονοπάτι Wnt/ β-κατενίνης, και άρα προκαλώντας με έμμεσο τρόπο EMT [ww, 48]. Ο κύριος τρόπος όμως, με τον οποίο συμμετέχει στην εκδήλωση αυτής της μορφής, είναι με την διακοπή σηματοδότησης του AR. Η PCDH7, είναι επίσης άμεσα συνδεδεμένη με την εκδήλωση του NEPC, καθώς είναι υπερεκφρασμένη σε όγκους αυτού του τύπου. Ο τρόπος με τον οποίο συμμετέχει, είναι η ενίσχυση της σηματοδότησης μονοπατιών όπως το ERK/AKT, τα οποία μπορούν στο καρκίνο να προωθούν την μετατροπή σε μεσεγχυματικό ιστό [www]. Επιπλέον, στο δίκτυο ρυθμιστικό ρόλο έχει μια σειρά από lncRNAs (AC0005609.16-20), τα οποία συνδέονται με την πλειοψηφία των πρωτοκαδερινών. Ο ρόλος πολλών lncRNAs στο NEPC, είναι επίσης γεγονός. Ένα χαρακτηριστικό παράδειγμα, είναι το lncRNA **LINC00261**. Πιθανώς οι συσχετίσεις των AC0005609 με τα γονίδια της πρωτοκαδερίνης να αφορούν διαφορετικό μηχανισμό, καθώς η δράση του **LINC00261** στον NEPC ενεργοποιεί μονοπάτια που υποστηρίζουν την επιβίωση και την ανάπτυξη των καρκινικών κυττάρων [45].

Παράλληλα, κοινό γνώρισμα των τριών σταδίων ήταν ο υψηλός εμπλουτισμός μονοπατιών που σχετίζονται τον μεταβολισμό λιπιδίων και την επεξεργασία λιποπρωτεϊνών. Στα στάδια 2 και 3, τα μονοπάτια αφορούν κυρίως τα χυλομικρά και την VLDL, ενώ στο στάδιο 4 εμφανίζονται κυρίως διεργασίες στην HDL και στις λιποπρωτεΐνες του πλάσματος. Το δίκτυο APOE σχηματίζεται και στα 3 στάδια. Τα γονίδια APOE και APOC2, που το αποτελούν δέχονται ρύθμιση από 2 μη κωδικά γονίδια, τα APOC1P1 και APOC4-APOC2. Σχεδόν όλα από αυτά αναφέρονται στην βιβλιογραφία ως γονίδια σχετιζόμενα με τον καρκίνο, και ορισμένα συγκεκριμένα για τον καρκίνο του προστάτη. Τα γονίδια APOE και APOC2 έχουν βασικό ρόλο στη ρύθμιση της μεταφοράς των λιπιδίων και της λιποπρωτεϊνικής. Το APOE αποτελεί απολιποπρωτεΐνη-κλειδί, η οποία συμμετέχει στον μεταβολισμό των λιποπρωτεϊνών μέσω της σύνδεσής της με σωματίδια χυλομικρών, VLDL, IDL και HDL, με προτίμηση στις πρώτες 2 κατηγορίες [50]. Παράλληλα, εμπλέκεται στην αναγνώριση και την πρόσληψη των λιποπρωτεϊνών από τα κύτταρα μέσω αλληλεπίδρασης με τους υποδοχείς LDLR και VLDLR. Επίσης, το APOE συμμετέχει στην απομάκρυνση των χυλομικρών. Από την άλλη, το APOC2 είναι επίσης βασικός ρυθμιστής του μεταβολισμού των λιπιδίων, καθώς δρα ως ενεργοποιητής της λιποπρωτεϊνικής λιπάσης (LPL), η οποία είναι υπεύθυνη για την υδρόλυση των τριγλυκεριδίων στις λιποπρωτεΐνες. Σε φυσιολογικές συνθήκες, το APOC2 εντοπίζεται κυρίως στις HDL. Συνεπώς, οι δράσεις και των 2 γονιδίων ταυτίζονται με τα ευρήματα των αναλύσεων οντολογίας. Το APOE έχει χαρακτηριστεί ως άμεσα εμπλεκόμενο γονίδιο στον καρκίνο

του προστάτη και αποτελεί δείκτη κακής πρόγνωσης. Συγκεκριμένα, έχει διαπιστωθεί ότι γενετικοί πολυμορφισμοί στο γονίδιο ApoE μπορούν να οδηγήσουν σε ελαττωματική απομάκρυνση της χοληστερόλης, με συνέπεια τη συσσώρευση της εντός των κυττάρων και την προώθηση της επιθετικότητας των όγκων μέσω της ενεργοποίησης του μορίου caveolin-1 (Cav-1) [51]. Παράλληλα, το APOE είναι υπερεκφρασμένο σε προχωρημένα στάδια και σχετίζεται με όγκους ανεξάρτητους στην σηματοδότηση AR [50]. Το APOC2, έχει αποδειχθεί πως μπορεί να λειτουργήσει ως χρήσιμος βιοδείκτης στον καρκίνο του παγκρέατος [52] και το APOC1P1 συμμετέχει στην μείωση της απόπτωσης των κυττάρων στον καρκίνο του μαστού [53]. Ενδιαφέρον παρουσιάζει η μετάβαση από επεξεργασία των χυλομικρών και της VLDL στα στάδια 2 και 3, σε επεξεργασία της HDL. Είναι γεγονός ότι η μπορεί να σχηματιστεί από τα χυλομικρά και την VLDL κατά την λιπόλυση [54]. Η HDL έχει αποδειγμένα πολύ σημαντικό ρόλο στο στάδιο 4. Μελέτες έχουν δείξει ότι μπορεί να επάγει τον πολλαπλασιασμό και την μετάσταση καρκινικών κυττάρων που εμφανίζουν σηματοδότηση ανεξάρτητη του AR. Αυτό συμβαίνει λόγω της αλληλεπίδρασης της HDL με διαφορετικές πρωτεΐνες όπως η S1P και η ABCA1, που προάγουν μεταγωγή σήματος για πολλαπλασιασμό, μετάσταση, ανάπτυξη και ανεξαρτητοποίηση από ανδρογόνα [55].

Συμπερασματικά, η μεθοδολογία έχει παράγει αποτελέσματα που ακολουθούν τις πληροφορίες ήδη δημοσιευμένων μελετών. Πολλά από τα γονίδια που εντοπίστηκε ότι σχηματίζουν δίκτυα, μελώνται ως σημαντικοί παράγοντες στην εξέλιξη του καρκίνου του προστάτη. Ιδιαίτερο ενδιαφέρον παρουσιάζει το δίκτυο των πρωτοκαδερινών στα πιο προχωρημένα στάδια του όγκου. Λόγω του τόσο μεγάλου αριθμού γονιδίων, μένει να αποδειχθεί και εργαστηριακά η συσχέτιση τους με τον καρκίνο του προστάτη. Με αυτόν τον τρόπο μπορεί να βρεθεί και μια πιο υψηλή συσχέτιση με την νευροενδοκρινική μορφή του. Με εξαίρεση την δραματική αύξηση των θετικών συσχετίσεων, η οποία παρατηρήθηκε μετά την διατήρηση μόνο των στατιστικά σημαντικών ζευγών και για την οποία οι εξηγήσεις που δόθηκαν αποτελούν μια σειρά από λογικές υποθέσεις, δεν υπάρχει κάποιο άλλο κομμάτι της έρευνας που να μην δικαιολογείται από την βιβλιογραφία. Μελλοντικά, η ίδια μεθοδολογία θα μπορούσε να εφαρμοστεί σε δεδομένα που αφορούν άλλες μορφές καρκίνου, ώστε να ελεγχθεί αν τα δεδομένα που παράγονται ακολουθούν και σε αυτές τις περιπτώσεις την βιβλιογραφία. Επιπλέον, ο κώδικας που χρησιμοποιήθηκε, θα μπορούσε να αυτοματοποιηθεί ακόμα περισσότερο σε ορισμένα σημεία. Στόχος είναι, μόνο με την εισαγωγή των δεδομένων, να παράγεται αυτόματα το σύνολο των γραφημάτων, θερμικών χαρτών, δικτύων κ.τ.λ σε ένα γραφικό περιβάλλον μέσω της R Shiny.

## 5. Αποθετήριο δεδομένων και κώδικα

Τα δεδομένα λήφθηκαν από τις κοινοπραξίες TCGA και GTEX και ο κώδικας από το αντίστοιχο project είναι διαθέσιμος στο GitHub ([https://github.com/JimFotia/BSc\\_Thesis](https://github.com/JimFotia/BSc_Thesis))

## 6. Βιβλιογραφία

- [1] Sana Belkahla et al., “Advances and Development of Prostate Cancer, Treatment, and Strategies: A Systemic Review,” *Frontiers in Cell and Developmental Biology* 10 (September 9, 2022): 991330, <https://doi.org/10.3389/fcell.2022.991330>.
- [2] Freddie Bray et al., “Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” *CA: A Cancer Journal for Clinicians* 74, no. 3 (May 2024): 229–63, <https://doi.org/10.3322/caac.21834>.
- [3] Jana McHugh et al., “Prostate Cancer Risk in Men of Differing Genetic Ancestry and Approaches to Disease Screening and Management in These Groups,” *British Journal of Cancer* 126, no. 10 (June 1, 2022): 1366–73, <https://doi.org/10.1038/s41416-021-01669-3>.
- [4] American Cancer Society. Prostate cancer causes, risk factors, and prevention. 2023. Available at: <https://www.cancer.org/cancer/prostate-cancer/causes-risks-prevention.html>
- [5] Mamello Sekhoacha et al., “Prostate Cancer Review: Genetics, Diagnosis, Treatment Options, and Alternative Approaches,” *Molecules* 27, no. 17 (September 5, 2022): 5730, <https://doi.org/10.3390/molecules27175730>.
- [6] Qintao Ge et al., “Molecular Classifications of Prostate Cancer: Basis for Individualized Risk Stratification and Precision Therapy,” *Annals of Medicine* 55, no. 2 (December 12, 2023): 2279235, <https://doi.org/10.1080/07853890.2023.2279235>.
- [7] Adam Abeshouse et al., “The Molecular Taxonomy of Primary Prostate Cancer,” *Cell* 163, no. 4 (November 2015): 1011–25, <https://doi.org/10.1016/j.cell.2015.10.025>.
- [8] Elisabeth A. Messner et al., “The Androgen Receptor in Prostate Cancer: Effect of Structure, Ligands and Spliced Variants on Therapy,” *Biomedicines* 8, no. 10 (October 15, 2020): 422, <https://doi.org/10.3390/biomedicines8100422>.
- [9] Christopher E Barbieri et al., “Exome Sequencing Identifies Recurrent SPOP, FOXA1 and MED12 Mutations in Prostate Cancer,” *Nature Genetics* 44, no. 6 (June 2012): 685–89, <https://doi.org/10.1038/ng.2279>.
- [10] Mona Teng et al., “Pioneer of Prostate Cancer: Past, Present and the Future of FOXA1,” *Protein & Cell* 12, no. 1 (January 2021): 29–38, <https://doi.org/10.1007/s13238-020-00786-8>.
- [11] Mark S. Litwin and Hung-Jui Tan, “The Diagnosis and Treatment of Prostate Cancer: A Review,” *JAMA* 317, no. 24 (June 27, 2017): 2532, <https://doi.org/10.1001/jama.2017.7248>.
- [12] Ganji Lakshmi Varaprasad et al., “Recent Advances and Future Perspectives in the Therapeutics of Prostate Cancer,” *Experimental Hematology & Oncology* 12, no. 1 (September 22, 2023): 80, <https://doi.org/10.1186/s40164-023-00444-9>.
- [13] Philip Posdzich et al., “Metastatic Prostate Cancer—A Review of Current Treatment Options and Promising New Approaches,” *Cancers* 15, no. 2 (January 11, 2023): 461, <https://doi.org/10.3390/cancers15020461>.
- [14] Xingyu Liu et al., “Cis-Regulatory Elements in Mammals,” *International Journal of Molecular Sciences* 25, no. 1 (December 26, 2023): 343, <https://doi.org/10.3390/ijms25010343>.
- [15] Anne I. Krämer and Christoph Handschin, “How Epigenetic Modifications Drive the Expression and Mediate the Action of PGC-1 $\alpha$  in the Regulation of Metabolism,” *International Journal of Molecular Sciences* 20, no. 21 (October 31, 2019): 5449, <https://doi.org/10.3390/ijms20215449>.
- [16] Jonathan A. Beagan and Jennifer E. Phillips-Cremins, “On the Existence and Functionality of Topologically Associating Domains,” *Nature Genetics* 52, no. 1 (January 2020): 8–16, <https://doi.org/10.1038/s41588-019-0561-1>.
- [17] Moray J Campbell, “Tales from Topographic Oceans: Topologically Associated Domains and Cancer,” *Endocrine-Related Cancer* 26, no. 11 (November 2019): R611–26, <https://doi.org/10.1530/ERC-19-0348>.



- [18] Juan J. Tena and José M. Santos-Pereira, "Topologically Associating Domains and Regulatory Landscapes in Development, Evolution and Disease," *Frontiers in Cell and Developmental Biology* 9 (July 6, 2021): 702787, <https://doi.org/10.3389/fcell.2021.702787>.
- [19] Frank Emmert-Streib, Matthias Dehmer, and Benjamin Haibe-Kains, "Gene Regulatory Networks and Their Applications: Understanding Biological and Medical Problems in Terms of Networks," *Frontiers in Cell and Developmental Biology* 2 (August 19, 2014), <https://doi.org/10.3389/fcell.2014.00038>.
- [20] Helen S. Long et al., "Making Sense of the Linear Genome, Gene Function and TADs," *Epigenetics & Chromatin* 15, no. 1 (December 2022): 4, <https://doi.org/10.1186/s13072-022-00436-9>.
- [21] Andreas Kloetgen et al., "Three-Dimensional Chromatin Landscapes in T Cell Acute Lymphoblastic Leukemia," *Nature Genetics* 52, no. 4 (April 2020): 388–400, <https://doi.org/10.1038/s41588-020-0602-9>.
- [22] Rebeca San Martin et al., "Chromosome Compartmentalization Alterations in Prostate Cancer Cell Lines Model Disease Progression," *Journal of Cell Biology* 221, no. 2 (February 7, 2022): e202104108, <https://doi.org/10.1083/jcb.202104108>.
- [23] Shuang G. Zhao et al., "Integrated Analyses Highlight Interactions between the Three-Dimensional Genome and DNA, RNA and Epigenomic Alterations in Metastatic Prostate Cancer," *Nature Genetics* 56, no. 8 (August 2024): 1689–1700, <https://doi.org/10.1038/s41588-024-01826-3>.
- [24] Chen Y, Chen L, Lun ATL, Baldoni PL, Smyth GK (2024). edgeR 4.0: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets.
- [25] Schauburger P, Walker A (2024). \_openxlsx: Read, Write and Edit xlsx Files\_. R package version 4.2.7.1, <<https://CRAN.R-project.org/package=openxlsx>>.
- [26] Hadley Wickham (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software*, 21(12), 1-20. URL <http://www.jstatsoft.org/v21/i12/>.
- [27] Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Golemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." *Journal of Open Source Software*, 4(43), 1686. doi:10.21105/joss.01686 <<https://doi.org/10.21105/joss.01686>>.
- [28] Harrell Jr F (2024). \_Hmisc: Harrell Miscellaneous\_. R package version 5.2-1, <<https://CRAN.R-project.org/package=Hmisc>>.
- [29] Lawrence M, Huber W, Pag`es H, Aboyoun P, Carlson M, et al. (2013) Software for Computing and Annotating Genomic Ranges. *PLoS Comput Biol* 9(8): e1003118. doi:10.1371/journal.pcbi.1003118
- [30] Wilke C (2024). \_ggridges: Ridgeline Plots in 'ggplot2'\_. R package version 0.5.6, <<https://CRAN.R-project.org/package=ggridges>>.
- [31] Simon Garnier, Noam Ross, Robert Rudis, Antônio P. Camargo, Marco Sciaini, and Cédric Scherer (2024). viridis(Lite) - Colorblind-Friendly Color Maps for R. viridis package version 0.6.5
- [32] Kolde R (2019). \_pheatmap: Pretty Heatmaps\_. R package version 1.0.12, <<https://CRAN.R-project.org/package=pheatmap>>.
- [33] Neuwirth E (2022). \_RColorBrewer: ColorBrewer Palettes\_. R package version 1.1-3, <<https://CRAN.R-project.org/package=RColorBrewer>>.
- [34] Xiaopei Zhang et al., "Mechanisms and Functions of Long Non-Coding RNAs at Multiple Regulatory Levels," *International Journal of Molecular Sciences* 20, no. 22 (November 8, 2019): 5573, <https://doi.org/10.3390/ijms20225573>.
- [35] Ryan C. Pink and David R.F. Carter, "Pseudogenes as Regulators of Biological Function," ed. Mark A. Lindsay and Sam Griffiths-Jones, *Essays in Biochemistry* 54 (May 3, 2013): 103–12, <https://doi.org/10.1042/bse0540103>.
- [36] Baikang Pei et al., "The GENCODE Pseudogene Resource," *Genome Biology* 13, no. 9 (September 5, 2012): R51, <https://doi.org/10.1186/gb-2012-13-9-r51>.

- [37] Moray J Campbell, "Tales from Topographic Oceans: Topologically Associated Domains and Cancer," *Endocrine-Related Cancer* 26, no. 11 (November 2019): R611–26, <https://doi.org/10.1530/ERC-19-0348>
- [38] Taj Mohammad et al., "Differential Gene Expression and Weighted Correlation Network Dynamics in High-Throughput Datasets of Prostate Cancer," *Frontiers in Oncology* 12 (June 1, 2022): 881246, <https://doi.org/10.3389/fonc.2022.881246>.
- [39] Elena A. Pudova et al., "Gene Expression Changes and Associated Pathways Involved in the Progression of Prostate Cancer Advanced Stages," *Frontiers in Genetics* 11 (January 21, 2021): 613162, <https://doi.org/10.3389/fgene.2020.613162>.
- [40] Eilon Krashin et al., "Thyroid Hormones and Cancer: A Comprehensive Review of Preclinical and Clinical Studies," *Frontiers in Endocrinology* 10 (February 13, 2019): 59, <https://doi.org/10.3389/fendo.2019.00059>.
- [41] Sepehr Torabinejad et al., "The Androgen-Thyroid Hormone Crosstalk in Prostate Cancer and the Clinical Implications," *European Thyroid Journal* 12, no. 3 (March 17, 2023): e220228, <https://doi.org/10.1530/ETJ-22-0228>.
- [42] Xinyu Wu et al., "Lipid Metabolism in Prostate Cancer," *American Journal of Clinical and Experimental Urology* 2, no. 2 (2014): 111–20.
- [43] Chandan Kumar-Sinha et al., "Elevated  $\alpha$ -Methylacyl-CoA Racemase Enzymatic Activity in Prostate Cancer," *The American Journal of Pathology* 164, no. 3 (March 2004): 787–93, [https://doi.org/10.1016/S0002-9440\(10\)63167-7](https://doi.org/10.1016/S0002-9440(10)63167-7).
- [44] Chan Chung et al., "Odorant Receptors in Cancer," *BMB Reports* 55, no. 2 (February 28, 2022): 72–80, <https://doi.org/10.5483/BMBRep.2022.55.2.010>.
- [45] Joshua Altschuler, Jennifer A. Stockert, and Natasha Kyprianou, "Non-Coding RNAs Set a New Phenotypic Frontier in Prostate Cancer Metastasis and Resistance," *International Journal of Molecular Sciences* 22, no. 4 (February 20, 2021): 2100, <https://doi.org/10.3390/ijms22042100>.
- [46] Yokota, Shinnichi, et al. "Identification of the Cluster Control Region for the Protocadherin- $\beta$  Genes Located beyond the Protocadherin- $\gamma$  Cluster." *The Journal of Biological Chemistry*, vol. 286, no. 36, 9 Sept. 2011, pp. 31885-31895. <https://doi.org/10.1074/jbc.M111.245605>.
- [47] Terry, Stephane, et al. "Protocadherin-PC promotes androgen-independent prostate cancer cell growth." *The Prostate*, vol. 66, no. 11, 2006, pp. 1100-1113. <https://doi.org/10.1002/pros.20494>.
- [48] Eva Slabáková et al., "Regulation of Neuroendocrine-like Differentiation in Prostate Cancer by Non-Coding RNAs," *Non-Coding RNA* 7, no. 4 (December 2, 2021): 75, <https://doi.org/10.3390/ncrna7040075>.
- [49] Shishodia, Gauri, Sweaty Koul, and Hari K. Koul. "Protocadherin 7 is overexpressed in castration resistant prostate cancer and promotes aberrant MEK and AKT signaling." *The Prostate*, vol. 79, no. 14, 2019, pp. 1683-1693. <https://doi.org/10.1002/pros.23898>.
- [50] Liwen Ren et al., "Apolipoproteins and Cancer," *Cancer Medicine* 8, no. 16 (November 2019): 7032–43, <https://doi.org/10.1002/cam4.2587>.
- [51] Ifere, Godwin O., et al. "Apolipoprotein E gene polymorphism influences aggressive behavior in prostate cancer cells by deregulating cholesterol homeostasis." *International Journal of Oncology*, vol. 43, no. 4, 2013, pp. 1002-1010. <https://doi.org/10.3892/ijo.2013.2057>.
- [52] Yingcheng He et al., "Apolipoproteins: New Players in Cancers," *Frontiers in Pharmacology* 13 (November 25, 2022): 1051280, <https://doi.org/10.3389/fphar.2022.1051280>.
- [53] X-H Liao et al., "Long Intergenic Non-Coding RNA APOC1P1-3 Inhibits Apoptosis by Decreasing  $\alpha$ -Tubulin Acetylation in Breast Cancer," *Cell Death & Disease* 7, no. 5 (May 26, 2016): e2236–e2236, <https://doi.org/10.1038/cddis.2016.142>.
- [54] Gotto, Antonio M., Jr. "Interrelationship of triglycerides with lipoproteins and high-density lipoproteins." *The American Journal of Cardiology*, vol. 66, no. 6, 4 Sept. 1990, pp. A20-A23.

[55] Kotani, Kazuhiko, et al. "High-Density Lipoprotein and Prostate Cancer: An Overview." *Journal of Epidemiology*, vol. 23, no. 5, 2013, pp. 313-319. <https://doi.org/10.2188/jea.JE20130006>.