

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΪΑΤΡΙΚΗ»



ΣΥΣΤΗΜΑΤΙΚΗ ΑΝΑΣΚΟΠΗΣΗ ΚΑΙ ΚΑΤΑΣΚΕΥΗ
ΓΟΝΙΔΙΑΚΟΥ ΠΡΟΦΙΛ ΓΙΑ ΤΗΝ ΠΡΟΒΛΕΨΗ ΤΗΣ
ΠΝΕΥΜΟΝΙΚΗΣ ΑΡΤΗΡΙΑΚΗΣ ΥΠΕΡΤΑΣΗΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
ΜΑΡΙΑ ΚΑΜΠΑ

ΕΠΙΒΛΕΠΩΝ
ΜΠΑΓΚΟΣ ΠΑΝΤΕΛΕΗΜΩΝ

ΛΑΜΙΑ
ΦΕΒΡΟΥΑΡΙΟΣ 2021

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΪΑΤΡΙΚΗ»



ΣΥΣΤΗΜΑΤΙΚΗ ΑΝΑΣΚΟΠΗΣΗ ΚΑΙ ΚΑΤΑΣΚΕΥΗ
ΓΟΝΙΔΙΑΚΟΥ ΠΡΟΦΙΛ ΓΙΑ ΤΗΝ ΠΡΟΒΛΕΨΗ ΤΗΣ
ΠΝΕΥΜΟΝΙΚΗΣ ΑΡΤΗΡΙΑΚΗΣ ΥΠΕΡΤΑΣΗΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
ΜΑΡΙΑ ΚΑΜΠΑ

ΕΠΙΒΛΕΠΩΝ
ΜΠΑΓΚΟΣ ΠΑΝΤΕΛΕΗΜΩΝ

ΛΑΜΙΑ
ΦΕΒΡΟΥΑΡΙΟΣ 2021

ΣΥΣΤΗΜΑΤΙΚΗ ΑΝΑΣΚΟΠΗΣΗ ΚΑΙ ΚΑΤΑΣΚΕΥΗ
ΓΟΝΙΔΙΑΚΟΥ ΠΡΟΦΙΛ ΓΙΑ ΤΗΝ ΠΡΟΒΛΕΨΗ ΤΗΣ
ΠΝΕΥΜΟΝΙΚΗΣ ΑΡΤΗΡΙΑΚΗΣ ΥΠΕΡΤΑΣΗΣ

Τριμελής Επιτροπή

Όνοματεπώνυμο

Βαθμίδα

Μπάγκος Παντελεήμων
(επιβλέπων)

Καθηγητής-Κοσμήτορας του Τμήματος
Πληροφορικής με Εφαρμογές στη
Βιοϊατρική του Πανεπιστημίου Θεσσαλίας

Χατζηγεωργίου Άρτεμις

Καθηγήτρια του Τμήματος Πληροφορικής
με Εφαρμογές στη Βιοϊατρική του
Πανεπιστημίου Θεσσαλίας

Μπράλιου Γεωργία

Επίκουρη Καθηγήτρια στο Τμήμα
Πληροφορικής με Εφαρμογές στη
Βιοϊατρική του Πανεπιστημίου Θεσσαλίας

«Υπεύθυνη Δήλωση μη λογοκλοπής και ανάληψης προσωπικής ευθύνης»

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, και γνωρίζοντας τις συνέπειες της λογοκλοπής, δηλώνω υπεύθυνα και ενυπογράφως ότι η παρούσα εργασία με τίτλο «Συστηματική Ανασκόπηση και Κατασκευή Γονιδιακού Προφίλ για την Πρόβλεψη της Πνευμονικής Αρτηριακής Υπέρτασης» αποτελεί προϊόν αυστηρά προσωπικής εργασίας και όλες οι πηγές από τις οποίες χρησιμοποίησα δεδομένα, ιδέες, φράσεις, προτάσεις ή λέξεις, είτε επακριβώς (όπως υπάρχουν στο πρωτότυπο ή μεταφρασμένες) είτε με παράφραση, έχουν δηλωθεί κατάλληλα και ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής.

Η Δηλούσα

Ημερομηνία

Κάμπα Μαρία

2/2/2021

Πίνακας Περιεχομένων

Πρόλογος	1
Περίληψη.....	3
Abstract	4
Συνοπτομογραφίες.....	5
Κεφάλαιο 1 - Εισαγωγή	7
1.1 Υπολογιστική Βιολογία.....	7
1.2 Πνευμονική Υπέρταση.....	8
1.2.1 Διαγνωστικά και Επιδημιολογικά Χαρακτηριστικά της Πνευμονικής Υπέρτασης	8
1.2.2 Οι Κύριες Ομάδες της Πνευμονικής Υπέρτασης.....	8
1.2.3 Τα Βασικά Συμπτώματα της Πνευμονικής Υπέρτασης.....	9
1.3 Πνευμονική Αρτηριακή Υπέρταση	9
1.3.1 Διαγνωστικά και Επιδημιολογικά Χαρακτηριστικά της Πνευμονικής Αρτηριακής Υπέρτασης	9
1.3.2 Μορφές Πνευμονικής Αρτηριακής Υπέρτασης.....	11
1.3.3 Συμπτώματα Πνευμονικής Αρτηριακής Υπέρτασης.....	11
Κεφάλαιο 2 – Μικροσυστοιχίες DNA	13
2.1 Το Κεντρικό Δόγμα της Βιολογίας.....	13
2.2 Γονιδιακή Έκφραση	13
2.3 Οι Μικροσυστοιχίες DNA.....	15
2.3.1 Τύποι Μικροσυστοιχιών	15
2.3.2 Διαδικασία Πειράματος Μικροσυστοιχιών.....	16
2.3.3 Βάσεις Δεδομένων Μικροσυστοιχιών.....	18
2.3.4 Εργαλεία Λογισμικού για Ανάλυση Δεδομένων Γονιδιακής Έκφρασης	19
2.4 Τεχνικές Ανάλυσης Δεδομένων Γονιδιακής Έκφρασης από DNA Μικροσυστοιχίες	20
2.4.1 Μέθοδος t-test.....	20
2.4.2 Μέθοδος Αναδειγματοληψίας Bootstrap.....	21
2.4.3 Μπεϋζιανές Μέθοδοι.....	22
2.4.4 Άλλες Μέθοδοι	23

Κεφάλαιο 3 – Μετα-ανάλυση	25
3.1 Συστηματική Ανασκόπηση και Μετα-ανάλυση	25
3.2 Μοντέλα Σταθερών και Τυχαίων Επιδράσεων	25
3.2.1 Μοντέλο Σταθερών Επιδράσεων	26
3.2.2 Μοντέλο Τυχαίων Επιδράσεων	27
3.3 Έλεγχος Ετερογένειας	28
3.4 Σφάλμα Μεροληψίας	28
3.5 Μετα-ανάλυση Δεδομένων Μικροσυστοιχιών DNA	29
3.5.1 Τεχνικές Μετα-ανάλυσης Δεδομένων Γονιδιακής Έκφρασης από DNA Μικροσυστοιχίες	30
3.5.1.1 Η τυποποιημένη διαφορά των μέσων σαν μέγεθος επίδρασης	30
3.5.1.2 Άλλες μέθοδοι	31
3.5.2 Πολλαπλές συγκρίσεις και μέθοδοι διόρθωσης σφαλμάτων Τύπου I	32
3.6 Σκοπός της Εργασίας	34
Κεφάλαιο 4 – Προτεινόμενη Μεθοδολογία της Παρούσας Εργασίας	35
4.1 Συστηματική Ανασκόπηση	35
4.2 Επεξεργασία Πινάκων Γονιδιακής Έκφρασης	35
4.3 Στατιστική Ανάλυση και Μετα-ανάλυση Δεδομένων Γονιδιακής Έκφρασης	37
4.4 Ανάλυση Εμπλουτισμού των Διαφορικά Εκφραζόμενων Γονιδίων	38
Κεφάλαιο 5 – Αποτελέσματα	39
5.1 Αποτελέσματα Συστηματικής Ανασκόπησης	39
5.2 Αποτελέσματα Στατιστικού Ελέγχου	41
5.3 Αποτελέσματα μετά από χρήση Εργαλείων Εμπλουτισμού	48
5.3.1 Ανίχνευση Βιολογικών Διεργασιών με χρήση του bioCompendium	48
5.3.2 Κατασκευή Δικτύου Αλληλεπίδρασης Πρωτεΐνης – Πρωτεΐνης μέσω της STRING	51
5.4 Διασταύρωση των Ευρημάτων μέσω άλλων Βάσεων Δεδομένων	60
Κεφάλαιο 6 – Συμπεράσματα & Συζήτηση	63
Βιβλιογραφία - Αναφορές	67
ΠΑΡΑΡΤΗΜΑ	73

Πρόλογος

Η παρούσα διπλωματική εργασία εκπονήθηκε στο πλαίσιο του Διατμηματικού Προγράμματος Μεταπτυχιακών Σπουδών του Πανεπιστημίου Θεσσαλίας με τίτλο «Πληροφορική και Υπολογιστική Βιοϊατρική». Η εργασία πραγματοποιήθηκε υπό την επίβλεψη του κ. Μπάγκου Παντελή, Καθηγητή – Κοσμήτορα στο τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική του Πανεπιστημίου Θεσσαλίας. Κύριο αντικείμενο της εργασίας ήταν η εφαρμογή εργαλείων υπολογιστικής βιολογίας πάνω σε δεδομένα γονιδιακής έκφρασης με σκοπό την αναγνώριση νέων μοριακών μηχανισμών που εμπλέκονται στην έναρξη ή εξέλιξη της ασθένειας της πνευμονικής αρτηριακής υπέρτασης.

Στο σημείο αυτό θα ήθελα να ευχαριστήσω θερμά όλους όσους συνέβαλαν στην ολοκλήρωση της διπλωματικής μου εργασίας. Αρχικά οφείλω να εκφράσω τις θερμές μου ευχαριστίες στον κ. Μπάγκο Παντελή, με τον οποίο συνεργάζομαι για δεύτερη συνεχή φορά σε ακαδημαϊκό επίπεδο και μου έχει επιδείξει εμπιστοσύνη και καθοδήγηση. Θέλω ακόμη να ευχαριστήσω την κ. Μπράλιου Γεωργία, επίκουρη καθηγήτρια του Π.Θ. για τις στοχευμένες παρατηρήσεις της επί της εργασίας μου, καθώς και για τις αξιόλογες συμβουλές της. Ανάλογες θερμές ευχαριστίες θέλω να απευθύνω στην καθηγήτρια κ. Αρτέμιδα Χατζηγεωργίου, η οποία δέχτηκε να είναι μέλος της εξεταστικής μου επιτροπής και κριτική αναγνώστρια της διπλωματικής εργασίας μου.

Επίσης, θέλω να εκφράσω τις ευχαριστίες μου στην κ. Βέννου Κωνσταντίνα, ερευνήτρια και κάτοχο διδακτορικού διπλώματος του Π.Θ., για την πολύτιμη βοήθειά της σχετικά με το τεχνικό μέρος της εργασίας. Επιπλέον, επιθυμώ να ευχαριστήσω τη συνάδελφό μου κ. Ιωάννα Ραχμάνη, για τη φιλότιμη συνεισφορά της στην εργασία μου όποτε την χρειάστηκα.

Τέλος θέλω να ευχαριστήσω από καρδιάς την οικογένειά μου και τους φίλους μου για την ηθική συμπαράσταση και υποστήριξη που μου έδειξαν όλο αυτό το διάστημα.

Περίληψη

Η πνευμονική αρτηριακή υπέρταση είναι μια σπάνια καρδιοπνευμονική νόσος παγκοσμίως, η οποία επιδεινώνεται με την πάροδο του χρόνου και είναι απειλητική για τη ζωή. Τα αιμοφόρα αγγεία των πνευμόνων στενεύουν και μπλοκάρονται με αποτέλεσμα η καρδιά να αναγκάζεται να δουλέψει πιο σκληρά για να αντλήσει αίμα μέσω των πνευμόνων. Αν και μέχρι στιγμής δεν υπάρχει θεραπεία, έχουν γίνει προσπάθειες για τη διευκρίνιση της παθολογίας της νόσου. Ωστόσο, οι μοριακοί μηχανισμοί της παραμένουν ασαφείς.

Στην παρούσα εργασία χρησιμοποιήθηκαν πέντε μελέτες από πειράματα μικροσυστοιχιών DNA [GSE117261 GSE113439 GSE48149 GSE15197 GSE79786] μέσα από τη βάση δεδομένων Gene Expression Omnibus για την αποσαφήνιση των υποψήφιων γονιδίων που εμπλέκονται στην έναρξη και εξέλιξη της πνευμονικής αρτηριακής υπέρτασης. Η συγκέντρωση των συνόλων δεδομένων μικροσυστοιχιών έγινε με την τεχνική της συστηματικής ανασκόπησης σύμφωνα με το πρότυπο PRISMA.

Μετά από επεξεργασία των πινάκων γονιδιακής έκφρασης για ενημέρωση των στοιχείων μεταξύ ανιχνευτών – γονιδίων, πραγματοποιήθηκε μετα-ανάλυση βάση του μοντέλου τυχαίων επιδράσεων, εφαρμόζοντας ως μέγεθος επίδρασης την τυποποιημένη διαφορά μέσω των όρων του Cohen ακολουθούμενη από διόρθωση του Hedges. Η ανάλυση έγινε σε περιβάλλον STATA 10.1. Με χρήση της μεθόδου διόρθωσης FDR των Benjamini και Hochberg για έλεγχο σφαλμάτων τύπου I, ανακτήθηκαν 195 διαφορεικά εκφραζόμενα γονίδια (ΔΕΓ) σε ποσοστό σημαντικότητας 1%, τα οποία στη συνέχεια χρησιμοποιήθηκαν για ανάλυση εμπλουτισμού. Από τα 195 γονίδια, τα 149 γονίδια υποεκφράζονταν και τα υπόλοιπα 46 υπερεκφράζονταν.

Τα δεδομένα ΔΕΓ χρησιμοποιήθηκαν για την κατασκευή δικτύου αλληλεπίδρασης πρωτεΐνης – πρωτεΐνης και η ανάλυση πραγματοποιήθηκε με τη χρήση της βάσης STRING 11.0. Επιπρόσθετα, χρησιμοποιήθηκε η πλατφόρμα bioCompendium, μέσω της οποίας ανακτήθηκαν τα KEGG μονοπάτια στα οποία εμπλέκονται τα ΔΕΓ της μετα-ανάλυσης.

Τα κυριότερα βιοχημικά μονοπάτια KEGG που βρέθηκαν εμπλουτισμένα και μπορούν να συσχετιστούν με τη νόσο ΠΑΥ αφορούν το μονοπάτι μεταβολισμού του αραχιδονικού οξέος, το μονοπάτι υποδοχέων ανοσοσφαιρινών, το μονοπάτι διαλογής πρωτεϊνών στο λυσόσωμα και στη μεσολαβούμενη από κλαθρίνη ενδοκύττωση, το μονοπάτι συστολής του αγγειακού λείου μυός και η οδός σηματοδότησης του αγγειακού ενδοθηλιακού αυξητικού παράγοντα. Η ανάλυση εμπλουτισμού της STRING παρουσίασε μικρά υποδίκτυα πρωτεϊνών που συμμετέχουν σε μονοπάτια μεταγωγής σήματος μέσω ενδοκύττωσης με κυστίδια κλαθρίνης, σε μονοπάτια μεταγωγής σήματος μέσω πρωτεϊνικών υποδοχέων GPCRs, σε οδούς μεταγωγής σήματος μέσω των ερυθροκυττάρων και σε σηματοδότηση μέσω του συστήματος ουβικιτίνης-πρωτεασώματος.

Συμπερασματικά, τα ΔΕΓ που προσδιορίζονται στην παρούσα μελέτη μπορούν να χρησιμοποιηθούν για την καλύτερη κατανόηση των μοριακών μηχανισμών που εμπλέκονται στην έναρξη και κατά την εξέλιξη της νόσου ΠΑΥ. Επιπλέον, τα δεδομένα που προέκυψαν μπορούν χρησιμοποιηθούν στον εντοπισμό νέων βιοδεικτών, αλλά και στη διάγνωση και θεραπεία της πνευμονικής αρτηριακής υπέρτασης.

Abstract

Pulmonary arterial hypertension is a rare cardiopulmonary disease worldwide that gets worse through time and can be life-threatening. Blood vessels in the lungs are narrowed and blocked. As a result, the heart must work harder in order to pump blood through the lungs. Even though there is yet no cure, efforts have been made to elucidate the pathology of pulmonary arterial hypertension. However, its molecular mechanisms remain unclear.

In the present study, five microarray datasets [GSE117261 GSE113439 GSE48149 GSE15197 GSE79786] from Gene Expression Omnibus database were used for the elucidation of candidate genes involved in the initiation and progression of pulmonary arterial hypertension. The collection of microarray datasets was performed according to the PRISMA guidelines for Systematic Reviews.

After processing the gene expression matrices and mapping gene identifiers to probes, a random effects meta-analysis was performed. The standardized mean difference proposed by Cohen was used as the effect size, followed by Hedges' *g* correction. The analysis took place on a STATA 10.1 environment. To control the Type I errors, the FDR correction procedure by Benjamini and Hochberg was applied. One hundred and ninety-five (195) differentially expressed genes (DEGs) were identified at a 1% significance level, including 149 downregulated and 46 upregulated genes that were subsequently used for function enrichment analysis.

These DEGs data were utilized in the construction of a protein-protein interaction network using STRING 11.0 tool. In addition, bioCompendium platform was used for detecting KEGG enriched pathways in which these genes are involved.

The DEGs were found to be enriched in 11 KEGG pathways including arachidonic acid metabolism, immunoglobulin receptor pathway, lysosome process and clathrin-mediated endocytosis, vascular smooth muscle contraction and vascular endothelial growth factor mechanisms. On the other hand, STRING PPI network analysis revealed small subnetworks in which DEGs are involved in signal transduction through the clathrin-mediated endocytosis system, GPCR signaling pathways, signaling pathways of red blood cells and the ubiquitin-proteasome system.

In conclusion, the DEGs identified in the present study can be used to better understand the molecular mechanisms involved in the onset and progression of PAH disease. In addition, the resulting data can be used to identify potential biomarkers, but also in the diagnosis and treatment of pulmonary arterial hypertension.

Συντομογραφίες

ΠΥ	Πνευμονική Υπέρταση
ΠΑΥ	Πνευμονική Αρτηριακή Υπέρταση
ΡΑΗ	Pulmonary Arterial Hypertension
GEO	Gene Expression Omnibus Database
ΔΕΓ	Διαφορικά Εκφραζόμενα Γονίδια
DEG	Differentially Expressed Gene
FDR	False Discovery Rate
PPI Network	Protein-Protein Interaction Network
VEGF	Vascular Endothelial Growth Factor
GPCR	G-Protein Coupled Receptor
BMPR	Bone Morphogenetic Protein Receptor
GWAS	Genome Wide Association Study

Κεφάλαιο 1 - Εισαγωγή

1.1 Υπολογιστική Βιολογία

Με το ξεκίνημα του 21^{ου} αιώνα ολοκληρώνεται το έργο για την αλληλούχιση του ανθρώπινου γονιδιώματος και μαζί με αυτό ξεκινά μια νέα εποχή για την γονιδιωματική. Μεγάλο μέρος των μελετών αφιερώνονται στην μελέτη γονιδιακής έκφρασης από μικροσυστοιχίες (Schena et al., 1998). Παράλληλα, οι ερευνητές εστιάζουν το ενδιαφέρον τους στην πρωτεομική, η οποία αναφέρεται στο σύνολο των πρωτεϊνών που παράγονται από τα κύτταρα και τους ιστούς του ανθρώπινου οργανισμού (Tyers & Mann, 2003). Η επανάσταση στην ανάλυση μεταγραφώματος έρχεται με την τεχνική αλληλούχισης νέας γενιάς (RNA-seq) λίγο αργότερα, φέρνοντας καταγιστικούς ρυθμούς στη συγκέντρωση δεδομένων γονιδιακής έκφρασης και τη σύγκριση γονιδιωμάτων (Wang et al., 2009). Γρήγορο ενδιαφέρον αποκτά ταυτόχρονα το κεφάλαιο της προσωποποιημένης ιατρικής, η οποία υπόσχεται στοχευμένες θεραπείες, οι οποίες βασίζονται στο μοναδικό γενετικό αποτύπωμα (Vogelstein et al., 2010). Αυτή η εκθετική αύξηση δεδομένων έδωσε στη συνέχεια το έναυσμα για την ίδρυση εξειδικευμένων βάσεων δεδομένων και οντολογιών για την καλύτερη κατανομή της πληροφορίας (Μπάγκος, 2015).

Κατά συνέπεια, αφενός μεν η παραγωγή ψηφιακών δεδομένων στον τομέα της σύγχρονης βιολογίας πραγματοποιείται πλέον με ασύλληπτες ταχύτητες, αλλά η αξιολόγηση και σύγκριση ακόμη και ενός μικρού συνόλου δεδομένων σε σχέση με ένα μεγαλύτερο σύνολο απαιτεί τη χρήση γρήγορων και αποτελεσματικών υπολογιστικών μεθόδων. Για παράδειγμα, η αναζήτηση στοίχισης ακολουθιών σε βάσεις δεδομένων για εύρεση ομοιοτήτων μεταξύ αλληλουχιών πυροδότησε το ενδιαφέρον στις αρχές της δεκαετίας του '90 για τη δημιουργία των πρώτων αλγορίθμων επίλυσης βιολογικού προβλήματος με χρήση στατιστικής κατανομής (Altschul et al., 1990, Can, 2014).

Γίνεται αντιληπτό ότι, η ανάθεση πολύπλοκων βιολογικών ερωτημάτων χρήζουν επίλυσης μέσω διεπιστημονικών κλάδων πέραν των μεθοδολογιών που προσφέρει η βιολογία. Ο τομέας που αναπτύσσεται ραγδαίως την τελευταία εικοσαετία και έχει συνδεθεί με αυτό του είδους τα βιολογικά προβλήματα χαρακτηρίζεται ως «Υπολογιστική Βιολογία» (Νικολάου & Χουβαρδάς, 2015).

Η υπολογιστική βιολογία δύναται να αποδοθεί ως το διεπιστημονικό γνωστικό αντικείμενο που συνδυάζει την επιστήμη της βιολογίας, την επιστήμη των υπολογιστών, την πληροφορική, τα μαθηματικά, τη στατιστική (Luscombe et al., 2001) και αρκετούς ακόμη κλάδους (Anandaram, 2017) με στόχο την ανάλυση ποικίλων βιολογικών δεδομένων και την οργάνωση πληροφορίας σε μεγάλη κλίμακα.

1.2 Πνευμονική Υπέρταση

1.2.1 Διαγνωστικά και Επιδημιολογικά Χαρακτηριστικά της Πνευμονικής Υπέρτασης

Η πνευμονική υπέρταση είναι μια νόσος που διακρίνεται από υψηλή αρτηριακή πίεση στους πνεύμονες. Επηρεάζει περίπου το 1% του παγκόσμιου πληθυσμού και είναι πιο συχνή σε ηλικίες άνω των 65 ετών. Χαρακτηρίζεται αιμοδυναμικά από μέση πνευμονική αρτηριακή πίεση (μΠΑΠ) μεγαλύτερη ή ίση των 25 χιλιοστών στήλης υδραργύρου ($\mu\text{ΠΑΠ} \geq 25\text{mmHg}$) σε κατάσταση ηρεμίας μετρημένη με δεξιό καρδιακό καθετηριασμό, ενώ η πίεση στα πνευμονικά τριχοειδή ή τον αριστερό κόλπο παραμένει μικρότερη από 15mmHg. Διάφορα δεδομένα υποστηρίζουν ότι η φυσιολογική μΠΑΠ κατά την ανάπαυση κυμαίνεται $14 \pm 3\text{mmHg}$. Το ανώτερο όριο φυσιολογικής μΠΑΠ δεν πρέπει να θεωρείται αυθαίρετα ότι είναι έως 24 mmHg, καθώς παραμένει ένα ζήτημα υπό συζήτηση (Simonneau et al., 2019).

1.2.2 Οι Κύριες Ομάδες της Πνευμονικής Υπέρτασης

Ο Παγκόσμιος Οργανισμός Υγείας έχει κατηγοριοποιήσει την πνευμονική υπέρταση σε πέντε κύριες ομάδες κλινικών καταστάσεων. Τα κριτήρια κατηγοριοποίησης της κάθε ομάδας πνευμονικής υπέρτασης αφορούν τους παθοφυσιολογικούς μηχανισμούς που προκαλούν την εκάστοτε ασθένεια, την κλινική παρουσίαση, τα αιμοδυναμικά χαρακτηριστικά των ασθενών της ομάδας και τη φαρμακευτική αγωγή που χορηγείται στους ασθενείς (Ryan et al., 2012). Είναι σημαντικό ανάλογα με τα συμπτώματα που έχει ο ασθενής, να συγκαταλέγεται η κλινική περίπτωση του στην κατάλληλη ομάδα πνευμονικής υπέρτασης έτσι ώστε να λάβει τις απαραίτητες εξετάσεις και έγκαιρη αντιμετώπιση της ασθένειάς του (Hoepfer et al., 2017).

Οι επικαιροποιημένες κατηγορίες πνευμονικές υπέρτασης σύμφωνα με τον ΠΟΥ είναι:

Ομάδα Α: Πνευμονική Αρτηριακή Υπέρταση (*Pulmonary Arterial Hypertension*)

Σε αυτήν την ομάδα συγκαταλέγονται η ιδιοπαθής ΠΑΥ, η κληρονομήσιμη ΠΑΥ, η ΠΑΥ που οφείλεται σε παθήσεις του συνδετικού ιστού, η ΠΑΥ λόγω χρήσης ανορεξιογόνων ή τοξικών ουσιών, η ΠΑΥ που οφείλεται σε HIV, κ.ά.

Ομάδα Β: Πνευμονική Υπέρταση οφειλόμενη σε παθήσεις της αριστερής πλευράς της καρδιάς (*Pulmonary Hypertension due to Left Heart Disease*)

Είναι η πιο κοινή μορφή ΠΥ. Η αριστερή κοιλία της καρδιάς δυσλειτουργεί και επιστρέφει αίμα πίσω στους πνεύμονες, προκαλώντας αύξηση πνευμονικής πίεσης.

Ομάδα Γ: Πνευμονική Υπέρταση οφειλόμενη σε παθήσεις των πνευμόνων (*Pulmonary Hypertension due to Lung Disease*)

Περιλαμβάνει ΠΥ λόγω χρόνιας πνευμονικής νόσου ή/ και υποξίας (χαμηλά επίπεδα οξυγόνου). Συμπεριλαμβάνονται στην συγκεκριμένη ομάδα η νόσος ΧΑΠ (Χρόνια Αποφρακτική Πνευμονοπάθεια/ COPD) και το Εμφύσημα (Emphysema).

Ομάδα Δ: Πνευμονική Υπέρταση λόγω χρόνιων θρόμβων αίματος στους πνεύμονες (*Pulmonary Hypertension due to Chronic Blood Clots in the Lungs*)

Σε αυτή την ομάδα συγκαταλέγεται η Χρόνια Θρομβοεμβολική Νόσος (Chronic Thromboembolic Pulmonary Hypertension/ CTEPH).

Ομάδα Ε: Πνευμονική Υπέρταση οφειλόμενη σε άγνωστες αιτίες (*Pulmonary Hypertension Due to Unknown Causes*)

Σε αυτή την ομάδα η ΠΥ εμφανίζεται με δευτερογενή ρόλο χωρίς να είναι κατανοητά τα αίτια, σε ασθένειες τύπου αιματολογικών νόσων όπως η δρεπανοκυτταρική αναιμία, σε συγγενείς καρδιοπάθειες και σε ορισμένες μεταβολικές διαταραχές.

1.2.3 Τα Βασικά Συμπτώματα της Πνευμονικής Υπέρτασης

Τα κύρια συμπτώματα Πνευμονικής Υπέρτασης είναι:

- ❖ Προοδευτική δύσπνοια
- ❖ Κόπωση
- ❖ Εξάντληση
- ❖ Ζάλη ή τάση για λιποθυμία
- ❖ Στηθάγχη
- ❖ Πρήξιμο στα πόδια και στην κοιλιακή χώρα
- ❖ Κυάνωση στο δέρμα και στα χείλη
- ❖ Αίσθημα ταχυπαλμίας ή ταχυκαρδίας

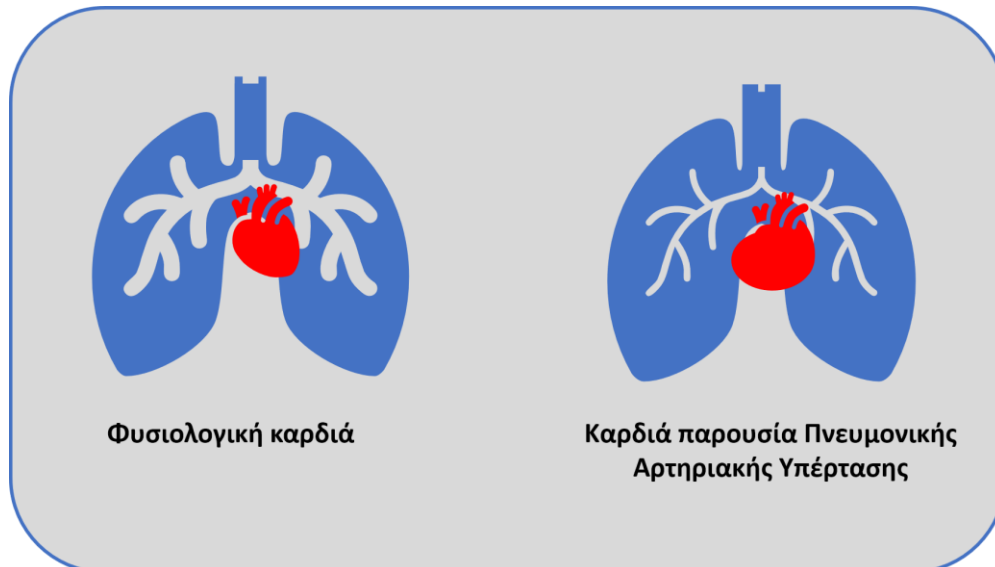
1.3 Πνευμονική Αρτηριακή Υπέρταση

1.3.1 Διαγνωστικά και Επιδημιολογικά Χαρακτηριστικά της Πνευμονικής Αρτηριακής Υπέρτασης

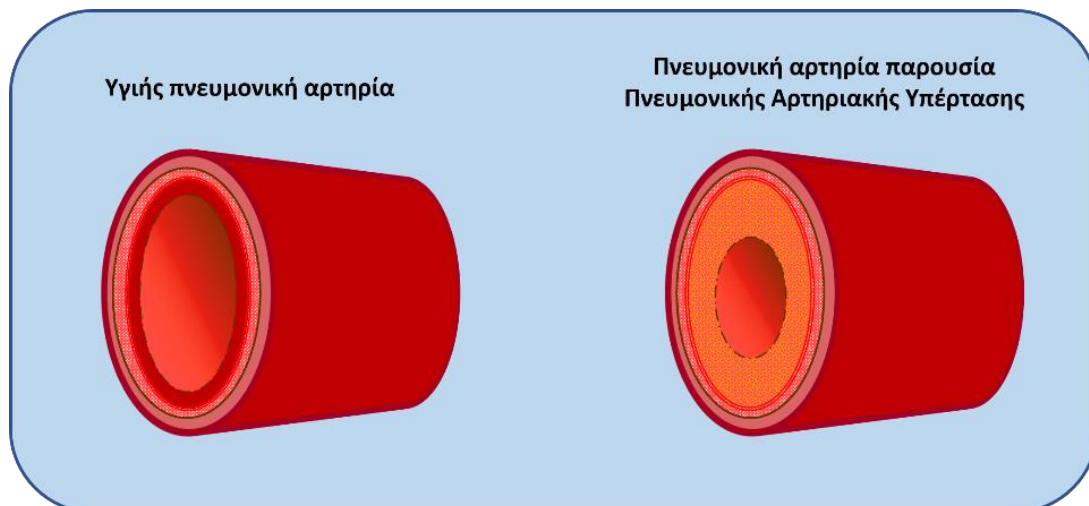
Η Πνευμονική Αρτηριακή Υπέρταση αποτελεί την πρώτη ομάδα Πνευμονικής Υπέρτασης. Είναι μια σοβαρή νόσος του πνευμονικού αγγειακού συστήματος, για την οποία, δυστυχώς, δεν υπάρχει θεραπεία. Επηρεάζει τόσο το κυκλοφορικό σύστημα των πνευμονικών αρτηριών, όσο και την ορθή λειτουργία της δεξιάς κοιλίας της καρδιάς με αποτέλεσμα να αυξάνεται σταδιακά η πίεση των πνευμονικών αρτηριών και η πνευμονική αγγειακή αντίσταση. Προσβάλλει τις αρτηρίες των πνευμόνων προκαλώντας σταδιακή στένωση έως και απόφραξη. Τα συμπτώματα της ΠΑΥ οδηγούν τους ασθενείς σε προοδευτική ανεπάρκεια της δεξιάς κοιλίας της καρδιάς, η οποία είναι και η κύρια αιτία θανάτου της ΠΑΥ. Ορίζεται αιμοδυναμικά ως (Prins & Thenappan, 2016):

- η παρατεταμένη αύξηση της μέσης πνευμονικής αρτηριακής πίεσης > 25mmHg στην ηρεμία και μεγαλύτερη από 30mmHg κατά την άσκηση
- πίεση ενσφήνωσης πνευμονικών τριχοειδών \leq 15mmHg
- πνευμονικές αγγειακές αντιστάσεις > 3Wood Units

Η συχνότητα εμφάνισης της νόσου κυμαίνεται από 2.0 έως 7.6 περιπτώσεις ανά εκατομμύρια ενήλικες ετησίως, ενώ ο επιπολασμός από 11 έως 26 περιπτώσεις ανά εκατομμύρια ενήλικες. Επίσης, η συχνότητα εμφάνισης της ΠΑΥ στις γυναίκες είναι τέσσερις φορές μεγαλύτερη απ' ότι στους άντρες (Thenappan et al., 2018).



Εικόνα 1.2 (αριστερά) Ποιοτική αναπαράσταση για τη φυσιολογική κατάσταση της καρδιάς. (δεξιά) Η εξαιρετικά υψηλή πίεση στις αρτηρίες από την καρδιά προς τους πνεύμονες προκαλεί σταδιακά τη διακοπή της φυσιολογικής λειτουργίας στη δεξιά πλευρά της καρδιάς.



Εικόνα 1.1 (αριστερά) Ποιοτική αναπαράσταση της φυσιολογικής κατάστασης των πνευμονικών αιμοφόρων αγγείων. (δεξιά) Η στένωση των αιμοφόρων αγγείων των πνευμόνων επηρεάζει τη φυσιολογική κυκλοφορία του αίματος στους πνεύμονες.

1.3.2 Μορφές Πνευμονικής Αρτηριακής Υπέρτασης

Υπάρχουν διαφορετικοί τύποι πνευμονικής αρτηριακής υπέρτασης. Οι κυριότεροι είναι:

1. Ιδιοπαθής ΠΑΥ

Η ιδιοπαθής πνευμονική αρτηριακή υπέρταση προκαλείται χωρίς να είναι γνωστά τα αίτια εμφάνισής της νόσου.

2. Κληρονομήσιμη ΠΑΥ

Η κληρονομήσιμη πνευμονική αρτηριακή υπέρταση οφείλεται σε γενετικούς παράγοντες. Συγκεκριμένα, κάποια μετάλλαξη σε ένα γονίδιο είναι υπεύθυνη για την εμφάνιση της νόσου σε έναν ασθενή και το πρόβλημα να είναι κληρονομικής φύσεως. Οι πιο γνωστές γενετικές μεταλλάξεις που έχουν συνδεθεί κατά καιρούς με την κληρονομήσιμη ΠΑΥ αναφέρονται στο γονίδιο BMPR2. Οι ερευνητές έχουν διαπιστώσει ότι το γονίδιο BMPR2 (Bone Morphogenetic Protein Receptor 2) είναι υπεύθυνο για τη σύνθεση διαμεμβρανικών πρωτεϊνών υπεύθυνες για τη δομή αγγειακών κυττάρων λείων μυών και συνάμα ρυθμίζει τον αριθμό των κυττάρων σε ιστούς (Lane et al., 2000). Η λειτουργίες παρόμοιων και μη γονιδίων όπως αυτή του BMPR2, τα καθιστούν σημαντικό παράγοντα ως προς την αντιστροφή της νόσου για την βελτίωση της υγείας των ανθρώπων με ΠΑΥ (Austin & Loyd, 2014, Morrell et al., 2019).

3. ΠΑΥ σχετιζόμενη με άλλες ασθένειες

- αγγειακή νόσος του συνδετικού ιστού
- συγγενής καρδιοπάθεια
- πυλαία υπέρταση
- λοίμωξη από HIV
- φάρμακα/τοξίνες
- σχιστοσωμίαση

1.3.3 Συμπτώματα Πνευμονικής Αρτηριακής Υπέρτασης

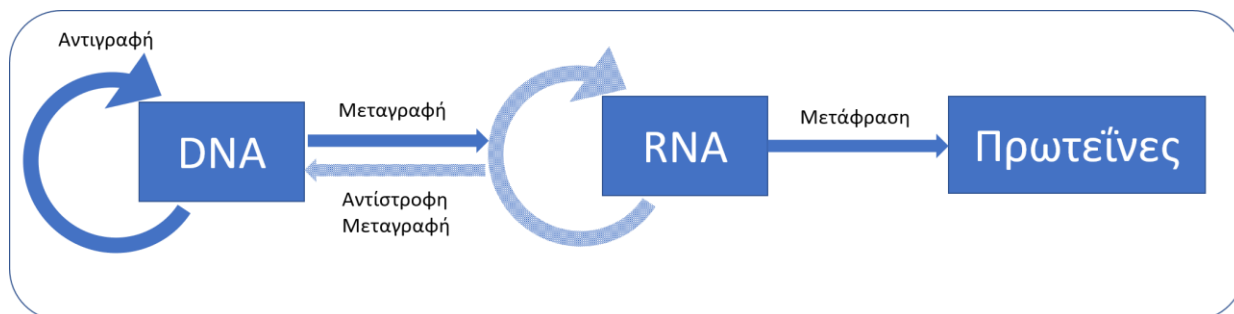
Τα κυριότερα συμπτώματα πνευμονικής αρτηριακής πίεσης είναι τα ακόλουθα:

- ❖ προοδευτικά επιδεινούμενη δύσπνοια
- ❖ κόπωση
- ❖ ζάλη
- ❖ βήχας
- ❖ πόνος στο θώρακα
- ❖ οίδημα στα άκρα
- ❖ κυάνωση
- ❖ δεξιά καρδιακή ανεπάρκεια σε συνδυασμό με χαμηλή αρτηριακή πίεση

Κεφάλαιο 2 – Μικροσυστοιχίες DNA

2.1 Το Κεντρικό Δόγμα της Βιολογίας

Σε όλα τα κύτταρα των ζωντανών ευκαρυωτικών οργανισμών περιέχεται η γενετική πληροφορία, η οποία είναι αποτυπωμένη σε συγκεκριμένα τμήματα του DNA. Τα τμήματα αυτά χαρακτηρίζονται ως γονίδια. Η γενετική πληροφορία μεταβιβάζεται στα θυγατρικά κύτταρα μέσα από μια καθορισμένη σειρά λειτουργιών. Το κεντρικό δόγμα της βιολογίας περιγράφει την κατεύθυνση μέσα από την οποία το DNA «κυλάει» προς τις πρωτεΐνες. Για την μεταβίβαση της γενετικής πληροφορίας απαραίτητη προϋπόθεση είναι ο αναδιπλασιασμός του DNA που βρίσκεται στο εσωτερικό του πυρήνα κάθε κυττάρου του οργανισμού. Η αντιγραφή του DNA επιτυγχάνεται σύμφωνα με την αρχή της συμπληρωματικότητας των βάσεων με τη βοήθεια του ενζύμου DNA πολυμεράση. Στη συνέχεια παρουσιάζεται η λειτουργία της μεταγραφής, όπου στα τμήματα του DNA που περιέχεται η γενετική πληροφορία ανοίγει η διπλή έλικα και αρχίζει η σύνθεση μονόκλωνων μορίων RNA με τη βοήθεια του ενζύμου RNA πολυμεράση. Το αγγελιαφόρο RNA (mRNA) είναι ένα από τα είδη μορίων RNA που παράγονται με τη λειτουργία της μεταγραφής και φέρει την πληροφορία για τη σύνθεση πρωτεϊνών που αποτελεί και το τελευταίο στάδιο της ροής της γενετικής πληροφορίας, την μετάφραση. Πλέον, το τμήμα mRNA που παράχθηκε από τη μεταγραφή μεταφράζεται σε μια πολυπεπτιδική αλυσίδα μέσα από διαδοχικά στάδια (Αλεπόρου-Μαρίνου και συν., 2012) (Καψάλης και συν., 2013).



Εικόνα 2.1 Το κεντρικό δόγμα της βιολογίας.

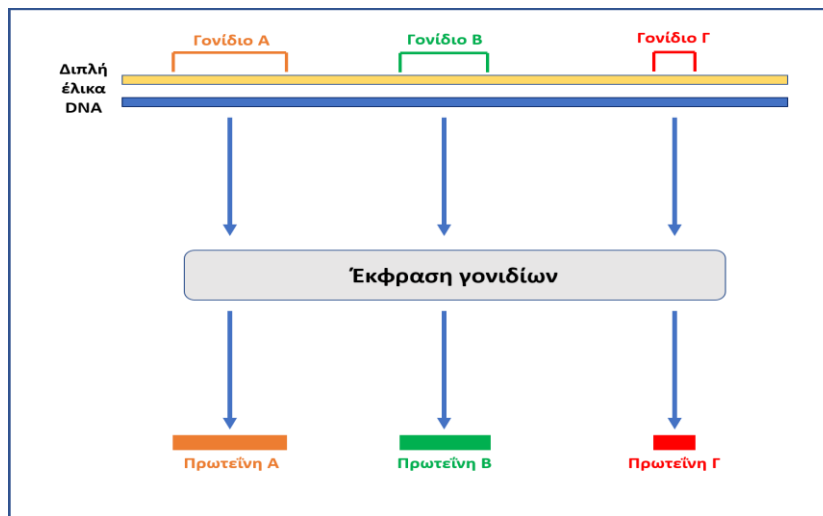
2.2 Γονιδιακή Έκφραση

Γίνεται γνωστό σε αυτό το σημείο ότι τα γονίδια είναι τα νουκλεοτιδικά μόρια που φέρουν την γενετική πληροφορία και είναι υπεύθυνα για την σύνθεση πρωτεϊνών. Οι πρωτεΐνες στη συνέχεια αναλαμβάνουν να επιτελέσουν λειτουργίες για την ομαλή λειτουργία του οργανισμού. Οι τρεις κύριες συνιστώσες, DNA, RNA και πρωτεΐνες, αποτελούν τα δομικά συστατικά για τη μελέτη γενετικής πληροφορίας που υπάγεται σε ένα οργανισμό.

Κάθε κύτταρο του ανθρώπινου οργανισμού έχει διαφορετικά επίπεδα έκφρασης γονιδίων (Εικόνα 2.2) (Alberts, 2002). Για παράδειγμα, γνωρίζουμε ότι στα κύτταρα του παγκρέατος παρουσιάζεται σε μεγάλη έκφραση το γονίδιο που συνθέτει την ινσουλίνη.

Γνωρίζουμε επίσης ότι σημαντικές αυξομειώσεις στην έκφραση γονιδίων έχουν συσχετιστεί με την παρουσία παθολογικών καταστάσεων (Schadt et al., 2005, Heller et al., 1997). Η ρύθμιση της γονιδιακής έκφρασης, δηλαδή ποια γονίδια υπερεκφράζονται και ποια όχι σε ορισμένους ιστούς ή καταστάσεις, αποτελούν μία από τις σημαντικότερες διερευνήσεις της σύγχρονης βιολογίας για την βελτίωση της ανθρώπινης υγείας (Lesk, 2019).

Μέχρι πρότινος, ο αριθμός των γονιδίων που μεταφράζεται σε πρωτεΐνες υπολογιζόταν ότι είναι περίπου 22.000 (Lander et al., 2001) και αντιπροσωπεύουν μόλις το 1% του ανθρώπινου γονιδιώματος. Γρήγορα έγινε αντιληπτό ότι η αναπτυξιακή και φυσιολογική πολυπλοκότητα του ανθρώπινου οργανισμού δεν θα μπορούσε να εξηγηθεί αποκλειστικά από τον αριθμό των γονιδίων που κωδικοποιούν πρωτεΐνες. Οι επιστήμονες ανακάλυψαν ότι περισσότερο από το 80% αυτού του μη γονιδιακού συστατικού του γονιδιώματος, το οποίο κάποτε θεωρήθηκε «ανεπιθύμητο DNA», στην πραγματικότητα έχει ρόλο στη ρύθμιση της δραστηριότητας συγκεκριμένων γονιδίων (Dogini et al., 2014). Έτσι λοιπόν, η προσπάθεια κατανόησης του υπόλοιπου 99% του ανθρώπινου γονιδιώματος συμβάλλει σημαντικά στην ανακάλυψη νέων εναλλακτικών λύσεων ως προς τη θεραπεία ανθρώπινων ασθενειών.



Εικόνα 2.2 Σχηματική ποιοτική απεικόνιση της γονιδιακής έκφρασης.

2.3 Οι Μικροσυστοιχίες DNA

Ο κλάδος της γονιδιωματικής που ασχολείται με την γονιδιακή έκφραση επιτρέπει στους ερευνητές να χρησιμοποιούν ειδικές βιοτεχνολογικές συσκευές, τις DNA microarrays ή DNA chips (μικροσυστοιχίες DNA) για την ταυτόχρονη αλληλούχιση πολλών ακολουθιών DNA (Knudsen, 2011). Από την εισαγωγή τους περίπου πριν από μια εικοσαετία (Schena et al., 1995), η τεχνολογία των μικροσυστοιχιών έχει βρει ευρεία χρήση στους τομείς της μοριακής γενετικής και της λειτουργικής γονιδιωματικής. Η εφαρμογή τους έχει βοηθήσει στην κατανόηση υποκείμενων βιολογικών μηχανισμών, στην αναγνώριση νέων γονιδίων που σχετίζονται με ασθένειες, στον έλεγχο για την ανταπόκριση σε φάρμακα από ομάδες ασθενών και στην πρόβλεψη των αποτελεσμάτων μιας νόσου. Πρόκειται για ένα πολύ μικρό γυάλινο πλακίδιο των 1,28cm² με περίπου 20.000 κηλίδες (spot) στις οποίες ακινητοποιούνται ανιχνευτές (probes).

Όπως αναφέρθηκε προηγουμένως, ένα γονίδιο εκφράζεται με το να μεταγραφεί από DNA σε μονόκλωνο mRNA και τέλος να μεταφραστεί σε πρωτεΐνη (κεντρικό δόγμα). Η έκφραση mRNA αντιπροσωπεύει τις δυναμικές πτυχές των κυττάρων. Για αυτό το λόγο χρησιμοποιείται το mRNA και όχι το DNA σαν δείκτης έκφρασης. Η τεχνολογία της DNA μικροσυστοιχίας επιτρέπει την μέτρηση της έκφρασης mRNA με έναν τρόπο υψηλής απόδοσης.

2.3.1 Τύποι Μικροσυστοιχιών

Οι μικροσυστοιχίες μπορούν να κατασκευαστούν χρησιμοποιώντας διάφορες τεχνολογίες με κύριες εκπροσώπους πλατφορμών μικροσυστοιχιών τις εταιρείες Affymetrix, Agilent και Illumina.

- Στις *επισημασμένες μικροσυστοιχίες (spotted microarrays)*, οι ανιχνευτές οι οποίοι είναι ολόκληροι κλώνοι mRNAs που αντιστοιχούν σε cDNAs, «επισημαίνονται» στην επιφάνεια του πλακιδίου. Το πλεονέκτημα των συγκεκριμένων μικροσυστοιχιών είναι ότι μπορούν να ανιχνεύσουν άγνωστες αλληλουχίες οργανισμών για τους οποίους υπάρχει περιορισμένη πληροφορία ως προς το γονιδίωμα. Οι επισημασμένες μικροσυστοιχίες μπορούν να κατασκευαστούν κατά την επιλογή του εκάστοτε ερευνητή.
- Στις *μικροσυστοιχίες ολιγονουκλεοτιδίων (oligonucleotide microarrays)* οι ανιχνευτές είναι σύντομες αλληλουχίες των 20 ή 70 νουκλεοτιδίων οι οποίες είναι σχεδιασμένες να είναι συμπληρωματικές σε τμήματα της αλληλουχίας στόχου, έτσι ώστε ένα γονίδιο να αντιπροσωπεύεται από ένα σύνολο ανιχνευτών αντί για ένα μόνο ανιχνευτή.
- Οι *μικροσυστοιχίες σύνθεσης in situ (in situ oligonucleotide arrays)* έχουν μικρά τμήματα 25-μερών ολιγονουκλεοτιδικών ανιχνευτών οι οποίοι κατασκευάζονται πάνω στην γυάλινη επιφάνεια της μικροσυστοιχίας με χρήση της τεχνολογίας της φωτολιθογραφίας σε συνδυασμό με ολιγονουκλεοτιδική χημεία στερεάς φάσης. Η τεχνολογία αυτή σε σύγκριση με τις προηγούμενες δύο είναι ότι προσφέρει πολύ μεγαλύτερο εύρος ανιχνευτών (~50.000 ανιχνευτές) σε ένα μόλις τσιπ μικροσυστοιχίας.
- Οι *μικροσυστοιχίες δύο καναλιών (dual channel)* ή δύο χρωμάτων τυπικά υβριδοποιούνται με cDNA που παρασκευάζεται από δύο δείγματα προς σύγκριση (ασθενής ιστός/υγιής

ιστός). Τα δείγματα επισημαίνονται με δύο διαφορετικές φθορίζουσες ουσίες (Cy3 και Cy5) οι οποίες εκπέμπουν διαφορετικές εντάσεις σήματος.

- Στις μονοκάναλες μικροσυστοιχίες (*single channel*) τα δείγματα προς σύγκριση επισημαίνονται με μία φθορίζουσα χρωστική. Οι σχετικές εντάσεις σήματος για κάθε ανιχνευτή ή σύνολο ανιχνευτών αντικατοπτρίζουν το επίπεδο έκφρασης της επισημασμένης αλληλουχίας στόχου.

Η επιλογή της κατάλληλης πλατφόρμας μικροσυστοιχίας πριν το ξεκίνημα ενός πειράματος εξαρτάται από το κόστος κατασκευής, τη διαθεσιμότητα για τα υπό ανάλυση είδη, την κάλυψη που προσφέρει ως προς το γονιδίωμα του οργανισμού που πρόκειται να εξετασθεί, την εγκυρότητα και διαθεσιμότητα εργαλείων λογισμικού για ανάλυση εικόνας, κ.ά.

2.3.2 Διαδικασία Πειράματος Μικροσυστοιχιών

Έχοντας ως γνώμονα τη *διατύπωση του βιολογικού ερωτήματος* ξεκινά η πειραματική διαδικασία. Το ερώτημα μπορεί να αφορά, λόγου χάριν, τη διαφοροποίηση της γονιδιακής έκφρασης σε δύο ομάδες πληθυσμού πριν και μετά τη χορήγηση ενός φαρμάκου.

Αρχικά το mRNA απομονώνεται από τους αρχικούς ιστούς ή κύτταρα αίματος των ομάδων που συμμετέχουν στο πείραμα. Κατόπιν, πραγματοποιείται αντίστροφη μεταγραφή από το δείγμα mRNA που προέκυψε σε συμπληρωματικά τμήματα cDNA (complementary DNA) και γίνεται η επισήμανσή τους με φθορίζουσα χρωστική. Κάθε πλακίδιο μικροσυστοιχίας έχει πάνω στην επιφάνειά της χιλιάδες ανιχνευτές (probes), οι οποίοι αντικατοπτρίζουν συμπληρωματικές αλληλουχίες γνωστών γονιδίων ή τμημάτων DNA. Τα επισημασμένα τμήματα cDNA προσδένονται πάνω στους ανιχνευτές, ενώ όσα δεν καταφέρνουν να προσδεθούν, απομακρύνονται (*υβριδοποίηση*).

Με τη χρήση δέσμης φωτός laser αναγνωρίζονται οι υβριδοποιημένες αλληλουχίες. Ανάλογα με την ένταση φωτός και με το χρώμα των ανιχνευτών, ποσοτικοποιείται η γονιδιακή έκφραση (*σάρωση*). Όταν το χρώμα είναι έντονα πράσινο σημαίνει ότι το δείγμα cDNA στο συγκεκριμένο ανιχνευτή υβριδοποιήθηκε σε μεγάλη ποσότητα στην κατάσταση ελέγχου (υγιής κατάσταση), ενώ αν είναι κόκκινο υβριδοποιήθηκε σε μεγάλη ποσότητα στην κατάσταση ενδιαφέροντος (κατάσταση νόσου). Αν είναι κίτρινο, υβριδοποιήθηκε σε ίση κατανομή και στις δύο καταστάσεις. Τέλος, αν το χρώμα κατά την οπτικοποίηση της μικροσυστοιχίας είναι μαύρο στα σημεία των ανιχνευτών, τότε δεν έχει υβριδοποιηθεί δείγμα cDNA (Εικόνα 2.3).

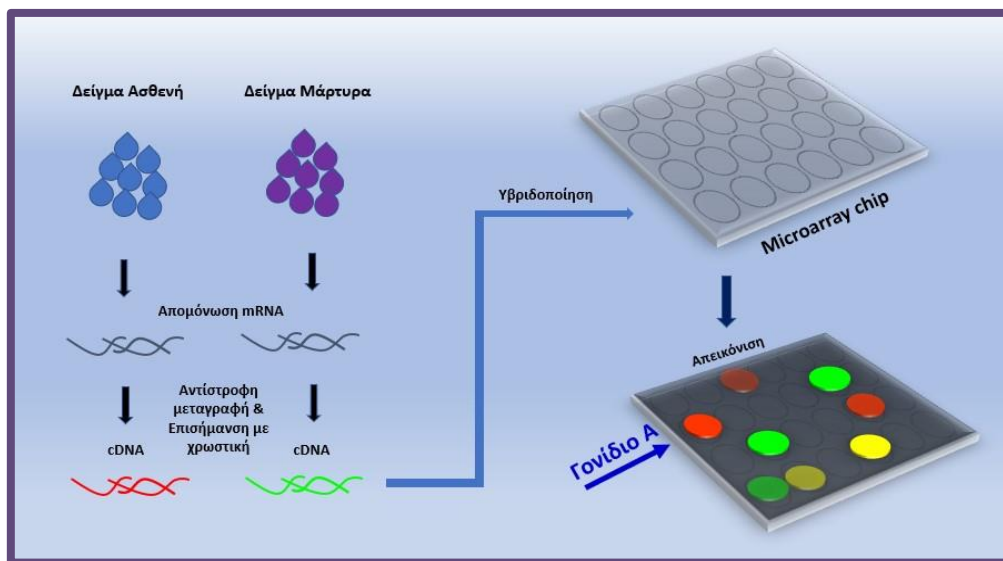
Ύστερα από την σάρωση, προκύπτει η τελική εικόνα μετά από συνένωση των δύο πινάκων χαρτογράφησης για τα δύο διαφορετικά χρώματα φθορισμού τα οποία αντιστοιχούν στα δείγματα αναφοράς και ελέγχου αντίστοιχα. Η επεξεργασία γίνεται με ειδικά προγράμματα και στοχεύει στο φιλτράρισμα για την εξάλειψη θορύβου και για διόρθωση σήματος της μικροσυστοιχίας.

Αμέσως μετά ακολουθεί η *κανονικοποίηση (normalization)* των δεδομένων (Quackenbush, 2002) η οποία εκτελείται για να προσαρμόζονται τα τελικά δεδομένα της μικροσυστοιχίας σε τυχόν διαφοροποιήσεις εξαιτίας των διαφορετικών κατασκευαστικών τεχνολογιών ή σε τυχόν λάθη που προέκυψαν κατά τη διαδικασία του πειράματος. Με την

κανονικοποίηση εξισορροπούνται τα δεδομένα μικροσυστοιχιών και ελαχιστοποιούνται τα συστηματικά σφάλματα (bias). Με αυτόν τον τρόπο καθίσταται εφικτή η σύγκριση δεδομένων από διαφορετικές πλατφόρμες μικροσυστοιχιών. Οι πιο γνωστές μέθοδοι για κανονικοποίηση δεδομένων γονιδιακής έκφρασης είναι:

- ❖ Ολική κανονικοποίηση γύρω από σταθερή τιμή
- ❖ Τυποποίηση με z-score (z-standardization)
- ❖ Κανονικοποίηση ποσοστημορίων (quantile normalization)
- ❖ Τοπική κανονικοποίηση μέσω σταθμισμένης εξομάλυνσης. (Locally weighted scatterplot smoothing, LOESS)

Απόρροια της προεπεξεργασίας των εικόνων και της κανονικοποίησης των δεδομένων της μικροσυστοιχίας είναι ο *πίνακας γονιδιακής έκφρασης*, ο οποίος περιέχει τιμές έκφρασης για κάθε ανιχνευτή. Κατόπιν, ξεκινά η ανάλυση του πίνακα γονιδιακής έκφρασης με στόχο την ταυτοποίηση των διαφορικά εκφραζόμενων γονιδίων ανάμεσα στις κλάσεις που έχουν οριστεί (π.χ. ασθενείς/μάρτυρες). Κατόπιν γίνεται ομαδοποίηση των γονιδίων με βάση τα επίπεδα έκφρασής τους. Τελευταίο στάδιο είναι η εφαρμογή προγνωστικών δεικτών για μια κατάσταση (λ.χ. ασθένεια) που εξετάζεται και για την αποφυγή της.

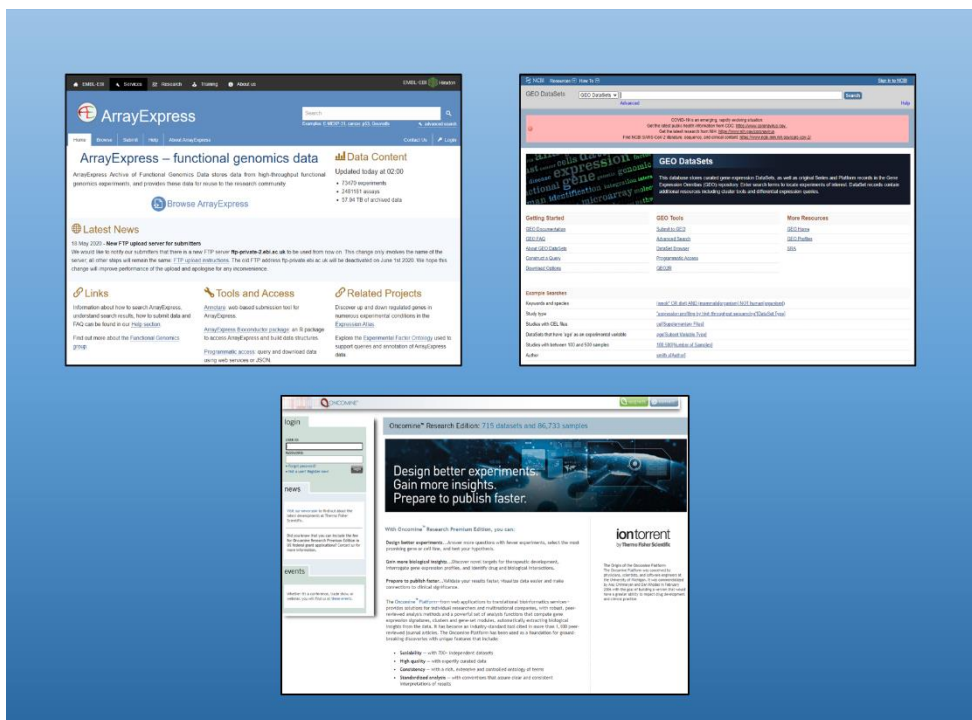


Εικόνα 2.3 Ποιοτική αναπαράσταση του διαγράμματος ροής ενός πειράματος cDNA μικροσυστοιχίας.

2.3.3 Βάσεις Δεδομένων Μικροσυστοιχιών

Υπάρχουν τρεις κύριες βάσεις για αναζήτηση πάνω σε δεδομένα μικροσυστοιχιών. Αυτές είναι:

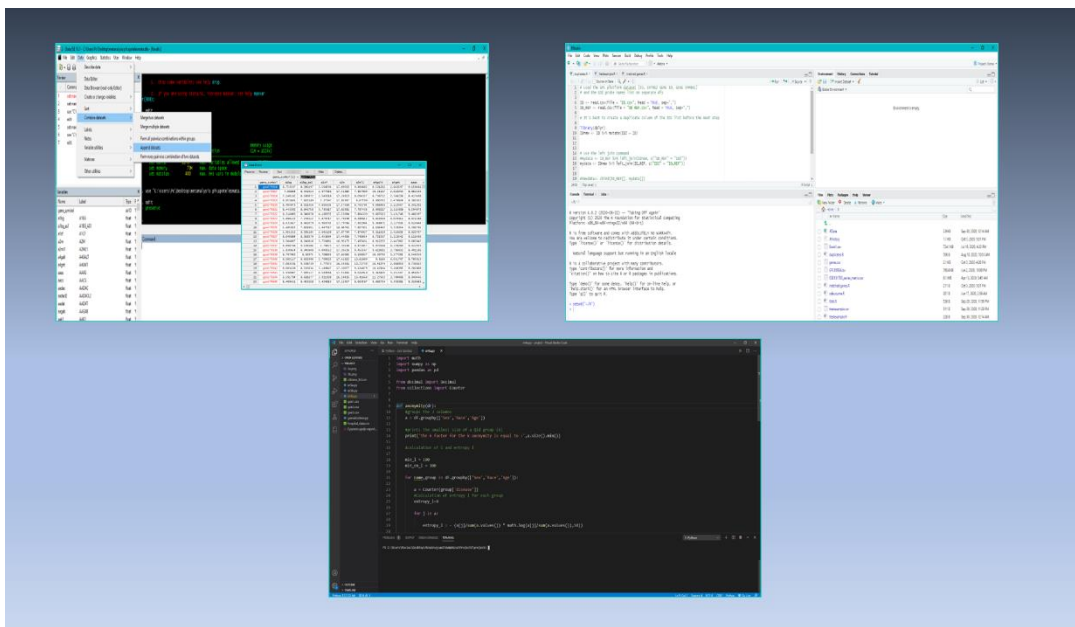
- ❖ **Gene Expression Omnibus (GEO):** Βάση δεδομένων του NCBI που παρέχει δεδομένα γονιδιακής έκφρασης.
<http://www.ncbi.nlm.nih.gov/geo/>
- ❖ **Array Express:** Δημόσια βάση δεδομένων μικροσυστοιχιών η οποία διατηρείται στο Ευρωπαϊκό Ινστιτούτο Βιοπληροφορικής EBI.
<http://www.ebi.ac.uk/arrayexpress/>
- ❖ **ONCOMINE:** Βάση δεδομένων που περιέχει πειράματα μικροσυστοιχιών που αφορούν διάφορους τύπους καρκίνου. Επίσης παρέχει στο χρήστη εργαλεία διαχείρισης των δεδομένων για την αποδοτικότερη εύρεση των επιθυμητών πειραμάτων και γονιδίων.
<http://www.oncomine.org/>



Εικόνα 2.4 Βάσεις δεδομένων που παρέχουν δεδομένα από πειράματα μικροσυστοιχιών.

2.3.4 Εργαλεία Λογισμικού για Ανάλυση Δεδομένων Γονιδιακής Έκφρασης

- ♦ Το *STATA* είναι ένα πακέτο λογισμικού και χρησιμοποιείται με αγορά άδειας χρήσης. Αν και το συγκεκριμένο λογισμικό προσφέρεται κυρίως για έτοιμες στατιστικές μεθόδους, δίνει επίσης τη δυνατότητα στο χρήστη να κατασκευάσει νέες συναρτήσεις και να εφαρμόσει στα σετ δεδομένων προς ανάλυση τη μεθοδολογία που επιθυμεί. Ο χρήστης μπορεί να δημιουργήσει εκτελέσιμα αρχεία με επέκταση *.ado* ως συναρτήσεις και να τα καλεί μέσα από τη γραμμή εντολών του *STATA* με τη βοήθεια επέκτασης *.do* αρχείων.
- ♦ Η γλώσσα *R* έχει κερδίσει επίσης έδαφος στον τομέα της ανάλυσης δεδομένων μέσα στους ακαδημαϊκούς κύκλους. Είναι μια γλώσσα ανοικτού κώδικα και δίνει στο χρήστη τη δυνατότητα να εκτελεί στατιστικές αναλύσεις σε συνδυασμό με οπτικοποίηση των αποτελεσμάτων. Ο χρήστης χρησιμοποιεί πακέτα εντολών, τα οποία είναι ειδικά σχεδιασμένα για την ανάλυση που θέλει να υλοποιήσει, όπως για παράδειγμα πακέτα για την επεξεργασία πινάκων από δεδομένα γονιδιακής έκφρασης. Η πλατφόρμα *R Studio* είναι ένα περιβάλλον ανάπτυξης για τις εντολές σε γλώσσα *R* και προσφέρει ένα σύνολο ολοκληρωμένων εργαλείων.
- ♦ Η γλώσσα *Python* είναι μια δυναμική γλώσσα προγραμματισμού υψηλού επιπέδου. Προσφέρει βιβλιοθήκες λογισμικού για το χειρισμό και την ανάλυση δεδομένων πινάκων και πολυδιάστατων συστοιχιών. Εν συγκρίσει με την *R*, η *Python* προσφέρεται για πολλά πεδία εφαρμογών εκτός της επιστήμης των δεδομένων.



Εικόνα 2.5 Οι τρεις διαφορετικές πλατφόρμες των στατιστικών εργαλείων. Πάνω αριστερά το περιβάλλον χρήστη STATA, πάνω δεξιά το περιβάλλον R Studio για εκτέλεση εντολών σε γλώσσα R, κάτω το περιβάλλον Visual Studio Code για εκτέλεση εντολών σε γλώσσα Python.

2.4 Τεχνικές Ανάλυσης Δεδομένων Γονιδιακής Έκφρασης από DNA Μικροσυστοιχίες

Μέσα από τους πίνακες γονιδιακής έκφρασης κύριος στόχος είναι η ανίχνευση των διαφορικά εκφραζόμενων γονιδίων υπό διαφορετικές συνθήκες. Τα σύνολα δεδομένων μικροσυστοιχιών χαρακτηρίζονται συνήθως από μικρό αριθμό δειγμάτων, μεγάλο αριθμό μεταβλητών και από παρουσία θορύβου. Ως εκ τούτου, αρκετές στατιστικές μέθοδοι έχουν προταθεί για τον αποτελεσματικό χειρισμό αυτών των συνόλων δεδομένων (Dudoit et al., 2002, Kontou et al., 2018). Παρακάτω αναφέρονται ορισμένες από τις πιο γνωστές μεθόδους ανάλυσης δεδομένων μικροσυστοιχιών.

2.4.1 Μέθοδος t-test

Για την ανάλυση γονιδιακών δεδομένων ανάμεσα σε δύο ανεξάρτητες ομάδες (ασθενείς/μάρτυρες), μία παραμετρική μέθοδος υπολογισμού των διαφορικά εκφραζόμενων γονιδίων είναι το *t test*. Το *t test* είναι ένα κριτήριο το οποίο αξιολογεί τη διαφορά μέσω των όρων για δεδομένους βαθμούς ελευθερίας. Χρησιμοποιείται όταν δεν είναι γνωστή η διακύμανση της δειγματοληπτικής κατανομής και γι' αυτό το λόγο πρέπει να υπολογιστεί από τη διακύμανση των τιμών του δείγματος.

Υπάρχουν δύο περιπτώσεις αξιολόγησης της διαφοράς μέσω των όρων ανάμεσα σε δύο ανεξάρτητα σύνολα με *t* κατανομή.

Περίπτωση Α (Student's t-test): Υποθέτουμε ότι οι κατανομές συχνοτήτων των δύο ομάδων είναι κανονικές με μέσους όρους μ_α και μ_μ και αντιστοίχως κοινή διακύμανση s^2 .

$H_0: \mu_\alpha = \mu_\mu$ | Ο μέσος όρος των τιμών της πρώτης ομάδας ισούται με τον μέσο όρο της δεύτερης.

$H_A: \mu_\alpha \neq \mu_\mu$ | Ο μέσος όρος των τιμών της πρώτης ομάδας δεν ισούται με τον μέσο όρο της δεύτερης.

Υπολογίζεται το μέγεθος επίδρασης (t εκτίμηση) $t_{stat} = \frac{(\mu_\alpha - \mu_\mu)}{SE}$.

SE ορίζεται το τυπικό σφάλμα (standard error) από τον παρακάτω τύπο

$$SE = s_p \sqrt{\frac{1}{n_\alpha} + \frac{1}{n_\mu}}, \text{ όπου } n_\alpha \text{ και } n_\mu, \text{ το μέγεθος δείγματος της κάθε ομάδας.}$$

Το μέγεθος s_p αντιπροσωπεύει τη συγκεντρωτική τυπική απόκλιση και προκύπτει από τον τύπο της συγκεντρωτικής διακύμανσης s_p^2 ως εξής

$$s_p^2 = \frac{(n_\alpha - 1)s_\alpha^2 + (n_\mu - 1)s_\mu^2}{(n_\alpha - 1) + (n_\mu - 1)}$$

Γίνεται σύγκριση του εκτιμητή *t* με τη νέα εκτίμηση που προκύπτει συναρτήσει του ορίου για απόρριψη ή αποδοχή της H_0 σε συνδυασμό με $df = n_\alpha + n_\mu - 2$ βαθμούς ελευθερίας.

Περίπτωση Β (Welch's t-test): Υποθέτουμε ότι οι κατανομές συχνοτήτων των δύο ομάδων είναι κανονικές με μέσους όρους μ_α και μ_μ και με αντίστοιχες διακυμάνσεις s_α^2 και s_μ^2 .

$H_0: \mu_\alpha = \mu_\mu$ | Ο μέσος όρος των τιμών της πρώτης ομάδας ισούται με τον μέσο όρο της δεύτερης.

$H_A: \mu_\alpha \neq \mu_\mu$ | Ο μέσος όρος των τιμών της πρώτης ομάδας δεν ισούται με τον μέσο όρο της δεύτερης.

Υπολογίζεται το μέγεθος επίδρασης (t εκτίμηση) $t_{stat} = \frac{(\mu_\alpha - \mu_\mu)}{SE_{\mu_\alpha - \mu_\mu}}$

SE ορίζεται το τυπικό σφάλμα (standard error) από τον παρακάτω τύπο

$$SE_{\mu_\alpha - \mu_\mu} = \sqrt{\frac{s_\alpha^2}{n_\alpha} + \frac{s_\mu^2}{n_\mu}}, \text{ όπου } n_\alpha \text{ και } n_\mu \text{ το μέγεθος δείγματος της κάθε ομάδας.}$$

Ο τύπος για τον υπολογισμό των βαθμών ελευθερίας στην προκειμένη περίπτωση δίνεται από τον τύπο

$$df = \frac{\left(\frac{s_\alpha^2}{n_\alpha} + \frac{s_\mu^2}{n_\mu} \right)^2}{\left(\frac{s_\alpha^2}{n_\alpha} \right)^2 / (n_\alpha - 1) + \left(\frac{s_\mu^2}{n_\mu} \right)^2 / (n_\mu - 1)}$$

2.4.2 Μέθοδος Αναδειγματοληψίας Bootstrap

Ο έλεγχος t-test χρησιμοποιείται για τον υπολογισμό του εκτιμητή t και στη συνέχεια για την ταυτοποίηση των διαφορικά εκφραζόμενων γονιδίων ανάμεσα σε δύο σύνολα (ασθενείς/μάρτυρες). Ένα μειονέκτημα αυτής της μεθοδολογίας είναι ότι αρκετές φορές σε κάθε γονίδιο αντιστοιχούν πολύ λίγες παρατηρήσεις (μικρό μέγεθος δείγματος), με αποτέλεσμα να είναι άγνωστη η δειγματική κατανομή ή ακόμη ο υπολογισμός του τυπικού σφάλματος για τον εκτιμητή t να μην είναι προσεγγιστικός. Για να επιλυθεί αυτό το πρόβλημα χρησιμοποιείται η τεχνική αναδειγματοληψίας Bootstrap.

Η τεχνική bootstrap είναι μια στατιστική μέθοδος αναδειγματοληψίας κατά την οποία εκτιμάται η κατανομή του αρχικού δείγματος (Efron & Tibshirani, 1994, Efron, 1982). Με τη βοήθεια της αναδειγματοληψίας bootstrap προσδιορίζεται ο νέος εκτιμητής για τη διαφορά μέσων τιμών ανάμεσα στα δύο σύνολα του δείγματος (ασθενείς/ μάρτυρες), η τυπική απόκλιση των μέσων τιμών για κάθε νέα αναδιανομή του αρχικού δείγματος και ορίζεται το τυπικό σφάλμα κατά προσέγγιση του αρχικού δείγματος για Β επαναλήψεις.

Η συγκεκριμένη μέθοδος έχει αξιολογηθεί να προσφέρει εξίσου τα ίδια αποτελέσματα στις τιμές των p-values που εμφανίζονται ως στατιστικά σημαντικές σε μελέτες γονιδιακής έκφρασης όταν ο αριθμός επαναλήψεων υπερβαίνει τις 200. Παρ' όλα αυτά, παραμένει μια αρκετά χρονοβόρα διαδικασία και απαιτεί υπολογιστική δύναμη από μηχανικής πλευράς.

Παράδειγμα αναδειγματοληψίας bootstrap

Έστω δείγμα μεγέθους $N=9$ ανεξάρτητων μεταβλητών (var) M και L με βάρη (n)

$$\begin{bmatrix} \text{var} & M & M & M & M & L & L & L & L & L \\ n & n_1 & n_2 & n_3 & n_4 & n_5 & n_6 & n_7 & n_8 & n_9 \end{bmatrix}$$

$$H_0: |\bar{y}_m - \bar{y}_l| = 0, \quad H_1: |\bar{y}_m - \bar{y}_l| \neq 0$$

Διαφορά μέσων παρατηρηθέντος δείγματος $\bar{x}_{obs} = |\bar{y}_m - \bar{y}_l|$

Αναδειγματοληψία Bootstrap για B επαναλήψεις

$$\text{Δείγμα \#1} \quad \begin{bmatrix} \text{var} & M & M & M & M & L & L & L & L & L \\ n & n_3 & n_2 & n_2 & n_1 & n_6 & n_7 & n_9 & n_5 & n_8 \end{bmatrix}$$

• Υπολογισμός διαφοράς μέσων $\bar{x}_1 = |\bar{y}_m - \bar{y}_l|$

$$\text{Δείγμα \#2} \quad \begin{bmatrix} \text{var} & M & M & M & M & L & L & L & L & L \\ n & n_1 & n_2 & n_4 & n_4 & n_9 & n_9 & n_7 & n_8 & n_6 \end{bmatrix}$$

• Υπολογισμός διαφοράς μέσων $\bar{x}_2 = |\bar{y}_m - \bar{y}_l|$

⋮ ⋮

$$\text{Δείγμα \#B} \quad \begin{bmatrix} \text{var} & M & M & M & M & L & L & L & L & L \\ n & n_1 & n_3 & n_3 & n_1 & n_7 & n_5 & n_8 & n_7 & n_9 \end{bmatrix}$$

• Υπολογισμός διαφοράς μέσων $\bar{x}_B = |\bar{y}_m - \bar{y}_l|$

$$\text{Υπολογισμός του Standard error} \quad SE_{\bar{x}} = \sqrt{\frac{s_m^2}{m} + \frac{s_l^2}{l}}$$

$$\text{Υπολογισμός του p-value} \quad p - \text{value} = \frac{\sum_{i=1}^B (\bar{x}_i \geq \bar{x}_{obs})}{B}$$

2.4.3 Μπεϋζιανές Μέθοδοι

Οι μέθοδοι που αναλύονται παραπάνω αφορούν την κλασική στατιστική. Στρέφονται γύρω από την εκτίμηση άγνωστων παραμέτρων και την εκτέλεση στατιστικών δοκιμών προκειμένου να γίνει αποδεκτή ή να απορριφθεί μια μηδενική υπόθεση. Μία εναλλακτική προσέγγιση στις στατιστικές μεθόδους είναι η Μπεϋζιανή προσέγγιση. Οι μπεϋζιανές μέθοδοι απορρέουν από την φιλοσοφία των πιθανοτήτων και συγκεκριμένα την ερμηνεία της πιθανότητας ως αβεβαιότητα και όχι ως συχνότητα. Η Μπεϋζιανή στατιστική αποδίδει κατανομές πιθανότητας στις παραμέτρους ενδιαφέροντος. Για παράδειγμα, αν η αβεβαιότητα για ένα συγκεντρωτικό μέγεθος επίδρασης, με αντίστοιχο συγκεντρωτικό τυπικό σφάλμα, αναπαρασταθεί με τη μορφή κανονικής κατανομής με επίκεντρο την καλύτερη σημειακή εκτίμηση, τότε με διάστημα Μπεϋζιανής εκτίμησης μπορεί να καλυφθεί η βεβαιότητα για αυτό το σύνολο.

Μια Μπεϋζιανή ανάλυση ξεκινά με την ανάθεση μιας εκ των προτέρων κατανομής (*a priori probability distribution*) σε κάθε άγνωστη παράμετρο και δεν λαμβάνει υπόψη τις πληροφορίες από τα παρατηρηθέντα δεδομένα. Εφαρμόζεται στη συνέχεια το θεώρημα του Bayes για να ληφθεί υπόψη η επίδραση των παρατηρούμενων δεδομένων. Έτσι λαμβάνεται μια κατανομή των παραμέτρων μετά την παρακολούθηση των δεδομένων και ονομάζεται εκ των υστέρων κατανομή (*posterior probability distribution*). Σε αντίθεση με τα διαστήματα εμπιστοσύνης των κλασικών μεθόδων, η Μπεϋζιανή ανάλυση παράγει διαστήματα αξιοπιστίας (*credible intervals*).

Στον έλεγχο υποθέσεων, η Μπεϋζιανή ανάλυση πραγματοποιείται βασιζόμενη στον Μπεϋζιανό παράγοντα (*Bayes factor*) για τη σύγκριση της μηδενικής με την εναλλακτική υπόθεση για δύο υποψήφια παραμετρικά σύνολα. Ανάλογα την τιμή που λαμβάνει ο Μπεϋζιανός παράγοντας, κρίνεται ποιο από τα δύο σύνολα υπερτερεί του άλλου. Ο κλειστός τύπος του Μπεϋζιανού παράγοντα, ο οποίος μπορεί να γραφτεί με όρους κατανομής ανάμεσα σε δύο δείγματα (*two-sample t-test*) υπό μια μηδενική και μια εναλλακτική υπόθεση, έχει προταθεί στο παρελθόν (Rouder et al., 2009, Gönen et al., 2005).

Άλλες μέθοδοι βασίζονται στην μεθοδολογία της Μαρκοβιανής αλυσίδας Μόντε Κάρλο (*Markov Chain Monte Carlo, MCMC*), λαμβάνοντας δείγματα από την εκ των υστέρων κατανομή. Αυτές οι μέθοδοι ξεκινούν από αρχικές τιμές για τις παραμέτρους και στη συνέχεια, σύμφωνα με ορισμένες λειτουργίες πρότασης και αποδοχής, προτείνονται και γίνονται αποδεκτές νέες τιμές (Kruschke, 2013).

2.4.4 Άλλες Μέθοδοι

- ✚ Μια ανάλογη μέθοδος αναδειγματοληψίας είναι το *permutation test*. Το permutation test, σε αντίθεση με το bootstrap, «ανακατεύει» όλες τις αρχικές παρατηρήσεις του δείγματος για P επαναλήψεις, ανεξάρτητα σε ποιο σύνολο ανήκουν.
- ✚ Για τους λόγους που αναφέρθηκαν πιο πάνω, το t-test δεν είναι η πιο αξιόπιστη μέθοδος ανάλυσης δεδομένων γονιδιακής έκφρασης. Εναλλακτικές μέθοδοι έχουν προταθεί, οι οποίες παρουσιάζουν μια εμπειρική Μπεϋζιανή προσέγγιση σε συνδυασμό με κανονικοποιημένη δοκιμή t-test (*Penalised t-test*).
- ✚ Μια άλλη εμπειρική Μπεϋζιανή προσέγγιση είναι η γενικευμένη μέθοδος των Lönnstedt και Speed από τον S. Smyth, κατά την οποία εφαρμόζονται εμπειρικά Μπεϋζιανά γραμμικά μοντέλα για την εκτίμηση της διαφορικής έκφρασης γονιδίων (*linear models for microarray data, limma*).
- ✚ Από τις πρώτες μεθοδολογίες για εκτίμηση διαφορικής έκφρασης γονιδίων ήταν η αλλαγή διπλώματος (*fold change, FC*). Σήμερα, η αλλαγή διπλώματος προτείνεται εφόσον ικανοποιούνται και επιπλέον κριτήρια, όπως για παράδειγμα σε συνδυασμό μέσω αξιολόγησης με p-value.
- ✚ Η μέθοδος *Rank Product* είναι μια μη παραμετρική μέθοδος για ανίχνευση ΔΕΓ σε πειράματα μικροσυστοιχιών. Τα γονίδια ταξινομούνται σύμφωνα με την αλλαγή διπλώματος και στη συνέχεια πραγματοποιείται ανάλυση χωριστά για τα υπερεκφραζόμενα και για τα υποεκφραζόμενα γονίδια.

Κεφάλαιο 3 – Μετα-ανάλυση

3.1 Συστηματική Ανασκόπηση και Μετα-ανάλυση

Η μετα-ανάλυση είναι η διαδικασία κατά την οποία πληροφορίες από διάφορες ανεξάρτητες μελέτες που εξετάζουν ένα κοινό πρόβλημα συνενώνονται για να διεξαχθεί ένας στατιστικός έλεγχος (Normand, 1999). Ο όρος «μετα-ανάλυση» εμφανίστηκε αρχικά στο πεδίο της ψυχολογίας και ύστερα συναντήθηκε στα πεδία των κοινωνικών επιστημών, της εκπαίδευσης, της ιατρικής κ.ό.κ. (Glass, 1976).

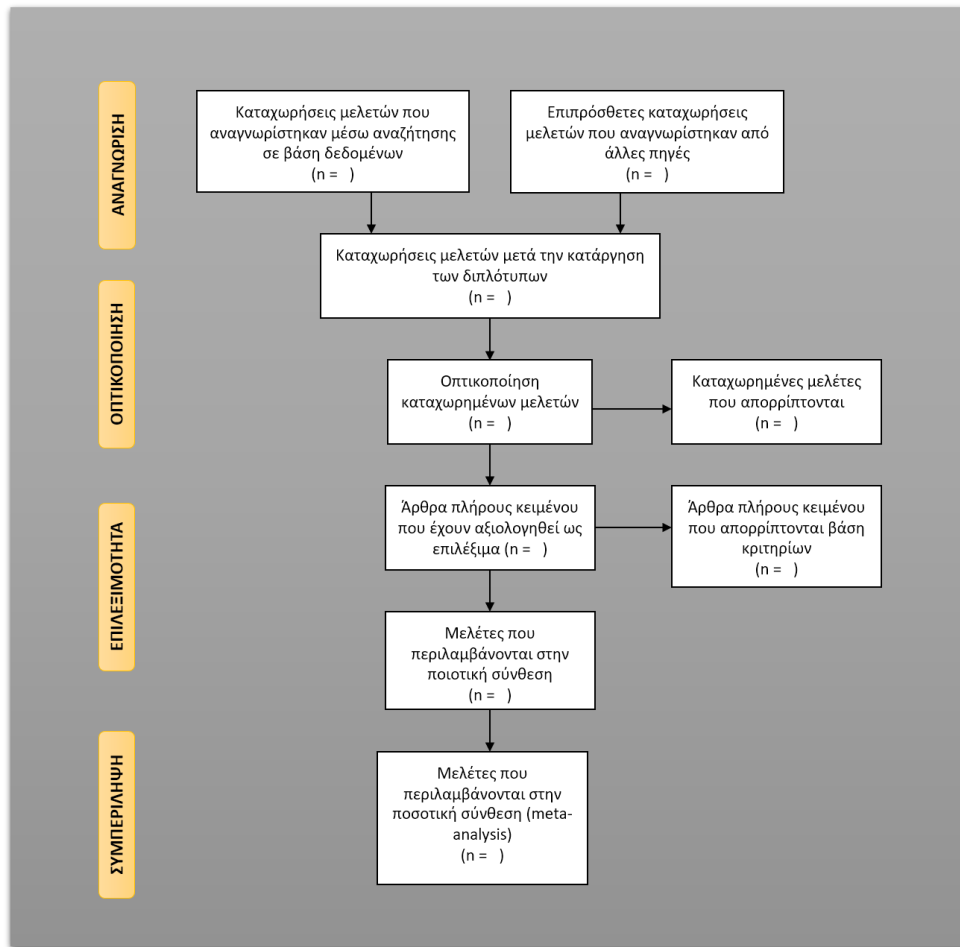
Για τη σωστή διεξαγωγή μιας μετα-ανάλυσης έχει προταθεί μεθοδολογία από ποικίλες πηγές για την έγκυρη συλλογή και επεξεργασία δεδομένων (Moher et al., 2009, Liberati et al., 2009). Στη μεθοδολογία που παρουσιάζεται παρακάτω αναφέρονται τα 6 βασικά βήματα για το σχεδιασμό και την υλοποίηση μιας συστηματικής ανασκόπησης (systematic review) η οποία βοηθά στην ορθή σύνθεση των μελετών για μετα-ανάλυση. Τα βήματα 3 έως 5 έχουν αναπαρασταθεί βάση βιβλιογραφίας με τη μορφή ενός διαγράμματος ροής. Το διάγραμμα αυτό αποτυπώνει όλα τα στάδια από την αρχική συλλογή μελετών μέχρι την τελική διαλογή τους. Το διάγραμμα έχει χαρακτηριστεί από τη βιβλιογραφία ως πρότυπο PRISMA (PRISMA statement) και η μορφή του φαίνεται στην εικόνα 3.1.

Τα βήματα που διέπουν μια συστηματική ανασκόπηση ακολουθούν το παρακάτω πρότυπο:

1. Διατύπωση ερευνητικού ερωτήματος
2. Καθορισμός κριτηρίων αποδοχής και απόρριψης μελετών
3. Αναζήτηση βιβλιογραφίας
4. Αξιολόγηση και επιλογή μελετών
5. Συλλογή δεδομένων από μελέτες που συγκεντρώθηκαν
6. Στατιστική ανάλυση (μετα-ανάλυση)

3.2 Μοντέλα Σταθερών και Τυχαίων Επιδράσεων

Οι περισσότερες μετα-αναλύσεις βασίζονται σε δύο κύρια στατιστικά μοντέλα. Στο μοντέλο σταθερών επιδράσεων (*fixed effects model*) και στο μοντέλο τυχαίων επιδράσεων (*random effects model*). Πριν ξεκινήσει η μετα-ανάλυση συνηθίζεται να ορίζεται ένα μέγεθος επίδρασης. Κατόπιν, επιλέγεται το μοντέλο επίδρασης πάνω στο οποίο θα διεξαχθεί ο στατιστικός έλεγχος.



Εικόνα 3.1 Διάγραμμα ροής για αναζήτηση μελετών στη βιβλιογραφία στο πλαίσιο της συστηματικής ανασκόπησης (Moher et al., 2009, Liberati et al., 2009)

3.2.1 Μοντέλο Σταθερών Επιδράσεων

Για το μοντέλο σταθερών επιδράσεων λαμβάνεται υπόψη ένα πραγματικό μέγεθος επίδρασης για όλες τις συγκεντρωμένες μελέτες που εξετάζονται ενώ θεωρείται ότι οι διαφορές που παρατηρούνται οφείλονται σε σφάλμα δειγματοληψίας. Έστω E το παρατηρούμενο μέγεθος επίδρασης για k ανεξάρτητες μελέτες. Η σχέση που περιγράφει το μοντέλο σταθερών επιδράσεων είναι η ακόλουθη (Borenstein, 2009).

$$E_i = \mu + e_i, e_i \sim N(0, \sigma_i^2) \text{ για } i = 1, 2, \dots, k$$

, όπου μ η μέση τιμή, σ^2 η πληθυσμιακή διακύμανση και e το σφάλμα δειγματοληψίας της κάθε μελέτης. Κατά συνέπεια, το βάρος κάθε μελέτης περιγράφεται από τη σχέση

$$W_{i \text{ fixed}} = \frac{1}{V_{E_i}}, \text{ όπου } V_{E_i} \text{ η διακύμανση της κάθε μελέτης (within-study variance).}$$

Στη συνέχεια, υπολογίζεται ο σταθμισμένος μέσος όρος M των παρατηρούμενων μεγεθών επίδρασης για κάθε μελέτη με τη μέθοδο της μέγιστης πιθανοφάνειας. Ο τύπος του είναι

$M_F = \frac{\sum_{i=1}^k W_{i_{fixed}} E_i}{\sum_{i=1}^k W_{i_{fixed}}}$. Η προκύπτουσα διακύμανση του σταθμισμένου μέσου όρου δίνεται από την σχέση $V_{M_F} = \frac{1}{\sum_{i=1}^k W_{i_{fixed}}}$ με αντίστοιχο τυπικό σφάλμα $SE_{M_F} = \sqrt{V_{M_F}}$.

Στο στατιστικό πρόγραμμα STATA η προεπιλεγμένη μέθοδος για το μοντέλο σταθερών επιδράσεων καλείται με την εντολή `fixed` και εφαρμόζει τη μέθοδο αντίστροφης διακύμανσης.

3.2.2 Μοντέλο Τυχαίων Επιδράσεων

Απεναντίας, σύμφωνα με το μοντέλο τυχαίων επιδράσεων, το μέγεθος επίδρασης διαφέρει από μελέτη σε μελέτη. Μέσα σε αυτό το πλαίσιο, το παρατηρούμενο μέγεθος επίδρασης της κάθε μελέτης E_i προέρχεται από μία κατανομή με πραγματικό μέγεθος επίδρασης ρ_i και πληθυσμιακή διακύμανση σ_i^2 . Κάθε μελέτη θεωρείται ότι έχει αντληθεί από κάποιο σύνολο υπερπληθυσμού με μέση τιμή μ και διακύμανση τ^2 . Η σχέση που περιγράφει το μοντέλο τυχαίων επιδράσεων είναι η ακόλουθη (Borenstein, 2009).

$$E_i = \rho_i + e_i, e_i \sim N(0, \sigma_i^2)$$

$$\rho_i = \mu + \varepsilon_i, \varepsilon_i \sim N(0, \tau^2) \text{ για } i = 1, 2, \dots, k$$

Από την παραπάνω περιγραφή προκύπτει ότι το παρατηρούμενο μέγεθος επίδρασης κάθε μελέτης επηρεάζεται όχι μόνο από παράγοντες που οφείλονται σε σφάλμα δειγματοληψίας, αλλά και στο πραγματικό μέγεθος επίδρασης. Η επιπρόσθετη παράμετρος που ορίζεται στην προκειμένη περίπτωση συμβολίζεται με τ^2 και είναι η διακύμανση του πραγματικού μεγέθους επίδρασης μεταξύ των μελετών (*between-studies variance*). Για τον προσδιορισμό του εκτιμητή τ^2 έχουν προταθεί διάφορες μεθοδολογίες οι οποίες αναλύονται στην επόμενη παράγραφο.

Σύμφωνα με το μοντέλο τυχαίων επιδράσεων το βάρος που ανατίθεται σε κάθε μελέτη ισούται με $W_{i_{random}} = \frac{1}{V_{E_i} + T^2}$, όπου T^2 ο εκτιμητής για την διακύμανση μεταξύ των μελετών. Ο σταθμισμένος μέσος όρος M ορίζεται από τον τύπο $M_R = \frac{\sum_{i=1}^k W_{i_{random}} E_i}{\sum_{i=1}^k W_{i_{random}}}$. Η διακύμανση δίνεται από τον τύπο $V_{M_R} = \frac{1}{\sum_{i=1}^k W_{i_{random}}}$ και το τυπικό σφάλμα $SE_{M_R} = \sqrt{V_{M_R}}$.

Στο στατιστικό πρόγραμμα STATA η προεπιλεγμένη μέθοδος για το μοντέλο τυχαίων επιδράσεων καλείται με την εντολή `random` και εφαρμόζει τη μέθοδο των DerSimonian και Laird, με εκτίμηση της ετερογένειας σε συνδυασμό με τη μέθοδο αντίστροφης διακύμανσης.

3.3 Έλεγχος Ετερογένειας

Αν και στο μοντέλο τυχαίων επιδράσεων η εμφάνιση ετερογένειας είναι συνυφασμένη με την παράμετρο τ^2 , στο μοντέλο σταθερών επιδράσεων ο βαθμός ετερογένειας στις εκτιμήσεις υπάρχει λόγω σφάλματος δειγματοληψίας εντός των μελετών.

Ο πιο συνηθισμένος τρόπος ένδειξης της ετερογένειας είναι μια στατιστική δοκιμή, η οποία συχνά περιγράφεται ως δοκιμασία χ^2 ή Cochran's Q (Cochran, 1954). Ο τύπος ορίζεται ως

$$Q = \sum_{i=1}^k W_i (E_i - M)^2$$

, όπου W_i είναι το βάρος μελέτης, E_i το μέγεθος επίδρασης και M ο σταθμισμένος μέσος όρος.

Η πιο γνωστή μέθοδος προσδιορισμού της παραμέτρου τ^2 είναι η μέθοδος των ροπών των DerSimonian και Laird (DerSimonian & Laird, 1986), η οποία είναι και η προεπιλεγμένη προσέγγιση σε αρκετές συναρτήσεις λογισμικού. Ο υπολογισμός προκύπτει από τη σχέση

$$T^2 = \frac{Q - df}{F}$$

, όπου $Q = \sum_{i=1}^k (W_{i_{random}} E_i^2) - \frac{(\sum_{i=1}^k W_{i_{random}} E_i)^2}{\sum_{i=1}^k W_{i_{random}}}$, $df = k - 1$ και

$$F = \sum_{i=1}^k W_{i_{random}} - \frac{\sum_{i=1}^k W_{i_{random}}^2}{\sum_{i=1}^k W_{i_{random}}}$$

Ο Higgins και οι συνεργάτες (2003) πρότειναν την χρήση του εκτιμητή I^2 , που είναι ο λόγος της πραγματικής ετερογένειας προς τη συνολική διακύμανση μεταξύ των εκτιμήσεων του παρατηρούμενου αποτελέσματος. Ο εκτιμητής I^2 κάνει χρήση του Q του Cochran και παίρνει τιμές από 0 έως 100%. Οι τιμές της τάξης των 25%, 50% και 75% θα μπορούσαν να θεωρηθούν χαμηλές, μέτριες και υψηλές, αντίστοιχα σε επίπεδο έκφρασης ετερογένειας.

$$I^2 = \left(\frac{Q - df}{Q} \right) * 100\%$$

3.4 Σφάλμα Μεροληψίας

Το σφάλμα μεροληψίας (bias) είναι το φαινόμενο κατά το οποίο η συνολική εκτίμηση της πρόβλεψης επηρεάζεται από μελέτες που δεν έχουν συμπεριληφθεί στον στατιστικό έλεγχο. Υπάρχουν αρκετοί λόγοι για την συμπερίληψη σφάλματος μεροληψίας στην τελική εκτίμηση.

✚ **Σφάλμα μεροληψίας λόγω δημοσίευσης (Publication bias)**

Το σφάλμα μεροληψίας λόγω δημοσίευσης είναι η προκατάληψη από τους εκδότες ακαδημαϊκών περιοδικών να προτιμούν να δημοσιεύουν μελέτες που αναφέρουν

στατιστικά σημαντικά αποτελέσματα, παρά μελέτες που αναφέρουν μη στατιστικά σημαντικά αποτελέσματα. Ως αποτέλεσμα, δεν συμπεριλαμβάνονται όλες οι σχετικές μελέτες σε μια μετα-ανάλυση και ενδέχεται να λείπουν δεδομένα.

✚ **Σφάλμα μεροληψίας λόγω γλώσσας κειμένου (Language bias)**

Είναι πιθανότερο να αναζητηθούν κείμενα τα οποία είναι γραμμένα στα αγγλικά και μπορούν να βρεθούν μέσα από βάσεις δεδομένων που υποστηρίζουν την αγγλική γλώσσα.

✚ **Σφάλμα μεροληψίας λόγω διαθεσιμότητας κειμένου (Availability/Cost bias)**

Συμπεριλαμβάνονται οι μελέτες οι οποίες είναι εύκολα προσβάσιμες στον αναλυτή χωρίς οικονομική επιβάρυνση ή χαμηλού κόστους. Σημαντικό ρόλο παίζει σε ποια επιστημονικά περιοδικά είναι εγγεγραμμένο το Πανεπιστήμιο του εκάστοτε ερευνητή/τριας.

✚ **Σφάλμα μεροληψίας λόγω επανάληψης (Duplication bias)**

Είναι πιθανό να δημοσιευτούν πολλές παρόμοιες μελέτες που αναφέρουν στατιστική σημαντικότητα.

3.5 Μετα-ανάλυση Δεδομένων Μικροσυστοιχιών DNA

Η μετα-ανάλυση είναι φθηνή διαδικασία και κερδίζει έδαφος στο χώρο της βιοστατιστικής όλο και πιο συχνά, καθώς η αύξηση των δεδομένων στο κομμάτι της γονιδιοματικής από πειράματα μικροσυστοιχιών έχει κάνει την εφαρμογή της πλέον απαραίτητη. Με τη χρήση της ενισχύεται η στατιστική δύναμη για τη λήψη μιας πιο ακριβούς εκτίμησης γύρω από το πλήθος γονιδίων που παρουσιάζουν διαφορά στην έκφραση παρουσία μιας κατάστασης. Η διεξαγωγή μιας μετα-ανάλυσης σε δεδομένα από πειράματα σε μικροσυστοιχίες παρουσιάζεται παρακάτω σε 5 βασικά στάδια (Ramasamy et al., 2008).

Βήμα 1: Επιλογή κατάλληλων σετ δεδομένων από μικροσυστοιχίες

Το πρώτο και κυριότερο στάδιο είναι να αποσαφηνιστεί σωστά ο στόχος του ερευνητικού έργου.

Βήμα 2: Απόκτηση δεδομένων από τις μελέτες

Ορίζονται κριτήρια επιλεξιμότητας. Πραγματοποιείται αναζήτηση βιβλιογραφίας σε βάσεις δεδομένων γονιδιακής έκφρασης υπό συγκεκριμένους όρους και ανακτώνται σχετικές μελέτες.

Βήμα 3: Προεπεξεργασία συνόλων δεδομένων από διαφορετικές πλατφόρμες

Για να είναι δυνατή η ανάλυση όλων των δεδομένων, πρέπει τα αρχεία από κάθε μελέτη να αποθηκεύονται με τη μορφή ενός πίνακα γονιδιακής έκφρασης (*gene expression data matrix*). Σε περίπτωση που τα πειράματα έχουν διεξαχθεί σε διαφορετικές πλατφόρμες μικροσυστοιχιών, προτείνεται βάση βιβλιογραφίας να προεπεξεργασθούν τα σύνολα δεδομένων των πινάκων με αλγοριθμικό τρόπο.

Βήμα 4: Αντιστοιχία ένα-προς-ένα μεταξύ ανιχνευτών και γονιδίων

Ένα βασικό πρόβλημα που έγκειται στα δεδομένα γονιδιακής έκφρασης είναι ότι λόγω της κύριας κατασκευής της μικροσυστοιχίας πολλοί ανιχνευτές είναι συνδεδεμένοι με την

αλληλουχία ενός γονιδίου. Αυτό παρουσιάζεται σαν πρόβλημα «πολλά προς πολλά» μέσα από τον πίνακα γονιδιακής έκφρασης. Ο χρήστης οφείλει να επιλύσει το πρόβλημα με το να κάνει μια τροποποίηση στα δεδομένα του πίνακα με περιγραφική στατιστική. Η σχέση μεταξύ γονιδίων και ανιχνευτών πρέπει να είναι της μορφής «ένα προς ένα» στο τελικό σετ δεδομένων.

Βήμα 5: Επιλογή μιας τεχνικής μετα-ανάλυσης

Τέλος, ο χρήστης μπορεί να εστιάσει στην ανάλυση των συνεχών δεδομένων που συγκέντρωσε, όπου στόχος είναι ο προσδιορισμός των γονιδίων που εκφράζονται διαφορετικά μεταξύ δύο συνθηκών.

3.5.1 Τεχνικές Μετα-ανάλυσης Δεδομένων Γονιδιακής Έκφρασης από DNA Μικροσυστοιχίες

Οι στατιστικές μέθοδοι για μετα-ανάλυση γονιδίων με διαφορετική έκφραση μπορούν να κατηγοριοποιηθούν σε τρεις κύριες ομάδες (Hong & Breitling, 2008, Kontou et al., 2018):

- ❖ τις μεθόδους που βασίζονται σε κάποιο μέγεθος επίδρασης,
- ❖ τις μεθόδους που συνδυάζουν p-values και
- ❖ τις μεθόδους που συνδυάζουν βαθμούς (ranks).

3.5.1.1 Η τυποποιημένη διαφορά των μέσων σαν μέγεθος επίδρασης

Όταν οι περισσότερες μετα-αναλύσεις επεξεργάζονται δεδομένα μελετών στις οποίες γίνεται αναφορά στο μέσο όρο ανάμεσα σε ομάδες και υπολογίζονται τυπικές αποκλίσεις, τα προτιμώμενα μεγέθη επίδρασης που χρησιμοποιούνται είναι η «ακατέργαστη» μέση διαφορά (*raw mean difference*), η τυποποιημένη μέση διαφορά (*standardized mean difference*) ή ο λόγος απόκρισης (*response ratio*).

Το πλεονέκτημα της τυποποιημένης διαφοράς μέσων είναι ότι μετατρέπει όλα τα μεγέθη επίδρασης σε μία κοινή μέτρηση και επιτρέπει στο χρήστη να συμπεριλάβει αποτελέσματα από διαφορετικές κλίμακες στην ίδια σύνθεση (Borenstein et al., 2011).

Ο υπολογισμός της τυποποιημένης διαφοράς μέσων ανάμεσα σε δύο ανεξάρτητες ομάδες με διαφορετικές διακυμάνσεις προκύπτει ως εξής

$$d = \frac{X_{case} - X_{control}}{S_{pooled}}$$

, όπου X_{case} και $X_{control}$ οι μέσοι όροι των ανεξάρτητων ομάδων κάθε μελέτης και S_{pooled} η συγκεντρωτική τυπική απόκλιση μεταξύ των δύο ομάδων. Η τυπική απόκλιση ορίζεται από τον τύπο

$$S_{pooled} = \sqrt{\frac{(n_{case} - 1)S_{case}^2 + (n_{control} - 1)S_{control}^2}{n_{case} + n_{control} - 2}}$$

, όπου n_{case} και $n_{control}$ το μέγεθος δείγματος της κάθε ομάδας και S_{case} και $S_{control}$ οι τυπικές αποκλίσεις αντίστοιχα των δύο ομάδων.

Η δειγματοληπτική εκτίμηση της τυποποιημένης μέσης διαφοράς ονομάζεται συχνά Cohen's d στην ερευνητική σύνθεση (Cohen, 1987).

Η διακύμανση του d προκύπτει από τον τύπο

$$Var_d = \frac{n_{case} + n_{control}}{n_{case}n_{control}} + \frac{d^2}{2(n_{case} + n_{control})}$$

Το τυπικό σφάλμα του d είναι η τετραγωνική ρίζα του Var_d και δίνεται από τον τύπο

$$SE_d = \sqrt{Var_d}$$

Προκύπτει ότι η εκτίμηση d προσδίδει μεροληπτικούς υπολογισμούς σε μικρά σύνολα δεδομένων. Το συστηματικό σφάλμα μπορεί να αποφευχθεί με διόρθωση, η οποία αναφέρεται στη βιβλιογραφία ως διόρθωση του Hedges (Hedges, 1981). Για την μετατροπή του d σε g του Hedges (*Hedges' g*), χρησιμοποιείται ένας παράγοντας διόρθωσης J . Ο τύπος για τον παράγοντα διόρθωσης δίνεται από τον τύπο

$$J = 1 - \frac{3}{4df - 1}$$

, όπου df είναι οι βαθμοί ελευθερίας. Στην περίπτωση δύο ανεξάρτητων συνόλων υπολογίζονται από τον τύπο n_1+n_2-2 .

Στη συνέχεια προκύπτει ο διορθωμένος εκτιμητής g

$$g = J * d$$

, η διακύμανση Var_g

$$Var_g = J^2 * Var_d$$

και το τυπικό σφάλμα SE_g

$$SE_g = \sqrt{Var_g} .$$

3.5.1.2 Άλλες μέθοδοι

- ✚ Η μέθοδος Rank Product έχει ήδη παρουσιαστεί στην παράγραφο με τις τεχνικές ανάλυσης δεδομένων γονιδιακής έκφρασης. Χρησιμοποιεί το FC για να ταξινομήσει τα γονίδια και υπολογίζει τα προϊόντα κατάταξης (ranks) σε δείγματα και μελέτες. Μια παρόμοια μέθοδος, η *Rank Sum*, χρησιμοποιεί το άθροισμα των βαθμών κατάταξης, και όλοι οι υπόλοιποι υπολογισμοί παραμένουν ίδιοι.
- ✚ Μια σχετική μέθοδος, που ονομάζεται *METRADISC*, βασίζεται στην ίδια αρχή, όμως είναι πιο γενικευμένη. Η κατάταξη σε κάθε μελέτη πραγματοποιείται με οποιαδήποτε διαθέσιμη

μέθοδο (FC, t-test, κ.λπ.) και στη συνέχεια υπολογίζεται η μέση κατάταξη ενός συγκεκριμένου γονιδίου σε όλες τις μελέτες. Ο συγκεντρωτικός μέσος όρος μπορεί να είναι σταθμισμένος ή μη.

- ✚ Μια άλλη κατηγορία μεθόδων που είναι δημοφιλής στη μετα-ανάλυση μελετών μικροσυστοιχιών περιλαμβάνει το συνδυασμό τιμών p-value. Ο Fisher είχε σημειώσει ότι δεδομένου ότι οι τιμές p-value από N ανεξάρτητα δείγματα είναι ομοιόμορφες τυχαίες μεταβλητές, το άθροισμα του λογαρίθμου τους θα ακολουθεί μια κατανομή χ^2 με $2N$ βαθμούς ελευθερίας. Αν και αυτή η μέθοδος αποτελεί μια ελκυστική εφαρμογή για τα συγκεκριμένα σετ δεδομένων, παρουσιάζει σημαντικά προβλήματα σχετικά με το συνδυασμό μεγεθών επίδρασης.

3.5.2 Πολλαπλές συγκρίσεις και μέθοδοι διόρθωσης σφαλμάτων Τύπου I

Υποθέτοντας ότι η μηδενική υπόθεση ισχύει σε περίπτωση που 10.000 γονίδια δοκιμαστούν σε επίπεδο σημαντικότητας 5% με δοκιμή t test, 500 γονίδια μπορούν να κηρυχθούν σημαντικά, τυχαία και μόνο. Το μειονέκτημα αυτής της μεθόδου λοιπόν, όσον αφορά την εφαρμογή της σε μεγάλα δείγματα δεδομένων από μικροσυστοιχίες είναι ότι οι εκτιμήσεις διακύμανσης μπορεί να αποκλίνουν από γονίδια με πολύ μικρή διακύμανση. Αυτά τα γονίδια συνδέονται με μια μεγάλη t εκτίμηση και επιλέγονται ψευδώς ως διαφορικά εκφρασμένα (false positives). Τα σφάλματα αυτού του είδους ονομάζονται σφάλματα τύπου 1 στον στατιστικό έλεγχο υποθέσεων. Για αυτό το λόγο χρησιμοποιούνται μέθοδοι διόρθωσης των τιμών p-value ύστερα από τη διεξαγωγή στατιστικού ελέγχου σε μεγάλα σύνολα δεδομένων.

Υπάρχουν δύο κύριες κατηγορίες μεθόδων. Εκείνες που ελέγχουν τον «family-wise» ρυθμό σφάλματος (family-wise error rate) και εκείνες που ελέγχουν τον ρυθμό ψευδών ανακαλύψεων (false discovery rate). Ορισμένες μέθοδοι διόρθωσης των εσφαλμένων στατιστικά σημαντικών p-value όπως έχουν προταθεί από βιβλιογραφικές αναφορές παρουσιάζονται παρακάτω.

1) Μέθοδος Bonferroni

Η διόρθωση Bonferroni είναι από τις πιο απλές μεθόδους FWER διόρθωσης λαθών. Δοθέντος του επιπέδου σημαντικότητας π.χ. $\alpha=0.05$ για 10.000 διαφορετικά τεστ (αριθμός γονιδίων), το νέο επίπεδο σημαντικότητας διαμορφώνεται σε $\alpha_{\text{new}}=\alpha/10.000=5\times 10^{-6}$. Εύκολα αντιλαμβάνεται κανείς ότι η συγκεκριμένη μέθοδος διόρθωσης είναι αρκετά συντηρητική καθώς επιστρέφει πολύ λίγα πραγματικά σημαντικά αποτελέσματα. Με άλλα λόγια, πολλά στατιστικά σημαντικά αποτελέσματα δε δηλώνονται σημαντικά ενώ είναι (Perneger, 1998). Ο τύπος που προσδιορίζει το νέο επίπεδο σημαντικότητας για N αριθμό μεταβλητών (γονίδια) είναι:

$$p_{\text{corr}} = \frac{\alpha}{N}$$

2) Μέθοδος Holm-Bonferroni (step-down Bonferroni)

Ο Holm (Holm, 1979) πρότεινε μια εναλλακτική μορφή της μεθόδου διόρθωσης Bonferroni με έλεγχο FWER. Οι τιμές των p-value ταξινομούνται από το μικρότερο στο μεγαλύτερο και συγκρίνονται με ακολουθία επιπέδων σημαντικότητας. Για επίπεδο σημαντικότητας $\alpha = p_i = 0.05$, όπου i η βαθμίδα των p-value σε αύξουσα ταξινόμηση και N ο αριθμός μεταβλητών (γονίδια), το νέο επίπεδο σημαντικότητας διαμορφώνεται από τον τύπο:

$$p_{corr(i)} = \frac{p_i}{N - i + 1}$$

3) Μέθοδος Šidák

Η διόρθωση Šidák (Šidák, 1967) προκύπτει επίσης με έλεγχο FWER υποθέτοντας ότι οι μεμονωμένες δοκιμές είναι ανεξάρτητες. Δοθέντων N διαφορετικών μεταβλητών (γονίδια) και αρχική θεώρηση κατωφλιού σημαντικότητας $\alpha = p_i = 0.05$, απορρίπτονται οι αρχικές μηδενικές υποθέσεις H_0 όπου το p-value του πίνακα έχει τιμή μικρότερη από την εξίσωση του τύπου:

$$p_{corr(i)} = 1 - (1 - p_i)^{1/N}$$

4) Μέθοδος Holland & Copenhaver

Η μέθοδος των Holland και Copenhaver (Holland & Copenhaver, 1988) επηρεασμένη από τη μέθοδο του Šidák προτείνει όπως και στην προηγούμενη μεθοδολογία την ταξινόμηση αρχικά σε αύξουσα σειρά των p-values. Θεωρώντας ως αρχικό κατώφλι την τιμή $\alpha = p_i = 0.05$, το νέο επίπεδο σημαντικότητας για την ερμηνεία όλων των p-value του πίνακα κατά αύξουσα σειρά τροποποιείται ως εξής:

$$p_{corr(i)} = 1 - (1 - p_i)^{1/(N-i+1)}$$

,όπου N ο αριθμός των μεταβλητών (γονίδια) και i ο αύξων αριθμός του εκάστοτε p-value.

5) Μέθοδος των Benjamini & Hochberg (step-up Simes)

Η διόρθωση με έλεγχο FDR όπως προτάθηκε από τους Benjamini και Hochberg (Benjamini & Hochberg, 1995) επεκτείνεται ως προς στο θεώρημα του Simes (Simes, 1986). Είναι μια μέθοδος που συλλέγει τα ποσοστά σφαλμάτων των ψευδώς σημαντικών δεδομένων σε δοκιμές μηδενικής υπόθεσης κατά τη διεξαγωγή πολλαπλών συγκρίσεων. Η συγκεκριμένη μέθοδος προσφέρει μεγαλύτερο υπολογιστικό έλεγχο ως προς την ανίχνευση ψευδώς σημαντικών μεταβλητών (γονιδίων) σε σχέση με τις προαναφερθείσες μεθοδολογίες. Στη μέθοδο FDR ταξινομούνται αρχικά οι τιμές των p-value σε αύξουσα σειρά. Το νέο προσαρμοσμένο επίπεδο σημαντικότητας για N μεταβλητές και για $i = 1 \dots N$ η σειρά κατάταξης των p-values σε αύξουσα ταξινόμηση, δίνεται από τον τύπο:

$$p_{corr(i)} = \frac{i}{N} p_i$$

3.6 Σκοπός της Εργασίας

Σκοπός της παρούσας εργασίας είναι η εύρεση γονιδίων που εκφράζονται διαφορετικά ως προς την ασθένεια της πνευμονικής αρτηριακής υπέρτασης και αν μπορούν να χρησιμοποιηθούν σαν παράγοντες πρόβλεψης για τη συγκεκριμένη ασθένεια.

Πρώτο στάδιο είναι η συγκέντρωση δεδομένων γονιδιακής έκφρασης από δείγμα πνεύμονα μέσα από πειράματα μικροσυστοιχιών DNA ανάμεσα σε ασθενείς με πνευμονική αρτηριακή υπέρταση και μάρτυρες. Θα πραγματοποιηθεί αναζήτηση μέσα από τις βάσεις δεδομένων Gene Expression Omnibus και ArrayExpress. Η συγκέντρωση των μελετών θα γίνει με τη βοήθεια της συστηματικής ανασκόπησης σύμφωνα με το πρότυπο PRISMA και με εφαρμογή κριτηρίων αξιολόγησης.

Επόμενο στάδιο είναι η διεξαγωγή μετα-ανάλυσης. Εφαρμόζοντας το μοντέλο τυχαίων επιδράσεων και λαμβάνοντας ως μέγεθος επίδρασης την τυποποιημένη διαφορά των μέσων όπως προτάθηκε από τον Cohen, θα υπολογιστούν οι μέσοι όροι των δύο ανεξάρτητων συνόλων από κάθε πίνακα γονιδιακής έκφρασης και στη συνέχεια με διόρθωση του Hedges θα ανακτηθούν οι διορθωμένες τιμές και τα αντίστοιχα τυπικά σφάλματα. Με εφαρμογή της παραπάνω μεθοδολογίας επιτυγχάνεται η σύγκριση δεδομένων γονιδιακής έκφρασης μεταξύ πολλαπλών πλατφορμών από πειράματα μικροσυστοιχιών. Οι παραπάνω αναλύσεις θα πραγματοποιηθούν τόσο σε μεμονωμένο επίπεδο για κάθε μελέτη ξεχωριστά που θα συγκεντρωθεί, όσο και στο σύνολο όλων μαζί ταυτόχρονα.

Στη συνέχεια, τα ΔΕΓ που θα προκύψουν από τη συγκεντρωτική μετα-ανάλυση, θα χρησιμοποιηθούν παρακάτω ως είσοδοι σε εργαλεία εμπλουτισμού. Τα εργαλεία που θα χρησιμοποιηθούν είναι η βάση STRING και η πλατφόρμα bioCompendium.

Το δίκτυο πρωτεΐνης-πρωτεΐνης που θα προκύψει από την STRING σε συνδυασμό με τα μονοπάτια KEGG που θα ανακτηθούν για τα ΔΕΓ γονίδια από την bioCompendium, θα προσδώσουν πλούσιο υλικό για περαιτέρω συζήτηση και συμπεράσματα πάνω στην ασθένεια της πνευμονικής αρτηριακής υπέρτασης.

Κεφάλαιο 4 – Προτεινόμενη Μεθοδολογία της Παρούσας Εργασίας

4.1 Συστηματική Ανασκόπηση

Η συστηματική ανασκόπηση πραγματοποιήθηκε χρησιμοποιώντας ως είσοδο στη βάση δεδομένων GEO και Arrayexpress ένα σύνολο λέξεων-κλειδιών (queries) με στόχο την συγκέντρωση των μελετών που αφορούν την διεξαγωγή πειραμάτων από μικροσυστοιχίες cDNA ανάμεσα σε ασθενείς με πνευμονική αρτηριακή υπέρταση και μάρτυρες. Η αναζήτηση στις βάσεις δεδομένων διεξήχθη στις 22/04/2020.

Οι λέξεις κλειδιά που εισήχθησαν στο search box της κάθε βάσης ξεχωριστά ήταν οι εξής:

("Primary pulmonary hypertension" OR "PPH" OR "Pulmonary Arterial Hypertension" OR "PAH" OR "Pulmonary Hypertension" OR "PH" OR "Pulmonary Hypertension Due to Left Heart Disease" OR "Pulmonary Hypertension Due to Lung Disease" OR "Pulmonary Hypertension Due to Chronic Blood Clots in the Lungs" OR "Pulmonary Hypertension Due to Unknown Causes")

Στη συνέχεια, οι μελέτες που ανακτήθηκαν έπειτα από συστηματική ανασκόπηση πέρασαν από συγκεκριμένα στάδια αξιολόγησης και απόρριψης. Εφαρμόζοντας κριτήρια επιλεξιμότητας συγκεντρώθηκε ο τελικός αριθμός μελετών προς ανάλυση. Τα κριτήρια που εφαρμόστηκαν στρέφονταν γύρω από εκείνες τις μελέτες που περιείχαν δεδομένα σε ανθρώπινο ιστό, περιείχαν δείγματα ανάμεσα σε ασθενείς με ΠΑΥ και μάρτυρες και ήταν πειράματα πάνω σε μικροσυστοιχίες cDNA. Απορρίφθηκαν όλες οι υπόλοιπες μελέτες που δεν ήταν πειράματα πάνω σε μικροσυστοιχίες cDNA, περιείχαν δείγματα από ιστό σε διαφορετικό οργανισμό από τον ανθρώπινο, δεν ήταν μελέτες ασθενών-μαρτύρων ή ήταν μελέτες άλλου βιολογικού τύπου.

Επόμενο στάδιο ήταν η επεξεργασία των πινάκων γονιδιακής έκφρασης για κάθε μία από τις μελέτες που συγκεντρώθηκαν προς ανάλυση. Οι πίνακες γονιδιακής έκφρασης για κάθε μελέτη με κωδικό GSExxx βρίσκονται αναρτημένοι σε αντίστοιχους συνδέσμους στη βάση GEO με το ακρωνύμιο “series matrix file”.

4.2 Επεξεργασία Πινάκων Γονιδιακής Έκφρασης

Οι πίνακες γονιδιακής έκφρασης περιέχουν κάποια προβλήματα ως προς τα δεδομένα που παρουσιάζουν.

- ✓ Αφενός μεν, πρέπει η στήλη των ProbeIDs του πίνακα γονιδιακής έκφρασης να βρίσκεται σε αντιστοιχία με τη στήλη των ProbeIDs της πλατφόρμας (GPL) της microarray που χρησιμοποιήθηκε. Με αυτόν τον τρόπο κάθε ProbeID του πίνακα αντιστοιχίζεται με το

αντίστοιχο ProbeID όπως έχει οριστεί από την κατασκευή του microarray chip. Αν κάποιος ProbeID δεν βρίσκει αντιστοιχία τότε διαγράφεται.

- ✓ Αφετέρου δε, η λίστα με τα ονόματα των γονιδίων GENE_SYMBOL που βρίσκονται σε αντιστοιχία με τη λίστα ENTREZ_GENE_ID της βάσης Gene πρέπει να ενημερωθεί και να αλλάξουν οι παλιές ονομασίες γονιδίων στις καινούριες. Με αυτόν τον τρόπο η ανάλυση που θα ακολουθήσει σε περαιτέρω στάδια θα παρέχει ενημερωμένα δεδομένα για την τρέχουσα περίοδο.

Η επεξεργασία των πινάκων γονιδιακής έκφρασης πραγματοποιήθηκε με τη χρήση φύλλων εργασίας Excel. Το πρώτο βήμα επεξεργασίας έγινε με χρήση της γραμμής εργαλείων *Styles > Conditional Formatting > Highlight Cell Rules > Duplicate Values*. Με αυτόν τον τρόπο έγινε αντιστοίχιση των ProbeIDs μεταξύ πίνακα γονιδιακής έκφρασης και πλατφόρμας.

Στη συνέχεια, για την ενημέρωση των στοιχείων του πίνακα από τις στήλες GENE_SYMBOL και ENTREZ_GENE_ID χρησιμοποιήθηκε το αρχείο *Homo_sapiens.gene_info* το οποίο βρίσκεται αναρτημένο στο σύνδεσμο https://ftp.ncbi.nlm.nih.gov/gene/DATA/GENE_INFO/Mammalia/. Χρησιμοποιώντας τα ενημερωμένα στοιχεία του αρχείου της βάσης και τη συνάρτηση VLOOKUP του Excel πραγματοποιήθηκε η ενημέρωση των στοιχείων του πίνακα γονιδιακής έκφρασης.

Ο τελικός πίνακας γονιδιακής έκφρασης με ενημερωμένα στοιχεία για κάθε μελέτη που συγκεντρώθηκε έχει τη μορφή που φαίνεται στην ακόλουθη εικόνα.

ID	GENE_NAMES	GSM3290067	GSM3290068	GSM3290069	GSM3290070	GSM3290071	GSM3290072	GSM3290073	GSM3290074
7896754	SEPTIN7P13	5.90473	6.10208	6.42231	5.1607	5.70061	6.07588	6.93247	6.42393
7896759	LINC01128	4.55273	4.84461	4.80529	4.88309	5.1026	4.91048	5.81596	5.48804
7896761	SAMD11	3.46496	3.79232	3.61386	3.80648	3.76464	3.66815	3.69958	3.70442
7896779	KLHL17	4.66527	4.42225	4.23017	4.24995	4.23209	4.27803	4.26371	4.18945
7896798	PLEKHN1	4.08055	4.03726	3.8933	3.9736	3.95942	4.01871	4.02769	3.92104
7896817	ISG15	3.94129	4.06235	4.19495	4.24744	4.32919	4.29188	4.92728	3.8238
7896822	AGRN	5.37598	5.14281	5.6874	5.63432	5.78044	5.60937	5.81703	5.25783
7896859	MIR200B	3.66578	3.4863	3.66705	3.60517	3.61726	3.40832	3.57048	3.4875
7896861	MIR200A	3.34637	3.51946	3.19362	3.29542	3.1383	3.20264	3.19565	3.29508
7896863	MIR429	3.70554	3.67206	3.7703	3.64163	3.75647	3.75991	3.67504	3.9962
7896865	TLL10	3.96284	3.94696	4.05533	3.87473	3.87511	3.97099	4.21342	4.26785
7896878	B3GALT6	4.12305	4.16109	4.3861	4.36575	4.18645	4.37292	4.16719	4.12419
7896882	SCNN1D	3.7026	3.43943	3.66404	3.68311	3.55603	3.66077	3.66914	3.67457
7896908	PUSL1	4.42183	4.39859	4.40659	4.5127	4.48402	4.073	4.46211	4.34598
7896917	CPTP	4.20062	4.23479	4.42353	4.36477	4.28172	4.27359	4.14278	4.07241
7896921	TAS1R3	3.87353	3.85055	3.90016	3.78024	3.68241	3.73008	3.7013	3.79904
7896929	VWA1	5.01012	4.83021	5.29785	5.33013	5.11232	5.09098	5.04013	4.77602
7896952	ATAD3A	3.91255	3.987	4.43251	4.6043	4.26367	4.42264	4.12101	4.14576
7896985	MIB2	4.00023	4.05292	4.27653	4.23367	4.02471	4.05105	4.23575	3.99115
7897026	CALML6	3.42986	3.38158	3.41916	3.38206	3.41558	3.57826	3.33169	3.27714
7897034	GABRD	3.5962	3.33433	3.5843	3.6081	3.90739	3.58879	3.63672	3.49779
7897044	PRKCZ	7.14753	6.66883	6.66653	6.61841	6.72316	6.81273	6.78708	6.47588
7897068	SKI	6.52582	6.3206	6.40039	6.38353	6.55928	6.38946	6.62603	6.41688
7897078	RER1	8.29468	8.50801	8.69109	8.52198	8.59631	8.51468	8.15204	8.28708
7897089	PLCH2	4.39217	3.86343	4.09346	3.78415	4.1404	4.01089	4.0986	4.18059
7897119	PRXL2B	4.1809	4.21006	4.49318	4.53973	4.30187	4.89794	4.50949	4.41121
7897128	ACTRT2	3.41633	3.38561	3.42741	3.61063	3.54876	3.46405	3.48569	3.44126
7897132	PRDM16	3.83539	3.97683	4.16434	3.98033	3.89028	4.09907	4.08012	4.12134
7897154	ARHGEF16	3.78744	3.75763	3.69087	3.76395	3.64098	3.58363	3.78131	3.7758
7897172	TPRGLL	7.35852	6.89138	6.92019	6.91586	6.73699	7.00275	6.91377	6.83896

Εικόνα 4.1 Απόσπασμα πίνακα γονιδιακής έκφρασης μετά από την αρχική επεξεργασία.

Επιπρόσθετα, προστέθηκε η γραμμή case_control στο τέλος του πίνακα με ένδειξη 0 ή 1 (0=μάρτυρας, 1=ασθενής). Ο πίνακας κάθε μελέτης ταξινομήθηκε βάση της στήλης gene_names. Αμέσως επόμενο βήμα είναι η εισαγωγή δεδομένων κάθε μελέτης ξεχωριστά στο πρόγραμμα STATA για περαιτέρω επεξεργασία και ανάλυση.

Σε αυτό το σημείο έτρεξε για κάθε μελέτη ξεχωριστά η εντολή collapse του STATA. Η εντολή αυτή εντοπίζει τις διπλότυπες καταστάσεις των gene_names –κάποιες

μέθοδος FDR των Benjamini και Hochberg. Από τις πέντε αυτές μεθόδους διατηρήθηκαν τα δεδομένα που επιστράφηκαν ως στατιστικά σημαντικά σε επίπεδο 1% μόνο της μεθόδου FDR.

4.4 Ανάλυση Εμπλουτισμού των Διαφορικά Εκφραζόμενων Γονιδίων

Μόλις ολοκληρώθηκαν οι στατιστικοί έλεγχοι, τελευταίο στάδιο ήταν η χρήση εργαλείων εμπλουτισμού για την ολοκλήρωση της παρούσας μελέτης. Τα εργαλεία εμπλουτισμού που χρησιμοποιήθηκαν ήταν η βάση bioCompendium για αναγνώριση των μονοπατιών KEGG στα οποία συμμετέχουν τα ΔΕΓ και η βάση String (11.0) για την παραγωγή δικτύου αλληλεπίδρασης πρωτεϊνών – πρωτεϊνών.

Η bioCompendium (<https://biocompendium.uni.lu/>) είναι μια διαδικτυακή πλατφόρμα υψηλής απόδοσης για ανάλυση δεδομένων. Είναι σχεδιασμένη για να λαμβάνει ως είσοδο λίστες γονιδίων ή πρωτεϊνών και στη συνέχεια να προσφέρει ένα ευρύ φάσμα βιολογικών πληροφοριών.

Η βάση δεδομένων KEGG συνδέεται μέσω της πλατφόρμας bioCompendium και απαντά σε βιοχημικά μονοπάτια, μοριακές σχέσεις, μεταβολικές διεργασίες και γενικότερα προσφέρει ολοκληρωμένες πληροφορίες με την απεικόνιση βιολογικών δικτύων γύρω από μια λίστα γονιδίων.

Η STRING (Search Tool for the Retrieval of INteracting Genes/proteins, <https://string-db.org/>) είναι μια ολοκληρωμένη βάση δεδομένων των γνωστών και καταγεγραμμένων αλληλεπιδράσεων πρωτεϊνών-πρωτεϊνών. Οι αλληλεπιδράσεις περιλαμβάνουν άμεσες (φυσικές) και έμμεσες (λειτουργικές) συσχετίσεις και προέρχονται από υπολογιστική πρόβλεψη, από μεταφορά γνώσεων μεταξύ οργανισμών και από αλληλεπιδράσεις που συγκεντρώνονται από άλλες (πρωτογενείς) βάσεις δεδομένων.

Ένα αρχείο με επέκταση .txt που περιείχε τη λίστα με τα ΔΕΓ που προέκυψαν από τη μετα-ανάλυση δόθηκε ως είσοδος στην πλατφόρμα bioCompendium με επιπρόσθετη επιλογή ID/Document Type: ascii. Στη συνέχεια επιλέχθηκε το πεδίο «Pathway Information» για να εμφανιστούν τα μονοπάτια της βάσης KEGG στα οποία συμμετέχουν τα γονίδια.

Στη βάση String ακολουθήθηκε μια σειρά επιλογών από το πεδίο «*multiple proteins*» για να παραχθεί το δίκτυο αλληλεπιδράσεων πρωτεϊνών με βάση τα στατιστικά σημαντικά γονίδια που προέκυψαν παραπάνω. Από τις βασικές ρυθμίσεις επιλέχθηκε η εμφάνιση του πλήρους δικτύου, η μοριακή αντίδραση ως ερμηνεία των ακμών μεταξύ κόμβων και για βαθμό ορίου εμπιστοσύνης το 0,400.

Μια τελευταία κίνηση ήταν η διασταύρωση των διαφορικά εκφραζόμενων γονιδίων που προέκυψαν μέσα από τη μετα-ανάλυση με βάσεις δεδομένων ευρυγονιδιωματικών μελετών (GWAS) ή κλασικών μελετών γενετικής συσχέτισης. Αυτό γίνεται για να φανεί αν και κατά πόσο τα ΔΕΓ της μετα-ανάλυσης εμπίπτουν με όσα δεδομένα έχουν συγκεντρωθεί τα τελευταία χρόνια γύρω από την ασθένεια της πνευμονικής αρτηριακής υπέρτασης. Οι βάσεις δεδομένων που χρησιμοποιήθηκαν ήταν η GWAS Catalog και η DisGeNET.

Κεφάλαιο 5 – Αποτελέσματα

5.1 Αποτελέσματα Συστηματικής Ανασκόπησης

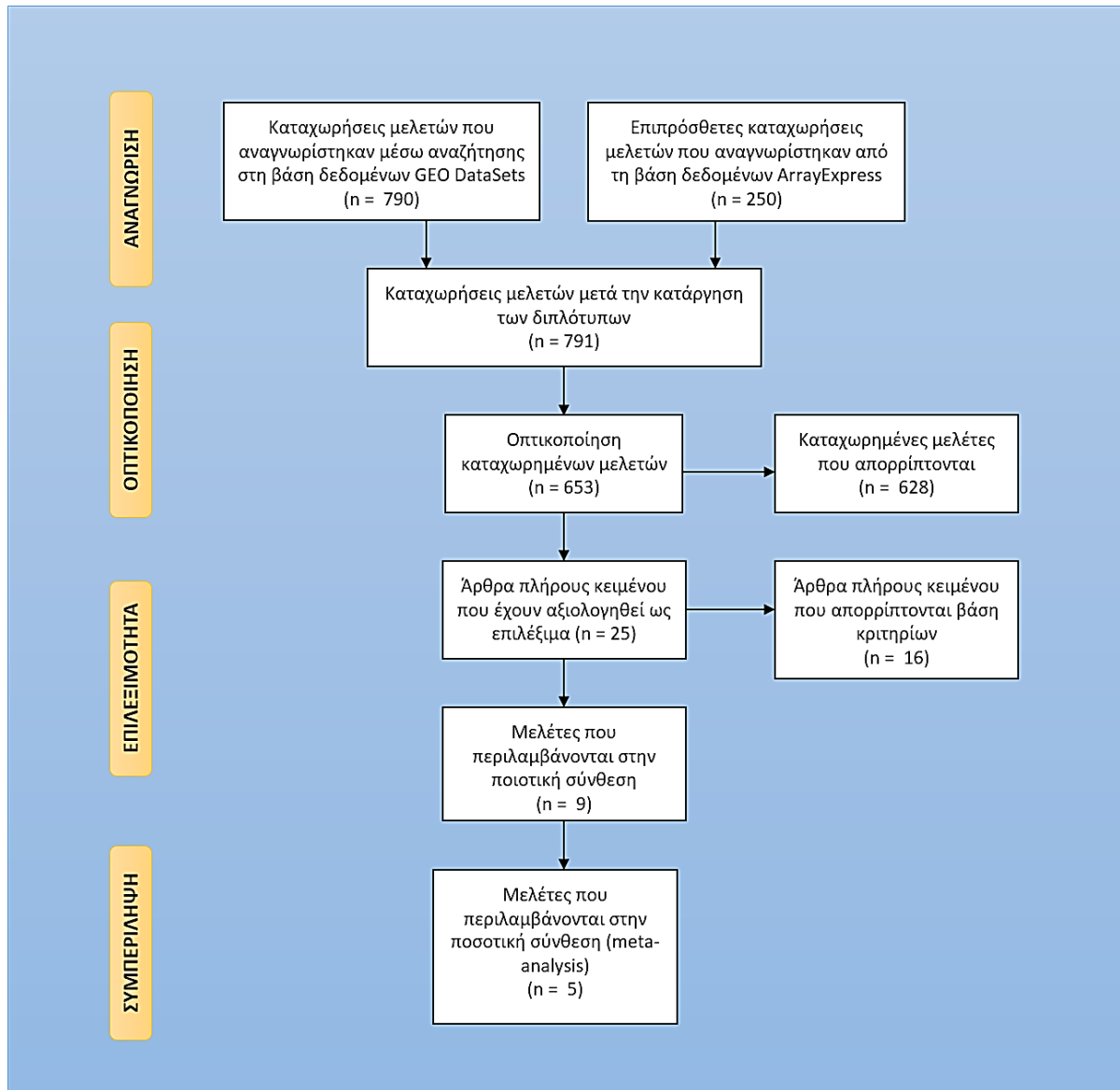
Οι λέξεις-κλειδιά που δόθηκαν ως είσοδοι στο κουτί αναζήτησης στις βάσεις GEO DataSets και ArrayExpress επέστρεψαν συγκεκριμένο αριθμό αποτελεσμάτων. Πιο αναλυτικά

- ❖ Για την βάση GEO DataSets μετά την είσοδο του βασικού ερωτήματος, επιλέχθηκε από την αριστερή στήλη το πεδίο *Entry type* > *Series* και από τη δεξιά στήλη το πεδίο *Top organisms* > *Homo sapiens*. Τα αποτελέσματα της αναζήτησης περιορίστηκαν και κατέληξαν να είναι το σύνολο 790.
- ❖ Για την βάση ArrayExpress τα αποτελέσματα μετά την είσοδο του ερωτήματος κατέληξαν να είναι το σύνολο 250.

Για τη συγκέντρωση μελετών πάνω στην ασθένεια της ΠΑΥ συμπεριλήφθηκαν 5 μελέτες, οι οποίες εξέταζαν την ασθένεια από δείγματα σε ιστούς πνευμόνων ανάμεσα σε υγιείς και νοσούντες με ΠΑΥ. Οι μελέτες αυτές είναι οι GSE117261, GSE113439, GSE48149, GSE15197, GSE79786.

Παρακάτω παρουσιάζεται το διάγραμμα ροής για συστηματική ανασκόπηση (Εικόνα 5.1). Το διάγραμμα αναφέρεται σε μελέτες που συγκεντρώθηκαν και εξέταζαν ιστούς πνεύμονα ανάμεσα σε ασθενείς και μάρτυρες ως προς την ασθένεια της πνευμονικής αρτηριακής υπέρτασης και βρέθηκαν να είναι στο σύνολο 5.

Επιπλέον, παρουσιάζεται οι τελικός πίνακας με συγκεντρωτικά δεδομένα όπως χρησιμοποιήθηκαν στο επόμενο βήμα που ήταν η μετα-ανάλυση (Πίνακας 5.1). Ο πίνακας παρουσιάζει τις 5 μελέτες που προσφέρουν δεδομένα γονιδιακής έκφρασης όπως προέκυψαν από το διάγραμμα ροής.



Εικόνα 5.1 Διάγραμμα ροής για συστηματική ανασκόπηση για την ασθένεια της πνευμονικής αρτηριακής υπέρτασης.

Πίνακας 5.1 Μελέτες που συγκεντρώθηκαν για μετα-ανάλυση για την ασθένεια της πνευμονικής αρτηριακής υπέρτασης.

GSE	CASES	CONTROLS	PLATFORM NAME	PLATFORM	TOTAL GENES	PMID
GSE117261	58	25	Affymetrix Human Gene 1.0 ST Array	GPL6244	18562	-
GSE113439	14	11	Affymetrix Human Gene 1.0 ST Array	GPL6244	17669	30963672
GSE48149	8	9	Illumina HumanRef-8 v3.0 expression beadchip	GPL16221	18187	21360508
GSE15197	18	13	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F	GPL6480	18396	20081107
GSE79786	3	3	Illumina HumanHT-12 V4.0 expression beadchip	GPL10904	13766	27456128

5.2 Αποτελέσματα Στατιστικού Ελέγχου

Όπως αναφέρθηκε παραπάνω πραγματοποιήθηκαν δύο μορφές στατιστικού ελέγχου.

- ✓ Στατιστική ανάλυση στα σύνολα δεδομένων κάθε μελέτης του πίνακα 1
- ✓ Μετα-ανάλυση για όλες τις μελέτες συγκεντρωτικά του πίνακα 1

Η στατιστική ανάλυση πραγματοποιήθηκε για κάθε μελέτη ξεχωριστά του πίνακα 5.1. Από τον συγκεκριμένο έλεγχο προέκυψαν οι στατιστικά σημαντικές μετρήσεις για κάθε μελέτη σε επίπεδο σημαντικότητας 1% και σε επίπεδο σημαντικότητας 1% με διόρθωση FDR. Τα αποτελέσματα για κάθε μελέτη παρουσιάζονται στον παρακάτω πίνακα (Πίνακας 5.2).

Πίνακας 5.2 Στατιστικός έλεγχος δεδομένων γονιδιακής έκφρασης για κάθε μελέτη.

GSE	Σημαντικά γονίδια σε επίπεδο σημαντικότητας 1%	Σημαντικά γονίδια με διόρθωση FDR σε επίπεδο σημαντικότητας 1%
GSE117261	3618	2008
GSE113439	8991	8165
GSE48149	1304	6
GSE15197	5450	3427
GSE79786	960	0

Ο δεύτερος στατιστικός έλεγχος αφορούσε τη διεξαγωγή μετα-ανάλυσης στα σύνολα των μελετών του πίνακα 1.

Από το σετ μελετών του πίνακα 5.1 επιστράφηκαν 1099 στατιστικά σημαντικά αποτελέσματα και ύστερα από τη μέθοδο διόρθωσης FDR μόλις 195 σε επίπεδο σημαντικότητας 1%. Από τα 195 γονίδια που επιστράφηκαν ως διαφορετικά εκφραζόμενα, τα 149 γονίδια υποεκφράζονταν και τα υπόλοιπα 46 υπερεκφράζονταν. Τα αποτελέσματα της μετα-ανάλυσης παρουσιάζονται σε δύο ξεχωριστούς πίνακες. Ο πρώτος (Πίνακας 5.3) απεικονίζει τα 46 ΔΕΓ που βρέθηκαν να υπερεκφράζονται και ο δεύτερος (Πίνακας 5.4) τα 149 ΔΕΓ που βρέθηκαν να υποεκφράζονται μαζί με τις αντίστοιχες εκτιμήσεις (g , SE_g , z , p -value).

Πίνακας 5.3 Τα 46 ΔΕΓ που υπερεκφράζονταν μετά από μετα-ανάλυση.

GENE	g	SE_g	z	p-value
ARL17A	1.794674	0.410412	4.372858	1.23E-05
ATP6V0CP3	1.628504	0.399671	4.074615	4.61E-05
BORCS8	1.558943	0.395405	3.942653	8.06E-05
CA2	0.823829	0.195656	4.210597	2.55E-05
CNOT6L	0.958255	0.212221	4.515369	6.32E-06
CT47A11	1.562328	0.395609	3.949173	7.84E-05
DDX17	0.643325	0.161888	3.973891	7.07E-05
DKFZP434A062	1.687888	0.403422	4.183928	2.87E-05
EPB42	0.688786	0.166015	4.148927	3.34E-05
EPHX2	0.864812	0.17581	4.919012	8.70E-07
FGF14_AS2	1.597304	0.39774	4.015952	5.92E-05
FOXN3_AS1	1.644613	0.400678	4.104572	4.05E-05
FREM1	1.061731	0.218743	4.853791	1.21E-06
GAREM1	0.949819	0.169899	5.590496	2.26E-08
GBP5	0.812676	0.167306	4.857417	1.19E-06
HBD	1.043961	0.170932	6.107458	1.01E-09
HERC5	0.626571	0.158066	3.963978	7.37E-05
IL12RB2	1.123402	0.211597	5.309158	1.10E-07
ITGB3	0.654902	0.158343	4.135975	3.53E-05
JMJD7	1.915059	0.418652	4.574341	4.78E-06
KRTAP2_1	1.692826	0.403738	4.19288	2.75E-05
LBX2_AS1	1.578463	0.396588	3.980113	6.89E-05
LGALS1	0.868438	0.198696	4.370683	1.24E-05
LINC00654	1.558266	0.395364	3.941346	0.000081
LOC646938	1.711747	0.404957	4.226986	2.37E-05
LPL	0.841796	0.170923	4.925004	8.44E-07
LRRC31	0.750582	0.166052	4.520167	6.18E-06
MIR99AHG	0.829208	0.167957	4.937019	7.93E-07
MYCNOS	1.858521	0.414736	4.481212	7.42E-06
MYOM1	0.705346	0.165733	4.255918	2.08E-05
NIFK_AS1	1.587352	0.39713	3.99706	6.41E-05

PCOLCE2	0.675236	0.17217	3.92191	8.78E-05
PDE8B	0.958432	0.162611	5.894013	3.77E-09
PIEZO2	0.943659	0.16937	5.57159	2.52E-08
PTPRD	0.893397	0.176423	5.063958	4.11E-07
RAVER2	0.728704	0.159089	4.580492	4.64E-06
RIMKLB	0.76936	0.159317	4.829126	1.37E-06
SLC4A1	0.854601	0.203027	4.209289	2.56E-05
SMIM1	2.270467	0.445001	5.102164	3.36E-07
TCAF1	0.784547	0.184252	4.258019	2.06E-05
TCF4	0.815565	0.198651	4.105523	4.03E-05
TMEM8B	0.830011	0.177725	4.670191	3.01E-06
TOB1_AS1	1.825299	0.412473	4.425257	9.63E-06
UBL7_AS1	1.575056	0.39638	3.973598	7.08E-05
SNORD56B	0.968009	0.196343	4.9302	8.21E-07
SNORD53B	0.902403	0.20898	4.318141	1.57E-05

Πίνακας 5.4 Τα 149 ΔΕΓ που υποεκφράζονται μετά από μετα-ανάλυση.

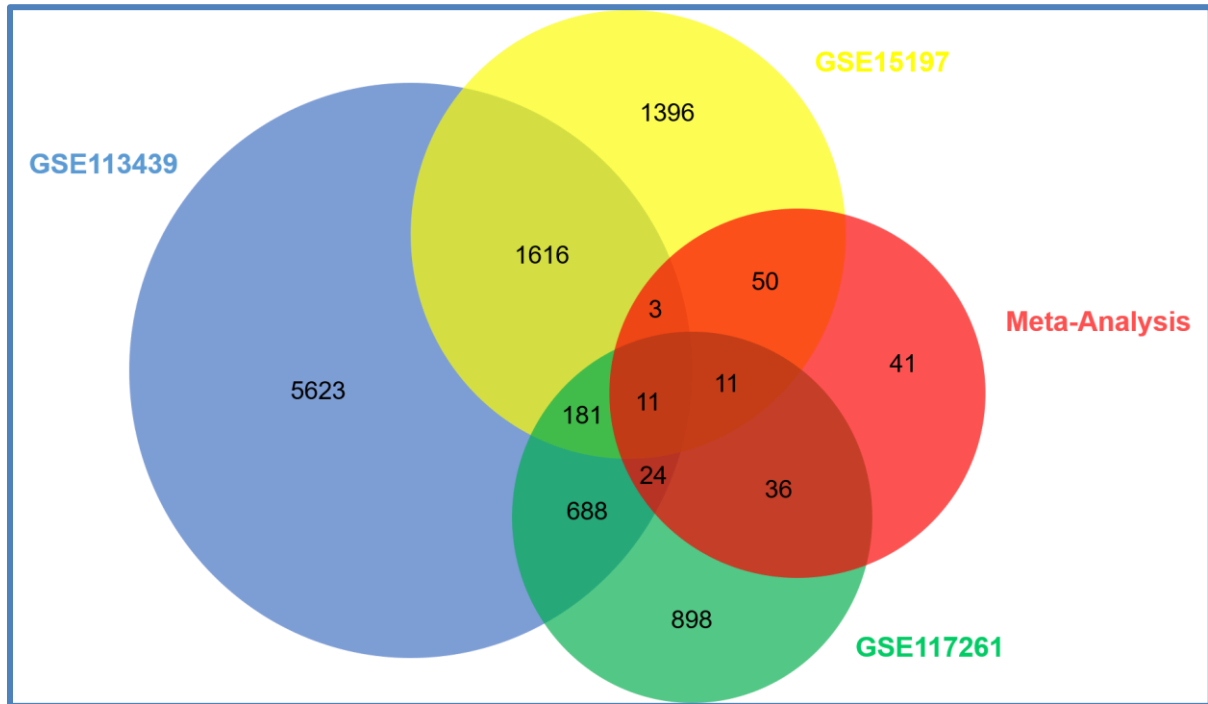
GENE	g	se_g	z	p-value
ABITRAM	-1.384	0.303506	-4.56005	5.11E-06
ACYPI	-0.77014	0.15926	-4.83572	1.33E-06
ADGRG1	-0.78517	0.17482	-4.49132	7.08E-06
ADRA1A	-1.07861	0.272443	-3.95902	7.53E-05
ALG8	-0.77036	0.178567	-4.31411	0.000016
AP2M1	-0.88625	0.161092	-5.50151	3.77E-08
APOBEC2	-0.84421	0.167561	-5.0382	4.70E-07
AQP1	-1.22298	0.22264	-5.49309	3.95E-08
AQP5	-0.67002	0.165457	-4.04952	5.13E-05
ARHGAP9	-0.77814	0.166589	-4.67103	3.00E-06
ATG4A	-1.09508	0.181634	-6.02903	1.65E-09
BCL7B	-0.72866	0.173584	-4.19771	0.000027
BPIFA1	-0.92738	0.168927	-5.48982	4.02E-08
BPIFB1	-0.75775	0.16791	-4.51285	6.40E-06
BVES	-1.02856	0.164409	-6.25608	3.95E-10
C1D	-0.64233	0.157707	-4.07295	4.64E-05
C7ORF31	-0.68562	0.158399	-4.32842	0.000015
CACNA2D3	-0.72793	0.165855	-4.38896	1.14E-05
CCDC84	-1.85356	0.414396	-4.47291	7.72E-06
CCNYL1	-1.27286	0.272024	-4.67921	2.88E-06
CDC42EP2	-0.76566	0.159205	-4.80928	1.51E-06
CEP170B	-0.72685	0.174376	-4.16828	3.07E-05
CHAMP1	-1.01234	0.240785	-4.20434	2.62E-05
CHTOP	-0.87256	0.179873	-4.85098	1.23E-06

CIAO2A	-0.74827	0.18925	-3.95386	7.69E-05
CLDN10	-0.96457	0.170258	-5.66533	1.47E-08
CLEC5A	-0.99627	0.18779	-5.30522	1.13E-07
CLPTM1	-0.91483	0.166865	-5.48244	4.19E-08
CLTA	-1.07773	0.165986	-6.49293	8.42E-11
COLCA2	-1.78139	0.409526	-4.34989	1.36E-05
COPS6	-0.66153	0.158815	-4.16538	3.11E-05
COPS7A	-0.83952	0.161009	-5.21408	1.85E-07
COX1	-2.11942	0.433454	-4.88962	1.01E-06
CTNS	-1.03578	0.246592	-4.20039	2.66E-05
CYB5A	-0.71871	0.159035	-4.51917	6.21E-06
CYP4Z1	-1.47006	0.307837	-4.77545	1.79E-06
CYTH4	-1.16506	0.236075	-4.93511	8.01E-07
CYTOR	-1.37468	0.30309	-4.53555	5.75E-06
DAAM2	-0.68229	0.159344	-4.28189	1.85E-05
DHTKD1	-0.89549	0.161755	-5.5361	3.09E-08
DMAC2	-1.2274	0.266985	-4.59726	4.28E-06
DOLK	-1.07761	0.247931	-4.34643	1.38E-05
DTD2	-0.69566	0.169583	-4.10218	4.09E-05
EEF2KMT	-0.72318	0.166594	-4.34093	1.42E-05
EEPD1	-0.73942	0.166145	-4.45047	8.57E-06
EHD1	-0.82103	0.192681	-4.26111	2.03E-05
EMP3	-0.87342	0.161664	-5.40269	6.57E-08
FAHD1	-0.75625	0.15906	-4.75449	1.99E-06
FAM13A_AS1	-1.92827	0.419579	-4.59573	4.31E-06
FAM24B	-1.19717	0.266209	-4.4971	6.89E-06
FDPSP2	-1.51511	0.194655	-7.78358	7.11E-15
GIMAP6	-1.385	0.297315	-4.65835	3.19E-06
GJA4	-0.70147	0.166128	-4.22245	2.42E-05
GNG10	-1.10915	0.261403	-4.24306	0.000022
GOLGA6A	-1.14439	0.277934	-4.11751	3.83E-05
GPR141	-0.79453	0.166931	-4.75963	1.94E-06
GPR89B	-1.29611	0.291925	-4.43987	9.00E-06
GRB2	-0.91419	0.161667	-5.6548	1.56E-08
H3C13	-1.63372	0.399996	-4.08433	4.42E-05
HCCS	-0.74599	0.173892	-4.28996	1.79E-05
HHEX	-0.75256	0.15904	-4.73187	2.22E-06
HHIPL2	-0.69163	0.165952	-4.16764	3.08E-05
HS3ST3B1	-0.6467	0.157503	-4.10597	4.03E-05
HTATIP2	-1.12919	0.210249	-5.37071	7.84E-08
IFITM2	-1.19621	0.291444	-4.10442	4.05E-05
IL10RB	-1.16726	0.275159	-4.24212	2.21E-05
ITGA10	-0.6187	0.157445	-3.92964	8.51E-05
KCNIP3	-1.02277	0.234407	-4.36322	1.28E-05

LAPTM5	-0.69104	0.158671	-4.35519	1.33E-05
LILRA2	-1.29623	0.256929	-5.04508	4.53E-07
LILRB5	-0.7273	0.159111	-4.57101	4.85E-06
LINC00667	-1.77755	0.409271	-4.34322	0.000014
LINC00938	-1.68916	0.403503	-4.18623	2.84E-05
LINC01000	-1.70869	0.404759	-4.2215	2.43E-05
MBP	-0.80217	0.159767	-5.02087	5.14E-07
MICB	-0.66108	0.157773	-4.19006	2.79E-05
MIR100HG	-2.05577	0.428739	-4.79493	1.63E-06
MLF2	-0.903	0.211834	-4.26279	2.02E-05
MRPL9	-0.68775	0.158327	-4.34388	0.000014
MYD88	-0.75906	0.159058	-4.77221	1.82E-06
NAAA	-0.71283	0.158964	-4.48425	7.32E-06
NECAP1	-0.71288	0.158843	-4.48793	7.19E-06
NEDD8	-1.05588	0.260476	-4.05367	5.04E-05
NES	-0.81981	0.161075	-5.08964	3.59E-07
NEURL1	-0.66963	0.165591	-4.0439	5.26E-05
NOD2	-0.73563	0.166427	-4.42014	9.86E-06
OR7E14P	-1.65344	0.401234	-4.12089	3.77E-05
PCSK7	-0.62687	0.158854	-3.94621	7.94E-05
PDE1C	-0.76781	0.159561	-4.81199	1.49E-06
PLCD3	-1.38294	0.250776	-5.51466	3.49E-08
PLXNA1	-0.83988	0.161183	-5.21074	1.88E-07
POM121L9P	-1.91193	0.418433	-4.56925	4.89E-06
PON2	-1.02506	0.237173	-4.32201	1.55E-05
POU2F2	-0.81152	0.167517	-4.84443	1.27E-06
PPM1L	-0.7279	0.165836	-4.38928	1.14E-05
PPP1R18	-1.14057	0.186588	-6.11275	9.79E-10
PPP4C	-0.88044	0.224137	-3.92811	8.56E-05
PQBP1	-0.9425	0.210606	-4.4752	7.63E-06
PRICKLE4	-1.23722	0.265397	-4.66176	3.14E-06
PSMD9	-0.97821	0.239726	-4.08053	4.49E-05
PYM1	-0.91752	0.173826	-5.27835	1.30E-07
RAB1F	-0.98549	0.171965	-5.73075	1.00E-08
RAC1	-0.76574	0.179738	-4.26031	2.04E-05
RAPGEFL1	-0.94691	0.232024	-4.08109	4.48E-05
RHOG	-1.2351	0.301512	-4.09636	0.000042
RIPOR3	-0.87222	0.180206	-4.8401	1.30E-06
RIT2	-0.64764	0.165539	-3.9123	9.14E-05
RNASE6	-0.64978	0.164861	-3.94142	0.000081
ROCK1P1	-2.04308	0.42781	-4.77567	1.79E-06
S1PR1	-1.19317	0.246905	-4.83249	1.35E-06
SDHAF3	-0.73248	0.165848	-4.41655	0.00001
SDHC	-0.76275	0.15949	-4.78247	1.73E-06

SEC61B	-0.68692	0.169975	-4.04127	5.32E-05
SLC12A8	-0.72027	0.158976	-4.53071	5.88E-06
SLC35B1	-0.7902	0.160418	-4.92587	8.40E-07
SLC3A2	-1.27262	0.226182	-5.62655	1.84E-08
SLC8B1	-0.90348	0.226898	-3.98188	6.84E-05
SLX4	-0.86705	0.193738	-4.47537	7.63E-06
SMPD1	-1.18691	0.29798	-3.98317	0.000068
SNHG32	-1.15155	0.294269	-3.91327	9.11E-05
SNX3	-1.14794	0.253405	-4.53004	5.90E-06
SRPX	-0.75783	0.186028	-4.07375	4.63E-05
STK40	-0.88059	0.161858	-5.44049	5.31E-08
TCEAL1	-0.83549	0.175947	-4.74853	2.05E-06
TCEAL4	-1.21109	0.207164	-5.84605	5.03E-09
TINAGL1	-0.85887	0.161059	-5.33265	9.68E-08
TINF2	-0.79718	0.160159	-4.97743	6.44E-07
TMBIM6	-0.70997	0.159181	-4.46014	8.19E-06
TMED7_TICAM2	-1.93644	0.420154	-4.60887	4.05E-06
TMEM106A	-0.86897	0.172041	-5.05095	4.40E-07
TMEM26	-0.78924	0.166912	-4.72848	2.26E-06
TNNT2	-0.88784	0.213703	-4.15455	3.26E-05
TOR1A	-0.8015	0.15982	-5.01498	5.30E-07
TSEN15	-0.82136	0.160268	-5.12495	2.98E-07
TVP23B	-1.37817	0.303119	-4.54662	5.45E-06
VBP1	-0.75257	0.174624	-4.30967	1.63E-05
VPS52	-0.8439	0.172976	-4.87871	1.07E-06
WASHC5	-1.21457	0.192734	-6.30183	2.94E-10
ZNF674_AS1	-1.55108	0.394931	-3.92747	8.58E-05
ZNF728	-1.26305	0.319898	-3.9483	7.87E-05
ZNF792	-0.68146	0.158519	-4.29891	1.72E-05
DRGX	-0.7426	0.177789	-4.17686	2.96E-05
ERVMER34_1	-0.80434	0.188313	-4.2713	1.94E-05
TCTEX1D4	-0.84772	0.180313	-4.70141	2.58E-06
MIR548F5	-6.38256	1.312326	-4.86355	1.15E-06
LOC339803	-1.16804	0.250983	-4.65389	3.26E-06
MIR126	-0.81658	0.207031	-3.94423	8.01E-05
PVALEF	-0.96716	0.228445	-4.23366	0.000023
SNORA2A	-0.91408	0.208696	-4.37998	1.19E-05

Για την σύγκριση των αποτελεσμάτων μεταξύ στατιστικού ελέγχου σε κάθε μελέτη μεμονωμένα και της συγκεντρωτικής μετα-ανάλυσης, παρουσιάζεται παρακάτω ένα διάγραμμα Venn (Hulsen et al., 2008) (Εικόνα 5.2). Στο διάγραμμα συμμετέχουν τα στατιστικά σημαντικά αποτελέσματα σε επίπεδο σημαντικότητας 1% κατά FDR των μελετών GSE15197, GSE113439 και GSE117261 (3427, 8165 και 2008 ΔΕΓ αντιστοίχως), καθώς και τα αποτελέσματα των 195 ΔΕΓ όπως προέκυψαν από την μετα-ανάλυση.



Εικόνα 5.2 Διάγραμμα Venn για την απεικόνιση των αποτελεσμάτων μεταξύ στατιστικού ελέγχου για κάθε μελέτη ξεχωριστά και για τη συγκεντρωτική μετα-ανάλυση.

Σύμφωνα με το διάγραμμα απεικονίζεται η διαφορά ως προς τον προσδιορισμό των ΔΕΓ μέσα από τη διαδικασία της μετα-ανάλυσης και μέσα από μεμονωμένους στατιστικούς ελέγχους. Σαράντα-ένα γονίδια αναγνωρίστηκαν ως διαφορεικά εκφραζόμενα με μετα-ανάλυση (41 από τα 195 στο σύνολο) και όχι με μεμονωμένους ελέγχους σε κάποια από τις υπόλοιπες τρεις μελέτες. Αξιοσημείωτο είναι επίσης το γεγονός ότι 181 κοινά ΔΕΓ αναγνωρίστηκαν μέσα από τις τρεις διαφορετικές μελέτες, ενώ η μεθοδολογία της μετα-ανάλυσης ενσωμάτωσε γονίδια τα οποία είναι διαφορεικά εκφρασμένα σε κάθε μελέτη μεμονωμένα.

Σύμφωνα με την παρούσα μεθοδολογία για μετα-ανάλυση ομαδοποιούνται τα πραγματικά διαφορετικά εκφραζόμενα γονίδια γύρω από όλες τις μελέτες. Αυτό δε σημαίνει απαραίτητα ότι θα συμπεριληφθεί η τομή τους ή μέρος αυτής στην ομαδοποίηση. Έτσι ενισχύεται η ευρωστία και η αξιοπιστία αυτής της μεθόδου.

5.3 Αποτελέσματα μετά από χρήση Εργαλείων Εμπλουτισμού

5.3.1 Ανίχνευση Βιολογικών Διεργασιών με χρήση του bioCompendium

Τα 195 ΔΕΓ που αφορούν την ασθένεια της πνευμονικής αρτηριακής υπέρτασης χρησιμοποιήθηκαν ως είσοδοι στην πλατφόρμα bioCompendium, όπου ανιχνεύθηκαν 11 KEGG βιοχημικά μονοπάτια στα οποία συμμετέχουν τα γονίδια αυτά και τα κυριότερα από αυτά αναλύονται παρακάτω. Το όριο σημαντικότητας ήταν το 0,05. Τα ευρήματα παρουσιάζονται στον πίνακα 5.5. Στο παρόν σημείο να αναφερθεί ότι πραγματοποιήθηκε αναζήτηση και στην βάση δεδομένων PANTHER για διερεύνηση τυχόν βιοχημικών μονοπατιών, όμως δεν επιστράφηκε κάποια στατιστικά σημαντική πληροφορία για οποιοδήποτε από τα 195 ΔΕΓ.

- A. Τα γονίδια RAC1, JMJD7 και GRB2 βρέθηκαν να εμπλέκονται στο μονοπάτι υποδοχέων των ανοσοσφαιρινών (FcεRI signaling pathway). Οι υποδοχείς αυτοί εκφράζονται σε σιτευτικά κύτταρα και συνδέονται με την απόκριση σε αλλεργικές αντιδράσεις. Πρόσφατες μελέτες δείχνουν ότι τα σιτευτικά κύτταρα ρυθμίζουν ενεργά τη φλεγμονή και μπορούν να παίξουν σημαντικό ρόλο σε αναπνευστικές ασθένειες. Η αποδόμηση των σιτευτικών κυττάρων συνεπάγεται απελευθέρωση κυτοκινών, οι οποίες δρουν ως φλεγμονώδεις μεσολαβητές για την αποκατάσταση της ομοιόστασης στα πνευμονικά αγγεία (Erfjält, 2014). Τα αιμοφόρα αγγεία είναι η κύρια οδός για τα φλεγμονώδη κύτταρα να ταξιδέψουν στον τόπο τραυματισμού. Συνεπώς, η φλεγμονή και η αγγειογένεση είναι πολύπλοκες διαδικασίες που είναι στενά συνδεδεμένες (Pullamsetti et al., 2011).
- B. Το γονίδιο COX1 βρέθηκε να εμπλέκεται στο μονοπάτι αραχιδονικού οξέος. Κατά το μεταβολισμό του αραχιδονικού οξέως συνθέτονται δύο προϊόντα φλεγμονώδους δράσεως, οι κυκλοοξυγενάσες (προσταγλανδίνες και θρομβοξάνες) και οι λιποοξυγενάσες (λευκοτριένες). Η κυκλοοξυγενάση COX1 (cyclooxygenase-1/Cytochrome c oxidase subunit 1) μετατρέπει το αραχιδονικό οξύ σε προσταγλανδίνη, η οποία με τη σειρά της ενεργοποιεί παράγωγα. Ένα σημαντικό παράγωγο που απελευθερώνεται από τα ενδοθηλιακά κύτταρα αιμοφόρων αγγείων είναι η προστακυκλίνη. Η συγκεκριμένη ορμόνη αποτελεί ισχυρό αγγειοδιασταλτικό, οδηγεί σε συστολή του πνευμονικού αγγείου και έχει συσχετιστεί με την αντιμετώπιση καρδιοαναπνευστικών νόσων, όπως η ΠΑΥ (Mitchell et al., 2015).
- Γ. Το γονίδιο JMJD7 από την άλλη, συμμετέχει επίσης στα βιοχημικά μονοπάτια αλινολενικού οξέως και λινολεϊκού οξέως, στοιχεία που συνθέτουν το αραχιδονικό οξύ. Η πρωτεΐνη JMJD7 (Jumonji C domain-containing 7) ανήκει στην οικογένεια πρωτεϊνών Jumonji, οι οποίες συμπεριφέρονται κυρίως ως επιγενετικοί ρυθμιστές που απομεθυλιώνουν ιστόνες. Η συγκεκριμένη μικρή πρωτεΐνη (316 αμινοξέα) έχει ταυτοποιηθεί ως πρωτεάση που διασπά τις μεθυλιωμένες ιστόνες αργινίνης.

Το γονίδιο JMJD7 βρίσκεται πολύ κοντά στο γονίδιο PLA2G4B που κωδικοποιεί μια φωσφολιπάση (Phospholipase A2 group IVB). Αν και υπάρχουν αρκετά επιστημονικά κενά, μελέτες έχουν προτείνει την φωσφωλιπάση A2 (PLA2) ως αντιφλεγμονώδη δείκτη στην φαρμακολογία πάνω σε χρόνιες ασθένειες του αναπνευστικού συστήματος (Pniewska & Pawliczak 2013).

Η σύντηξη των δύο γονιδίων JMJD7-PLA2G4B (χρωμόσωμα 15) κωδικοποιεί μια νέα πρωτεΐνη, την JMJD7-PLA2G4B Readthrough. Χάρη σε αυτό το συνδυασμό, είναι εφικτή η απελευθέρωση αραχιδονικού οξέως και όλα τα εν συνεχεία παράγωγά του. Αξίζει να σημειωθεί, ότι η JMJD7-PLA2G4B εμφανίζεται στο βιολογικό σύστημα του αγγειακού ενδοθηλιακού αυξητικού παράγοντα (VEGF, Vascular Endothelial Growth Factor) και στο σύστημα συστολής αγγειακών λείων μυών.

- Δ. Τέσσερις πρωτεΐνες γονιδίων εμπλέκονται στην οδό των λυσοσωμάτων. Πιο συγκεκριμένα, η πρωτεΐνη του γονιδίου LPTM5 (Lysosomal-associated transmembrane protein 5) λειτουργεί ως μεταφορέας λιγασών ουβικιτίνης από το σύμπλεγμα Golgi προς τον ενδο-λυσοσωμικό χώρο. Η πρωτεΐνη του CTNS (Cystinosis) μετακινεί τα αποδιαταγμένα τμήματα πρωτεϊνών έξω από το λυσόσωμα. Η πρωτεΐνη του γονιδίου SMPD1 (acid sphingomyelinase) εντοπίζεται μέσα στο λυσόσωμα και είναι υπεύθυνη για την μετατροπή λιπιδίων σφιγγομυελίνης σε άλλους τύπους λιπιδίων. Η κλαθρίνη (Clathrin light chain A, CLTA) είναι μια σημαντική πρωτεΐνη που εμπλέκεται στην μεταφορά κυστιδίων με ενδοκύττωση. Η εξαρτημένη από κλαθρίνη ενδοκύττωση αποτελεί σημαντικό μηχανισμό για την εσωτερίκευση υποδοχέων μορφογενετικής πρωτεΐνης των οστών (Bone Morphogenetic Protein Receptor 2, BMPR2), μιας ομάδας πρωτεϊνών που ανήκουν στην υπερ-οικογένεια του αυξητικού παράγοντα μετασχηματισμού (Transforming Growth Factor β , TGF β) και είναι κρίσιμοι παράγοντες στην διαφοροποίηση και ανάπτυξη των πνευμονικών ενδοθηλιακών κυττάρων και των κυττάρων λείων μυών των πνευμονικών αρτηριών (Goumans et al., 2018).
- Ε. Το γονίδιο που κωδικοποιεί τον $\alpha 1$ -αδρενεργικό υποδοχέα (ADRA1A) παρουσιάζεται στο σύστημα συστολής του αγγειακού λείου μυός. Η σηματοδότηση $\alpha 1$ -αδρενοϋποδοχέων προκαλεί τη συστολή των κυττάρων λείου μυός, ελέγχοντας έτσι την αρτηριακή πίεση. Μέχρι στιγμής, επιλεκτικοί αποκλειστές μόνο των β -αδρενεργικών υποδοχέων χρησιμοποιούνται ευρέως ως θεραπευτικοί παράγοντες σε καρδιαγγειακές νόσους, συμπεριλαμβανομένων της υπέρτασης και της καρδιακής ανεπάρκειας (Ahles & Engelhardt, 2014).
- ΣΤ. Από τα κυριότερα σηματοδοτικά μονοπάτια που αναγνωρίστηκαν μέσα από τη βάση KEGG είναι εκείνο του αγγειακού ενδοθηλιακού αυξητικού παράγοντα (Vascular Endothelial Growth Factor, VEGF). Ο συγκεκριμένος παράγοντας παίζει σημαντικό ρόλο στην αγγειογένεση, ρυθμίζοντας τον πολλαπλασιασμό των αγγειακών ενδοθηλιακών κυττάρων των πνευμόνων. Διαταραχή της έκφρασης αυτού του παράγοντα οδηγεί σε αναστολή ανάπτυξης ενδοθηλιακών κυττάρων και εν τέλει στην απόπτωσή τους, ενώ χρήζει ιδιαίτερης διερεύνησης η συμβολή του ως προς την παθογένεση πνευμονικών νόσων (Voelkel et al., 2006). Η Rac1 (Ras-related C3 botulinum toxin substrate 1) ανήκει στην οικογένεια των Rho GTPασών και ενδεχομένως επηρεάζει την έκφραση του υποδοχέα 2 του αγγειακού ενδοθηλιακού αυξητικού παράγοντα (VEGFR2) στα ανθρώπινα ενδοθηλιακά κύτταρα (Meissner et al., 2009).

Πίνακας 5.5 Μονοπάτια KEGG στα οποία συμμετέχουν συγκεκριμένα διαφορικά εκφραζόμενα γονίδια ως προς την πνευμονική αρτηριακή υπέρταση (↑ υπερέκφραση, ↓ υποέκφραση).

KEGG Pathway ID	KEGG Pathway Name	Adjusted P-Value	Gene Name
hsa04664	Fc epsilon RI signaling pathway	1.58E-02	RAC1 (↓) JMJD7 (↑) GRB2 (↓)
hsa00590	Arachidonic acid metabolism	1.77E-02	COX1 (↓) JMJD7 (↑)
hsa05016	Huntington's disease	1.77E-02	CLTA (↓) SDHC (↓) AP2M1 (↓) COX1 (↓)
hsa03018	RNA degradation	1.77E-02	C1D (↓)
hsa04142	Lysosome	1.77E-02	CTNS (↓) CLTA (↓) LAPTM5 (↓) SMPD1 (↑)
hsa04270	Vascular smooth muscle contraction	1.77E-02	ADRA1A (↓) JMJD7 (↑)
hsa00592	alpha-Linolenic acid metabolism	1.95E-02	JMJD7 (↑)
hsa05210	Colorectal cancer	1.95E-02	RAC1 (↓) TCF4 (↑) GRB2 (↓)
hsa04370	VEGF signaling pathway	2.54E-02	RAC1 (↓) JMJD7 (↑)
hsa00591	Linoleic acid metabolism	3.87E-02	JMJD7 (↑)
hsa04912	GnRH signaling pathway	4.40E-02	JMJD7 (↑) GRB2 (↓)

5.3.2 Κατασκευή Δικτύου Αλληλεπίδρασης Πρωτεΐνης – Πρωτεΐνης μέσω της STRING

Τα 195 διαφορεικά εκφραζόμενα γονίδια που προέκυψαν από την παραπάνω μετα-ανάλυση χρησιμοποιήθηκαν ως είσοδο στη βάση δεδομένων STRING 11.0 για την κατασκευή ενός δικτύου αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης (*protein-protein interaction network*).

Το δίκτυο αλληλεπίδρασης πρωτεΐνης - πρωτεΐνης όπως προέκυψε από την είσοδο των 195 ΔΕΓ της μετα-ανάλυσης για τις πέντε μελέτες πάνω στην πνευμονική αρτηριακή υπέρταση παρουσιάζεται στην εικόνα 5.3. Από τα 195 γονίδια αναγνωρίστηκαν τα 160. Οι πρωτεΐνες των γονιδίων αναπαρίστανται με τη μορφή κόμβων, με αριθμό ακμών 71 και βαθμό ορίου εμπιστοσύνης το 0,400. Η τιμή εμπλουτισμού p-value του δικτύου είναι 0.484, το οποίο υποδηλώνει πως το δίκτυο δεν περιέχει στατιστικά σημαντικές αλληλεπιδράσεις μεταξύ των πρωτεϊνών. Για όσα γονίδια δεν έχει καταγραφεί σύνδεση με άλλες πρωτεΐνες από εκείνες που δόθηκαν ως είσοδοι βάση βιβλιογραφίας παρουσιάζονται ως μη συνδεδεμένοι κόμβοι. Για την καλύτερη κατανόηση του δικτύου απεικονίζονται μόνο οι πρωτεΐνες των γονιδίων που παρουσιάζουν σύνδεση μέσω ακμών στην εικόνα 5.4.

Οι συνδέσεις μεταξύ των κόμβων του δικτύου παρουσιάζονται με έγχρωμες ακμές. Κάθε ακμή αναπαριστά μια διαφορετική αλληλεπίδραση ή συσχέτιση μεταξύ ζευγαριών κόμβων.

- ✓ Η μαύρη ακμή δείχνει συνέκφραση μεταξύ πρωτεϊνών.
- ✓ Η κίτρινη ακμή εμφανίζεται όταν υπάρχει βιβλιογραφική αναφορά σε περιοδικό από κοινού και για τις δύο πρωτεΐνες που είναι σε σύνδεση.
- ✓ Η ροζ ακμή εμφανίζεται όταν έχει προηγηθεί πειραματικός/βιοχημικός προσδιορισμός για τη σύνδεση των πρωτεϊνών.
- ✓ Η μπλε ακμή εμφανίζεται όταν έχει προηγηθεί συσχέτιση από επιμελημένες βάσεις δεδομένων.
- ✓ Η πράσινη ακμή αντιπροσωπεύει πρωτεΐνες γονιδίων με κοινά λειτουργικά χαρακτηριστικά.
- ✓ Η κόκκινη ακμή αναφέρεται σε ομολογία στο επίπεδο της αμινοξικής ακολουθίας μεταξύ πρωτεϊνών.

Οι συνδεδεμένες πρωτεΐνες όπως απεικονίζονται στην εικόνα 5.3 παρουσιάζουν 3 βασικά υποδίκτυα.

1. Το πρώτο υποδίκτυο περιλαμβάνει τα γονίδια COPS7A, COPS6, TOR1A και NEDD8. Η πρωτεΐνες που κωδικοποιούνται από τα γονίδια COPS7A (COP9 signalosome complex subunit 7a) και COPS6 (COP9 signalosome complex subunit 6) αποτελούν μέρος ενός συμπλέγματος, που είναι υπεύθυνο για τη ρύθμιση της οδού σύζευξης με ουβικιτίνη. Η πρωτεΐνη NEDD8 συμπεριφέρεται όπως η ουβικιτίνη. Η πρωτεΐνη TOR1A (Torsin-1A) είναι πρωτεΐνη συνοδός του ενδοπλασματικού δικτύου. Αναγνωρίζει τις αναδιπλωμένες πρωτεΐνες που ορίζονται για αποικοδόμηση και τις μετατοπίζει μέσα στο κυτταρόπλασμα.

2. Το δεύτερο υποδίκτυο αναφέρεται στα γονίδια GRB2, S1PR1, RAC1, RHOG και ARHGAP9. Κύριο παρακλάδι αυτού του υποδικτύου είναι το σύμπλεγμα των NECAP1, AP2M1 και CLTA.

Η πρωτεΐνη GRB2 (Growth factor receptor-bound protein 2) αποτελεί πρωτεΐνη προσαρμολογία. Η S1PR1 (Sphingosine 1-phosphate receptor) είναι υποδοχέας που συνδέεται με G-πρωτεΐνες. Η σηματοδότησή της οδηγεί μεταξύ άλλων σε ενεργοποίηση της RAC1 και MAP κινασών. Η RAC1 (Ras-related C3 botulinum toxin substrate) όπως και η RHOG (Rho-related GTP-binding protein) αποτελούν πρωτεΐνες δέσμευσης GTP. Η πρωτεΐνη του γονιδίου ARHGAP9 (Rho GTPase-activating protein 9) δρα σαν μετατροπέας των ενεργών συμπλόκων GTP με πρωτεΐνες της Rho οικογένειας σε ανενεργά σύμπλοκα GDP. Οι προαναφερθείσες λειτουργίες πρωτεϊνών εμφανίζονται στα πλαίσια σηματοδοτικών μορίων, όπως του επιδερμικού αυξητικού παράγοντα (EGF), που ενισχύει τον πολλαπλασιασμό των κυττάρων, καθώς και στο σύστημα μεταβίβασης σήματος από ακολουθία MAP-κινασών. Το μονοπάτι MAP-κινασών αποτελεί μια περίπλοκη σηματοδοτική ακολουθία, η οποία ενδέχεται να επιφέρει ακόμη και μεταβολές στην έκφραση των γονιδίων. Ο υποδοχέας S1PR1 παρουσιάζεται σε σύνδεση με την RAC1, ενώ παράλληλα έχει καταγραφεί ότι το μονοπάτι S1PR1/RAC1 είναι σημαντικός παράγοντας της πνευμονικής αγγειογένεσης και σηματοδότησης των ενδοθηλιακών κυττάρων (Todorova et al., 2017).

Η πρωτεΐνη του NECAP1 (Adaptin ear-binding coat-associated protein 1) συμμετέχει στην ενδοκυττάρωση. Η AP2M1 (AP-2 complex subunit mu) αποτελεί μέρος υποδοχέα και εμπλέκεται σε εξαρτώμενη από κλαθρίνη ενδοκύττωση. Η κλαθρίνη ή αλλιώς CLTA (Clathrin light chain A) αποτελεί κύρια πρωτεΐνη σε μονοπάτια κυστιδιακής μεταφοράς πρωτεϊνών. Πρωτεΐνες-φορτία (cargo proteins) ενσωματώνονται σε κυστίδια που περιβάλλονται από κλαθρίνη και στη συνέχεια προορίζονται για σύντηξη με μεμβράνες-στόχους.

3. Το τρίτο υποδίκτυο αναφέρεται στα γονίδια AQP1, CA2, HBD, EPB42 και SLC4A1.

Η ακουαπορίνη ή AQP1 (Aquaporin-1) σχηματίζει κανάλια αποκλειστικά για τη μεταφορά νερού. Σημαντικές μοριακές λειτουργίες της συγκεκριμένης πρωτεΐνης είναι η διαμεμβρανική μεταφορά διοξειδίου του άνθρακα και οξειδίου του αζώτου (Endeward et al., 2006, Herrera et al., 2006). Βιολογικές διεργασίες που χρήζουν παρατήρησης είναι η κυτταρική απόκριση στην υποξία και στο νιτρικό οξύ. Η CA2 (Carbonic anhydrase 2) εμπλέκεται επίσης σε βιολογικές διεργασίες μεταφοράς διοξειδίου του άνθρακα όπως και στην οδό σηματοδότησης της αγγειοτενσίνης. Η SLC4A1 (Band 3 anion transport protein) λειτουργεί ως δομική πρωτεΐνη της μεμβράνης των ερυθροκυττάρων και ως μεταφορέας ανταλλαγής ανιόντων. Η πρωτεΐνη της αιμοσφαιρίνης ή HBD (Hemoglobin subunit delta) είναι η πρωτεΐνη του αίματος και συμμετέχει στη μεταφορά οξυγόνου από τον πνεύμονα στους διάφορους περιφερειακούς ιστούς. Η EPB42 (Erythrocyte membrane protein band 4.2) είναι πρωτεΐνη δέσμευσης ATP και πιθανώς παίζει σημαντικό ρόλο στη ρύθμιση του σχήματος των ερυθροκυττάρων.

Μεταξύ των προαναφερθέντων υποδικτύων παρεμβάλλονται διάφορες πρωτεΐνες, οι οποίες είτε λειτουργούν ως σύνδεσμοι μεταξύ των υποδικτύων είτε ως παράγωγα αυτών. Ορισμένες εξ αυτών είναι:

- Η πρωτεΐνη ITGB3 (Integrin beta-3) αποτελεί σύνδεσμο ανάμεσα στις πρωτεΐνες SLC4A1 και RAC1. Είναι διαμεμβρανικός υποδοχέας γλυκοπρωτεΐνης για δομικά συμπλέγματα του εξωκυττάριου στρώματος και μεταξύ άλλων μοριακών λειτουργιών της προσφέρεται για τη δέσμευση με τον υποδοχέα 2 του αγγειακού ενδοθηλιακού αυξητικού παράγοντα (VEGF) (Somanath et al., 2009). Ως γνωστόν, ο παράγοντας ανάπτυξης του αγγειακού ενδοθηλίου διεγείρει την αγγειογένεση και αυξάνει την αγγειακή διαπερατότητα (Ferrara et al., 2003).
- Η πρωτεΐνη PLXNA1 (Plexin-A1) αποτελεί εξωκυττάρια σηματοδοτική πρωτεΐνη της οικογένειας των Semaphorins και αλληλεπιδρά επίσης με την RAC1. Η PLXNA1 έχει συνδεθεί με την απορρύθμιση των λείων μυϊκών ινών των αεραγωγών σε άτομα με άσθμα, το οποίο στη συνέχεια συνδέεται με την εμφάνιση της χρόνιας αποφρακτικής πνευμονοπάθειας (Movassagh et al., 2016).
- Οι πρωτεΐνες PSMD9 (26S proteasome non-ATPase regulatory subunit 9) και SEC61B (Protein transport protein Sec61 subunit beta) απορρέουν από το πρώτο υποδίκτυο πρωτεϊνών που περιγράφηκε παραπάνω και συμμετέχουν στην μεσολαβούμενη σηματοδότηση από ιντερλευκίνη IL-1 και στη μεταφορά πολυπεπτιδίων αντίστοιχα. Πιο συγκεκριμένα, το πρωτεάσωμα 26S είναι ένας μεγάλος «καταστροφέας», όπου πρωτεΐνες σηματοδοτημένες με ουβικιτίνη αποικοδομούνται. Τυχόν δυσλειτουργία του μονοπατιού ουβικιτίνης-πρωτεασώματος ενδεχομένως να επηρεάζει την ανάπτυξη πνευμονικών αρτηριακών κυττάρων λείου μυός και να συμβάλλει δυνητικά στην ανάπτυξη πνευμονικής αρτηριακής υπέρτασης (Li et al., 2011).
- Από το δεύτερο υποδίκτυο συνδέονται μέσω της RAC1 σε σειρά πρώτα η MYD88 (Myeloid differentiation primary response protein MyD88) και ύστερα η NOD2 (Nucleotide-binding oligomerization domain-containing protein 2). Η MyD88 συμμετέχει στην ενδοκυττάρια σηματοδότηση είτε μέσω σύζευξης με υποδοχείς Toll-like είτε με υποδοχέα της Ιντερλευκίνης 1. Συμμετέχει στην ανοσοαπόκριση και τη φλεγμονώδη απόκριση. Η NOD2 ανήκει στην οικογένεια υποδοχέων NOD (Nod-like Receptors, NLRs) και παίζει επίσης ρόλο στην ανοσοαπόκριση και ενεργοποιεί παράγωγα αντιφλεγμονώδους αντίδρασης. Αν και η λειτουργία του NOD2 δεν έχει διερευνηθεί αρκετά πάνω στα πνευμονικά αγγειακά κύτταρα λείου μυός, ευρήματα σε ζωικά μοντέλα έχουν δείξει ότι παίζει σημαντικό ρόλο στην απόκριση των συγκεκριμένων κυττάρων σε καταστάσεις υποξίας, και κατά συνέπεια μπορεί να είναι ένας πολλά υποσχόμενος στόχος για θεραπεία πνευμονικών αγγειακών παθήσεων (Kwon et al., 2018).
- Από το δεύτερο υποδίκτυο απαντώνται μετά την S1PR1 οι πρωτεΐνες GNG (Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-10) και η ADRA1A (Alpha-1A adrenergic receptor). Η GNG είναι G-πρωτεΐνη δέσμευσης νουκλεοτιδίων γουανίνης που εμπλέκεται ως μετατροπέας διαμεμβρανικού σήματος (αντικατάσταση GDP με GTP). Ο άλφα-αδρενεργικός υποδοχέας ADRA1A μεσολαβεί σε συνδυασμό με G-πρωτεΐνες για σηματοδότηση ERK στα καρδιακά μυοκύτταρα. Στις βιολογικές διεργασίες της συγκεκριμένης πρωτεΐνης περιλαμβάνονται η έκκριση επινεφρίνης-

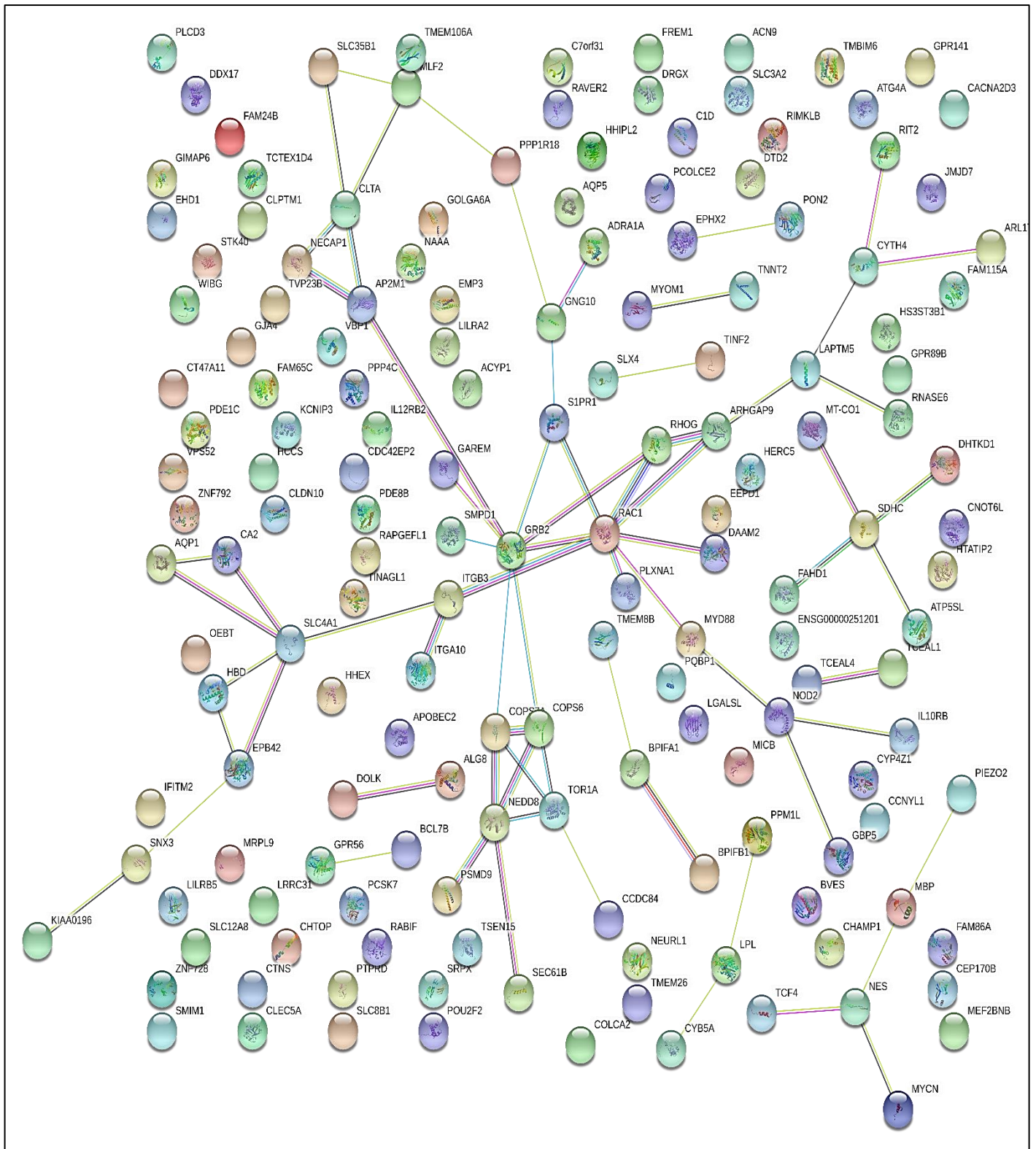
νορεπινεφρίνης που εμπλέκονται στη ρύθμιση της συστηματικής αρτηριακής πίεσης και της αγγειοσυστολής.

- Παρουσιάζεται επίσης μια διασταύρωση πέντε γονιδίων εκτός των υποδικτύων που αναφέρθηκαν πιο πάνω. Τα γονίδια αυτά είναι MT-CO1 (COX1), FAHD1, DHTKD1 και DMAC2 και σημείο αναφοράς το SDHC. Τα συγκεκριμένα γονίδια και τα παράγωγά τους συμμετέχουν σε μεταβολικές διεργασίες των μιτοχονδρίων (Κύκλος του Krebs, μεταφορά ηλεκτρονίων, οξειδωτική φωσφορυλίωση).
- Οι πρωτεΐνες LAPTM5, RNASE6 και CYTH4 εμφανίζονται σε σύνδεση αμέσως μετά την ARHGAP9 και δείχνουν να εμπλέκονται σε διεργασίες των λυσοσωμάτων.
- Μια μικρή ομάδα γονιδίων γραμμικά συνδεδεμένα είναι το τρίπτυχο TCF4, NES και MYCN. Εκτός των TCF4 και MYCN που δρουν ως μεταγραφικοί παράγοντες, η πρωτεΐνη Nestin έχει καταγραφεί πρόσφατα να παρουσιάζει υψηλά ποσοστά έκφρασης σε πνευμονικές αρτηρίες σε ποντίκια με ΠΑΥ (Saboor et al., 2016, Bhagwani et al., 2019).

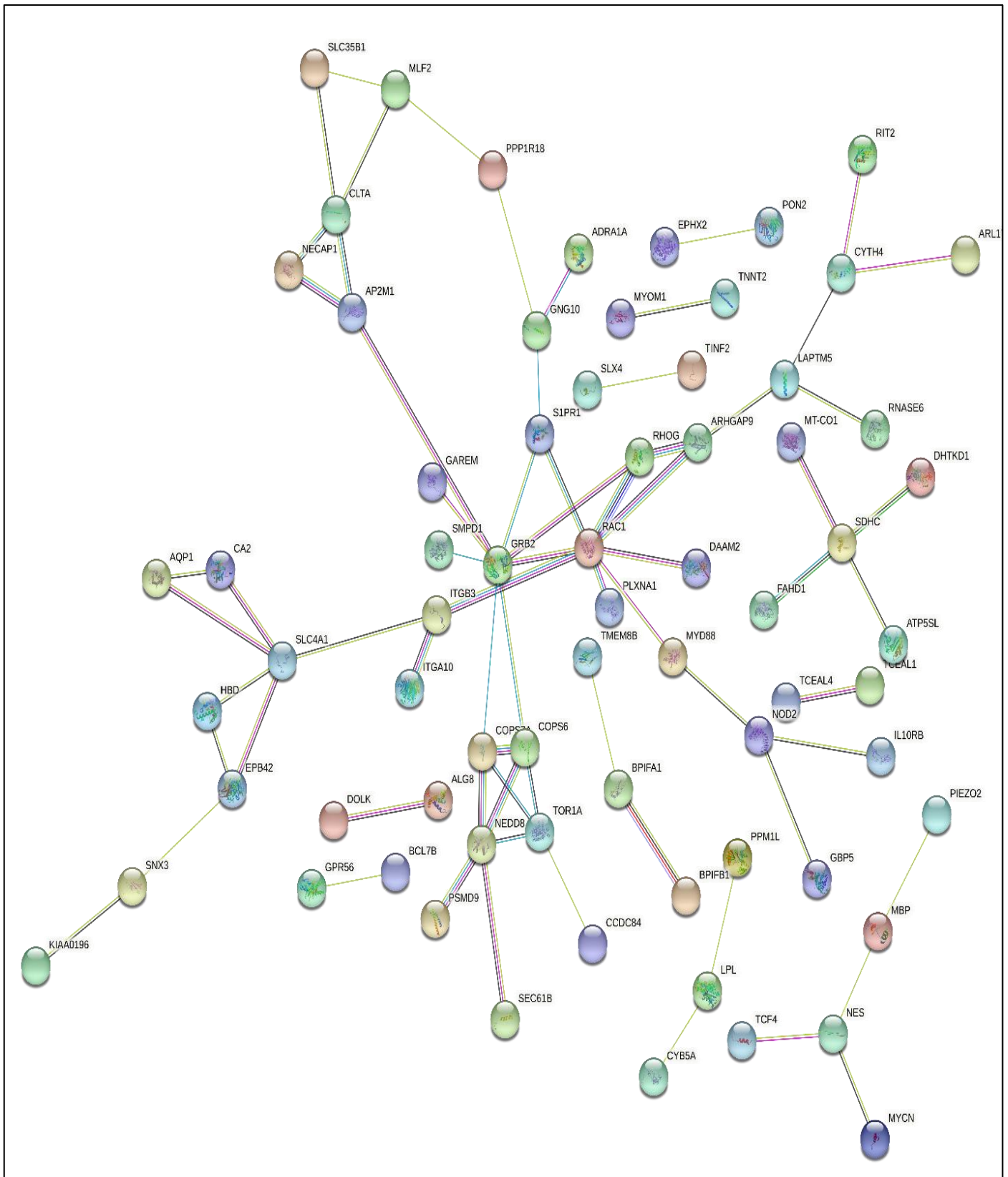
Η ανάλυση εμπλουτισμού μέσω της βάσης STRING προσέδωσε GO όρους ως προς 2 μονοπάτια βιολογικής διεργασίας (biological process) και 3 μονοπάτια που συνδέονται με τη βάση Reactome. Τα γονίδια τα οποία παρεμβάλλονται σε αυτά τα μονοπάτια μπορεί να τα διακρίνει κανείς μέσα από το δίκτυο αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης (Εικόνα 5.3-4) καθώς και αναλυτικότερα στον πίνακα 5.6. Τα περισσότερα από αυτά έχουν ήδη αναλυθεί στις παραπάνω παραγράφους και επιβεβαιώνονται με τα δεδομένα του πίνακα 5.6. Διακρίνονται δύο βασικά βιολογικά μονοπάτια μέσα από τον πίνακα 5.6. Το πρώτο αφορά την μεταγωγή σήματος μέσω των ερυθροκυττάρων και το δεύτερο τη μεταγωγή σήματος μέσω ενδοκύττωσης με κυστίδια κλαθρίνης.

Επιπρόσθετα, παρουσιάζεται ο πίνακας 5.7 των διαφορετικά εκφραζόμενων γονιδίων που αναγνωρίστηκαν ως κομβικά σημεία μέσα από το δίκτυο αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης της STRING και έχουν τη μεγαλύτερη συνδεσιμότητα μεταξύ άλλων πρωτεϊνών σε όριο εμπιστοσύνης 0,400. Απεικονίζεται με πρόσημο (+/-) η διαφορά στην έκφραση του γονιδίου όπως προέκυψε μέσα από τη μετα-ανάλυση.

Τέλος, παρουσιάζεται ο πίνακας 5.8 στον οποίο αναπαριστώνται τα ονόματα των 160 γονιδίων από τα 195 στο σύνολο, που αναγνωρίστηκαν μέσα από τη βάση STRING και συντελούν στη δημιουργία δικτύου αλληλεπίδρασης πρωτεϊνών σχετικά με την ασθένεια της πνευμονικής αρτηριακής υπέρτασης.



Εικόνα 5.3 Το συνολικό δίκτυο αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης (PPI network) των 195 ΔΕΓ για την πνευμονική αρτηριακή υπέρταση. Από τα 195 γονίδια αναγνωρίστηκαν τα 160 με αριθμό ακμών 71. Οι γραμμές με διαφορετικά χρώματα δείχνουν τα διαφορετικά είδη των αλληλεπιδράσεων. Ως όριο εμπιστοσύνης έχει οριστεί το 0,400.



Εικόνα 5.4 Δίκτυο αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης (PPI network) που κωδικοποιείται από 70 ΔΕΓ ως προς την ασθένεια της πνευμονικής αρτηριακής υπέρτασης. Στο παρόν δίκτυο δεν παρουσιάζονται οι ασύνδετοι κόμβοι.

Πίνακας 5.6 Στατιστικά σημαντικά βιολογικά μονοπάτια που προέκυψαν στην STRING μέσα από το δίκτυο αλληλεπιδράσεων πρωτεϊνών-πρωτεϊνών για τα ΔΕΓ της πνευμονικής αρτηριακής υπέρτασης.

Biological Process (GO)	Term description	adjusted p-value	Protein labels
GO:0015669	gas transport	0.0479	AQP1,AQP5,CA2,HBD
GO:0015670	carbon dioxide transport	0.0479	AQP1,AQP5,CA2
Reactome Pathways	Term description	adjusted p-value	Protein labels
HSA-8856825	Cargo recognition for clathrin-mediated endocytosis	0.0129	AP2M1,CLTA,COPS6,COPS7A,NECAP1,NEDD8,TOR1A
HSA-1247673	Erythrocytes take up oxygen and release carbon dioxide	0.0225	AQP1,CA2,SLC4A1
HSA-1237044	Erythrocytes take up carbon dioxide and release oxygen	0.032	AQP1,CA2,SLC4A1

Πίνακας 5.7 Γονίδια με τη μεγαλύτερη συνδεσιμότητα στο δίκτυο αλληλεπιδράσεων πρωτεϊνών-πρωτεϊνών με όριο εμπιστοσύνης 0,400.

Γονίδιο	Όνομα Πρωτεΐνης που κωδικοποιεί (Uniprot)	Συνδεδεμένοι κόμβοι	Υπερ/υποέκφραση (+/-)
GRB2	Growth factor receptor-bound protein 2	8	-
RAC1	Ras-related C3 botulinum toxin substrate 1	8	-
SLC4A1	Band 3 anion transport protein	5	+
NEDD8	NEDD8	5	-
CLTA	Clathrin light chain A	4	-
SDHC	Succinate dehydrogenase cytochrome b560 subunit	4	-

Πίνακας 5.8 Λίστα με τα 160 ονόματα γονιδίων που αναγνωρίστηκαν μέσα από την ανάλυση εμπλουτισμού του δικτύου αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης της STRING.

Ονόματα γονιδίων που αναγνωρίστηκαν μέσα από το δίκτυο αλληλεπίδρασης πρωτεϊνών της STRING			
ACN9	DDX17	KCNIP3	RAC1
ACYPI	DHTKD1	KIAA0196	RAPGEFL1
ADRA1A	DOLK	LAPTM5	RAVER2
ALG8	DRGX	LGALSL	RHOG
AP2M1	DTD2	LILRA2	RIMKLB
APOBEC2	EEPD1	LILRB5	RIT2
AQP1	EHD1	LPL	RNASE6
AQP5	EMP3	LRRC31	S1PR1
ARHGAP9	ENSG00000251201 (TMED7-TICAM2)	MBP	SDHC
ARL17A	EPB42	MEF2BNB	SEC61B
ATG4A	EPHX2	MICB	SLC12A8
ATP5SL	FAHD1	MLF2	SLC35B1
BCL7B	FAM115A	MRPL9	SLC3A2
BPIFA1	FAM24B	MT-CO1 (COX1)	SLC4A1
BPIFB1	FAM65C	MYCN	SLC8B1
BVES	FAM86A	MYD88	SLX4
C1D	FREM1	MYOM1	SMIM1
C7orf31	GAREM	NAAA	SMPD1
CA2	GBP5	NECAP1	SNX3
CACNA2D3	GIMAP6	NEDD8	SRPX
CCDC84	GJA4	NES	STK40
CCNYL1	GNG10	NEURL1	TCEAL1
CDC42EP2	GOLGA6A	NOD2	TCEAL4
CEP170B	GPR141	OEBT	TCF4
CHAMP1	GPR56	PCOLCE2	TCTEX1D4
CHTOP	GPR89B	PCSK7	TINAGL1
CLDN10	GRB2	PDE1C	TINF2
CLEC5A	HBD	PDE8B	TMBIM6
CLPTM1	HCCS	PIEZO2	TMEM106A
CLTA	HERC5	PLCD3	TMEM26
CNOT6L	HHEX	PLXNA1	TMEM8B
COLCA2	HHIPL2	PON2	TNNT2
COPS6	HS3ST3B1	POU2F2	TOR1A
COPS7A	HTATIP2	PPM1L	TSEN15
CT47A11	IFITM2	PPP1R18	TVP23B
CTNS	IL10RB	PPP4C	VBP1
CYB5A	IL12RB2	PQBP1	VPS52
CYP4Z1	ITGA10	PSMD9	WIBG
CYTH4	ITGB3	PTPRD	ZNF728
DAAM2	JMJD7	RABIF	ZNF792

Τα ευρήματα της παραπάνω ανάλυσης εμπλουτισμού μέσω της βάσης STRING παρουσιάζουν μια γραμμική σύνδεση μεταξύ των γονιδίων AQP1, SLC4A1, ITGB3, RAC1, GRB2, S1PR1 GNG10 και ADRA1A. Τα γονίδια αυτά διακρίνονται για τη συμμετοχή τους σε βιολογικές διεργασίες αγγειογένεσης, στη ρύθμιση της συστηματικής αρτηριακής πίεσης και της αγγειοσυστολής, στη φλεγμονώδη απόκριση και σε διαμεμβρανική μεταφορά αερίων μέσω της οδού των ερυθροκυττάρων. Επιπρόσθετα τα γονίδια GRB2, RAC1, SLC4A1, NEDD8, CLTA και SDHC αποτελούν κομβικά σημεία του δικτύου αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης της συγκεκριμένης ανάλυσης εμπλουτισμού. Συνοψίζοντας, αναγνωρίστηκαν μικρά υποδίκτυα πρωτεϊνών που συμμετέχουν:

- i. σε μονοπάτια μεταγωγής σήματος μέσω ενδοκύττωσης με κυστίδια κλαθρίνης,
- ii. σε μονοπάτια μεταγωγής σήματος μέσω πρωτεϊνικών υποδοχέων GPCRs,
- iii. σε οδούς μεταγωγής σήματος μέσω των ερυθροκυττάρων,
- iv. σε σηματοδότηση μέσω του συστήματος ουβικιτίνης-πρωτεασώματος.

5.4 Διασταύρωση των Ευρημάτων μέσω άλλων Βάσεων Δεδομένων

Τελευταίο στάδιο της παρούσας εργασίας ήταν να γίνει έλεγχος διασταύρωσης των αποτελεσμάτων της μετα-ανάλυσης με βάσεις δεδομένων που συγκεντρώνουν ευρυγονιδιωματικές μελέτες (GWAS) ή μελέτες γενετικής συσχέτισης και τα αποτελέσματά τους ως προς την ασθένεια της πνευμονικής αρτηριακής υπέρτασης. Χρησιμοποιήθηκαν οι βάσεις δεδομένων GWAS Catalog και DisGeNET.

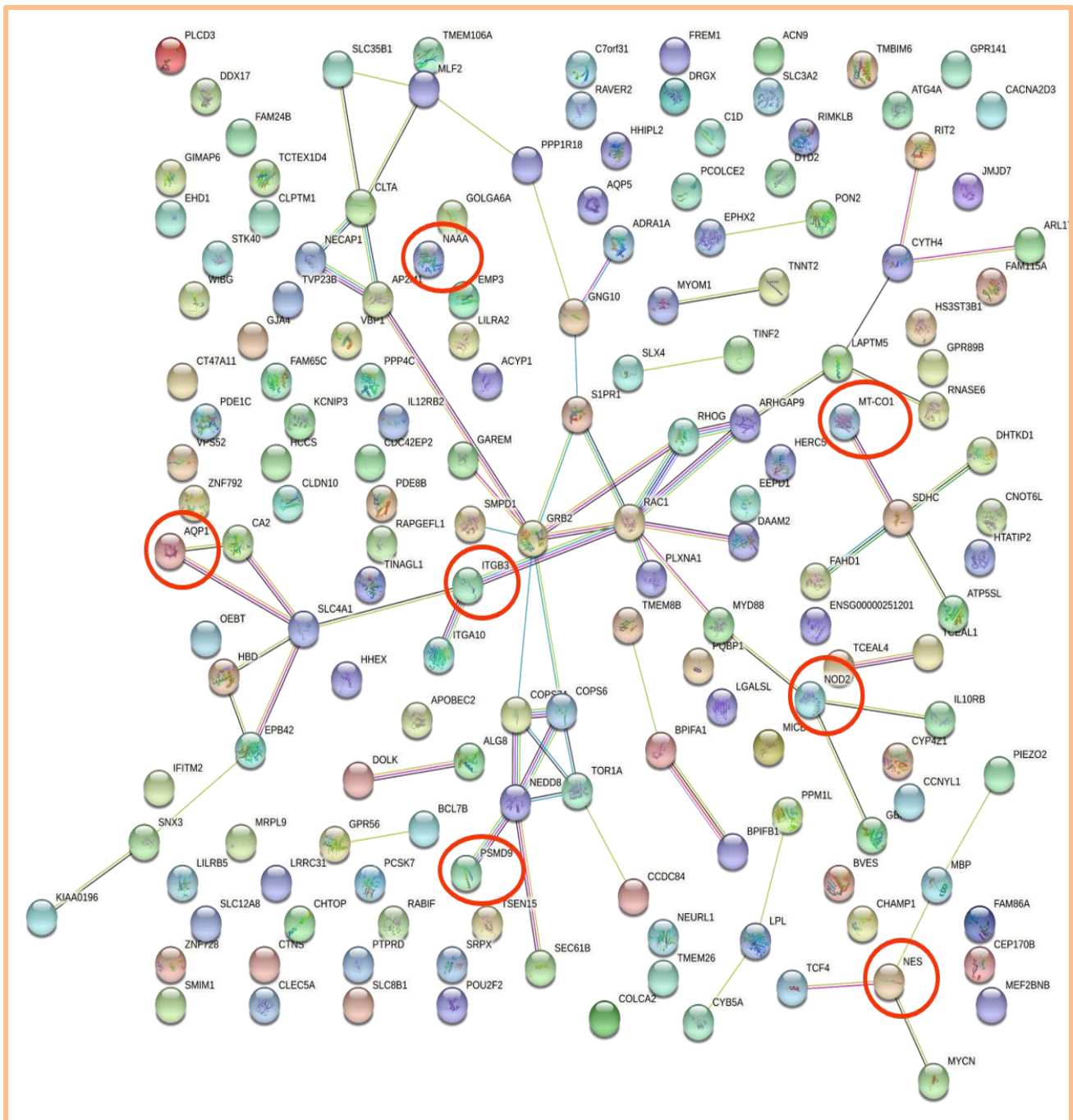
Μέσα από την βάση δεδομένων GWAS Catalog δεν βρέθηκε κάποια διασταύρωση με κάποιο από τα 195 διαφορετικά εκφραζόμενα γονίδια της μετα-ανάλυσης. Όμως, μέσα από τη βάση DisGeNET επιστράφηκαν 7 διασταυρούμενα γονίδια όπως καταγράφηκαν μέσα από τη μετα-ανάλυση και την ασθένεια της ΠΑΥ (Πίνακας 5.9). Τα γονίδια αυτά εμπλέκονται στο δίκτυο αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης της STRING είτε μέσα από διαδοχικές συνδέσεις είτε χώρια (Εικόνα 5.5).

- ♦ Το γονίδιο της κυκλοοξυγενάσης (COX1) ενεργοποιεί παράγωγα προστακυκλίνης και αποτελεί ισχυρό αγγειοδιασταλτικό των πνευμονικών αγγείων μέσω του μεταβολισμού του αραχιδονικού οξέως (βλ. §5.3.1).
- ♦ Δύο παράγωγα γονιδίων (AQP1, ITGB3) εμπλέκονται σε βιολογικές διεργασίες διαμεμβρανικής μεταφοράς διοξειδίου του άνθρακα και οξειδίου του αζώτου, καθώς και στη δέσμευση με τον υποδοχέα 2 του αγγειακού ενδοθηλιακού αυξητικού παράγοντα, αντίστοιχα (βλ. §5.3.2).
- ♦ Το πρωτεάσωμα 26S (PSMD9) συμμετέχει στο σύστημα ουβικιτίνης-πρωτεασώματος. Τυχόν δυσλειτουργία αυτού του μονοπατιού μπορεί να επηρεάσει τον πολλαπλασιασμό των κυττάρων λείου μυός των πνευμονικών αρτηριών και συνεπώς να παίζει ρόλο σε εύρεση θεραπείας για την ΠΑΥ (βλ. §5.3.2).
- ♦ Αρκετές μελέτες εξετάζουν τα υψηλά ποσοστά έκφρασης της πρωτεΐνης Nestin (NES) σε ενδοθηλιακά κύτταρα ζωικών μοντέλων παρουσία ΠΑΥ και δείχνουν ότι οι αγγειακές βλάβες του πνεύμονα οφείλονται στην υπερέκφραση της συγκεκριμένης πρωτεΐνης (βλ. §5.3.2).
- ♦ Πρόσφατες μελέτες σε ζωικά μοντέλα δείχνουν ότι η πρωτεΐνη του γονιδίου NOD2 παρουσιάζει υπερέκφραση σε πνευμονικά αρτηριακά κύτταρα λείου μυός σε καταστάσεις υποξίας και προτείνουν τη διερεύνηση του υποκείμενου μηχανισμού σε καταστάσεις πνευμονικών παθήσεων (βλ. §5.3.2).
- ♦ Η πρωτεΐνη του γονιδίου NAAA (N-Acylethanolamine acid amidase) δεν βρέθηκε να αλληλεπιδρά με κάποιο άλλο γονίδιο μέσα από την ανάλυση δικτύου της STRING. Αποτελεί ένζυμο μεταβολισμού της κυστεΐνης που εκφράζεται σε μεγάλο βαθμό σε μακροφάγα και περιφερειακούς ιστούς, συμπεριλαμβανομένων και των πνευμόνων. Έχει προταθεί ως στόχος για την ανάπτυξη νέων φαρμάκων για τη θεραπεία της φλεγμονής σε ποικίλες νόσους (Li et al., 2018).

Το γεγονός ότι η παρούσα εργασία ανίχνευσε συγκεκριμένα γονίδια για τα οποία υπάρχουν ήδη ενδεικτικές μελέτες ότι σχετίζονται με την πνευμονική αρτηριακή υπέρταση ή ότι χρήζουν διερεύνησης σε καρδιοαναπνευστικές νόσους, ενισχύει τα αποτελέσματα της συγκεκριμένης μεθοδολογίας, που είχε ως στόχο την εύρεση γονιδίων-στόχων και τα παράγωγά τους (πρωτεΐνες) που εμπλέκονται στην έναρξη ή στην εξέλιξη της ΠΑΥ.

Πίνακας 5.9 Πίνακας των 7 ΔΕΓ της μετα-ανάλυσης που διασταυρώθηκαν μέσα από τη βάση δεδομένων DisGeNET για την ασθένεια της ΠΑΥ.

Όνομα Γονιδίου	Πρωτεΐνη που κωδικοποιεί (Uniprot)
NES	Nestin
COX1 (MT-CO1)	Cytochrome c oxidase subunit 1
NOD2	Nucleotide-binding oligomerization domain-containing protein 2
AQP1	Aquaporin-1
ITGB3	Integrin beta-3
NAAA	N-acyl ethanolamine-hydrolyzing acid amidase
PSMD9	26S proteasome non-ATPase regulatory subunit 9



Εικόνα 5.5 Δίκτυο αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης των 195 ΔΕΓ για την πνευμονική αρτηριακή υπέρταση. Στο δίκτυο επισημαίνονται με κόκκινες ελλείψεις τα γονίδια που διασταυρώθηκαν ως πιθανοί μελλοντικοί βιοδείκτες για την ασθένεια ΠΑΥ μέσα από τη βάση δεδομένων DisGeNET.

Κεφάλαιο 6 – Συμπεράσματα & Συζήτηση

Η πνευμονική αρτηριακή υπέρταση (ΠΑΥ) είναι μια σπάνια και επικίνδυνη για την υγεία καρδιοπνευμονική διαταραχή, που χαρακτηρίζεται από πολλαπλασιασμό των πνευμονικών αγγειακών κυττάρων λείου μυός, καταστροφή των περιφερειακών αγγείων των πνευμόνων και ενίσχυση των φλεγμονωδών κυττάρων, με αποτέλεσμα την μη φυσιολογική αγγειοσυστολή, την αγγειακή αναδιαμόρφωση και δυσλειτουργία της δεξιάς κοιλίας της καρδιάς. Οι μοριακοί μηχανισμοί που εμπλέκονται στην εμφάνιση της νόσου ΠΑΥ παραμένουν ακόμη υπό διερεύνηση.

Στην παρούσα εργασία πραγματοποιήθηκε στατιστικός έλεγχος σε σύνολα δεδομένων από DNA μικροσυστοιχίες. Τα πειράματα αφορούσαν μελέτες που εξέταζαν την έκφραση γονιδίων πάνω στην ασθένεια της πνευμονικής αρτηριακής υπέρτασης. Στόχος της εργασίας ήταν να εντοπιστούν τα διαφορικά εκφραζόμενα γονίδια που μπορούν να χρησιμοποιηθούν ως παράγοντες πρόβλεψης για τον κίνδυνο εμφάνισης της πνευμονικής αρτηριακής υπέρτασης.

Διενεργήθηκε συστηματική ανασκόπηση σύμφωνα με το πρότυπο PRISMA μέσα από τις βάσεις δεδομένων Gene Expression Omnibus και ArrayExpress για συλλογή μελετών που αφορούσαν πειράματα μικροσυστοιχιών πάνω σε ιστούς πνεύμονα από δείγματα ασθενών-μαρτύρων. Συγκεντρώθηκαν 5 μελέτες προς ανάλυση για την ασθένεια της ΠΑΥ.

Μετά από επεξεργασία των πινάκων γονιδιακής έκφρασης για κάθε μελέτη με αντιστοίχιση των νέων ονομάτων γονιδίων για κάθε ανιχνευτή, διεξήχθη μετα-ανάλυση στο περιβάλλον STATA. Εφαρμόστηκε ως μέγεθος επίδρασης η τυποποιημένη διαφορά των μέσων όρων ανάμεσα στα δύο ανεξάρτητα σύνολα των ομάδων (ασθενείς/μάρτυρες) σε συνδυασμό με διόρθωση του Hedges. Ύστερα από εφαρμογή διορθωτικών μεθόδων για την αποφυγή σφαλμάτων τύπου I, η μετα-ανάλυση επέστρεψε 195 διαφορικά εκφραζόμενα γονίδια, όπου 149 γονίδια υποεκφράζονταν και τα υπόλοιπα 46 υπερεκφράζονταν. Τα αποτελέσματα εκφράζονται σε επίπεδο σημαντικότητας 1% μετά από διόρθωση με τη μέθοδο FDR των Benjamini και Hochberg.

Πραγματοποιήθηκαν αναλύσεις εμπλουτισμού με χρήση των εργαλείων STRING (11.0) και bioCompendium (KEGG Pathway Information).

Με το εργαλείο της bioCompendium έγινε χρήση του πεδίου της βάσης KEGG για ανίχνευση των βιοχημικών μονοπατιών στα οποία συμμετέχουν τα ΔΕΓ. Τα στατιστικά σημαντικά μονοπάτια της βάσης KEGG που ανιχνεύθηκαν ήταν στο σύνολο 11 και τα γονίδια που καταγράφηκαν για τη συμμετοχή τους στα βιοχημικά μονοπάτια της KEGG ως επί το πλείστον υποεκφράζονταν. Τα κυριότερα βιοχημικά μονοπάτια KEGG που βρέθηκαν εμπλουτισμένα και μπορούν να συσχετιστούν με την εκκίνηση ή την εξέλιξη της νόσου ΠΑΥ είναι:

- Το μονοπάτι μεταβολισμού του αραχιδονικού οξέος. Παράγωγα του αραχιδονικού οξέος, όπως η ορμόνη προστακυκλίνη, αποτελούν ισχυρά αγγειοδιασταλτικά και οδηγούν σε συστολή του πνευμονικού αγγείου.

- Το μονοπάτι υποδοχέων των ανοσοσφαιρινών. Οι υποδοχείς αυτοί εκφράζονται σε σιτευτικά κύτταρα και πρόσφατα καταγράφηκε ότι ρυθμίζουν ενεργά τη φλεγμονή και παίζουν ρόλο στην αποκατάσταση της ομοιόστασης στα πνευμονικά αγγεία.
- Το μονοπάτι διαλογής πρωτεϊνών στο λυσόσωμα και εν συνεχεία στη μεσολαβούμενη από κλαθρίνη ενδοκύττωση. Η εξαρτημένη από κλαθρίνη ενδοκύττωση αποτελεί σημαντικό μηχανισμό για την εσωτερίκευση υποδοχέων μορφογενετικής πρωτεΐνης των οστών (BMPR), που είναι κρίσιμοι παράγοντες στην διαφοροποίηση και ανάπτυξη των πνευμονικών ενδοθηλιακών κυττάρων και των κυττάρων λείων μυών των πνευμονικών αρτηριών.
- Το μονοπάτι συστολής του αγγειακού λείου μυός. Το γονίδιο που κωδικοποιεί τον α1-αδρενεργικό υποδοχέα (ADRA1A) παρουσιάστηκε στο σύστημα συστολής του αγγειακού λείου μυός.
- Η οδός σηματοδότησης του αγγειακού ενδοθηλιακού αυξητικού παράγοντα. Ο συγκεκριμένος παράγοντας (VEGF) παίζει σημαντικό ρόλο στην αγγειογένεση, ρυθμίζοντας τον πολλαπλασιασμό των αγγειακών ενδοθηλιακών κυττάρων των πνευμόνων. Διαταραχή της έκφρασης αυτού του παράγοντα οδηγεί σε αναστολή ανάπτυξης ενδοθηλιακών κυττάρων και εν τέλει στην απόπτωσή τους, ενώ χρίζει ιδιαίτερης διερεύνησης η συμβολή του ως προς την παθογένεση πνευμονικών νόσων

Με το εργαλείο της STRING επιστράφηκε δίκτυο αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης για τα 195 ΔΕΓ της πνευμονικής αρτηριακής υπέρτασης στο οποίο αναγνωρίστηκαν 160 γονίδια. Το δίκτυο παρουσίασε τρία κύρια υποδίκτυα.

- Το πρώτο υποδίκτυο αναφέρεται σε γονίδια που εμπλέκονται στο σύστημα ουβικιτίνης-πρωτεασώματος και είναι τα COPS7A, COPS6, TOR1A, NEDD8 και PSMD9. Μηχανισμοί αναστολής του πρωτεασώματος μπορούν να καταστείλουν την ανάπτυξη πνευμονικών αγγειακών κυττάρων λείου μυός και κατά συνέπεια να χρησιμεύσουν ως προς τη θεραπεία της ΠΑΥ.
- Το δεύτερο υποδίκτυο αναφέρεται στη μεταγωγή σήματος μέσω G-πρωτεϊνικών υποδοχέων (GPCRs) και αποτελείται από τα γονίδια GRB2, S1PR1, RAC1, RHOG και ARHGAP9. Τα πρωτεϊνικά παράγωγα των γονιδίων αυτών εμφανίζονται στα πλαίσια σηματοδοτικών μορίων, όπως του επιδερμικού αυξητικού παράγοντα (EGF), που ενισχύει τον πολλαπλασιασμό των κυττάρων, καθώς και στο σύστημα μεταβίβασης σήματος από ακολουθία MAP-κινασών. Κύριο παρακλάδι του δεύτερου υποδικτύου είναι το σύμπλεγμα των NECAP1, AP2M1 και CLTA και συμμετέχουν σε μονοπάτια μεταγωγής σήματος μέσω ενδοκύττωσης με κυστίδια κλαθρίνης. Η εξαρτημένη από κλαθρίνη ενδοκύττωση αποτελεί σημαντικό μηχανισμό για την εσωτερίκευση υποδοχέων μορφογενετικής πρωτεΐνης των οστών (BMP), η οποία είναι κρίσιμος παράγοντας στην διαφοροποίηση και ανάπτυξη των πνευμονικών ενδοθηλιακών κυττάρων και των κυττάρων λείων μυών των πνευμονικών αρτηριών.
- Το τρίτο υποδίκτυο αναφέρεται σε γονίδια που συμμετέχουν σε μεταγωγή σήματος μέσω της μεταφοράς οξυγόνου στα ερυθροκύτταρα και είναι τα AQP1, CA2, HBD,

EPB42 και SLC4A1. Αλλαγές στη συμπεριφορά των ερυθροκυττάρων μπορούν να προσδώσουν κλινικά συμπεράσματα για την ανάπτυξη ή την εξέλιξη της ΠΑΥ.

Τα γονίδια που παρουσίασαν μια γραμμική σύνδεση μεταξύ τους είναι τα AQP1, SLC4A1, ITGB3, RAC1, GRB2, S1PR1 GNG10 και ADRA1A. Τα γονίδια αυτά διακρίνονται για τη συμμετοχή τους σε βιολογικές διεργασίες αγγειογένεσης, στη ρύθμιση της συστηματικής αρτηριακής πίεσης και της αγγειοσυστολής, στη φλεγμονώδη απόκριση και σε διαμεμβρανική μεταφορά αερίων μέσω της οδού των ερυθροκυττάρων. Επιπρόσθετα τα γονίδια GRB2, RAC1, SLC4A1, NEDD8, CLTA και SDHC αποτελούν κομβικά σημεία του δικτύου αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης της συγκεκριμένης ανάλυσης εμπλουτισμού και επομένως να μπορούν να προσφέρουν νέες πληροφορίες ως προς την παθοβιολογία της ΠΑΥ.

Τελευταίο βήμα για την ολοκλήρωση της παρούσας εργασίας ήταν η διασταύρωση των ευρημάτων της μετα-ανάλυσης με άλλες βάσεις δεδομένων που συνδυάζουν αποτελέσματα ευρυγονιδιωματικών μελετών ή μελετών γενετικής επιδημιολογίας γύρω από την ασθένεια της ΠΑΥ. Χρησιμοποιήθηκαν οι βάσεις δεδομένων GWAS Catalog και DisGeNET. Μόνο η δεύτερη επέστρεψε διασταυρούμενα αποτελέσματα ως προς τα 195 ΔΕΓ της μετα-ανάλυσης για την ασθένεια της πνευμονικής αρτηριακής υπέρτασης. Τα γονίδια που έχουν καταχωρηθεί ως κρίσιμοι παράγοντες για αυτή τη νόσο μέσα στη βάση δεδομένων DisGeNET και παρουσιάζονται ως διαφορετικά εκφραζόμενα στην παρούσα μελέτη είναι τα NES, COX1 (MT-CO1), NOD2, AQP1, ITGB3, NAAA και PSMD9.

Να επισημανθεί ότι οι μελέτες γονιδιακής έκφρασης που χρησιμοποιήθηκαν για τον στατιστικό έλεγχο στρέφονταν γύρω από τον ιστό του πνεύμονα. Σε περίπτωση που η επικέντρωση της ανάλυσης ήταν γύρω από ένα διαφορετικό ιστό (π.χ. κύτταρα περιφερειακού αίματος) τα δεδομένα έκφρασης γονιδίων θα μπορούσαν να παρέχουν μια διαφορετική εικόνα γύρω από τις βιοχημικές οδούς στις οποίες συμμετέχουν και αντίστοιχα διαφορετικά πιθανά αίτια ανάπτυξης της νόσου.

Ιδιαίτερης παρατήρησης χρήζει το γεγονός, ότι τα ΔΕΓ που βρέθηκαν σε αυτή τη μελέτη, ή στο σύνολο του δικτύου που σχηματίζουν αλληλεπιδράσεις, δεν είναι στο γενικότερο σύνολό τους τα ίδια γονίδια που φέρουν πολυμορφισμούς και έχουν σε παλαιότερο χρόνο συσχετιστεί με τη νόσο μέσα από μελέτες γενετικής συσχέτισης ή GWAS (Rhodes et al., 2018). Εδώ φαίνεται και η δυναμική προσέγγιση της μετα-ανάλυσης όσον αφορά τα δεδομένα γονιδιακής έκφρασης και τη συνδρομή της ως προς την ανακάλυψη νέας γνώσης. Μπορεί τα γονίδια γενετικών πολυμορφισμών να πυροδοτήσουν την εκκίνηση ενός κακόβουλου σήματος, όμως τα γονίδια που προέκυψαν μέσα από μια ανάλυση μικροσυστοιχιών είναι πολύ πιο πιθανό να συμμετάσχουν σε μεταγενέστερα στάδια για την εξέλιξη της νόσου. Έτσι λοιπόν, η χρήση της μετα-ανάλυσης παραμένει αναγκαία όσο ποτέ άλλοτε για την εξαγωγή σημαντικών πληροφοριών που περιέχονται μέσα από τον πλέον τεράστιο όγκο δεδομένων γονιδιακής έκφρασης που έχουν παραχθεί και αποθηκευτεί σε δημόσια αποθετήρια.

Συμπερασματικά, η παρούσα μελέτη εντόπισε ΔΕΓ που μπορούν να συσχετίζονται με την έναρξη ή την εξέλιξη της πνευμονικής αρτηριακής υπέρτασης. Πρόσθετη έρευνα απαιτείται για τη διερεύνηση των 195 ΔΕΓ και των βιολογικών τους λειτουργιών με στόχο τον εντοπισμό νέων βιοδεικτών για έγκαιρη διάγνωση και θεραπεία της ΠΑΥ.

Βιβλιογραφία - Αναφορές

1. Ahles, A., & Engelhardt, S. (2014). Polymorphic variants of adrenoceptors: pharmacology, physiology, and role in disease. *Pharmacological reviews*, 66(3), 598-637.
2. Alberts, B. (2002). *Molecular Biology of the Cell: Hauptbd.* Garland.
3. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215(3), 403-410.
4. Anandaram, H. (2020). Role of Bioinformatics in Nanotechnology: An Initiation towards Personalized Medicine. In *Data Analytics in Medicine: Concepts, Methodologies, Tools, and Applications* (pp. 1875-1894). IGI Global.
5. Austin, E. D., & Loyd, J. E. (2014). The genetics of pulmonary arterial hypertension. *Circulation research*, 115(1), 189-202.
6. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1), 289-300.
7. Bhagwani, A. R., Hultman, S., Farkas, D., Moncayo, R., Dandamudi, K., Zadu, A. K., ... & Farkas, L. (2019). Endothelial cells are a source of Nestin expression in Pulmonary Arterial Hypertension. *Plos one*, 14(3), e0213890.
8. Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis.* John Wiley & Sons.
9. Can, T. (2014). Introduction to bioinformatics. In *miRNomics: MicroRNA Biology and Computational Analysis* (pp. 51-71). Humana Press, Totowa, NJ.
10. Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, 10(1), 101-129.
11. Cohen J. (1987) *Statistical Power Analysis for the Behavioral Sciences* (2nd ed). Lawrence Erlbaum Associates, Hillsdale NJ.
12. DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled clinical trials*, 7(3), 177-188.
13. Dogini, D. B., Pascoal, V. D., Avansini, S. H., Vieira, A. S., Pereira, T. C., & Lopes-Cendes, I. (2014). The new world of RNAs. *Genetics and molecular biology*, 37(1), 285-293.
14. Dudoit, S., Yang, Y. H., Callow, M. J., & Speed, T. P. (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica sinica*, 111-139.
15. Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans.* Society for industrial and applied mathematics.
16. Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap.* CRC press.

17. Endeward, V., Musa-Aziz, R., Cooper, G. J., Chen, L. M., Pelletier, M. F., Virkki, L. V., ... & Gros, G. (2006). Evidence that aquaporin 1 is a major pathway for CO₂ transport across the human erythrocyte membrane. *The FASEB Journal*, 20(12), 1974-1981.
18. Erjefält, J. S. (2014). Mast cells in human airways: the culprit?. *European Respiratory Review*, 23(133), 299-307.
19. Ferrara, N., Gerber, H. P., & LeCouter, J. (2003). The biology of VEGF and its receptors. *Nature medicine*, 9(6), 669-676.
20. Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher*, 5(10), 3-8.
21. Gönen, M., Johnson, W. O., Lu, Y., & Westfall, P. H. (2005). The Bayesian two-sample *t* test. *The American Statistician*, 59(3), 252-257.
22. Goumans, M. J., Zwijsen, A., Ten Dijke, P., & Bailly, S. (2018). Bone morphogenetic proteins in vascular homeostasis and disease. *Cold Spring Harbor Perspectives in Biology*, 10(2), a031989.
23. Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107-128.
24. Heller, R. A., Schena, M., Chai, A., Shalon, D., Bedilion, T., Gilmore, J., ... & Davis, R. W. (1997). Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proceedings of the National Academy of Sciences*, 94(6), 2150-2155.
25. Herrera, M., Hong, N. J., & Garvin, J. L. (2006). Aquaporin-1 transports NO across cell membranes. *Hypertension*, 48(1), 157-164.
26. Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj*, 327(7414), 557-560.
27. Hoeper, M. M., Ghofrani, H. A., Grünig, E., Klose, H., Olschewski, H., & Rosenkranz, S. (2017). Pulmonary hypertension. *Deutsches Ärzteblatt International*, 114(5), 73.
28. Holland, B. S., & Copenhaver, M. D. (1988). Improved Bonferroni-type multiple testing procedures. *Psychological Bulletin*, 104(1), 145.
29. Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65-70.
30. Hong, F., & Breitling, R. (2008). A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments. *Bioinformatics*, 24(3), 374-382.
31. Hulsen, T., de Vlieg, J., & Alkema, W. (2008). BioVenn—a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC genomics*, 9(1), 1-6.
32. Knudsen, S. (2011). *A biologist's guide to analysis of DNA microarray data*. John Wiley & Sons.
33. Kontou, P. I., Pavlopoulou, A., & Bagos, P. G. (2018). Methods of analysis and meta-analysis for identifying differentially expressed genes. In *Genetic Epidemiology* (pp. 183-210). Humana Press, New York, NY.

34. Kruschke, J. K. (2013). *Bayesian estimation supersedes the t test*. *Journal of Experimental Psychology: General*, 142(2), 573.
35. Kwon, M. Y., Hwang, N., Park, Y. J., Perrella, M. A., & Chung, S. W. (2018). *NOD2 deficiency exacerbates hypoxia-induced pulmonary hypertension and enhances pulmonary vascular smooth muscle cell proliferation*. *Oncotarget*, 9(16), 12671.
36. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., ... & Funke, R. (2001). *Initial sequencing and analysis of the human genome*.
37. Lane, K. B., Machado, R. D., Pauciulo, M. W., Thomson, J. R., Phillips, J. A., Loyd, J. E., ... & Trembath, R. C. (2000). *Heterozygous germline mutations in BMPR2, encoding a TGF- β receptor, cause familial primary pulmonary hypertension*. *Nature genetics*, 26(1), 81-84.
38. Lesk, A. (2019). *Introduction to bioinformatics*. Oxford university press.
39. Li, M., Dong, X., Liu, Y., Sun, X., Li, Z., & He, J. (2011). *Inhibition of ubiquitin proteasome function suppresses proliferation of pulmonary artery smooth muscle cells*. *Naunyn-Schmiedeberg's archives of pharmacology*, 384(6), 517-523.
40. Li, M., Dong, X., Liu, Y., Sun, X., Li, Z., & He, J. (2011). *Inhibition of ubiquitin proteasome function suppresses proliferation of pulmonary artery smooth muscle cells*. *Naunyn-Schmiedeberg's archives of pharmacology*, 384(6), 517-523.
41. Li, Y., Zhou, P., Chen, H., Chen, Q., Kuang, X., Lu, C., ... & Qiu, Y. (2018). *Inflammation-restricted anti-inflammatory activities of a N-acyl ethanolamine acid amidase (NAAA) inhibitor F215*. *Pharmacological research*, 132, 7-14.
42. Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., ... & Moher, D. (2009). *The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration*. *Journal of clinical epidemiology*, 62(10), e1-e34.
43. Luscombe, N. M., Greenbaum, D., & Gerstein, M. (2001). *What is bioinformatics? A proposed definition and overview of the field*. *Methods of information in medicine*, 40(04), 346-358.
44. Meissner, M., Michailidou, D., Stein, M., Hrgovic, I., Kaufmann, R., & Gille, J. (2009). *Inhibition of Rac1 GTPase downregulates vascular endothelial growth factor receptor-2 expression by suppressing Sp1-dependent DNA binding in human endothelial cells*. *Experimental dermatology*, 18(10), 863-869.
45. Mitchell, J. A., Ahmetaj-Shala, B., Kirkby, N. S., Wright, W. R., Mackenzie, L. S., Reed, D. M., & Mohamed, N. (2015). *Role of prostacyclin in pulmonary hypertension*. *Global Cardiology Science and Practice*, 2014(4), 53.
46. Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). *Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement*. *Annals of internal medicine*, 151(4), 264-269.

47. Morrell, N. W., Aldred, M. A., Chung, W. K., Elliott, C. G., Nichols, W. C., Soubrier, F., ... & Loyd, J. E. (2019). *Genetics and genomics of pulmonary arterial hypertension*. *European Respiratory Journal*, 53(1), 1801899.
48. Movassagh, H., Tatari, N., Shan, L., Koussih, L., Alsubait, D., Khattabi, M., ... & Gounni, A. S. (2016). *Human airway smooth muscle cell proliferation from asthmatics is negatively regulated by semaphorin3A*. *Oncotarget*, 7(49), 80238.
49. Normand, S. L. T. (1999). *Meta-analysis: formulating, evaluating, combining, and reporting*. *Statistics in medicine*, 18(3), 321-359.
50. Perneger, T. V. (1998). *What's wrong with Bonferroni adjustments*. *Bmj*, 316(7139), 1236-1238.
51. Pniewska, E., & Pawliczak, R. (2013). *The involvement of phospholipases A2 in asthma and chronic obstructive pulmonary disease*. *Mediators of inflammation*, 2013.
52. Prins, K. W., & Thenappan, T. (2016). *WHO Group I Pulmonary Hypertension: Epidemiology and Pathophysiology*. *Cardiology clinics*, 34(3), 363.
53. Pullamsetti, S. S., Savai, R., Janssen, W., Dahal, B. K., Seeger, W., Grimminger, F., ... & Schermuly, R. T. (2011). *Inflammation, immunological reaction and role of infection in pulmonary hypertension*. *Clinical Microbiology and Infection*, 17(1), 7-14.
54. Quackenbush, J. (2002). *Microarray data normalization and transformation*. *Nature genetics*, 32(4), 496-501.
55. Ramasamy, A., Mondry, A., Holmes, C. C., & Altman, D. G. (2008). *Key issues in conducting a meta-analysis of gene expression microarray datasets*. *PLoS Med*, 5(9), e184
56. Rhodes, C. J., Batai, K., Bleda, M., Haimel, M., Southgate, L., Germain, M., ... & Arora, A. (2018). *Genetic determinants of risk in pulmonary arterial hypertension: international case-control studies and meta-analysis*.
57. Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). *Bayesian t tests for accepting and rejecting the null hypothesis*. *Psychonomic bulletin & review*, 16(2), 225-237.
58. Ryan, J. J., Thenappan, T., Luo, N., Ha, T., Patel, A. R., Rich, S., & Archer, S. L. (2012). *The WHO classification of pulmonary hypertension: A case-based imaging compendium*. *Pulmonary circulation*, 2(1), 107-121.
59. Saboor, F., Reckmann, A. N., Tomczyk, C. U., Peters, D. M., Weissmann, N., Kaschtanow, A., ... & Mietens, A. (2016). *Nestin-expressing vascular wall cells drive development of pulmonary hypertension*. *European Respiratory Journal*, 47(3), 876-888.
60. Schadt, E. E., Lamb, J., Yang, X., Zhu, J., Edwards, S., GuhaThakurta, D., ... & Lum, P. Y. (2005). *An integrative genomics approach to infer causal associations between gene expression and disease*. *Nature genetics*, 37(7), 710-717.
61. Schena, M., Heller, R. A., Theriault, T. P., Konrad, K., Lachenmeier, E., & Davis, R. W. (1998). *Microarrays: biotechnology's discovery platform for functional genomics*. *Trends in biotechnology*, 16(7), 301-306.

62. Schena, M., Shalon, D., Davis, R. W., & Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270(5235), 467-470.
63. Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318), 626-633.
64. Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3), 751-754.
65. Simonneau, G., Montani, D., Celermajer, D. S., Denton, C. P., Gatzoulis, M. A., Krowka, M., ... & Souza, R. (2019). Haemodynamic definitions and updated clinical classification of pulmonary hypertension. *European Respiratory Journal*, 53(1), 1801913.
66. Somanath, P. R., Malinin, N. L., & Byzova, T. V. (2009). Cooperation between integrin $\alpha v \beta 3$ and VEGFR2 in angiogenesis. *Angiogenesis*, 12(2), 177-185.
67. Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., ... & Jensen, L. J. (2016). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic acids research*, gkw937.
68. Thenappan, T., Ormiston, M. L., Ryan, J. J., & Archer, S. L. (2018). Pulmonary arterial hypertension: pathogenesis and clinical management. *Bmj*, 360, j5492.
69. Todorova, D., Simoncini, S., Lacroix, R., Sabatier, F., & Dignat-George, F. (2017). Extracellular vesicles in angiogenesis. *Circulation research*, 120(10), 1658-1673.
70. Tyers, M., & Mann, M. (2003). From genomics to proteomics. *Nature*, 422(6928), 193-197.
71. Voelkel, N. F., Vandivier, R. W., & Tuder, R. M. (2006). Vascular endothelial growth factor in the lung. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 290(2), L209-L221.
72. Vogenberg, F. R., Barash, C. I., & Pursel, M. (2010). Personalized medicine: part 1: evolution and development into theranostics. *Pharmacy and Therapeutics*, 35(10), 560.
73. Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics*, 10(1), 57-63.
74. Αλεπόρου-Μαρίνου, Β., Αργυροκαστρίτης, Α., Κομητοπούλου, Α., Πιαλόγλου, Π. & Σγουρίτσα, Β. (1999). Βιολογία Θετικής Κατεύθυνσης Γ' τάξης Ενιαίου Λυκείου.
75. Καψάλης Α., Μπουρμπουχάκης Ι.-Ε., Περάκη Β. και Σαλαμαστράκης Σ. (2013) Βιολογία β' γενικού λυκείου Γενικής Παιδείας.
76. Μπάγκος, Π., 2015. Βιοπληροφορική. [ηλεκτρ. βιβλ.] Αθήνα:Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Διαθέσιμο στο: <http://hdl.handle.net/11419/5016>
77. Νικολάου, Χ., Χουβαρδάς, Π., 2015. Υπολογιστική βιολογία. [ηλεκτρ. βιβλ.] Αθήνα:Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Διαθέσιμο στο: <http://hdl.handle.net/11419/1577>

ΠΑΡΑΡΤΗΜΑ

Αρχείο .ado που πραγματοποιεί στατιστικό έλεγχο με τυποποιημένη μέση διαφορά και διόρθωση Hedges

Καλείται σαν συνάρτηση από το CL του STATA με όνομα myDL5

```
program define myDL5, rclass
    version 10.1
    syntax , x(varlist numeric max=1) Type(varlist numeric max=1)
    Study(varlist numeric max=1) [ Effect(string) ]

preserve

tempvar last
qui bysort `study': gen `last'=_n==_N
qui sum `study'
local max=r(max)
tempvar x0 x1 sd0 sd1 n0 n1 d sed m gmfm gmfm_1 J g seg
qui gen double `x0'=.
qui gen double `x1'=.
qui gen double `sd0'=.
qui gen double `sd1'=.
qui gen double `n0'=.
qui gen double `n1'=.
qui gen double `d'=.
qui gen double `sed'=.
qui gen double `m'=.
qui gen double `gmfm'=.
qui gen double `gmfm_1'=.
qui gen double `J'=.
qui gen double `g'=.
qui gen double `seg'=.

forvalues i=1(1)`max' {
    qui sum `x' if `study'==`i' & `type'==0
    qui replace `x0'=r(mean) if `study'==`i'
    qui replace `sd0'=r(sd) if `study'==`i'
    qui replace `n0'=r(N) if `study'==`i'

    qui sum `x' if `study'==`i' & `type'==1
    qui replace `x1'=r(mean) if `study'==`i'
    qui replace `sd1'=r(sd) if `study'==`i'
    qui replace `n1'=r(N) if `study'==`i'

    qui replace `m'=(`n1'+`n0'-2) if `study'==`i'
    qui replace `d'=(`x1' - `x0')/(sqrt( ( (`n1'- 1)*(`sd1')^2 +(`n0'-
1)*(`sd0')^2)/(`n1' + `n0' -2 ))) if `study'==`i'
    qui replace `sed'= sqrt( (`n1' + `n0')/(`n1'*`n0') +(`d')^2/(2*(`n1'+`n0')))
if `study'==`i'
    qui replace `gmfm' = exp((lgamma(`m'/2))) if `study'==`i'
    qui replace `gmfm_1'= exp((lgamma((`m'-1)/2))) if `study'==`i'
    qui replace `J' =`gmfm'/((sqrt(`m'/2))*`gmfm_1') if `study'==`i'
    qui replace `g' =`J'*`d' if `study'==`i'
    qui replace `seg'=(`J')^2*`sed' if `study'==`i'

** var(g)=J*var(d)
}

qui keep if `last'
qui keep if `x' !=.
```

Σελίδα 74

```

metan `g' `seg' , `effect' nograph
**metan `n1' `x1' `sd1' `n0' `x0' `sd0', `effect' nograph
**metan `n1' `x1' `sd1' `n0' `x0' `sd0',fixed nograph hedges
**metan `d' `sed' , random nograph
**list

restore

return scalar g=r(ES)
return scalar seg=r(seES)

end

```

Αρχείο .do

Εκτελείται μέσα από το CL του STATA

Επιστρέφει αρχείο με όνομα resultmyDL5.txt που περιέχει τρεις μεταβλητές (gene, g, seg)

```

set maxvar 32767
set memory 70m

set more off
file open meta using "C:\Users\...\resultmyDL5.txt", write append
file write meta "gene"
file write meta ","
file write meta "g"
file write meta ","
file write meta "seg" _n

foreach var of varlist albg - zcchc18{
    di ""
    di ""
    di in ye "Performing meta-analysis for `var' gene "
    di ""

    myDL5,x(`var') type(type) study(study) effect(random)

    file write meta "`var'"
    file write meta ","
    file write meta "`r(g)'"
    file write meta ","
    file write meta "`r(seg)'"_n
}

file close meta

```

Λοιπές εντολές που χρησιμοποιήθηκαν στο περιβάλλον STATA

```
//Εντολές που φτιάχνουν το z-score και το p-value του z-score μετά την μετα-ανάλυση  
//που επέστρεψε το αρχείο resultmyDL5.txt
```

```
generate z= g/ se  
generate p=2*(1-normal(abs( z )))
```

```
//Εντολές που εφαρμόζουν διορθωτικές μεθόδους για τα σφάλματα τύπου I σε επίπεδο  
//σημαντικότητας 1%
```

```
multproc, puncor(0.01) pval(p) meth(simes) rej(fdr)  
multproc, pval(p) meth(bonferroni) rej(bonf)  
multproc, pval(p) meth(sidak) rej(sidak)  
multproc, pval(p) meth(holm) rej(holm)  
multproc, pval(p) meth(holland) rej(holland)
```