



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

**ΜΕΤΑ-ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΓΟΝΙΔΙΑΚΗΣ ΕΚΦΡΑΣΗΣ**

**Κωνσταντίνα Βέννου**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**Υπεύθυνος**

**Παντελής Μπάγκος**

**Αναπληρωτής Καθηγητής**

**Λαμία, Νοέμβριος 2016**





**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ  
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ  
ΔΙΑΤΜΗΜΑΤΙΚΟ ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ  
ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΙΑΤΡΙΚΗ  
ΚΑΤΕΥΘΥΝΣΗ «ΥΠΟΛΟΓΙΣΤΙΚΗ ΙΑΤΡΙΚΗ ΚΑΙ  
ΒΙΟΛΟΓΙΑ»**

**ΜΕΤΑ-ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΓΟΝΙΔΙΑΚΗΣ ΕΚΦΡΑΣΗΣ**

**Κωνσταντίνα Βέννου**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**Επιβλέπων  
Παντελής Μπάγκος  
Αναπληρωτής Καθηγητής**

**Λαμία, Νοέμβριος 2016**

«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 «Μετα-ανάλυση δεδομένων γονιδιακής έκφρασης» 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 ΔΗΛΟΥΣΑ

Κωνσταντίνα Βέννου

Ημερομηνία

23/11/2016

Υπογραφή

## **ΜΕΤΑ-ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΓΟΝΙΔΙΑΚΗΣ ΕΚΦΡΑΣΗΣ**

**Κωνσταντίνα Βέννου**

### **Τριμελής Επιτροπή:**

Παντελής Μπάγκος, Αναπληρωτής Καθηγητής (επιβλέπων)

Βασίλειος Πλαγιανάκος, Αναπληρωτής Καθηγητής

Μαρία Αδάμ, Επίκουρος Καθηγητής



## ΠΕΡΙΕΧΟΜΕΝΑ

ΠΡΟΛΟΓΟΣ.....	8
ΠΕΡΙΛΗΨΗ .....	12
ABSTRACT.....	16
ΚΕΦΑΛΑΙΟ 1 <sup>ο</sup> : ΠΡΟΕΚΛΑΜΨΙΑ.....	20
1.1 ΟΡΙΣΜΟΣ .....	22
1.2 ΠΑΡΑΓΟΝΤΕΣ ΚΙΝΔΥΝΟΥ ΓΙΑ ΤΗΝ ΕΜΦΑΝΙΣΗ ΠΡΟΕΚΛΑΜΨΙΑΣ .....	23
1.3 ΕΠΙΠΛΟΚΕΣ .....	24
1.4 ΓΕΝΕΤΙΚΗ ΠΡΟΣΕΓΓΙΣΗ .....	25
1.5 ΠΡΟΛΗΨΗ ΚΑΙ ΘΕΡΑΠΕΥΤΙΚΗ ΑΝΤΙΜΕΤΩΠΙΣΗ.....	26
ΚΕΦΑΛΑΙΟ 2 <sup>ο</sup> : ΜΙΚΡΟΣΥΣΤΟΙΧΙΕΣ .....	28
2.1. ΕΙΣΑΓΩΓΗ.....	30
2.2 ΚΑΤΑΣΚΕΥΗ ΜΙΚΡΟΣΥΣΤΟΙΧΙΩΝ .....	31
2.3 ΠΕΙΡΑΜΑΤΙΚΗ ΔΙΑΔΙΚΑΣΙΑ ΤΩΝ ΜΙΚΡΟΣΥΣΤΟΙΧΙΩΝ.....	33
2.3.1 Η ιδέα του πειράματος, το βιολογικό ερώτημα.....	33
2.3.2 Ο τύπος του τσιπ των μικροσυστοιχιών .....	33
2.3.2.1 Μικροσυστοιχίες εκτύπωσης .....	34
2.3.2.2 Μικροσυστοιχίες σύνθεσης <i>in situ</i> .....	35
2.4 ΠΑΡΑΣΚΕΥΗ ΤΟΥ ΔΕΙΓΜΑΤΟΣ .....	36
2.4.1 Ιχνηθέτηση.....	37
2.4.2 Υβριδοποίηση.....	38
2.4.3 Σάρωση.....	38
2.4.4 Κανονικοποίηση (normalization) .....	40
2.4.4.1 Κανονικοποίηση ολική έντασης (total intensity normalization) .....	41
2.4.4.2 Lowess (LOcally WEighted Scatterplot Smoothing regression) κανονικοποίηση .....	42
2.5 ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ .....	43
2.5.1 Αλλαγή διπλώματος (fold-change, FC).....	43
2.5.2 Έλεγχος t (t-test).....	43
2.5.2.1 Μέθοδοι διόρθωσης <i>p-value</i> .....	44
Διόρθωση Bonferroni .....	44
Διόρθωση Bonferroni Step-Down (Holm) .....	45
Διόρθωση Holland .....	45
Benjamini and Hochberg False Discovery Rate (FDR).....	45
Διόρθωση Sidak .....	45

2.5.3 Μέθοδοι αναδειγματοληψίας .....	46
2.6 Βάσεις δεδομένων μικροσυστοιχιών .....	46
<b>ΚΕΦΑΛΑΙΟ 3<sup>ο</sup>: ΜΕΤΑ-ΑΝΑΛΥΣΗ .....</b>	<b>48</b>
3.1 ΕΙΣΑΓΩΓΗ.....	50
3.2 ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΜΕΤΑ-ΑΝΑΛΥΣΗΣ.....	50
3.3 ΕΠΙΛΟΓΗ ΣΤΑΤΙΣΤΙΚΟΥ ΜΟΝΤΕΛΟΥ .....	51
3.4 ΈΛΕΓΧΟΣ ΕΤΕΡΟΓΕΝΕΙΑΣ .....	54
3.5 ΠΡΟΒΛΗΜΑΤΑ ΒΙΒΛΙΟΓΡΑΦΙΑΣ.....	56
3.6 ΜΕΤΑ-ΑΝΑΛΥΣΗ ΣΕ ΔΕΔΟΜΕΝΑ ΜΙΚΡΟΣΥΣΤΟΙΧΙΩΝ .....	57
3.6.1 Μέθοδος βασισμένη στο t-test.....	57
3.6.2 Μέθοδος συνδυασμού των p-values .....	58
3.6.3 Μέθοδος υπολογισμού του γινομένου των βαθμών κατάταξης (rank product) .....	58
<b>ΚΕΦΑΛΑΙΟ 4<sup>ο</sup>: ΜΕΘΟΔΟΛΟΓΙΑ .....</b>	<b>60</b>
4.1 ΣΚΟΠΟΣ.....	62
4.2 ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ .....	62
4.3 ΚΑΤΑΓΡΑΦΗ ΔΕΔΟΜΕΝΩΝ .....	62
4.4 Η ΔΙΑΔΙΚΑΣΙΑ ΤΗΣ ΜΕΤΑ-ΑΝΑΛΥΣΗΣ ΜΕ ΤΗ ΧΡΗΣΗ ΤΟΥ ΣΤΑΤΙΣΤΙΚΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ STATA .....	65
<b>ΚΕΦΑΛΑΙΟ 5<sup>ο</sup>: ΑΠΟΤΕΛΕΣΜΑΤΑ - ΣΥΖΗΤΗΣΗ.....</b>	<b>66</b>
5.1 ΑΠΟΤΕΛΕΣΜΑΤΑ.....	68
5.2 ΣΥΖΗΤΗΣΗ.....	79
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ.....</b>	<b>82</b>
<b>ΠΑΡΑΡΤΗΜΑ .....</b>	<b>90</b>



## Πρόλογος



Η παρούσα διπλωματική εργασία εκπονήθηκε στα πλαίσια του διατμηματικού μεταπτυχιακού κύκλου σπουδών του Τμήματος Πληροφορικής με Εφαρμογές στη Βιοϊατρική της Σχολής Θετικών Επιστημών του Πανεπιστημίου Θεσσαλίας κατά τη διάρκεια του ακαδημαϊκού έτους 2015-2016, υπό την επίβλεψη του κυρίου Παντελή Μπάγκου, Αναπληρωτή Καθηγητή του Τμήματος Πληροφορικής με Εφαρμογές στη Βιοϊατρική του Πανεπιστημίου Θεσσαλίας.

Οφείλω ένα μεγάλο ευχαριστώ στους ανθρώπους που με βοήθησαν, με καθοδήγησαν και με στήριξαν στην προσπάθεια για την ολοκλήρωση της παρούσας εργασίας.

Πρώτα στον επιβλέπων της εργασίας κύριο Παντελή Μπάγκο, Αναπληρωτή Καθηγητή του Τμήματος Πληροφορικής με Εφαρμογές στη Βιοϊατρική του Πανεπιστημίου Θεσσαλίας, για την εμπιστοσύνη που μου έδειξε κατά την ανάθεση της διπλωματικής μου εργασίας, για την καθοδήγηση του καθώς και τις πολύτιμες συμβουλές του.

Θα ήθελα επίσης να ευχαριστήσω την κυρία Παναγιώτα Κοντού, υποψήφια διδάκτωρ του Τμήματος Πληροφορικής με Εφαρμογές στη Βιοϊατρική του Πανεπιστημίου Θεσσαλίας για την πολύτιμη βοήθειά της για την ολοκλήρωση της διπλωματικής εργασίας.

Θα ήταν παράληψη να μην ευχαριστήσω όλη την επιστημονική ομάδα του Εργαστήριο Μοριακής και Υπολογιστικής Βιολογίας και Γενετικής (ΕΜΥΒΓ) του Τμήματος Πληροφορικής με Εφαρμογές στη Βιοϊατρική του Πανεπιστημίου Θεσσαλίας και τους συναδέλφους μου για την εξαιρετική συνεργασία μας κατά τη διάρκεια εκπόνησης της παρούσας διπλωματικής εργασίας.

Τέλος θα ήθελα να ευχαριστήσω επίσης την οικογένειά μου για την διαρκή τους υποστήριξη σε όλα.



## Περίληψη



Η προεκλαμψία αποτελεί μια σοβαρή επιπλοκή της κύησης, με κλινικά συμπτώματα την παρουσία νέο-εμφανιζόμενης υπέρτασης και λευκωματουρίας, μετά την 20<sup>η</sup> εβδομάδα της κύησης και χαρακτηρίζεται από σημαντικά ποσοστά μητρικής και νεογνικής νοσηρότητας και θνησιμότητας. Υπάρχουν ενδείξεις ότι η προεκλαμψία θα μπορούσε να θεωρηθεί μια γενετική νόσος ωστόσο παραμένει αδιευκρίνιστος ο ακριβής τρόπος κληρονόμησης της. Είναι κοινώς αποδεκτό ότι η προεκλαμψία είναι γενετικά ετερογενές νόσημα, μέχρι σήμερα έχουν βρεθεί διάφορα υποψήφια γονίδια για την προεκλαμψία, τα οποία σχετίζονται με την πλακουντοποίηση, την ρύθμιση της αρτηριακής πίεσης, την αγγειακή διαμόρφωση και τη λειτουργία των ενδοθηλιακών κυττάρων. Στην συγκεκριμένη μελέτη έγινε μια προσπάθεια εντοπισμού γενετικών παραγόντων οι οποίοι οδηγούν στην εμφάνιση της προεκλαμψίας μέσω της στατιστικής μεθόδου της μετα-ανάλυσης πειραματικών δεδομένων μικροσυστοιχιών. Οι μικροσυστοιχίες παρέχουν τη δυνατότητα την ανάλυση της γονιδιακής έκφρασης πολλών γονιδίων ταυτόχρονα. Μελετώντας ανεξάρτητες έρευνες για την ίδια γενική υπόθεση, αποκτώνται πιο αντιπροσωπευτικές και ακριβείς εκτιμήσεις της διαφορικής έκφρασης γονιδίων. Στην παρούσα μελέτη η μέθοδος της μετα-ανάλυσης υλοποιήθηκε με μελέτες από την βάση δεδομένων GEO και με τη χρήση του στατιστικού προγράμματος STATA, όπου υπολογίστηκε ο αριθμός των στατιστικά σημαντικών γονιδίων με τη χρήση μεθόδων διόρθωσης (διόρθωση FDR, Bonferroni, Holm, Holland, Sidak). Από την μετα-ανάλυση προέκυψαν 783 γονίδια τα οποία εμφανίζουν στατιστικά σημαντική συσχέτιση με την εμφάνιση προεκλαμψίας, το οποίο δεν γινόταν εμφανές από τις μελέτες μεμονωμένα. Περαιτέρω έρευνα κρίνεται αναγκαία για τη διερεύνηση της συσχέτισης των γονιδίων που βρέθηκαν στην συγκεκριμένη μελέτη με την εμφάνιση της προεκλαμψίας.

**Λέξεις - κλειδιά:** Μετα-ανάλυση, Προεκλαμψία, Μικροσυστοιχίες, Γονιδιακή έκφραση





## **Abstract**



Preeclampsia is a serious complication of pregnancy, with clinical symptoms the presence of newly occurring hypertension and proteinuria after the 20<sup>th</sup> week of pregnancy and it is characterized by significant rates of maternal and neonatal morbidity and mortality. There is evidence that preeclampsia could be considered as a genetic disease, however, remains unclear the precise mode of inheritance. It is commonly accepted that preeclampsia is a heterogeneous genetically disease. There have been several candidate genes for preeclampsia, which related to placentation, blood pressure regulation, vascular formation and function of endothelial cells. This study was an attempt to identify genetic factors that could lead to the onset of preeclampsia, with the use of the statistical method of meta-analysis at microarrays data. Microarrays allow to analyze gene expression of many genes simultaneously. Studying independent studies on the same general assumption, provide more representative and accurate estimates of differential gene expression. In this study, the method of meta-analysis was carried out with studies from GEO database and with the use of the statistical program STATA, wherein the number calculated the statistically significant genes using correction methods (FDR correction, Bonferroni, Holm, Holland, Sidak). From the procedure of meta-analysis 783 genes have exhibit a statistically significant correlation with the occurrence of preeclampsia, something that was not evident in the individual studies. Further research is thought to be necessary to investigate the correlation of genes found in this study with the onset of preeclampsia.

**Key words:** Meta-analysis, Preeclampsia, Microarray, Gene expression



## **Κεφάλαιο 1<sup>ο</sup>:**

### **Προεκλαμψία**



## 1.1 Ορισμός

Η προεκλαμψία αποτελεί μια σοβαρή επιπλοκή της κύησης και χαρακτηρίζεται από σημαντικά ποσοστά μητρικής και νεογνικής νοσηρότητας και θνησιμότητας. Το πρόβλημα του συνδρόμου προεκλαμψίας ή τοξιναιμίας της κύησης είναι πολύ παλιό. Ο Ιπποκράτης, τον 4<sup>ο</sup> αιώνα προ Χριστού, στο κεφάλαιο «περί αφορισμών», απέδωσε τον όρο *προεκλαμψία* σαν ένα είδος αυτοδηλητηρίασης της εγκύου γυναικός (Μαμόπουλος 1988, Gursoy et al. 2002).

Αν και δεν υπάρχουν ακριβή στοιχεία θεωρείται ότι σε όλο τον κόσμο, το ποσοστό της προεκλαμψίας ανέρχεται στο 3-5% των κύσεων (Hutcheon, Lisonkova, and Joseph 2011). Η παθογένεια της προεκλαμψίας δεν έχει κατανοηθεί πλήρως και η μόνη αποτελεσματική θεραπεία έχει αποδειχθεί να είναι ο τοκετός.

Η διάγνωση της προεκλαμψίας στηρίζεται στην παρουσία νέο-εμφανιζόμενης υπέρτασης και λευκωματουρίας, μετά την 20<sup>η</sup> εβδομάδα της κύησης (Gifford et al. 2000). Ως υπέρταση χαρακτηρίζεται η τιμή της συστολικής αρτηριακής πίεσης  $\geq 140\text{mmHg}$  και της διαστολικής αρτηριακής πίεσης  $\geq 90\text{mmHg}$ , σε μετρήσεις που απέχουν τουλάχιστον 4-6 ώρες η μια από την άλλη, σε προηγουμένως νορμοτασική έγκυο. Η υπέρταση χαρακτηρίζεται σοβαρή, όταν η συστολική αρτηριακή πίεση είναι  $\geq 160\text{mmHg}$  ή διαστολική αρτηριακή πίεση είναι  $\geq 110\text{mmHg}$  (Hauth et al. 2000, Buchbinder et al. 2002). Η λευκωματουρία ορίζεται ως ποσότητα λευκώματος μεγαλύτερη από 0,3 g σε συλλογή 24ώρου (Gifford et al. 2000, Sibai 2003). Σε μερικές περιπτώσεις προεκλαμψία εμφανίζεται και χωρίς την παρουσία λευκωματουρίας. Στις περιπτώσεις αυτές η υπέρταση συνοδεύεται από επίμονη κεφαλαλγία, διαταραχές της όρασης, επιγαστραλγία ή άλγος του δεξιού υποχονδρίου με ναυτία και εμέτους, θρομβοπενία και παθολογικές τιμές των ηπατικών ενζύμων (Sibai 2003, Brown et al. 2000).

Η προεκλαμψία διακρίνεται σε δύο μορφές την ήπια και την βαρεία είτε σοβαρή, ανάλογα με τη βαρύτητά της. Βαρεία θεωρείται όταν παρατηρείται σοβαρή αρτηριακή υπέρταση (ΑΠ  $\geq 160\text{mmHg}/110\text{mmHg}$ ) σε συνδυασμό με λευκωματουρία, ή όταν η υπέρταση συνοδεύεται από υψηλού βαθμού λευκωματουρία ( $\geq 2$  g σε συλλογή ούρων 24ώρου) (Sibai 2003). Ως βαρεία, επίσης, χαρακτηρίζεται η προεκλαμψία, η οποία συνοδεύεται από πνευμονικό και γενικευμένο οίδημα (άνω άκρα και πρόσωπο), ολιγουρία, συμπτώματα από το Κεντρικό Νευρικό Σύστημα (κεφαλαλγία, διαταραχές όρασης, σκοτώματα, παροδική τύφλωση), θρομβοπενία,

εκλαμψία, αυξημένα ηπατικά ένζυμα συνδυαστικά με επίμονη επιγαστραλγία ή άλγος του δεξιού υποχονδρίου και καθυστέρηση της ενδομήτριας ανάπτυξης (Brown et al. 2000, Sibai 2003).

## 1.2 Παράγοντες κινδύνου για την εμφάνιση προεκλαμψίας

Η προεκλαμψία θεωρείται νόσος της πρώτης κύησης. Έχει βρεθεί ότι ιστορικό φυσιολογικής κύησης είτε προηγούμενης αποβολής (τεχνητής είτε αυτόματης) μειώνει τον κίνδυνο για προεκλαμψία, ενώ η αλλαγή συντρόφου καταργεί αυτή την προστατευτική δράση σε επόμενη κύηση (Safilas et al. 2003).

Οι κυριότερες αιτίες εμφάνισης προεκλαμψίας από την πλευρά της μητέρας είναι οι εξής:

- Προϋπάρχουσα λανθάνουσα χρόνια νεφρική νόσος
- Ιστορικό σακχαρώδη διαβήτη
- Ανοσολογικοί παράγοντες
- Χαμηλές κοινωνικοοικονομικές τάξεις (διαιτητικός παράγων)
- Χαμηλό ανάστημα (κάτω από 155 cm)
- Ακραίες ηλικίες αναπαραγωγής (κάτω των 17 και άνω των 35 ετών)
- Παχυσαρκία και αντίσταση στην ινσουλίνη
- Οικογενειακό ιστορικό προεκλαμψίας
- Γονιδιακή προδιάθεση
- Συστηματικά νοσήματα του κολλαγόνου
- Χαμηλό βάρος γέννησης της μητέρας
- Λοιμώξεις
- Ιστορικό θρομβοφιλίας
- Χρόνια ιδιοπαθής υπέρταση
- Περιορισμένη έκθεση στο σπέρμα του πατέρα είτε αλλαγή συντρόφου (νέος πατέρας)
- Υπερομοκυστεϊναμία
- Μαύρη φυλή (Sibai 2006).

Από την πλευρά του εμβρύου, πάλι οι κυριότερες αιτίες είναι:

- Πολύδυμη κύηση
- Αυξημένος όγκος πλακούντα



- Υδατιδώδης μύλη και
- Πλακουντικός ύδρωπας (από ασυμβατότητα Rhesus) (Roberts et al. 1996).

Η κύηση, φυσιολογική, θεωρείται στον οργανισμό ως μια κατάσταση φλεγμονής, κυρίως στο τρίτο τρίμηνο. Σύμφωνα με αυτό η προεκλαμψία θεωρείται ως μια ακραία έκφραση της συστηματικής φλεγμονής της μητέρας. Με αποτέλεσμα οποιοσδήποτε παράγοντας προκαλεί μεταβολή στην απάντηση της μητέρας στη «φλεγμονή» είναι δυνατόν να οδηγήσει σε προεκλαμψία (Redman and Sargent 2003).

### 1.3 Επιπλοκές

Οι οξείες επιπλοκές οι οποίες μπορεί να συμβούν όταν μια έγκυος εμφανίσει προεκλαμψία είναι οι παρακάτω.

Στην μητέρα είναι:

- Διάχυτη ενδαγγειακή πήξη, σύνδρομο HELLP (10-20%)
- Πνευμονικό οίδημα (2-5%)
- Αποκόλληση πλακούντα (1-4%)
- Οξεία νεφρική ανεπάρκεια (1-5%)
- Εκλαμψία (<1%)
- Ηπατική ανεπάρκεια ή αιμορραγία (<1%)
- Θάνατος (σπάνια)

Και από την πλευρά του εμβρύου είναι:

- Πρόωρος τοκετός (15-67%)
- Καθυστέρηση της εμβρυϊκής ανάπτυξης (10-25%)
- Περιγεννητικός θάνατος (1-2%)
- Υποξία, νευρολογική βλάβη (<1%)

Οι επιπτώσεις της προεκλαμψίας στη μητέρα και το έμβρυο εξαρτώνται από τη βαρύτητα της νόσου, την ηλικία της κύησης κατά την εμφάνιση της νόσου, την παρουσία ή όχι άλλων παθολογικών καταστάσεων στη μητέρα και την ποιότητα των παρεχόμενων ιατρικών υπηρεσιών. Σε ήπια προεκλαμψία μετά την 36<sup>η</sup> εβδομάδα της κύησης η έκβαση της μητέρα και του εμβρύου είναι συνήθως ευνοϊκή (Sibai 2003, Hauth et al. 2000). Σε αντίθεση σε κυήσεις που εμφανίζεται πριν την 33<sup>η</sup> εβδομάδα

της κύησης και σε κύηση με προϋπάρχοντα προβλήματα αυξάνονται σημαντικά η μητρική και εμβρυική θνησιμότητα (Sibai 2003, Zhang, Meikle, and Trumble 2003).

Επιπροσθέτως γυναίκες με ιστορικό προεκλαμψίας έχουν αυξημένο κίνδυνο εμφάνισης επιπλοκών στο καρδιαγγειακό σύστημα κατά τη διάρκεια της ζωής τους (Manten et al. 2007) και παρουσιάζουν μεγαλύτερο κίνδυνο εμφάνισης υποθυρεοειδισμού (Levine et al. 2009).

## 1.4 Γενετική προσέγγιση

Υπάρχουν ενδείξεις ότι η προεκλαμψία θα μπορούσε να θεωρηθεί μια γενετική νόσος, από την παρατήρηση ότι υπάρχει προδιάθεση για την εμφάνιση προεκλαμψίας σε γυναίκες των οποίων συγγενείς πρώτου βαθμού εμφάνισαν προεκλαμψία (Cooper, Brennecke, and Wilton 1993). Εκ παραλλήλου, όμοια, υψηλά ποσοστά εμφάνισης της προεκλαμψίας σε μονοζυγωτικά δίδυμα συγκριτικά με διζυγωτικά δίδυμα θεωρούνται αποδείξεις της γονιδιακής συμμετοχής στην εκδήλωση της προεκλαμψίας (Salonen Ros et al. 2000). Εντούτοις η συσχέτιση της προεκλαμψίας με χρωμοσωμικές ανωμαλίες του εμβρύου φανερώνουν ότι και ο γονότυπος του εμβρύου (πατέρας) ενδέχεται να επηρεάζει την εμφάνιση της προεκλαμψίας. Επιπλέον και οι περιβαλλοντικοί παράγοντες θεωρούνται από μελέτες ως υπαίτιοι για την εμφάνιση της προεκλαμψίας (Chappell and Morgan 2006, Serrano 2006).

Μολονότι είναι αποδεκτός ο ρόλος των γενετικών παραγόντων στην αιτιολογία της προεκλαμψίας, παραμένει αδιευκρίνιστος ο ακριβής τρόπος κληρονομής. Μια σχετικά πρόσφατη εκδοχή θεωρεί πως η προδιάθεση για προεκλαμψία αποτελεί μια σύνθετη αλληλεπίδραση μεταξύ ενός συνόλου γονιδίων μητρικής προέλευσης με περιβαλλοντολογικούς παράγοντες (Bodnar, Caton, and Roberts 2007) και εμβρυϊκούς γονότυπους (Walker 2000). Είναι πλέον κοινώς αποδεκτό ότι η προεκλαμψία είναι γενετικά ετερογενές νόσημα, το οποίο δεν κληρονομείται με τους απλούς (επικρατή/υπολειπόμενα γονίδια) νόμους του Mendel (Lachmeijer et al. 2001). Μέχρι σήμερα έχουν βρεθεί διάφορα υποψήφια γονίδια για την προεκλαμψία, τα οποία σχετίζονται με την πλακουντοποίηση, την ρύθμιση της αρτηριακής πίεσης, την αγγειακή διαμόρφωση και τη λειτουργία των ενδοθηλιακών κυττάρων (Chappell and Morgan 2006).

## 1.5 Πρόληψη και θεραπευτική αντιμετώπιση

Τις τελευταίες δύο δεκαετίες έχουν γίνει προσπάθειες εύρεσης μεθόδων, οι οποίες αποσκοπούν στη μείωση της συχνότητας εμφάνισης και βαρύτητας της προεκλαμψίας. Μελετήθηκαν αρκετές μη φαρμακευτικές παρεμβάσεις όπως τα συμπληρώματα πρωτεϊνών, ψευδαργύρου, μαγνησίου και ω3 λιπαρών οξέων, η σωματική άσκηση, η μείωση του σωματικού βάρους, η αύξηση είτε η μείωση πρόσληψης αλάτος, τα οποία αποδείχτηκε να μην έχουν κάποια σαφή προφυλακτική δράση (Makrides, Duley, and Olsen 2006). Η χορήγηση ασβεστίου, ωστόσο εμφανίστηκε να επηρεάζει τη συχνότητα πρόωρων τοκετών καθώς επίσης και τους νεογνικούς θανάτους από προεκλαμψία (Trumbo and Ellwood 2007). Μελέτες που έγιναν για τη λήψη βιταμινών C και E έδειξαν ότι δεν δρουν προληπτικά αντίθετα αυξάνουν τη συχνότητα πρόωρου τοκετού (Polyzos et al. 2007). Από την άλλη αρκετές τυχαιοποιημένες μελέτες έδειξαν ότι η χορήγηση ασπιρίνης δρα ευεργετικά σε γυναίκες η οποίες έχουν υψηλό κίνδυνο εμφάνισης προεκλαμψίας (βάση παθολογικών ευρημάτων σε συγκεκριμένες εξετάσεις και το ιστορικό τους) (Chiapparino et al. 2004, Ebrashy et al. 2005). Επιπρόσθετα η χρήση ηπαρίνης χαμηλού μοριακού βάρους εμφανίστηκε να δρα ευεργετικά σε γυναίκες με ιστορικό θρομβοφιλίας, νεφρική πάθησης είτε προεκλαμψίας σε προηγούμενη κύηση (Sergio et al. 2006).

Ένας από τους σημαντικότερους παράγοντες αντιμετώπισης της προεκλαμψίας είναι η σωστή προγεννητική φροντίδα ώστε να γίνει πρόωμη αναγνώριση των κλινικών συμπτωμάτων της νόσου, πριν εξελιχθεί σε σοβαρότερη μορφή. Έπειτα η επιλογή των σωστών θεραπευτικών χειρισμών εξαρτάται από την ηλικία της κύησης, τη βαρύτητα της νόσου και φυσικά την κατάσταση του εμβρύου. Κάποιοι από αυτούς μπορεί να είναι είτε ο πρόωρος τοκετός, είτε η χορήγηση αντιυπερτασικών φαρμάκων, είτε κορτικοστεροειδών (Sibai 2003, Gifford et al. 2000).

Τέλος στην προσπάθεια εύρεσης νέων θεραπευτικών μεθόδων ερευνάται η χορήγηση ουσιών είτε για την καταστολή συγκεκριμένων ανοσολογικών μονοπατιών που ενεργοποιούνται κατά την προεκλαμψία (Guo et al. 2000, Harkin et al. 2004) είτε κατά την απόπτωση του πλακούντα (Zimmermann, Bonzon, and Green 2001).

Φαίνεται λοιπόν επιτακτική η ανάγκη εύρεσης των γονιδιακών τόπων που επηρεάζουν την εμφάνιση προεκλαμψίας, ώστε γυναίκες με γενετική προδιάθεση, της

νόσου, να τους παρέχονται οι κατάλληλες ιατρικές υπηρεσίες και είτε να προλαμβάνεται είτε να αντιμετωπίζεται στην αρχή της εμφάνισής της κατά την οποία είναι πιο ήπια η βαρύτητα της νόσου.

## **Κεφάλαιο 2<sup>ο</sup>:**

### **Μικροσυστοιχίες**



## 2.1. Εισαγωγή

Όλοι οι ζωντανοί οργανισμών περιέχουν δεοξυριβονουκλεϊκό οξύ το γνωστό DNA, ένα μόριο που κωδικοποιεί όλες τις απαιτούμενες πληροφορίες για την ανάπτυξη και τη λειτουργία του κάθε οργανισμού. Το DNA είναι ένα φυσικό μόριο, το οποίο είναι ικανό να κωδικοποιεί πληροφορίες σε μια γραμμική δομή. Το κάθε κύτταρο εκφράζει πληροφορίες από διάφορα μέρη αυτής της δομής με μια σχέση εξάρτησης-περιεχομένου. Το DNA κωδικοποιεί τα γονίδια και τα ρυθμιστικά στοιχεία ελέγχοντας εάν τα γονίδια είναι ενεργά ή ανενεργά (on/off). Είναι γνωστό ότι όλα τα κύτταρα του ανθρώπινου σώματος περιέχουν το ίδιο DNA, ωστόσο υπάρχουν εκατοντάδες διαφορετικά είδη κυττάρων, το καθένα από τα οποία εκφράζει μια μοναδική γονιδιακή σύνθεση του DNA. Κατά μία έννοια το DNA θα μπορούσε να θεωρηθεί ότι βρίσκεται σε ένα αριθμό καταστάσεων. Οι μικροσυστοιχίες γνωστές και ως DNA τσιπ (chip) είναι ένα εργαλείο το οποίο χρησιμοποιείται για να διαβάσει τις καταστάσεις του DNA. Η νέα αυτή τεχνολογία αναπτύχθηκε στις αρχές της δεκαετίας του '90. Οι μικροσυστοιχίες είχαν μια μεταμορφωτική επίδραση στις βιολογικές επιστήμες. Όπως η εφεύρεση του τηλεσκοπίου άλλαξε τον τρόπο αντιλαμβανόμεστε το σύμπαν, έτσι και οι μικροσυστοιχίες άλλαξαν την τρόπο που αντιλαμβανόμεστε το γονιδίωμα.

Τα μοτίβα που λαμβάνονται από τα πειράματα των μικροσυστοιχιών έχουν βοηθήσει τους ερευνητές να κατανοήσουν γενετικούς μηχανισμούς, την εξέλιξη κάποιων νόσων (Miklos and Maleszka 2004, Breitling, Amtmann, and Herzyk 2004), την πρόβλεψη μοριακών λειτουργιών γονιδίων (Park, Kim, and Kim 2001, Spellman et al. 1998), την κατασκευή λειτουργικών μονοπατιών (Roberts et al. 2000), και τον εντοπισμό νέων γονιδίων ή παραλλαγές ματίσματος τους (Shoemaker et al. 2001).

Οι μικροσυστοιχίες μετρούν γεγονότα στο γονιδίωμα. Ένα γεγονός μπορεί να είναι η μεταγραφή ενός γονιδίου, η σύνδεση μιας πρωτεΐνης σε ένα τμήμα του DNA, η παρουσία είτε η απουσία μίας μετάλλαξης, η μεταβολή του αριθμού των αντιγράφων μίας θέσης (locus), μια αλλαγή στην κατάσταση μεθυλίωσης του DNA, είτε οποιοδήποτε κατάσταση είτε δραστηριότητα που σχετίζεται με το DNA είτε τα μόρια του ριβονουκλεϊκού οξέος (RNA). Ως γονιδιωματική ανάγνωση, οι μικροσυστοιχίες προσδιορίζουν εάν τα γεγονότα αυτά έχουν συμβεί (Seidel 2008).

Ο σκοπός των μικροσυστοιχιών είναι να μελετούν την έκφραση πολλών γονιδίων ταυτόχρονα σε απόκριση κάποιας βιολογικής διαταραχής. Η χρησιμότητα

τους έγκειται στο να ερευνά τις συγκεντρώσεις των μορίων σε ένα πολύπλοκο μείγμα με αποτέλεσμα να αποτελούν ένα ισχυρό αναλυτικό εργαλείο για μεγάλη γκάμα πειραμάτων. Η αρχή της μεθόδου που χρησιμοποιούν οι μικροσυστοιχίες βασίζεται στην αρχή της συμπληρωματικότητας των βάσεων των DNA και RNA (δηλαδή η Αδενίνη, A, δεσμεύεται με την Θυμίνη, T, στο DNA και με την Ουρακίλη, U, στο RNA και η Κυτοσίνη, C, δεσμεύεται με την Γουανίνη, G). Αυτή η ιδιότητα της συμπληρωματικότητας είναι και αυτό που χρησιμοποιείται για τη μέτρηση της γονιδιακής έκφρασης στις μικροσυστοιχίες. Δηλαδή μονόκλωνα μόρια συμπληρωματικά μεταξύ τους μπορούν να ενωθούν και να δημιουργήσουν ένα συμπληρωματικό δίκλωνο μόριο. Αυτή είναι η έννοια της υβριδοποίησης και είναι η βάση σε πολλές βιολογικές διεργασίες μέσα σε ένα κύτταρο καθώς επίσης και η βάση σε πολλά πειράματα Μοριακής Βιολογίας. Τα μόρια μπορούν να έχουν διαφορετική προέλευση (διαφορετικές πηγές) αλλά εάν είναι συμπληρωματικά θα υβριδιστούν.

## 2.2 Κατασκευή μικροσυστοιχιών

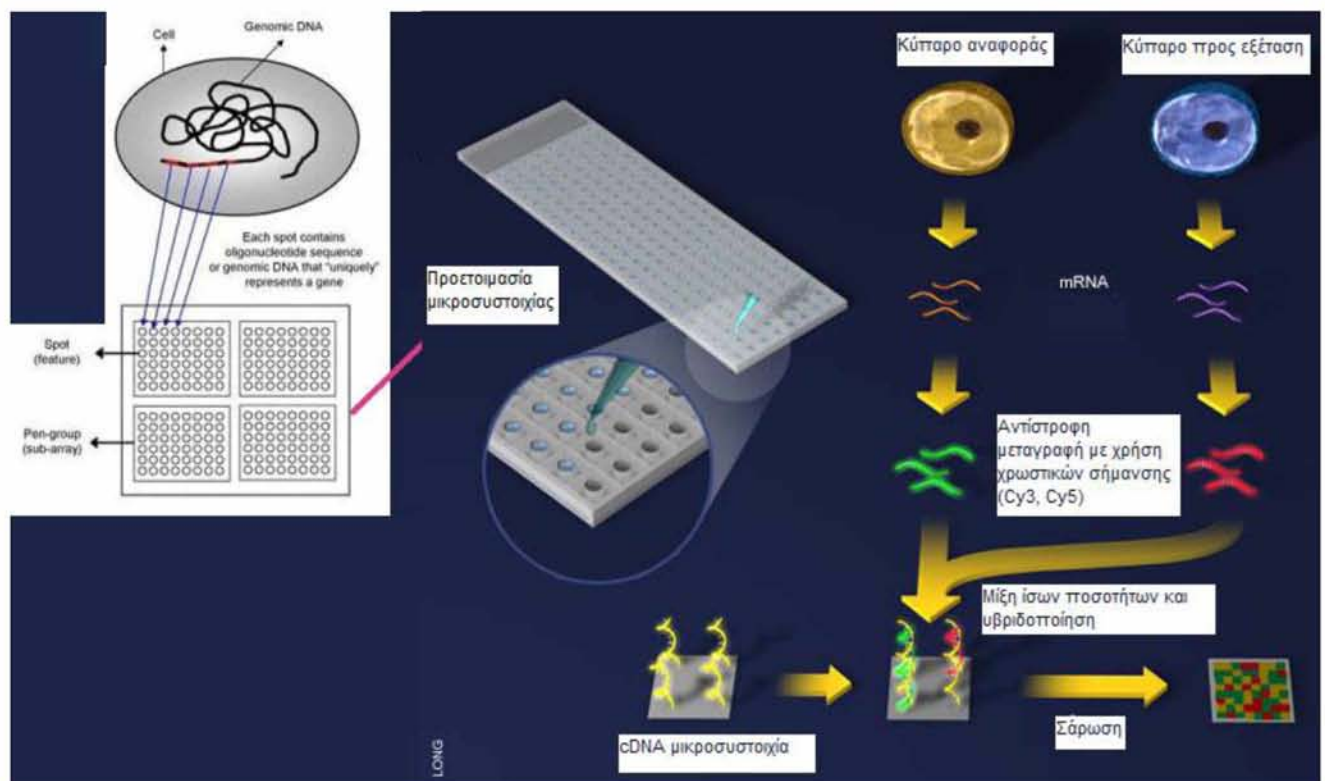
Οι μικροσυστοιχίες παρέχουν τη δυνατότητα μέτρησης πολλών γονιδίων με μια μέθοδο αντιστροφής και παραλληλισμού της ευρέως διαδεδομένης μεθόδου του στυπώματος (blot) κατά Northern. Αντί της ιχνηθέτησης αυτού που είναι γνωστό και διερεύνηση ενός πολύπλοκου μείγματος αγνώστων για να τονιστεί ένας μοναδικός στόχος, στη μέθοδο των μικροσυστοιχιών ιχνηθετούν (label) το πολύπλοκο μείγμα, το οποίο είναι με τη μορφή διαλύματος και χρησιμοποιούν μια δισδιάστατη επιφάνεια με γνωστά μόρια ή ιχνηθέτες σε διακεκριμένες θέσεις, ως ανάγνωση.

Για το λόγο αυτό οι μικροσυστοιχίες χρησιμοποιούνται πολύ, πλέον, για την ανάλυση της γονιδιακής έκφρασης, δηλαδή της ποσότητας του αγγελιοφόρου RNA (mRNA) είτε των πρωτεϊνών που έχουν παραχθεί από ένα συγκεκριμένο κύτταρο μια συγκεκριμένη στιγμή. Η πιο διαδεδομένη μεθόδός τους είναι η σύγκριση των επιπέδων έκφρασης ενός συνόλου γονιδίων συγκεκριμένων συνθηκών και ιδιοτήτων (κατάσταση A) με τα επίπεδα έκφρασης ίδιου συνόλου γονιδίων προερχόμενα από κύτταρο αναφοράς, το οποίο βρίσκεται υπό κανονικές συνθήκες (κατάσταση B).

Η δομή του τσιπ των μικροσυστοιχιών, είναι ένα γυάλινο πλακίδιο (chip), επάνω στο οποίο είναι συνδεδεμένες συγκεκριμένες αλληλουχίες ειδικά για συγκεκριμένα γονίδια, τους ανιχνευτές (probes), οι οποίοι βρίσκονται σε κάποιες σταθερές θέσεις σε μια κουκίδα (spot) του τσιπ. Μπορούν να υπάρχουν χιλιάδες



κουκίδες σε μια σειρά και η κάθε μια να σχετίζεται με ένα μοναδικό γονίδιο. Αυτές οι αλληλουχίες, κατόπιν, υποβάλλονται σε υβριδοποίηση με αντίγραφα νουκλεϊκών οξέων από βιολογικά δείγματα, τα οποία έχουν ιχνηθετηθεί με φθορίζουσες ουσίες, τους στόχους (targets) (Brazma and Vilo 2000, Lee et al. 2004). Μετά το στάδιο της υβριδοποίησης, οι κουκίδες στο υβριδοποιημένο τσιπ μικροσυστοιχιών διεγείρονται από ένα λέιζερ και σαρώνονται σε κατάλληλα μήκη κύματος για την ανίχνευση των φθορίζουσών ουσιών. Το μέγεθος του φθορισμού που εκπέμπεται κατά τη διέγερση μετατρέπεται σε μια ποσοτική τελική έξοδο η οποία αντιπροσωπεύει το σχετικό επίπεδο έκφρασης του γονιδίου αυτού, δηλαδή την ποσότητα των αντιγράφων του κάθε είδους RNA που υπάρχει στο υπό μελέτη δείγμα (Babu 2004).



**Εικόνα 2.1** Το πειραματικό πρωτόκολλο των μικροσυστοιχιών. Η προετοιμασία του πλακιδίου των μικροσυστοιχιών γίνεται με σταθεροποίηση των κατάλληλων αλληλουχιών στις κουκίδες. Η προετοιμασία του δείγματος γίνεται με απομόνωση βιολογικού υλικού (mRNA) και από του δύο τύπους των προς μελέτη κυττάρων, εν συνεχεία σημαίνονται με διαφορετικού χρώματος φθορίζουσες χρωστικές και υβριδοποιούνται στο πλακίδιο. Τέλος το τσιπ διεγείρεται με λέιζερ και σαρώνεται για να δώσει την προς επεξεργασία και μελέτη εικόνα. Αυτή η εικόνα ανακτήθηκε από την ιστοσελίδα [http://www.biorom.uma.es/contenido/av\\_bma/apuntes/T3/hibridac.html](http://www.biorom.uma.es/contenido/av_bma/apuntes/T3/hibridac.html)

## **2.3 Πειραματική διαδικασία των μικροσυστοιχιών**

Κατά την διεξαγωγή ενός πειράματος μικροσυστοιχιών πρέπει να ακολουθηθούν κάποια στάδια που θα δούμε περιγράψουμε πιο κάτω.

### ***2.3.1 Η ιδέα του πειράματος, το βιολογικό ερώτημα***

Η αρχική ιδέα του πειράματος δεν είναι άλλο από την αρχική διατύπωση του βιολογικού ερωτήματος, που ελπίζουμε να απαντηθεί στο τέλος του πειράματος, το οποίο πρέπει να διατυπώνεται προσεκτικά και με ακρίβεια πριν ξεκινήσει η διαδικασία ενός πειράματος μικροσυστοιχιών. Ανάλογα με τον τομέα που θα πραγματοποιηθεί η μελέτη διατυπώνονται τα ερωτήματα. Στον τομέα της ανακάλυψης φαρμάκων, για παράδειγμα, διατυπώνονται ερωτήματα για την απόκριση κυττάρων είτε ιστών ασθενών υπό την επίδραση συγκεκριμένων φαρμακευτικών ουσιών (Ramasamy et al. 2008). Φυσικά για την σωστή διατύπωση πρέπει να λαμβάνεται υπόψη το διαθέσιμο βιολογικό υλικό, η σταθερότητα του (για παράδειγμα το mRNA είναι πιο ασταθές από το DNA, αλλά πιο εύκολο στη συλλογή του), το μέγεθος της ποσότητάς του, διότι θα επηρεάσει τα αποτελέσματα του πειράματος.

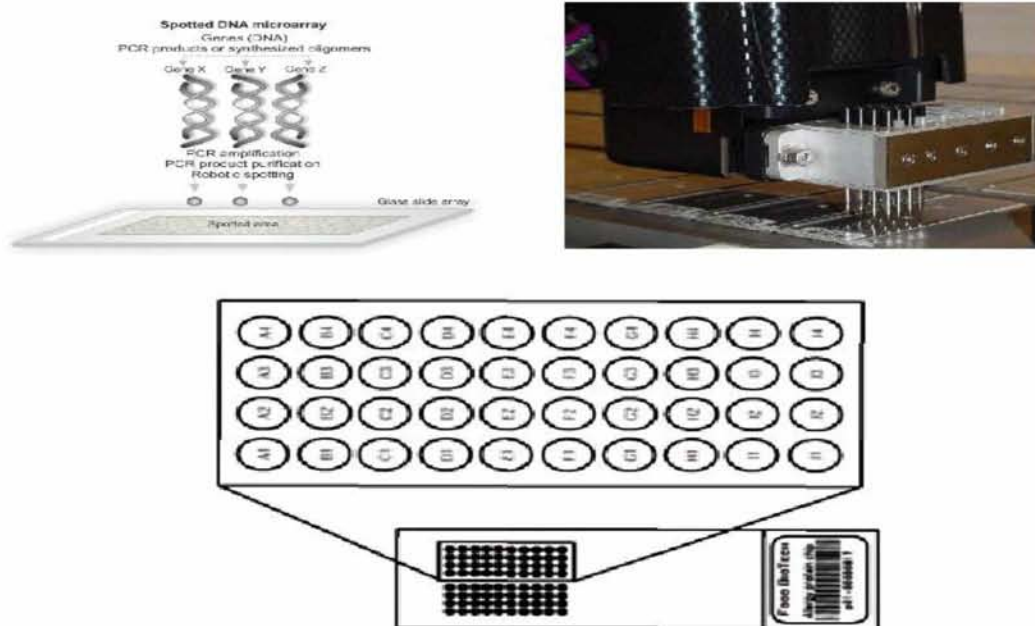
### ***2.3.2 Ο τύπος του τσιπ των μικροσυστοιχιών***

Με βάση το βιολογικό ερώτημα που έχει διατυπωθεί πρέπει να επιλεγθεί ο τύπος του τσιπ των μικροσυστοιχιών που θα χρησιμοποιηθεί στο πείραμα. Υπάρχουν κυρίως δύο είδη τεχνολογιών μικροσυστοιχιών σε ευρεία χρήση μεταξύ των περισσότερων εργαστηρίων. Στην μια κατηγορία ανήκουν οι μικροσυστοιχίες εκτύπωσης (spotted microarrays) στις οποίες οι ανιχνευτές εκτυπώνονται ρομποτικά στην επιφάνεια του τσιπ μετά την παρασκευή τους. Οι οποίες χωρίζονται σε μικροσυστοιχίες ολιγονουκλεοτιδίων και σε μικροσυστοιχίες συμπληρωματικού DNA (complementary DNA, cDNA). Στη δεύτερη κατηγορία ανήκουν οι μικροσυστοιχίες σύνθεσης επί τόπου (in situ). Η κάθε μια κατηγορία διαθέτει μοναδικά χαρακτηριστικά, τα οποία προσφέρουν πλεονεκτήματα αλλά και μειονεκτήματα σε ένα δεδομένο πλαίσιο. Παράλληλες μορφές μέτρησης DNA και RNA εξακολουθούν να μεταβάλλονται και να εξελίσσονται αλλά αυτές οι τρεις είναι οι πιο ευρέως χρησιμοποιούμενες επί του παρόντος (Seidel 2008).

### 2.3.2.1 Μικροσυστοιχίες εκτύπωσης

Οι μικροσυστοιχίες εκτύπωσης ήταν οι πρώτες ευρέως διαθέσιμες πλατφόρμες στα μέσα του '90 και εξακολουθούν να τυγχάνουν ευρεία χρήση. Το όνομα αυτού του τύπου των συστοιχίων προέρχεται από το γεγονός ότι οι ανιχνευτές, πρώτα συντίθενται και στη συνέχεια εκτυπώνονται στην επιφάνεια η οποία είναι παρόμοια με μια γυάλινη πλάκα μικροσκοπίου (Cheung et al. 1999). Οι αλληλουχίες των ανιχνευτών λαμβάνονται από δημόσιες πληροφορίες που είναι αποθηκευμένες σε βάσεις δεδομένων ακολουθιών, όπως είναι η Genbank (Benson et al. 2015) και είναι ειδικά επιλεγμένες ώστε να χαρακτηρίζεται όσο το δυνατόν πιο ειδικά το βιολογικό στοιχείο (γονίδιο, RNA, πρωτεΐνη ή οτιδήποτε άλλο) που πρόκειται να αντιπροσωπεύουν (Tomiuk and Hofmann 2001). Μόλις επιλεγούν οι αλληλουχίες για τους ανιχνευτές συντίθενται *in vitro*. Οι ανιχνευτές μπορεί να είναι είτε ολιγονουκλεοτίδια (20-70 νουκλεοτίδια) είτε cDNA προϊόν αλυσιδωτής αντίδρασης πολυμεράσης (PCR) (Stekel 2003).

Το κύριο πλεονέκτημα των μικροσυστοιχίων εκτύπωσης είναι το μεγάλο μέγεθος των ανιχνευτών που προσφέρει υψηλή ευαισθησία, βέβαια στον αντίποδα αυτό κάνει πιο εύκολο να συμβεί μη ειδική-υβριδοποίηση (cross hybridization) σε αντίθεση με τις πιο μικρές αλληλουχίες, προσδίδοντας μια σχετικά πιο χαμηλή ακρίβεια (Halperin, Buhot, and Zhulina 2004).



**Εικόνα 2.2 Παραγωγή μικροσυστοιχίων εκτύπωσης.** Οι μικροσυστοιχίες εκτύπωσης (στο κάτω μέρος της εικόνας), παράγονται από ένα ρομποτικό εκτυπωτή, (στο δεξιό άνω μέρος της εικόνας), οι ανιχνευτές προσυντιθέμενες αλληλουχίες (προϊόντα PCR, cDNA, ολιγονουκλεοτίδια) εκτυπώνονται επάνω σε επικαλυμμένες γυάλινες αντικειμενοφόρες πλάκες (στο άνω αριστερά μέρος της εικόνας). Κάθε σημείο στη συστοιχία αντιπροσωπεύει ένα συγκεκριμένο συνεχόμενο θραύσμα γονιδίου, δηλαδή 40-70 νουκλεοτίδια για συστοιχίες ολιγονουκλεοτίδιο, ή αρκετές εκατοντάδες νουκλεοτίδια για τα προϊόντα PCR.



### 2.3.2.2 Μικροσυστοιχίες σύνθεσης *in situ*

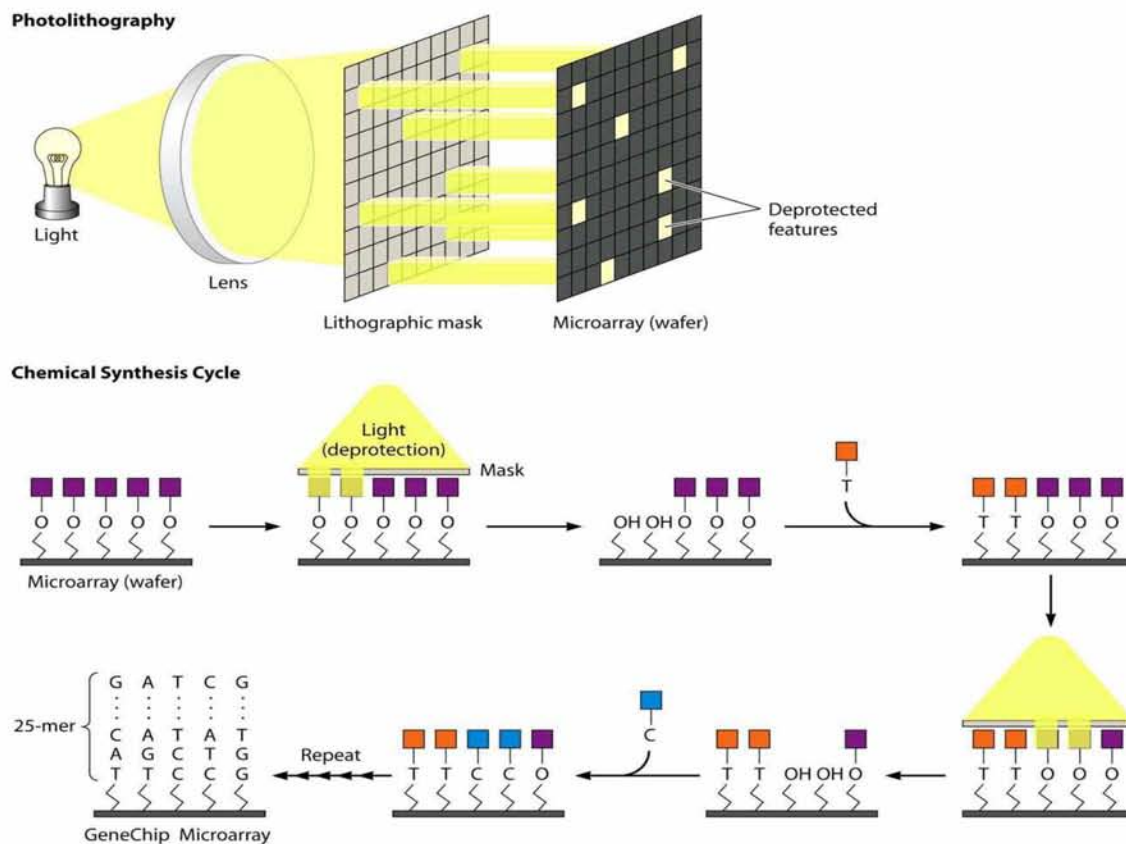
Αυτό το είδος πλατφόρμας αναπτύχθηκε και είναι η πιο εμπορική πλατφόρμα της εταιρίας Affymetrix. Το πλακίδιο αποτελείται σχετικά μικρά τμήματα 25-μερών ολιγονουκλεοτιδίων, τα οποία συντίθενται *in situ* στην γυάλινη επιφάνεια του τσιπ. Υπεριώδες φως διέρχεται μέσω μιας λιθογραφικής μάσκας που δρα ως φίλτρο είτε μεταδίδοντας είτε μπλοκάροντας το φως από την χημικά προστατευμένη επιφάνεια μικροσυστοιχίας (wafer). Η διαδοχική εφαρμογή των ειδικών λιθογραφικών масκών καθορίζει τη σειρά σύνθεσης της αλληλουχίας επί της επιφάνειας, έτσι ώστε η αλληλουχία των ολιγονουκλεοτιδίων να δημιουργηθεί κατά ένα νουκλεοτίδιο τη φορά, συζευγμένο χημικά επάνω στο προηγούμενο, σε καθορισμένη θέση επάνω στην επιφάνεια του τσιπ της μικροσυστοιχίας (Pease et al. 1994).

Αυτό συμβαίνει με το υπεριώδες φως, το οποίο απομακρύνει τις προστατευτικές ομάδες από την επιφάνεια της συστοιχίας, επιτρέποντας την προσθήκη ενός μόνο προστατευμένου νουκλεοτιδίου. Διαδοχικοί γύροι φωτοαποπροστασίας, μεταβάλλουν τα πρότυπα φιλτραρίσματος της μάσκας και προσθήκες μονών νουκλεοτιδίων δημιουργούν τη μορφή των ειδικών ανιχνευτών (25 νουκλεοτιδίων) των μικροσυστοιχιών.

Αντί να χρησιμοποιείται μία μόνο αλληλουχία για την ανίχνευση της έκφρασης του κάθε γονιδίου, όπως θα ήταν σύνηθες για μια συστοιχία εκτύπωσης, η Affymetrix χρησιμοποιεί ένα σύνολο ανιχνευτών για τη μέτρηση της έκφρασης ενός γονιδίου. Σετ ανιχνευτών περιέχουν δύο τύπους ανιχνευτών για να μετρήσουν τα γονίδια που μας ενδιαφέρουν, ανιχνευτές με τέλειο ταίριασμα (perfect match, PM) και με μη ταίριασμα (mismatch, MM). Ο σκοπός του ανιχνευτή με μη ταίριασμα είναι να λειτουργεί ως αρνητικός έλεγχος για την υβριδοποίηση. Ένα τυπικό σετ ανιχνευτών περιέχει 11 ανιχνευτές με τέλειο ταίριασμα και 11 ανιχνευτές με μη ταίριασμα.

Ακολουθώντας την Affymetrix και άλλες εταιρίες κατασκεύασαν μικροσυστοιχίες σύνθεσης *in situ*, όπως η Roche NimbleGen ([www.nimblegen.com](http://www.nimblegen.com)) και η Agilent Technologies ([www.agilent.com](http://www.agilent.com)). Με τη βασική διαφορά ότι και οι δύο εταιρίες χρησιμοποίησαν μεγαλύτερου μήκους ολιγονουκλεοτιδικών αλληλουχιών ανιχνευτές (από 60 έως 100 νουκλεοτίδια) και διαφορετικό τρόπο κατασκευής. Η NimbleGen χρησιμοποιεί φωτο-διαμεσολαβούμενη σύνθεση χωρίς τη χρήση λιθοφωτογραφικών масκών αλλά με χρήση εικονικών είτε ψηφιακών масκών (Singh-

Gasson et al. 1999), ενώ η Agilent για την σύνθεση των ανιχνευτών βασίζεται στην τεχνολογία ψεκασμού μελάνης (inkjet) (Carter et al. 2003).



**Εικόνα 2.3 Παραγωγή μικροσυστοιχιών με φωτολιθογραφία.** Περιγραφή της φωτολιθογραφικής σύνθεσης ολιγονουκλεοτιδίων σε μικροσυστοιχίες της Affymetrix. Υπεριώδες φως διέρχεται μέσω μιας λιθογραφικής μάσκας που δρα ως φίλτρο είτε μεταδίδοντας είτε μπλοκάροντας το φως από την χημικά προστατευμένη επιφάνεια μικροσυστοιχίας (wafer). Η διαδοχική εφαρμογή των ειδικών λιθογραφικών μασκών καθορίζει τη σειρά σύνθεσης της αλληλουχίας επί της επιφάνειας. (Κάτω) Ο κύκλος της χημικής σύνθεσης. Υπεριώδες φως απομακρύνει τις προστατευτικές ομάδες (τετράγωνα) από την επιφάνεια της συστοιχίας, επιτρέποντας την προσθήκη ενός μόνο προστατευμένου νουκλεοτιδίου όπως εκπλύνεται πάνω από τη μικροσυστοιχία. Διαδοχικοί γύροι φωτοαποπροστασίας, μεταβάλλουν τα πρότυπα φίλτραρίσματος της μάσκας και προσθήκες μονών νουκλεοτιδίων δημιουργούν τη μορφή των ειδικών ανιχνευτών (25 νουκλεοτιδίων) των μικροσυστοιχιών. Εικόνα από (Dalma-Weiszhausz et al. 2006)

## 2.4 Παρασκευή του δείγματος

Το πρώτο στάδιο σε ένα πείραμα μικροσυστοιχιών είναι η απομόνωση του βιολογικού υλικού, το οποίο θα χρησιμοποιηθεί ως στόχος. Κατά την ανάλυση γονιδιακού προφίλ έκφρασης χρειάζεται απομόνωση RNA από κύτταρα ή ιστούς. Εάν δεν ιχνηθετηθεί το mRNA, το οποίο αντιπροσωπεύει τη γονιδιακή έκφραση, στο

διάλυμα του ολικού RNA, μπορεί να απομονωθεί το mRNA και να ιχνηθετηθεί μεμονωμένα.

Στα θηλαστικά κύτταρα περιέχονται περίπου  $10^{-5}$   $\mu\text{g}$  RNA, 80-85% από το οποίο είναι ριβοσωμικό RNA (rRNA). Το περισσότερο από το υπόλοιπο RNA περιέχει μια ποικιλία από μεταφορικό RNA (tRNA) και μικροπυρηνικό RNA (snRNA). Τα παραπάνω RNA είναι συγκεκριμένου μεγέθους και αλληλουχίας και μπορούν να απομονωθούν επαρκώς με πολλές τεχνικές. Αντιθέτως, το mRNA, το οποίο αποτελεί το 1-5% του συνολικού RNA του κυττάρου, το μέγεθος και η αλληλουχία του ποικίλλει σύμφωνα την πρωτεΐνη που κωδικοποιεί. Το μεγαλύτερο μέρος όμως του mRNA ενός ευκαρυωτικού κυττάρου φέρει στο 3' άκρο του μια ουρά από κατάλοιπα πολυαδενυλικού οξέος (εώς και 150 Αδενίνες), η οποία επιτρέπει την απομόνωση του με χρωματογραφία συγγένειας σε ολιγοdT-cellulose. Ένα από τα βασικά προβλήματα στο χειρισμό του RNA είναι η ύπαρξη RNAσών στο χώρο, ένζυμα τα οποία βρίσκονται σε όλα τα κύτταρα και απελευθερώνονται κατά τη λύση τους. Τα ένζυμα αυτά καταστρέφουν το δεσμό μεταξύ της ριβόζης και της φωσφοδιεστερικής ομάδας και για αυτό το λόγο το RNA χρειάζεται ιδιαίτερα λεπτούς χειρισμούς.

#### **2.4.1 Ιχνηθέτηση**

Για την ιχνηθέτηση του δείγματος υπάρχουν διάφοροι τρόποι προσέγγισης (με αντίστροφη μεταγραφή, με χρήση RNA πολυμεράσης, με τη διαδικασία Eberwine κ.α.). Η πιο απλή και κλασσική μέθοδος ιχνηθέτησης είναι με αντίστροφη μεταγραφή από mRNA σε cDNA, με τη χρήση ιχνηθετιμένων με φθορίζουσες ουσίες νουκλεοτιδίων.

Στα πειράματα που χρησιμοποιούνται εκτυπωμένες μικροσυστοιχίες cDNA ή ολιγονουκλεοτιδίων χρησιμοποιούνται πάντα δύο δείγματα, ένα δείγμα ελέγχου και ένα δείγμα αναφοράς. Τα δείγματα mRNA αναφοράς και ελέγχου (για παράδειγμα ένας ιστός ο οποίος πάσχει από κάποια ασθένεια και ένας ιστός ο οποίος δεν πάσχει) μεταγράφονται σε cDNA ξεχωριστά, με τη χρήση του ενζύμου αντίστροφη μεταγραφάση και με νουκλεοτίδια με διαφορετική φθορίζουσα ουσία (Freeman, Robertson, and Vrana 2000).

Οι φθορίζουσες ουσίες που χρησιμοποιούνται περισσότερο είναι η κυανίνη3 (Cy3), η οποία προσδίδει το πράσινο και η κυανίνη5 (Cy5), η οποία προσδίδει το

κόκκινο. Η Cy3 συνήθως χρησιμοποιείται στην ιχνηθέτηση του δείγματος αναφοράς και η Cy5 συνήθως χρησιμοποιείται στην ιχνηθέτηση του δείγματος ελέγχου. Τα μόρια cDNA που δημιουργήθηκαν μεταφέρονται στα τσιπ των μικροσυστοιχιών ώστε να υβριδοποιηθούν με τους ανιχνευτές και στη συνέχεια να εκτιμηθεί η σχετική ποσότητα κάθε γονιδιακού μεταγράφου, στα δυο δείγματα, με την μέτρηση του σήματος φθορισμού των ιχνηθετών.

#### **2.4.2 Υβριδοποίηση**

Όπως αναφέρθηκε και προηγουμένως το επόμενο στάδιο του πειράματος είναι το στάδιο της υβριδοποίησης. Η υβριδοποίηση είναι η διαδικασία κατά την οποία τα μονόκλωνα μόρια των ιχνηθετημένων στόχων δημιουργούν δεσμούς υδρογόνου μεταξύ των συμπληρωματικών τους νουκλεωτιδίων, των μονόκλωνων μορίων των ανιχνευτών. Για να πραγματοποιηθεί η διαδικασία τα μόρια των στόχων, τα οποία βρίσκονται σε κατάλληλο ρυθμιστικό διάλυμα, τοποθετούνται επάνω στη διάταξη της μικροσυστοιχίας. Ακολουθώς οι ανιχνευτές μαζί με τους στόχους επωάζουν σε συγκεκριμένη θερμοκρασία για συγκεκριμένο χρονικό διάστημα.

Η θερμοκρασία υβριδοποίησης, η σύσταση της αλληλουχίας, το μήκος των στόχων, το μήκος των ανιχνευτών, η δευτεροταγής δομή, το pH, η συγκέντρωση των αλάτων και άλλα είναι παράγοντες που επηρεάζουν την αποτελεσματικότητα της υβριδοποίησης και τη συνοχή των διμερών που προκύπτουν. Πρέπει οι συνθήκες να είναι τέτοιες ώστε να ευνοούνται οι ειδικές υβριδοποιήσεις (specific hybridization) και να μην επιτρέπονται οι μη ειδικές (non specific hybridization), όπως για παράδειγμα μεταξύ των στόχων και του υποστρώματος. Ο πιο σημαντικός παράγοντας για αυτό τον έλεγχο είναι η θερμοκρασία του ρυθμιστικού διαλύματος υβριδοποίησης (Freeman, Robertson, and Vrana 2000). Μετά το πέρας της υβριδοποίησης γίνονται διαδοχικές πλύσεις των επιφανειών των μικροσυστοιχιών ώστε να αποβληθεί το διάλυμα της υβριδοποίησης και να απομακρυνθούν τυχόν μη ειδικές υβριδοποιήσεις, που ίσως να έχουν δημιουργηθεί. Με αυτό τον τρόπο το πλακίδιο είναι έτοιμο για το επόμενο στάδιο, την σάρωση από το ειδικό μηχάνημα.

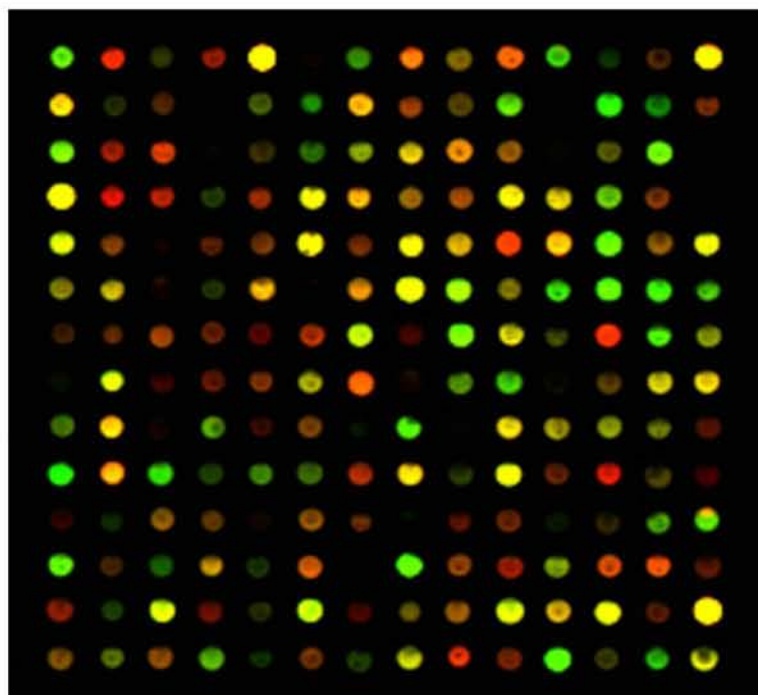
#### **2.4.3 Σάρωση**

Το στάδιο της σάρωσης είναι το στάδιο το οποίο μας δίνει τις μετρήσεις για το κάθε προς μελέτη δείγμα (αναφοράς και ελέγχου). Αυτό συμβαίνει με τη σάρωση της



επιφάνειας των μικροσυστοιχιών από ένα ειδικό μηχάνημα, τον σαρωτή (scanner), το οποίο μετατρέπει την ένταση σήματος συγκεκριμένων μηκών κύματος σε ψηφιακή εικόνα.

Ένας σαρωτής εκτελεί σάρωση των περιοχών των πλακιδίων και παράγει ένα ψηφιακό χάρτη, μια εικόνα, για κάθε χρώμα των εντάσεων φθορισμού για κάθε στοιχείο (pixel). Για ένα πείραμα με δύο δείγματα μελέτης (αναφοράς και ελέγχου) ο σαρωτής παράγει δύο εικόνες, μια για κάθε χρώμα φθορισμού (Kerr 2007). Οι χρησιμοποιούμενες φθορίζουσες ουσίες ιχνηθέτησης, απορροφούν και εν συνεχεία εκπέμπουν σε διαφορετικό φάσμα ακτινοβολίας. Οι φθορίζουσες ουσίες Cy3 και Cy5, που προαναφέραμε, εκπέμπουν στα 510-550 nm και 630-660 nm αντίστοιχα. Για το λόγο αυτό ο σαρωτής σαρώνει αρχικά το πλακίδιο με το ένα μήκος κύματος και εν συνεχεία με το άλλο δίνοντας έτσι ως αποτέλεσμα δυο εικόνες, οι οποίες με την κατάλληλη επεξεργασία μέσω προγραμμάτων συνενώνονται και παράγεται μια εικόνα. Η συνδυασμένη εικόνα επιτρέπει να αναγνωριστούν ποια γονίδια βρίσκονται σε μεγαλύτερη είτε μικρότερη έκφραση στο δείγμα ελέγχου συγκριτικά με το δείγμα αναφοράς.



Εικόνα 2.4 Συνδυασμένη εικόνα μικροσυστοιχίας. Πείραμα με χρήση δύο δειγμάτων (ελέγχου και αναφοράς) (Baird 2012).



Όπως φαίνεται στην Εικόνα 2.4 κάθε κουκίδα εμφανίζεται με ένα χρώμα, το οποίο εξαρτάται από την ποσότητα του δείγματος ελέγχου και του δείγματος αναφοράς.

High Red & Green = Yellow/White	High Red & Low Green = Red	Low Red & High Green = Green	Low Red & Low Green = Black
			

Εικόνα 2.5 Ερμηνεία των χρωμάτων που εμφανίζονται στις μικροσυστοιχίες (Baird 2012)

Πιο συγκεκριμένα όπως βλέπουμε και στην Εικόνα 2.5 οι ερμηνείες των χρωμάτων είναι:

- Όταν μια κουκίδα έχει χρώμα κίτρινο είτε λευκό, οι ποσότητες του δείγματος αναφοράς και του δείγματος ελέγχου είναι υψηλές και ίσες.
- Όταν μια κουκίδα έχει χρώμα κόκκινο, η ποσότητα του δείγματος που εκπέμπει στο κόκκινο μήκος κύματος (το δείγμα ελέγχου) είναι υψηλή και η ποσότητα του δείγματος που εκπέμπει στο πράσινο μήκος κύματος (το δείγμα αναφοράς) είναι χαμηλή.
- Όταν μια κουκίδα έχει χρώμα πράσινο, η ποσότητα του δείγματος που εκπέμπει στο κόκκινο μήκος κύματος (το δείγμα ελέγχου) είναι χαμηλή και η ποσότητα του δείγματος που εκπέμπει στο πράσινο μήκος κύματος (το δείγμα αναφοράς) είναι υψηλή.
- Τέλος όταν μια κουκίδα έχει χρώμα μαύρο, οι ποσότητες του δείγματος αναφοράς και του δείγματος ελέγχου είναι χαμηλές και ίσες είτε κανένα από τα δύο δείγματα δεν έχει υβριδοποιηθεί (Baird 2012).

#### 2.4.4 Κανονικοποίηση (normalization)

Η κανονικοποίηση των αποτελεσμάτων των μικροσυστοιχιών στοχεύει στην ελαχιστοποίηση των συστηματικών σφαλμάτων, τα οποία εντοπίζονται στα εκτιμώμενα επίπεδα έκφρασης των γονιδίων, με σκοπό να γίνονται ευκολότερα αντιληπτές οι βιολογικές διαφορές των δεδομένων και επίσης να είναι εφικτή η σύγκριση των επιπέδων έκφρασης σε δεδομένα τα οποία προέρχονται από

διαφορετικά τσιπ μικροσυστοιχιών. Μπορεί και ρυθμίζει τυχόν διαφορές στην ιχνηθέτηση, στην ανίχνευση αποδοτικότητας των φθορίζοντων ιχνηθετών και στις διαφορετικές ποσότητες των αρχικών δειγμάτων που χρησιμοποιήθηκαν στο πείραμα. Καθώς τέτοιου είδους προβλήματα μπορεί να προκαλέσουν μετατόπιση της μέσης αναλογίας των δύο φθορίζουσών ουσιών, που χρησιμοποιούνται, και να πρέπει να γίνει αναβαθμονόμηση των εντάσεων ώστε να αναλυθεί σωστά το πείραμα.

Οι κυριότερες τεχνικές οι οποίες χρησιμοποιούνται στην κανονικοποίηση δεδομένων γονιδιακής έκφρασης είναι οι εξής.

#### 2.4.4.1 Κανονικοποίηση ολική έντασης (total intensity normalization)

Η κανονικοποίηση ολικής έντασης βασίζεται στην παραδοχή ότι η ποσότητα του αρχικού mRNA είναι η ίδια και για τα δύο επισημασμένα δείγματα. Δέχεται δηλαδή ότι όλα τα μόρια είναι παρόμοιου μεγέθους και ότι σε κάθε δείγμα υπάρχει ίδιος αριθμός μορίων. Με αποτέλεσμα να δέχεται ότι αν κάποια γονίδια στο δείγμα είναι υπερεκφρασμένα, τότε υπάρχουν σε ίσο περίπου αριθμό γονίδια στο δείγμα που είναι υποεκφρασμένα. Λόγω του μεγάλου πλήθους γονιδίων στο πλακίδιο θεωρείται πως υβριδοποιείται ίσος αριθμός RNA και από τα δύο δείγματα, με αποτέλεσμα η συνολική ένταση φθορισμού για όλα τα στοιχεία στη συστοιχία να είναι ίδια και για τις δύο φθορίζουσες ουσίες. Με βάση αυτή την προσέγγιση μπορεί να υπολογιστεί ένας παράγοντας κανονικοποίησης (normalization factor) από τον τύπο:

$$N_{ολικ\acute{o}} = \frac{\sum_{g=1}^{N_{συστοιχ\acute{i}ας}} R_i}{\sum_{g=1}^{N_{συστοιχ\acute{i}ας}} G_i}$$

όπου  $R_i$  και  $G_i$  είναι οι τιμές έκφρασης του γονιδίου  $i$  του πειράματος  $n$  στο δείγμα ελέγχου και στο δείγμα αναφοράς αντίστοιχα και  $N_{συστοιχ\acute{i}ας}$  ο συνολικός αριθμός των γονιδίων που μελετώνται. Οι τιμές έκφρασης των γονιδίων για τα δύο δείγματα ορίζονται ως:

$$R'_i = R_i \quad \text{και} \quad G'_i = N_{ολικ\acute{o}} * G_i$$

και έτσι ο κανονικοποιημένος λόγος έκφρασης κάθε γονιδίου είναι:

$$T'_i = \frac{R'_i}{G'_i} = \frac{1}{N_{ολικ\acute{o}}} * \frac{R_i}{G_i}$$

και ρυθμίζει τη μέση αναλογία να είναι ίση με τη μονάδα (Quackenbush 2001).

#### 2.4.4.2 Lowess (LOcally WEighted Scatterplot Smoothing regression) κανονικοποίηση

Η Lowess κανονικοποίηση, γραμμική οπισθοδρόμηση τοπικών βαρών, προτείνεται ως μέθοδος κανονικοποίησης που μπορεί να εξαλείψει αποτελέσματα από τις λογαριθμικές αναλογίες, που οφείλονται στην ένταση σήματος (Hoheisel 2006).

Η Lowess κανονικοποίηση χρησιμοποιεί μια συνάρτηση βαρών η οποία κάνει λιγότερο σημαντική τη συμβολή των δεδομένων σημείων της μικροσυστοιχίας που απέχουν πολύ από κάθε σημείο (Quackenbush 2001). Καθώς υπάρχουν αναφορές, η οποίες δείχνουν ότι οι λογαριθμικές αναλογίες ενδέχεται να έχουν συστηματικές παρεκκλίσεις οι οποίες οφείλονται στην ένταση σήματος, πράγμα που εμφανίζεται συχνότερα σαν απόκλιση για τα σημεία που δίνουν χαμηλή ένταση σήματος (Stekel 2003).

Ο ευκολότερος τρόπος οπτικοποίησης παρεκκλίσεων που οφείλονται στην ένταση σήματος είναι η δημιουργία ενός διαγράμματος με άξονες το  $\log_2 \frac{R_i}{G_i}$  και το  $\log_{10} R_i * G_i$ , τα διαγράμματα αυτά ονομάζονται R-I (ratio-intensity) και φανερώνουν σε μεγάλο βαθμό τις παρεκκλίσεις στις λογαριθμικές αναλογίες. Η κανονικοποίηση Lowess ανιχνεύει τις αποκλίσεις στα R-I και τις διορθώνει κάνοντας τοπική γραμμική οπισθοδρόμηση σε συνάρτηση των  $\log_{10}$ (εντάσεων) εξαλείφοντας την υπολογισμένα πιο ταιριαστή μέση  $\log_2$ (αναλογία) από την πειραματικά παρατηρούμενη αναλογία για κάθε σημείο δεδομένου. Η Lowess κανονικοποίηση χρησιμοποιεί μια συνάρτηση βαρών που κάνει λιγότερο σημαντική τη συμβολή των δεδομένων σημείων της μικροσυστοιχίας που απέχουν πολύ από κάθε σημείο (Quackenbush 2001).

Αν θέσουμε  $x_i = \log_{10} R_i * G_i$  και  $y_i = \log_2 \frac{R_i}{G_i}$ , η Lowess κανονικοποίηση πρώτα υπολογίζει το  $y(x_k)$ , την εξάρτηση της  $\log_2$ (αναλογίας) από τις  $\log_{10}$ (εντάσεις) και στη συνέχεια χρησιμοποιεί αυτήν τη συνάρτηση, σημείο προς σημείο, για να διορθώσει τις τιμές της  $\log_2$ (αναλογίας) έτσι ώστε

$$\log_2(T'_i) = \log_2(T_i) - y(x_i) = \log_2(T_i) - \log_2(2^{y(x_i)})$$

Όπως και με άλλες μεθόδους κανονικοποίησης, μπορούμε να κάνουμε τον παραπάνω μετασχηματισμό και στις εντάσεις με ισοδύναμο αποτέλεσμα, όπου

$$G'_i = G_i * 2^{y(x_i)} \quad \text{και} \quad R'_i = R_i$$

## 2.5 Στατιστική Ανάλυση

Τα δεδομένα που λαμβάνονται από το τσιπ των μικροσυστοιχιών είναι ένας πίνακας της γονιδιακής έκφρασης, στον οποίο οι γραμμές αναπαριστούν τα γονίδια και οι στήλες τα δείγματα. Οι τιμές που υπάρχουν σε κάθε μια θέση, του πίνακα αυτού, αντιστοιχούν στις τιμές έκφρασης του συγκεκριμένου γονιδίου υπό τις συγκεκριμένες πειραματικές συνθήκες. Αυτές οι τιμές συνήθως είναι ο λογάριθμος με βάση το δύο του λόγου της έντασης για το ένα χρωματικό κανάλι (π.χ. κόκκινο) προς την ένταση για το δεύτερο χρωματικό κανάλι (π.χ. πράσινο).

Τα αποτελέσματα που λαμβάνονται από ένα σχετικά μικρό αριθμό δειγμάτων πειράματος μικροσυστοιχιών είναι αρκετά πολύπλοκα για να αναλυθούν και για το λόγο αυτό γίνεται χρήση στατιστικών εργαλείων, τα οποία αποσκοπούν στην εύρεση πιθανών σχέσεων μεταξύ των δεδομένων.

### 2.5.1 Αλλαγή διπλώματος (*fold-change, FC*)

Μια από τις πρώτες μεθόδους που χρησιμοποιήθηκαν στην στατιστική ανάλυση της γονιδιακής έκφρασης ήταν η μέθοδος αλλαγής διπλώματος. Συνήθως έχει ως οριακή τιμή την  $FC \pm 2$ . Η μέτρηση της αλλαγής του διπλώματος για κάθε γονίδιο περιγράφει τη διαφορά μιας συγκεκριμένης αρχικής τιμής στην τελική, αναπαριστώντας έτσι την έκφραση του γονιδίου. Με τον τρόπο αυτό εμφανίζεται εάν ένα γονίδιο υπερεκφράζεται είτε υπόεκφράζεται (Ramasamy et al. 2008).

### 2.5.2 Έλεγχος *t* (*t-test*)

Το *t-test* είναι μια στατιστική μέθοδος η οποία αξιολογεί αν οι μέσες τιμές δύο ομάδων έχουν στατιστικά σημαντικές διαφορές. Το *t-test* δύο δειγμάτων χρησιμοποιείται για την σύγκριση των μέσων τιμών δύο ανεξάρτητων πληθυσμών. Ελέγχεται έτσι εάν η διαφορά μέσων τιμών σε δύο ανεξάρτητους πληθυσμούς προέκυψε τυχαία ή όχι. Γίνεται χρήση της μηδενικής υπόθεσης η οποία θεωρεί ότι οι δύο μέσες τιμές είναι ίσες. Το στατιστικό *t* μπορεί να υπολογιστεί, υποθέτοντας ίσες διακυμάνσεις, ως εξής:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

όπου  $\mu_1 - \mu_2$  η διαφορά μέσων τιμών του πληθυσμού και  $\bar{X}_1 - \bar{X}_2$  η διαφορά μέσων τιμών του δείγματος και  $n_1, n_2$  ο πληθυσμός του δείγματος 1 και 2 αντίστοιχα

Η τιμή της ομαδοποιημένης διακύμανσης  $S_P$  δίνεται από τον τύπο

$$S_P = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

και το τυπικό σφάλμα  $SE$  (Standard error) δίνεται από τον τύπο

$$SE = \sqrt{\frac{S_P^2}{n_1} + \frac{S_P^2}{n_2}}$$

Στους στατιστικούς ελέγχους υποθέσεων ορίζεται μια υπόθεση  $H_0$ , μηδενική υπόθεση (null hypothesis), η οποία συνήθως θεωρεί ότι δεν υπάρχει σημαντική διαφορά μεταξύ των δύο ομάδων που μελετώνται, ενώ η  $H_1$ , εναλλακτική υπόθεση (alternative hypothesis), θεωρεί ότι υπάρχει διαφορά. Η τιμή  $p$  (p-value) είναι το μικρότερο επίπεδο σημαντικότητας ( $\alpha$ ) στο οποίο η μηδενική υπόθεση μπορεί να απορριφθεί. Το επίπεδο σημαντικότητας ορίζεται από τον ερευνητή αλλά συνήθως είναι είτε 5% ( $\alpha=0.05$ ) είτε 1% ( $\alpha=0.01$ ). Όσο πιο μικρή η p-value, τόσο μεγαλύτερη είναι η βεβαιότητα ότι απορρίπτεται η  $H_0$ . Η p-value κυμαίνεται ανάμεσα στο 0 και στο 1, κάθε γονίδιο εμφανίζει μια p-value που προσδιορίζει το κατά πόσο εμφανίστηκε τυχαία ως διαφορετικά εκφρασμένο, το οποίο στη συνέχεια χρησιμοποιείται για να γίνει έλεγχος σημαντικότητας του γονιδίου.

Στα πειράματα μικροσυστοιχιών υπάρχει όμως το πρόβλημα εμφάνισης πολλών εσφαλμένων αποτελεσμάτων ως σημαντικά (false positive) εξαιτίας του μεγάλου αριθμού γονιδίων που μελετώνται. Για το λόγο αυτό χρησιμοποιούνται οι μέθοδοι διόρθωσης της p-value (Fay and Gerow 2013).

#### 2.5.2.1 Μέθοδοι διόρθωσης p-value

- Διόρθωση Bonferroni

Με την μέθοδο διόρθωσης Bonferroni η p-value κάθε γονιδίου πολλαπλασιάζεται με το συνολικό αριθμό γονιδίων της μικροσυστοιχίας. Εάν η διορθωμένη τιμή  $p$  ενός γονιδίου βρίσκεται κάτω από το όριο σφάλματος τότε το γονίδιο αυτό είναι στατιστικά σημαντικό (Dudoit et al. 2002). Ο τύπος για τη διορθωμένη τιμή  $p$  ενός γονιδίου είναι

$$p_{cor(i)} = p_{(i)} * n$$

- Διόρθωση Bonferroni Step-Down (Holm)

Η μέθοδος διόρθωσης Bonferroni Step-Down (Holm) είναι παρόμοια με την μέθοδο Bonferroni, με την διαφορά ότι είναι πιο αυστηρή ως προς τη νέα p-value. Στην αρχή οι p-values των γονιδίων ταξινομούνται κατά αύξουσα σειρά. Στη συνέχεια η πρώτη p-value πολλαπλασιάζεται με το συνολικό αριθμό γονιδίων της μικροσυστοιχίας  $n$ . Οι επόμενες p-values διορθώνονται με τον ακόλουθο τύπο (Holm 1979)

$$p_{cor(i)} = p_{(i)} * (n - i)$$

- Διόρθωση Holland

Η μέθοδος διόρθωσης Holland είναι μια παραλλαγή της μεθόδου Holm. Αρχικά και σε αυτή οι p-values των γονιδίων ταξινομούνται κατά αύξουσα σειρά και στη συνέχεια η πρώτη p-value πολλαπλασιάζεται με το συνολικό αριθμό γονιδίων της μικροσυστοιχίας μείον ένα,  $n-1$ , και οι επόμενες p-values διορθώνονται με τον ακόλουθο τύπο (Holland and Copenhaver 1987)

$$p_{cor(i)} = p_{(i)} * (n - i + 1)$$

- Benjamini and Hochberg False Discovery Rate (FDR)

Η μέθοδος διόρθωσης False Discovery Rate (FDR) θεωρείται ισχυρότερη από τις παραπάνω καθώς μπορεί να αναγνωρίζει ποια γονίδια θεωρήθηκαν λανθασμένα μη στατιστικά σημαντικά και ποια θεωρήθηκαν λανθασμένα στατιστικά σημαντικά. Στην αρχή της μεθόδου οι p-values των γονιδίων ταξινομούνται κατά αύξουσα σειρά και στη συνέχεια η μεγαλύτερη p-value παραμένει ίδια και οι επόμενες p-values διορθώνονται με τον ακόλουθο τύπο (Benjamini and Hochberg 1995)

$$p_{cor(i)} = \frac{n}{n - i} * p_{(i)}$$

- Διόρθωση Sidak

Η μέθοδος διόρθωσης Sidak χρησιμοποιείται στους πολλαπλούς ελέγχους. Η διόρθωση στις p-values γίνεται με τον ακόλουθο τύπο (Šidák 1967)

$$p_{cor(i)} = 1 - (1 - p_{(i)})^{\frac{1}{n}}$$

### **2.5.3 Μέθοδοι αναδειγματοληψίας**

Οι μελέτες μικροσυστοιχίας συνήθως διαθέτουν μικρό μέγεθος δείγματος και μη κανονική κατανομή των τιμών έκφρασης των γονιδίων. Για το λόγο αυτό το t-test ακολουθείτε αρκετές φορές από τη μέθοδο Bootstrap, ώστε να παραχθούν εκ νέου και πιο αξιόπιστα το τυπικό σφάλμα και το διαστήματα εμπιστοσύνης.

Συγκεκριμένα η Bootstrap είναι μια στατιστική τεχνική η οποία βασίζεται στη μέθοδο της αναδειγματοληψίας και χρησιμοποιείται συχνά για την εξαγωγή συμπερασμάτων βασισμένων σε παραμετρικές υποθέσεις όταν αυτές είτε είναι υπό αμφισβήτηση είτε όταν τα παραμετρικά συμπεράσματα είναι αδύνατα είτε απαιτούνται πολύ περίπλοκοι τύποι για τον υπολογισμό λαθών. Στη μετα-ανάλυση πειραμάτων μικροσυστοιχιών είναι αποτελεσματική όταν η κατανομή ενός δείγματος είναι άγνωστη και το μέγεθος του δείγματος είναι μικρό. Η ακρίβεια της bootstrapping εξαρτάται από τον αριθμό των παρατηρήσεων στο αρχικό δείγμα και τον αριθμό των επαναλήψεων. Η μέθοδος ακολουθεί τρία βήματα, αναδειγματοληψία, υπολογισμό της κατανομής Bootstrap και εξαγωγή συμπερασμάτων με τη χρήση της κατανομής. Έστω ότι υπάρχει ένα σύνολο δεδομένων  $N$ . Αρχικά, κατά τον αλγόριθμο Monte Carlo δημιουργούνται χιλιάδες καινούργια δείγματα ίδιου μεγέθους με το αρχικό σύνολο δεδομένων έπειτα από τυχαία αναδειγματοληψία και επανατοποθέτηση από το αρχικό πραγματικό δείγμα. Εναλλακτικά πρέπει να υπολογιστούν  $N^N$  νέα πιθανά δείγματα. Λόγω της αναδειγματοληψίας, τα νέα δείγματα ενδέχεται να εμφανίζουν κάποιες τιμές μόνο μία φορά, περισσότερες από μία είτε και καθόλου. Στη συνέχεια υπολογίζεται για κάθε νέο δείγμα η τιμή που επιθυμείται (π.χ. η μέση τιμή) και εκτιμάται μια νέα κατανομή της τιμής, η Bootstrap κατανομή. Τέλος, χρησιμοποιώντας την κατανομή Bootstrap λαμβάνονται πληροφορίες για το σχήμα, το μέσο, το τυπικό σφάλμα και την τυπική απόκλιση της κατανομής του δείγματος (Efron and Tibshirani 1994).

### **2.6 Βάσεις δεδομένων μικροσυστοιχιών**

Με τη πάροδο των χρόνων και εξαιτίας της δημιουργίας νέων οικονομικότερων τσιπ μικροσυστοιχιών πραγματοποιούνται πολλά πειράματα ανάλυσης γονιδιακής έκφρασης, με αποτέλεσμα την ύπαρξη πολύ μεγάλου όγκου δεδομένων. Δεδομένα τα οποία χρειάζεται να αποθηκευτούν ώστε να μπορούν στην συνέχεια να αναλυθούν.



Για το λόγο αυτό δημιουργήθηκαν διάφορες βάσεις δεδομένων οι οποίες περιέχουν χιλιάδες πειράματα μικροσυστοιχιών.

Οι βάσεις δεδομένων μικροσυστοιχιών προσφέρουν την δυνατότητα καταχώρησης αποτελεσμάτων που προκύπτουν από πειράματα μικροσυστοιχιών καθώς επίσης και την ανάκτηση πειραματικών δεδομένων. Για τα δεδομένα που υπάρχουν για το κάθε πείραμα στην βάση δίνονται οι πληροφορίες σχετικά με το είδος των δεδομένων, τα γονίδια που μελετώνται, την πλατφόρμα μικροσυστοιχιών που χρησιμοποιήθηκε και διάφορες πληροφορίες σχετικά με τα δείγματα του πειράματος.

Οι πιο γνωστές και ευρέως χρησιμοποιούμενες βάσεις δεδομένων μικροσυστοιχιών είναι οι:

- Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>)
- Array Express (<http://www.ebi.ac.uk/arrayexpress/>)
- ONCOMINE (<https://www.oncomine.org/resource/login.html>)
- CIBEX (<http://cibex.nig.ac.jp/data/index.html>)



## **Κεφάλαιο 3<sup>ο</sup>:**

### **Μετα-ανάλυση**



### 3.1 Εισαγωγή

Η μετα-ανάλυση μπορεί να χαρακτηριστεί ως ένα στατιστικό εργαλείο που συνδυάζει και επεξεργάζεται αποτελέσματα ερευνών που μελετούν το ίδιο ερώτημα. Είναι ένα ισχυρό εργαλείο για την αξιολόγηση των επιδράσεων των υποψήφιων γονιδίων στους σύνθετους φαινότυπους.

Η ανάπτυξη αυτής της μεθοδολογίας ξεκινάει από την εποχή του Fisher, το 1920. Σαν όρος εμφανίστηκε στο χώρο της ψυχολογίας στα τέλη της δεκαετίας του '70, αλλά χρησιμοποιείται πλέον ευρέως στην επιδημιολογία και στις κλινικές δοκιμές. Τελευταία γίνεται δημοφιλής ως μέθοδος για την επίλυση προβλημάτων σε μελέτες γενετικού περιεχομένου.

Για το αντικείμενο που πρόκειται να ερευνηθεί θα πρέπει να γίνεται μια εκτενής αναζήτηση στη βιβλιογραφία ώστε να βρεθούν όλες οι διαθέσιμες μελέτες που να αναφέρονται στο θέμα που εξετάζεται (Normand 1999). Μόλις συγκεντρωθεί ένας αριθμός μελετών, που συνήθως πρέπει να είναι τουλάχιστον μεγαλύτερος από τρεις, καταγράφονται τα δεδομένα που θα χρησιμοποιηθούν στην μετα-ανάλυση. Η αξιοπιστία της μετα-ανάλυσης επηρεάζεται από την αξιοπιστία των συμπεριλαμβανόμενων στοιχείων, οπότε απαιτείται ιδιαίτερη προσοχή κατά την ανάκτηση και επεξεργασία τους.

Για την πραγματοποίηση της μετα-ανάλυσης πρέπει να προσδιοριστεί το μέγεθος επίδρασης. Μέγεθος επίδρασης μπορεί να είναι είτε η διαφορά των μέσων τιμών είτε ο συντελεστής συσχέτισης είτε ο σχετικός λόγος πιθανοτήτων (Odds Ratio, OR), ο οποίος δίνει την πιθανότητα ένα γεγονός να συμβεί προς την πιθανότητα να μη συμβεί.

Δύο μοντέλα μπορούν να χρησιμοποιηθούν για να συνδυάσουν τις μεμονωμένες εκτιμήσεις του μεγέθους επίδρασης:

- Το μοντέλο σταθερών επιδράσεων (Fixed effects model)
- Το μοντέλο τυχαίων επιδράσεων (Random effects model)

### 3.2 Πλεονεκτήματα μετα-ανάλυσης

Τα κυριότερα πλεονεκτήματα της μεθοδολογίας της μετα-ανάλυσης είναι τα εξής:

- Παρέχει μια αντικειμενική σύνθεση πολλών σχετικών πειραμάτων είτε μελετών για ένα συγκεκριμένο πρόβλημα έτσι ώστε να οδηγηθεί η έρευνα σε ένα μοναδικό συνολικό συμπέρασμα.

- Επιτρέπει την αξιολόγηση της μεροληπτικότητας των υπαρχουσών μελετών, με τον υπολογισμό του αριθμού των μελετών που οδήγησαν στο συμπέρασμα ύπαρξης μη σημαντικής επίδρασης (εξαιτίας αυτού του λόγου ενδέχεται να μην έχουν πάρει ευρεία δημοσιότητα) και οι οποίες είχαν πραγματοποιηθεί πριν ο ερευνητής καταλήξει στο γενικό συμπέρασμα που θα ανατρέψει την προηγούμενη πεποίθηση διαπιστώνοντας ότι υπάρχει κάποια σημαντική επίδραση.
- Επιτρέπει την αντικειμενική ερμηνεία της αξίας των μεθόδων οι οποίες είχαν χρησιμοποιηθεί σε περιορισμένες μελέτες, επιτρέποντας στον ερευνητή να δώσει σταθμικά βάρη στις στατιστικές συναρτήσεις ελέγχου των μελετών αυτών πριν συγκεντρώσει και συνθέσει τα αποτελέσματα.
- Δίνει στον ερευνητή την δυνατότητα να προσδώσει διαφορετική βαρύτητα στις υπάρχουσες μελέτες με βάση το μέγεθος του δείγματος της εκάστοτε μελέτης.
- Παρέχει δυνατότητες συνδυασμού των ελέγχων έτσι ώστε να αξιοποιηθούν τα αποτελέσματα των προηγούμενων ερευνών τα οποία χρησιμοποίησαν μόνο την στατιστική συνάρτηση  $t$  (π.χ. τύπος του Winner, Winner's formula) είτε εκείνες για τις οποίες μόνο οι  $p$ -values είχαν αναφερθεί ανεξάρτητα από την στατιστική συνάρτηση που είχε χρησιμοποιηθεί (π.χ. τύπος του Fisher, Fisher's formula).
- Ακόμη, δίνει την δυνατότητα στον ερευνητή να μετρήσει το μέγεθος της επίδρασης, δηλαδή την ποσότητα (σε τυποποιημένες μονάδες) κατά την οποία η πραγματική εναλλακτική υπόθεση (δηλαδή το αποτέλεσμα της επίδρασης) διαφέρει από την ελεγχθείσα μηδενική υπόθεση, ώστε να καθοριστεί πρακτικά η σημασία μιας διαπίστωσης η οποία είναι στατιστικά σημαντική (Πανάρετος and Ξεκαλάκη 2000)

### 3.3 Επιλογή στατιστικού μοντέλου

Η επιλογή του κατάλληλου στατιστικού μοντέλου το οποίο θα χρησιμοποιηθεί στην μετα-ανάλυση εξαρτάται από τον τύπο των δεδομένων των μελετών που χρησιμοποιούνται. Τα δεδομένα μπορεί να είναι δυαδικά ή συνεχή.

Στα συνεχή δεδομένα ως παράγοντας με τον οποίο θα γίνει η μετα-ανάλυση και η διακύμανση ( $s_i^2$ ) της κάθε μιας μελέτης, δηλαδή το μέγεθος επίδρασης  $\theta$ ,

χρησιμοποιείται η διαφορά μέσων τιμών ανάμεσα στα δείγματα ελέγχου και τα δείγματα αναφοράς (Normand 1999). Η διαφορά τους μπορεί να είναι είτε τυποποιημένη (standardized mean difference) είτε μη τυποποιημένη (un standardized mean difference).

Η τυποποιημένη διαφορά μέσων τιμών δίνεται από τον τύπο

$$d_i = \frac{\overline{x_{ti}} - \overline{x_{ci}}}{sd_i}$$

όπου  $\overline{x_{ci}}$  η μέση τιμή του δείγματος ελέγχου και  $\overline{x_{ti}}$  η μέση τιμή του δείγματος αναφοράς της μελέτης (*i*)

και η τυπική απόκλιση δίνεται από τον τύπο

$$sd_i = \sqrt{\frac{(n_{ti}-1)sd_{ti}^2 + (n_{ci}-1)sd_{ci}^2}{n_{ti} + n_{ci} - 2}}$$

Η μη τυποποιημένη διαφορά μέσων τιμών δίνεται από τον τύπο

$$d_i = \overline{x_{ti}} - \overline{x_{ci}}$$

όπου  $\overline{x_{ci}}$  η μέση τιμή του δείγματος ελέγχου και  $\overline{x_{ti}}$  η μέση τιμή του δείγματος αναφοράς της μελέτης (*i*) (Thakkestian et al. 2005).

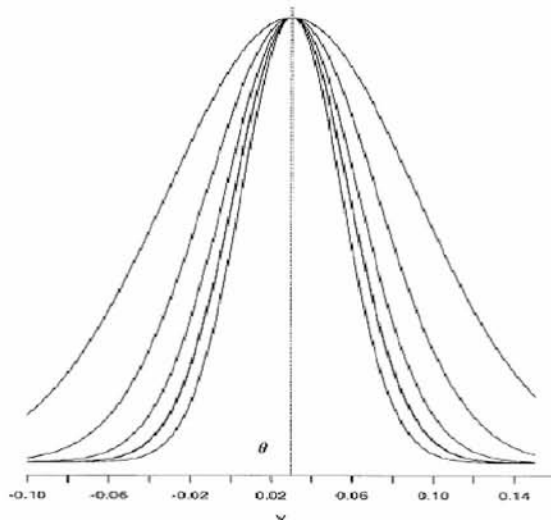
Υπάρχουν κάποιες αιτίες μεταβλητότητας οι οποίες πρέπει να λαμβάνονται υπόψη κατά τον συνδυασμό των επιμέρους μελετών (δειγματοληπτικό σφάλμα μεταξύ των μελετών, διακυμάνσεις μέσα στις ίδιες τις μελέτες και άλλα).

Τα δύο μοντέλα τα οποία χρησιμοποιούνται για να συνδυάσουν τις μεμονωμένες εκτιμήσεις του μεγέθους επίδρασης είναι:

- Μοντέλο σταθερών επιδράσεων (fixed effect model)

Θεωρείται ότι όλες οι ανεξάρτητες μελέτες ( $Y_i, i = 1, 2, 3 \dots n$ ) έχουν ένα μέσο κοινό μέγεθος επίδρασης  $\theta$  το οποίο είναι ανεξάρτητο του αριθμού των μελετών και της διακύμανσης αυτών ( $s_i^2, i = 1, 2, 3, \dots n$ ). Ο τύπος που εκφράζει αυτό το μοντέλο είναι

$$Y_i \sim N(\theta, s_i^2) \quad \text{για } i = 1, 2, 3, \dots, k$$

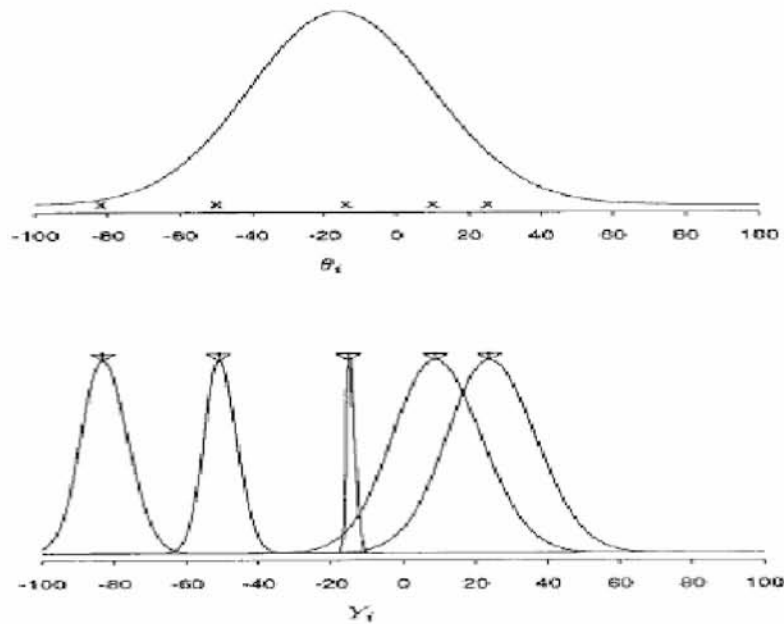


**Εικόνα 3.1 Το μοντέλο σταθερών επιδράσεων.** Η κατανομή 5 υποθετικών επιδράσεων χρησιμοποιώντας το μοντέλο. Κάθε δείγμα  $Y_i$  έχει ένα κοινό μέγεθος επίδρασης  $\theta$ . Η διαφορά ανάμεσα στις 5 μελέτες είναι η διαφορετική διακύμανση κάθε μελέτης  $s_i^2$ , δηλαδή πόσο καλά υπολογίζει κάθε μελέτη το  $\theta$  (Carlin 2000).

- Μοντέλο τυχαίων επιδράσεων (random effect model)

Σε αυτή την περίπτωση, υπάρχουν πραγματικές διαφορές ανάμεσα στα αποτελέσματα. Έτσι, η καθεμιά ανεξάρτητη μελέτη ( $Y_i$ ,  $i=1,2,3...n$ ) έχει ένα ξεχωριστό μέγεθος επίδρασης  $\theta_i$ ,  $i=1,2,3...n$  μια τιμή διακύμανσης ( $s_i^2$ ,  $i=1, 2, 3...n$ ). Το κάθε  $\theta_i$  θεωρείται ότι είναι δείγμα που προέρχεται από ένα σύνολο τιμών εκτιμητών με μέση τιμή  $\theta$  και διακύμανση  $\tau^2$ . Συνεπώς, σε αυτό το μοντέλο εισάγεται μια νέα μεταβλητή  $\tau^2$  που αντιπροσωπεύει την παρατηρούμενη ετερογένεια ανάμεσα στις ξεχωριστές μελέτες. Ο τύπος που εκφράζει αυτό το μοντέλο είναι

$$Y_i \sim N(\theta_i, s_i^2) \quad , \quad \theta_i \sim N(\theta, \tau^2)$$



**Εικόνα 3.2 Το μοντέλο τυχαίων επιδράσεων.** Κατανομή 5 υποθετικών δειγμάτων κάνοντας χρήση μοντέλου τυχαίων επιδράσεων. Κάθε μέγεθος επίδρασης  $\theta_i$  προέρχεται από τον υπερπληθυσμό με μέγεθος επίδρασης  $\theta$  και διακύμανση  $\tau^2$ . Στο παράδειγμα κάθε ένα από τα μεγέθη επίδρασης δημιούργησαν τα 5 αποτελέσματα των μοντέλων (Carlin 2000).

### 3.4 Έλεγχος ετερογένειας

Κατά την διεξαγωγή της μετα-ανάλυσης υπάρχουν κάποια σημεία που πρέπει να τα προσέχουμε. Ένα τέτοιο σημείο είναι ο έλεγχος του βαθμού ετερογένειας ανάμεσα στις μελέτες, διότι η ανομοιογένεια μπορεί να επηρεάσει τα αποτελέσματα της μετα-ανάλυσης με συνέπεια να προκύψουν λάθος συμπεράσματα. Για το λόγο αυτό η παρουσία ετερογένειας μεταξύ των μελετών προσδιορίζεται από το Q test του Cochran.

Το Q δίνεται από τον παρακάτω τύπο (Villanueva and Zavarsek 2004)

$$Q = \sum_i^k W_i (d_i - D)^2 \sim \chi_{k-1}^2$$

$$\text{με } D = \frac{\sum_{i=1}^k w_i d_i}{\sum_{i=1}^k w_i} \quad \text{και} \quad w_i = \frac{1}{\text{var}(d_i)}$$

Παρατηρείται ότι το  $Q$  ακολουθεί την κατανομή  $\chi^2$  με  $k-1$  βαθμούς ελευθερίας, όπου  $k$  ο αριθμός των μελετών και μπορεί να εφαρμοστεί για τον έλεγχο της ετερογένειας τόσο σε συνεχή όσο και σε διακριτά δεδομένα. Το  $Q$  test του Cochran αποδίδει καλύτερα όταν η μετα-ανάλυση περιλαμβάνει πολλές μελέτες.

Μια άλλη στατιστική μέθοδος που χρησιμοποιείται για τον υπολογισμό της ετερογένειας είναι το  $I^2$  test που βασίζεται στο  $Q$  test. Το  $I^2$  test υπολογίζεται από τον παρακάτω τύπο:

$$I^2 = \max \left( 0, \frac{Q - (k-1)}{Q} \right)$$

και παίρνει τιμές από 0 έως 100%. Τιμές κάτω από 25% φανερώνουν μια μικρή είτε αμελητέα ετερογένεια, ενώ τιμές που υπερβαίνουν το 50% υποδηλώνουν ότι η ετερογένεια αποτελεί σοβαρό πρόβλημα.

Ένας τρίτος εκτιμητής για την ύπαρξη ετερογένειας είναι το  $\tau^2$  το οποίο υπολογίζει τη μεταβλητότητα μεταξύ των δύο μελετών. Ο εκτιμητής  $\tau^2$  ορίστηκε αρχικά από τους DerSimonian και Laird με τον παρακάτω τύπο

$$\tau_{DL}^2 = \max \left\{ 0, \frac{Qw - (k-1)}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}} \right\}$$

Υψηλές τιμές των δεικτών ετερογένειας δείχνουν την ύπαρξη ετερογένειας ενώ όταν είναι ίσοι ή πλησιάζουν το μηδέν συμπεραίνουμε ότι δεν υπάρχει ετερογένεια μεταξύ των μελετών (DerSimonian and Laird 1986).

Ο υπολογισμός της ετερογένειας (στη μελέτη μας) γίνεται αυτόματα από στατιστικά πακέτα (στην παρούσα εργασία από το STATA) μέσω μιας εντολής η οποία υπολογίζει τους δείκτες ετερογένειας (`metan`). Για να εξεταστεί η ύπαρξη ετερογένειας θεωρούνται οι παρακάτω υποθέσεις:

- $H_0$  : Τα δείγματα είναι ομοιογενή
- $H_1$  : Τα δείγματα είναι ανομοιογενή



Χρησιμοποιώντας το μοντέλο τυχαίων επιδράσεων, η εντολή `metan` υλοποιεί τους παρακάτω τύπους για τον υπολογισμό της ετερογένειας (DerSimonian and Laird 1986)

$$\theta(\tau)_{\text{MLE}} = \frac{\sum_i W_i(\tau) Y_i}{\sum_k W_k(\tau)} \quad \text{με} \quad W_i(\tau) = \frac{1}{s_i^2 + \tau^2} \quad \text{και} \quad Y_i = \bar{x}_{1i} - \bar{x}_{2i}$$

Το αποτέλεσμα της `metan` θα δείξει αν ισχύει ή όχι η μηδενική υπόθεση ( $H_0$ ). Αν η  $H_0$  απορριφθεί τότε υπάρχει ετερογένεια.

### 3.5 Προβλήματα βιβλιογραφίας

Ένα δεύτερο σημείο που μπορεί να δημιουργήσει πρόβλημα στην διεξαγωγή της μετα-ανάλυσης είναι κατά την αναζήτηση της βιβλιογραφίας. Ένα βασικό πρόβλημα κατά την έρευνα της βιβλιογραφίας είναι η ξενόγλωσση βιβλιογραφία, διότι έρευνες οι οποίες πραγματοποιούνται σε μη αγγλόφωνες χώρες συνήθως δεν δημοσιεύονται σε περιοδικά της αγγλικής γλώσσας, τα οποία καταχωρούνται σε σημαντικές διεθνείς βιβλιογραφικές βάσεις δεδομένων. Τα σημαντικά αποτελέσματα μπορούν να δημοσιευθούν στα διεθνή περιοδικά, ενώ τα ασήμαντα αποτελέσματα εμφανίζονται στην τοπική βιβλιογραφία, με συνέπεια δεδομένα από τέτοιες μελέτες να χάνονται κατά την αναζήτηση της βιβλιογραφίας (Pan et al. 2005). Άλλο ένα πρόβλημα αποτελεί η λεγόμενη γκριζα βιβλιογραφία (*grey literature*), μελέτες οι οποίες πραγματοποιούνται σε παγκόσμια κλίμακα, δεν δημοσιεύονται σε περιοδικά και σε ιατρικές βάσεις δεδομένων στο διαδίκτυο, διότι τα αποτελέσματα τους είναι είτε μέτρια είτε αρνητικά. Έτσι, σε αυτές τις μελέτες δεν μπορούμε να έχουμε πρόσβαση, καθώς υπάρχουν μόνο σε έντυπη μορφή (Hopewell et al. 2007). Έτσι προκύπτει το συστηματικό σφάλμα δημοσίευσης (*publication bias*), από το γεγονός ότι μελέτες με «θετικά» ευρήματα δημοσιεύονται πιο γρήγορα, πιο εύκολα και σε περιοδικά μεγάλου κύρους, σχετικά με μελέτες που «απέτυχαν» να δείξουν μια συσχέτιση (αρνητικές μελέτες). Αν υπάρχει τέτοιο συστηματικό σφάλμα, κάθε αποτέλεσμα της μετά-ανάλυσης μπορεί να είναι επισφαλές (Petitti 1994). Τέλος άλλο ένα πρόβλημα που εμφανίζεται συχνά στην έρευνα της βιβλιογραφίας είναι το ονομαζόμενο «φαινόμενο του Πρωτέα» (*Proteus phenomenon*) (Ioannidis and Trikalinos 2005) και εμφανίζεται συχνά στη μοριακή επιδημιολογία. Παρατηρείται δηλαδή το φαινόμενο

πρώιμες δημοσιεύσεις που αφορούν τη συσχέτιση ενός γενετικού παράγοντα με μια ασθένεια να δείχνουν μια ισχυρή θετική είτε αρνητική συσχέτιση, πυροδοτώντας την επακόλουθη αύξηση μελετών στο ίδιο θέμα προκειμένου να επαληθευτούν και να αναπαραχθούν παρόμοια αποτελέσματα, είτε αυτή η αρχική συσχέτιση ισχύει είτε όχι. Για το λόγο αυτό καλό θα ήταν τα ευρήματα των μελετών αναφέρονται εκτενώς και επιπλέον οι αδημοσίευτες πληροφορίες να εκτιμούνται για τις διαφορές σε ανθρωπομετρικούς βιοχημικούς δείκτες συγκρινόμενους με τις δημοσιευμένες μελέτες και να χρησιμοποιούνται στις μετα-αναλύσεις.

### 3.6 Μετα-ανάλυση σε δεδομένα μικροσυστοιχιών

Έχουμε αναφέρει ότι οι μελέτες μικροσυστοιχιών χρησιμοποιούνται πλέον εκτενώς προσφέροντας αποτελέσματα της γονιδιακής έκφρασης χιλιάδων γονιδίων, τα οποία βέβαια χρειάζονται ανάλυση και διερεύνηση. Με την διεξαγωγή μετα-ανάλυσης σε δεδομένα μικροσυστοιχιών μπορεί να παραχθεί ένα τελικό συμπέρασμα στην έκφραση γονιδίων σε διάφορες καταστάσεις (διαφορική έκφραση γονιδίων σε φυσιολογικές και παθολογικές καταστάσεις).

Οι μέθοδοι που χρησιμοποιούνται για την μετα-ανάλυση δεδομένων μικροσυστοιχιών είναι κυρίως τρεις.

#### 3.6.1 Μέθοδος βασισμένη στο t-test

Μια ευρέως διαδεδομένη μέθοδος μετα-ανάλυσης δεδομένων μικροσυστοιχιών είναι η μέθοδος t-test. Σε αυτή τη μέθοδο ως μέγεθος επίδρασης χρησιμοποιείται η τυποποιημένη μέση διαφορά των τιμών έκφρασης των δειγμάτων (ελέγχου και αναφοράς) (Choi et al. 2003). Η οποία δίνεται από τον τύπο

$$d_i = \frac{\overline{x_{ti}} - \overline{x_{ci}}}{sd_i}$$

όπου  $\overline{x_{ci}}$  η μέση τιμή του δείγματος ελέγχου και  $\overline{x_{ti}}$  η μέση τιμή του δείγματος αναφοράς της μελέτης (*i*) και η τυπική απόκλιση δίνεται από τον τύπο

$$sd_i = \sqrt{\frac{(n_{ti}-1)sd_{ti}^2 + (n_{ci}-1)sd_{ci}^2}{n_{ti} + n_{ci} - 2}}$$

Λαμβάνοντας υπόψη ότι οι μελέτες μικροσυστοιχιών συνήθως διαθέτουν μικρό μέγεθος δείγματος σε κάθε ομάδα μελέτης και μη κανονική κατανομή των

τιμών έκφρασης των γονιδίων το t-test ακολουθείτε αρκετές φορές από μεθόδους αναδειγματοληψίας όπως είναι η μέθοδος Bootstrap και το Permutation test (έλεγχος μεταθέσεων), ώστε να παραχθούν εκ νέου και πιο αξιόπιστα το τυπικό σφάλμα και το διαστήματα εμπιστοσύνης.

### **3.6.2 Μέθοδος συνδυασμού των p-values**

Η μέθοδος συνδυασμού των p-values είναι μια εύχρηστη μέθοδος κατά την οποία μπορούν να συνδυαστούν οι τιμές των p-values από ανεξάρτητες μεταβλητές. Αρκεί να δίνονται οι τιμές των p-values όλων των γονιδίων που μελετώνται ανεξαρτήτως μεθόδου ανάλυσης. Η πιο συχνά χρησιμοποιούμενη μέθοδος είναι η μέθοδος αθροίσματος των λογαρίθμων των p-values. Σε αυτή τη μέθοδο ως μέγεθος επίδρασης χρησιμοποιείται το  $S_i$ , το οποίο δίνεται από τον τύπο

$$S_i = -2 \sum_{k=1}^k \log(p_{ik})$$

όπου  $S_i$ , είναι το άθροισμα των λογαρίθμων των p-values για κάθε γονίδιο  $i$  μεταξύ των μελετών  $k$  (Campain and Yang 2010).

Ωστόσο η μέθοδος συνδυασμού των p-values παρουσιάζει προβλήματα σχετικά με τον συνδυασμό των μεγεθών επίδρασης (όπως στην περίπτωση ελέγχου διαφορετικών μηδενικών υποθέσεων). Επιπλέον, με τη μέθοδο συνδυασμού της p-value, η κατεύθυνση της σύνδεσης δεν λαμβάνεται υπόψη, ως εκ τούτου όλες οι p-values πρέπει να είναι μονόπλευρες, αλλιώς τα υπέρ- και υπό- ρυθμιζόμενα γονίδια πρέπει να συνδυαστούν ξεχωριστά. Επιπροσθέτως αυτή η μέθοδος δεν μπορεί να ποσοτικοποιήσει το εύρος της συσχέτισης (δηλαδή το μέγεθος της επίδρασης, effect size) και δεν υπολογίζει την ετερογένεια μεταξύ των μελετών. Για το λόγο αυτό χρησιμοποιούνται διάφορες άλλες μέθοδοι οι οποίες συνδυάζουν, για παράδειγμα, τα ισοδύναμα z-scores αντί των p-values (Kontou, Pavlopoulou, and Bagos 2016 ).

### **3.6.3 Μέθοδος υπολογισμού του γινομένου των βαθμών κατάταξης (rank product)**

Η μέθοδος υπολογισμού του γινομένου των βαθμών κατάταξης (rank product) χρησιμοποιείται κυρίως για την μετα-ανάλυση δεδομένων γονιδιακής έκφρασης. Στην μέθοδο αυτή υπολογίζεται η αλλαγή πτύχωσης (fold change) για κάθε γονίδιο, το οποίο ισούται με το λόγος έκφρασης των γονιδίων του δείγματος ελέγχου ως προς το δείγμα αναφοράς και στη συνέχεια κατατάσσονται κατά φθίνουσα σειρά, ανάλογα με

την τιμή του λόγου των εκφράσεων, δίνοντας σε κάθε γονίδιο έναν αριθμό (βαθμό κατάταξης) ο οποίος αντιπροσωπεύει την κατάταξη του γονιδίου στην συνολική λίστα. Η διαδικασία αυτή γίνεται σε όλα τα γονίδια όλων και όλες τις μελέτες που χρησιμοποιούνται στην μετα-ανάλυση. Σε αυτή τη μέθοδο ως μέγεθος επίδρασης χρησιμοποιείται το γινόμενο των βαθμών κατάταξης, το οποίο προκύπτει από τον πολλαπλασιασμό των βαθμών κατάταξης του κάθε γονιδίου για όλες τις μελέτες, το οποίο δίνεται από τον τύπο

$$RP_g = \left( \prod_{i=1}^k r_{gik} \right)^{\frac{1}{k}}$$

όπου  $r_{gik}$  είναι η  $i$ -οστή θέση του γονιδίου  $g$  στη λίστα των γονιδίων και  $k$  το σύνολο των επαναλήψεων του γονιδίου στις μελέτες (Breitling et al. 2004).

Σε μια άλλη παραλλαγή της μεθόδου γίνεται χρήση του μέσου όρου των βαθμών κατάταξης των γονιδίων μεταξύ των μελετών ως μέγεθος επίδρασης, το οποίο δίνεται από τον τύπο

$$R^* = \frac{\sum_i^k R_i}{k}$$

(Zintzaras and Ioannidis 2008)

## **Κεφάλαιο 4°:**

### **Μεθοδολογία**



## 4.1 Σκοπός

Σκοπός της συγκεκριμένης εργασίας ήταν η μελέτη της γονιδιακής έκφρασης της προεκλαμψίας, με χρήση της διαδικασίας της μετα-ανάλυσης σε αποτελέσματα πειραμάτων μικροσυστοιχιών DNA. Όστε να μπορέσει να ερευνηθεί καλύτερα η συγκεκριμένη ασθένεια, για την οποία ακόμη δεν γνωρίζουν του γενετικούς παράγοντες που την προκαλούν είτε την επηρεάζουν.

## 4.2 Συλλογή Δεδομένων

Για την συλλογή των κατάλληλων για την προκείμενη έρευνα πραγματοποιήθηκε αναζήτηση βιβλιογραφίας στην διεθνή βάση μικροσυστοιχιών GEO. Η αναζήτηση έγινε χρησιμοποιώντας την λέξη-κλειδί ‘preeclampsia’ και ‘homo sapiens’, γιατί ζητούσαμε αποτελέσματα από πειράματα στο ανθρώπινο είδος. Επιλέχθηκαν μελέτες ασθενών-μαρτύρων ενώ αντίθετα όσες μελέτες ερευνούσαν διαφορετικό είδος (ζώα) και επεξεργάζοταν δείγματα microRNA απερρίφθησαν. Στη συνέχεια στις μελέτες που επιλέχθηκαν καταγράφηκαν οι βασικές τους πληροφορίες σε λογιστικά φύλλα Excel.

## 4.3 Καταγραφή δεδομένων

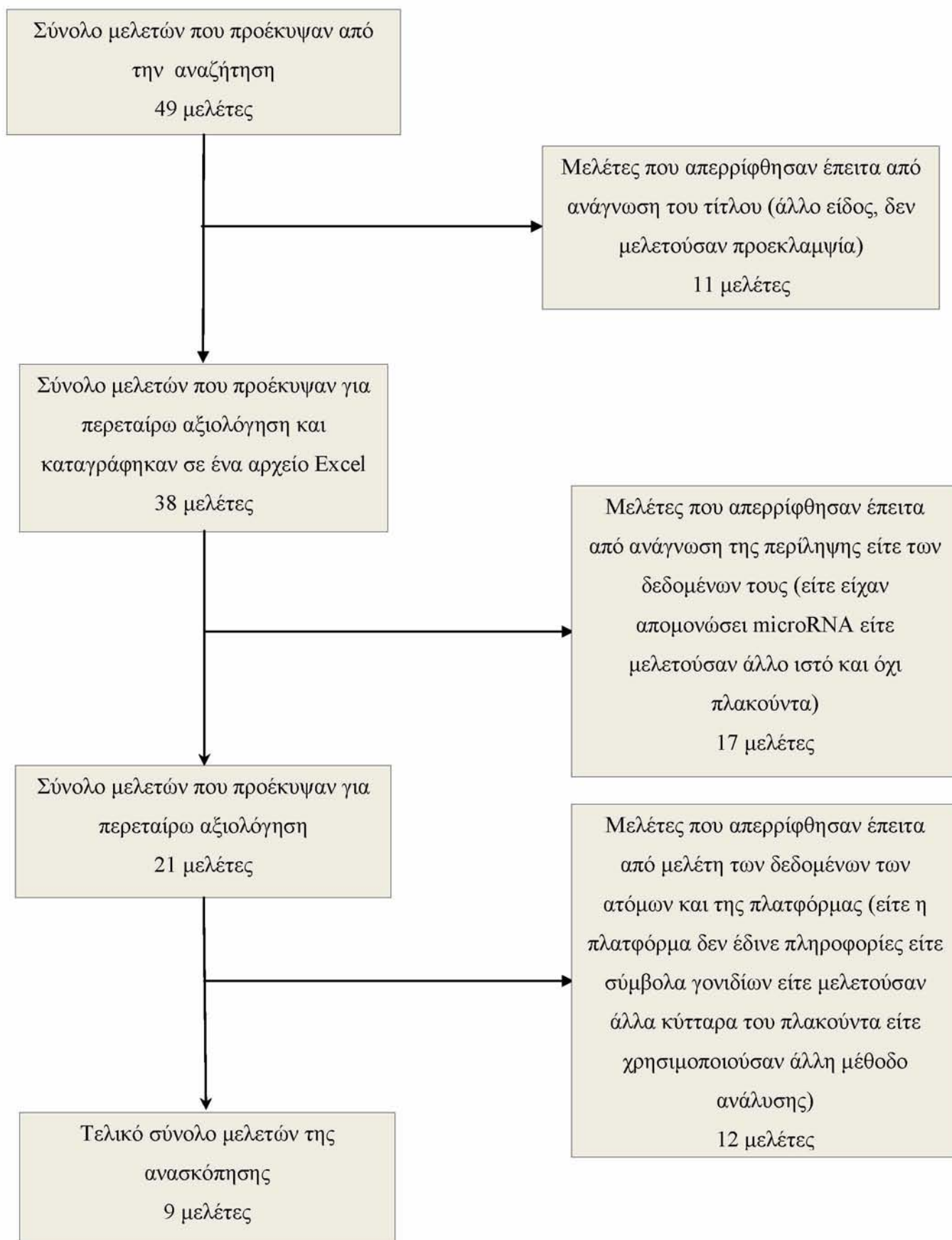
Στα λογιστικά φύλλα Excel καταχωρήθηκε ο κωδικός της μελέτης, το όνομα της εργασίας, το δείγμα ιστού που χρησιμοποιήθηκε, η μέθοδος ανάλυσης, PMID (εάν υπήρχε) και το μέγεθος του δείγματος των 49 αρχικών μελετών. Από εκεί επιλέχθηκαν οι μελέτες που χρησιμοποιούσαν το ίδιο δείγμα ιστού και την ίδια μέθοδο ανάλυσης. Στη συνέχεια με βάση τις πληροφορίες και τα δεδομένα που παρείχε η εκάστοτε μελέτη καταλήξαμε σε 9 μελέτες (όπως φαίνεται και στην Εικόνα 4.1 ).

Από τις μελέτες αυτές οι πληροφορίες που χρησιμοποιήθηκαν ήταν ο πίνακας γονιδιακής έκφρασης (series matrix file) και οι πληροφορίες της εκάστοτε πλατφόρμας μικροσυστοιχιών (GPL) που χρησιμοποιούσε η κάθε μελέτη. Ο πίνακας γονιδιακής έκφρασης παρείχε δεδομένα για την έκφραση των γονιδίων και γονιδιακών τόπων του κάθε ενός από τα δύο δείγματα, έπειτα από κανονικοποίηση είτε/και λογαρίθμηση. Η αντιστοίχιση του ονόματος των γονιδίων με τους

αντίστοιχους ανιχνευτές γίνεται με τη χρήση του αρχείου της χρησιμοποιούμενης πλατφόρμας.

Για την ανάλυση των δεδομένων χρησιμοποιήθηκε ο πίνακας γονιδιακής έκφρασης, στον οποίο έπειτα από επεξεργασία οι γραμμές αναπαριστούσαν τα σύμβολα των ονομάτων των γονιδίων (Gene Symbol) και οι στήλες τις τιμές έκφρασης των γονιδίων για τους ασθενείς και υγιείς (Case και Control). Επειδή πολλοί ανιχνευτές υβριδοποιούνται από πολλαπλά μετάγραφα του ίδιου γονιδίου, εμφανίζονταν εγγραφές για το ίδιο γονίδιο με διαφορετικές τιμές έκφρασης. Για αυτά τα γονίδια δημιουργούταν μια νέα γραμμή με τον μέσο όρο εμφάνισής τους με τη χρήση του στατιστικού προγράμματος STATA και της εντολής `collapse`.





**Εικόνα 4.1.** Απεικόνιση των βημάτων της αξιολόγησης και επιλογής των μελετών οι οποίες προέκυψαν από την βάση μικροσυστοιχιών GEO

## 4.4 Η διαδικασία της μετα-ανάλυσης με τη χρήση του στατιστικού προγράμματος STATA

Αρχικά εισάχθηκε στο πρόγραμμα STATA ο πίνακας γονιδιακής έκφρασης της κάθε μελέτης. Επιλέχθηκε το στατιστικό μοντέλο που θα χρησιμοποιούνταν στην μετα-ανάλυση, συγκεκριμένα πραγματοποιήθηκε το μοντέλο των τυχαίων επιδράσεων καθώς δίνεται διαφορετική βαρύτητα στο μέτρο επίδρασης από την εκάστοτε μελέτη και δεν υπάρχει η ίδια επίδρασης στον πληθυσμό. Επιπρόσθετα οι μεταβλητές από την κάθε μελέτη ήταν συνεχείς οπότε χρησιμοποιήθηκε ως μέσο επίδρασης η τυποποιημένη διαφορά μέσων τιμών μέσω του t-test. Ο τύπος που χρησιμοποιείται είναι

$$t = \frac{(\bar{x}_1 - \bar{x}_0) - (\mu_1 - \mu_0)}{sd_{x_1-x_0}}$$

όπου  $\bar{x}_1 - \bar{x}_0$  είναι η διαφορά μέσων τιμών του δείγματος,  $\mu_1 - \mu_0$  η διαφορά μέσων τιμών του πληθυσμού και  $sd_{x_1-x_0}$  το τυπικό σφάλμα που προκύπτει από τον τύπο

$$sd_{x_1-x_0} = \sqrt{\frac{sd_{pooled}^2}{n_1} + \frac{sd_{pooled}^2}{n_0}}$$

και το δίνεται από τον τύπο (Choi et al. 2003)

$$sd_{pooled} = \sqrt{\frac{(n_1-1)sd_1^2 + (n_0-1)sd_0^2}{n_1 + n_0 - 2}}$$

Στη συνέχεια πραγματοποιήθηκε η μέθοδος bootstrap, σε κάθε μια μελέτη ξεχωριστά για την ταυτοποίηση των διαφορετικά εκφρασμένων γονιδίων, διότι στις μελέτες μικροσυστοιχίας υπάρχει μικρό μέγεθος δείγματος συνήθως και μη κανονική κατανομή των τιμών έκφρασης. Στα δεδομένα ακολούθως υπολογίστηκαν νέες τιμές του τυπικού σφάλματος (standard error), νέες τιμές του p-value και του z test για κάθε γονίδιο. Στην συνέχεια συνενώθηκαν σε ένα αρχείο τα νέα δεδομένα των επιμέρους μελετών (νέες τιμές τυπικού σφάλματος κ.λ.π.) και πραγματοποιήθηκε μετα-ανάλυση. Έπειτα εφαρμόστηκαν διαδοχικά οι μέθοδοι διόρθωσης για τις p-values FDR, Holm, Holland, Bonferroni και Sidak με σκοπό να εντοπιστούν τα στατιστικά σημαντικά γονίδια. Με κριτήριο την p-value θεωρήθηκαν στατιστικά σημαντικά μόνο τα γονίδια που είχαν τιμή  $p < 0.01$ . Η χρήση της τιμής 0.01 γίνεται διότι το πλήθος των γονιδίων είναι πολύ μεγάλο και υπάρχει μεγαλύτερη πιθανότητα να θεωρηθεί κάτι λανθασμένα σημαντικό.

## **Κεφάλαιο 5<sup>ο</sup>:**

### **Αποτελέσματα - Συζήτηση**



## 5.1 Αποτελέσματα

Όπως έχει ήδη αναφερθεί υλοποιήθηκε η μέθοδος της μετα-ανάλυσης χρησιμοποιώντας το μοντέλο τυχαίων επιδράσεων, διότι υπήρχε ετερογένεια, και ως μέθοδος επίδρασης επιλέχθηκε η τυποποιημένη διαφορά των μέσων τιμών. Για την εκτίμηση της κατανομής της μέσης τιμής του δείγματος χρησιμοποιήθηκε το t-test. Επιπρόσθετα έγινε διόρθωση μέσω της εμπειρικής κατανομής της μεθόδου bootstrap, καθώς το δείγμα ήταν μικρό και η κατανομή ήταν μη κανονική (κάτι που φαίνεται στον Πίνακα 5.1 των μελετών οι οποίες χρησιμοποιήθηκαν στην μετα-ανάλυση). Έτσι υπολογίστηκαν με μεγαλύτερη ακρίβεια το τυπικό σφάλμα και το διάστημα εμπιστοσύνης και ακολούθησε μετα-ανάλυση με την κανονική κατανομή του effect size. Από τη μετα-ανάλυση δημιουργήθηκε ένα αρχείο το οποίο για κάθε γονίδιο έδινε τις τιμές του τυπικού σφάλματος (standard error), p-value και του z test. Τα γονίδια με  $p\text{-value} < 0.01$  θεωρήθηκαν στατιστικώς σημαντικά, όμως ο μεγάλος αυτός αριθμός γονιδίων φανερώνει την ύπαρξη false positive αποτελεσμάτων. Σε αυτές τις p-values εφαρμόστηκαν οι μέθοδοι διόρθωσης FDR, Bonferroni, Sidak, Holland και Holm. Με τη χρήση αυτών των μεθόδων διόρθωσης περιορίστηκαν περισσότερο τα στατιστικώς σημαντικά γονίδια, ο αριθμός των οποίων υπάρχει συγκεντρωτικά στον Πίνακα 5.2.

**Πίνακας 5.1** Οι μελέτες που χρησιμοποιήθηκαν στην μετα-ανάλυση, το σύνολο των ατόμων (υγιών και ασθενών) και των γονιδίων της κάθε μελέτης

GSE	Platform	Cases	Controls	Αριθμός γονιδίων	Αριθμός probes
GSE74341	Agilent-039494 SurePrint G3 Human GE v2 8x60K Microarray 039381	15	10	32080	54295
GSE54618	Illumina HumanHT-12 V4.0 expression beadchip	5	12	31331	43961
GSE30186	Illumina HumanHT-12 V4.0 expression beadchip	6	6	31424	44053
GSE25906	Illumina human-6 v2.0 expression beadchip	23	37	24612	30549
GSE24129	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript	8	8	19067	20808

	(gene) version]				
GSE14722 A	[HG-U133A] Affymetrix Human Genome U133A Array	12	11	12710	20857
GSE14722 B	[HG-U133B] Affymetrix Human Genome U133B Array	12	11	9785	15213
GSE10588	ABI Human Genome Survey Microarray Version 2	17	26	16467	18439
GSE4707	Agilent-012391 Whole Human Genome Oligo Microarray G4112A (Feature Number version)	10	4	18841	29085

**Πίνακας 5.2.** Ο αριθμός των στατιστικά σημαντικών γονιδίων με  $p\text{-value} < 0.01$  και των στατιστικά σημαντικών γονιδίων από την εφαρμογή της αντίστοιχης μεθόδου διόρθωσης.

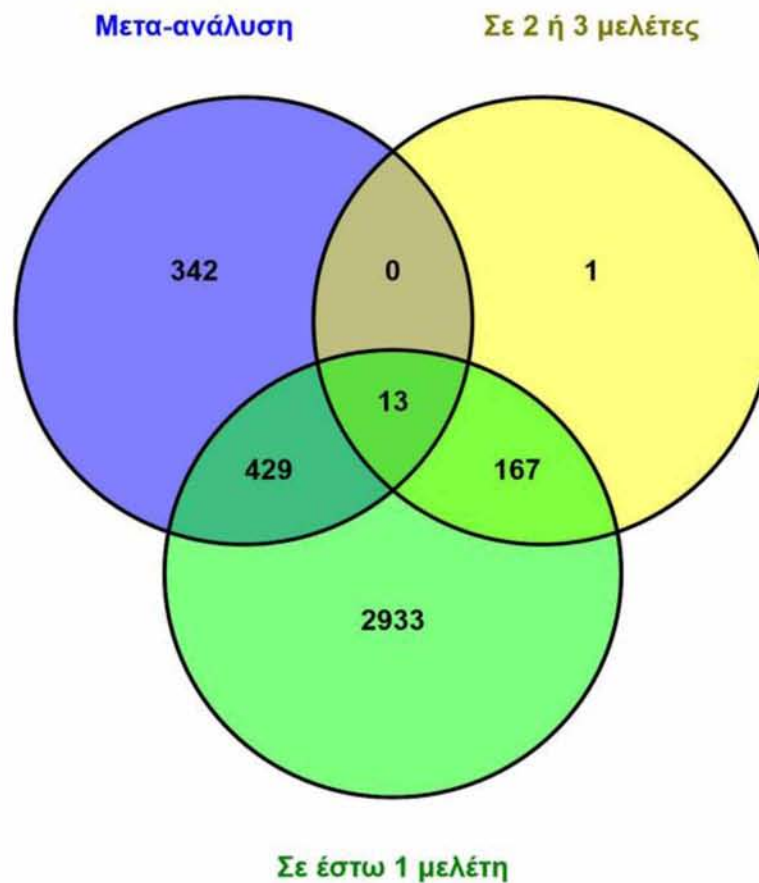
	$p\text{-value} < 0.01$	FDR (0.01)	Bonferroni	Sidak	Holm	Holland
Αριθμός στατιστικά σημαντικών γονιδίων	3573	783	196	196	196	196

Σε κάθε μελέτη ξεχωριστά χρησιμοποιήσαμε τη μέθοδο bootstrap και βρέθηκε ο αριθμός των στατιστικά σημαντικών γονιδίων της κάθε μελέτης, επιλέχθηκαν γονίδια με  $p\text{-value}$  μικρότερο του 0.01 και γονίδια τα οποία σε επίπεδο σημαντικότητας 1% κρίθηκαν σημαντικά από τη μέθοδο διόρθωσης FDR. Αυτός ο αριθμός γονιδίων συγκρίθηκε με τον αντίστοιχο αριθμό γονιδίων της ανάλυσης κάθε μελέτης ξεχωριστά. Όπως φαίνεται στον Πίνακα 5.3 η μέθοδος της μετα-ανάλυσης εντοπίζει κυρίως περισσότερα σημαντικά αποτελέσματα απ' ότι η κάθε μία μελέτη ξεχωριστά. Κάποιες μελέτες εμφανίζουν ελάχιστα είτε και μηδενικά στατιστικώς σημαντικά γονίδια και μια-δύο μελέτες, εμφανίζουν μεγαλύτερο αριθμό στατιστικώς σημαντικών γονιδίων.

**Πίνακας 5.3** Ο αριθμός των στατιστικά σημαντικών γονιδίων με p-value <0.01 και της μεθόδου διόρθωσης FDR που προκύπτουν από την μετα-ανάλυση όλων των μελετών συγκεντρωτικά και της κάθε μελέτης ξεχωριστά.

	Συγκεντρωτικά γονίδια όλων των μελετών στα οποία έγινε η μετα-ανάλυση	GSE74341	GSE25906	GSE10588	GSE4707	GSE24129	GSE54618	GSE30186	GSE14722 GPL96 U133A	GSE14722 GPL97 U133B
Αριθμός σημαντικών γονιδίων σε $p < 0.01$	3573	2265	1775	3090	880	205	71	465	192	106
Αριθμός σημαντικών γονιδίων κατά FDR	783	1832	352	1543	4	0	0	0	0	0

Τα γονίδια τα οποία θεωρούνται σημαντικά από τη μέθοδο διόρθωσης FDR σε επίπεδο σημαντικότητας 1% είναι αρκετά διαφορετικά σε κάθε μελέτη από αυτά που θεωρούνται σημαντικά από τη μέθοδο της μετα-ανάλυσης. Όπως φαίνεται από το διάγραμμα Venn που ακολουθεί.



**Εικόνα 5.1** Διάγραμμα Venn μεταξύ των στατιστικά σημαντικών γονιδίων της μετα-ανάλυσης και των μελετών GSE74341, GSE25906, GSE10588 και GSE4707 (οι οποίες διαθέτουν στατιστικώς σημαντικά γονίδια από τη μέθοδο διόρθωσης FDR σε επίπεδο σημαντικότητας 1%)

Είναι εμφανές ότι χωρίς την μετα-ανάλυση δεν θα θεωρούνταν στατιστικώς σημαντικά πολλά γονίδια τα οποία είναι, και γενικά είναι εμφανές ότι κάθε μελέτη θα θεωρούσε διαφορετικά γονίδια στατιστικώς σημαντικά είτε, συγκεκριμένα, σε κάποιες μελέτες δεν θα υπήρχαν καθόλου στατιστικώς σημαντικά γονίδια από τη μέθοδο διόρθωσης FDR σε επίπεδο σημαντικότητας 1% (όπως φαίνεται και στον Πίνακα 5.3).



**Πίνακας 5.4** Τα 50 στατιστικώς σημαντικά γονίδια από τη μέθοδο διόρθωσης FDR σε επίπεδο σημαντικότητας 1% που προέκυψαν από την μετα-ανάλυση

Σύμβολο Γονιδίου	p-value
PAPPA2	3.40E-16
LHB	2.40E-15
PPGB	2.00E-14
FGIF	5.40E-14
FKBP1C	5.40E-14
HKE2	9.20E-14
LRRC8	1.20E-13
ZNF651	1.40E-12
SLCO2A1	3.00E-12
CKLF6	4.50E-12
GBA	1.90E-11
MTX1	2.10E-11
CLDN1	8.50E-11
PTE1	1.90E-10
TMEM88B	2.20E-10
KIAA1764	2.60E-10
MXI1	3.10E-10
C6orf109	3.90E-10
ZCCHC2	4.50E-10
FBXL16	5.20E-10
IGSF8	5.60E-10
PPP1R35	9.70E-10
AKIP	1.50E-09
C13orf9	1.70E-09
SASH1	1.70E-09

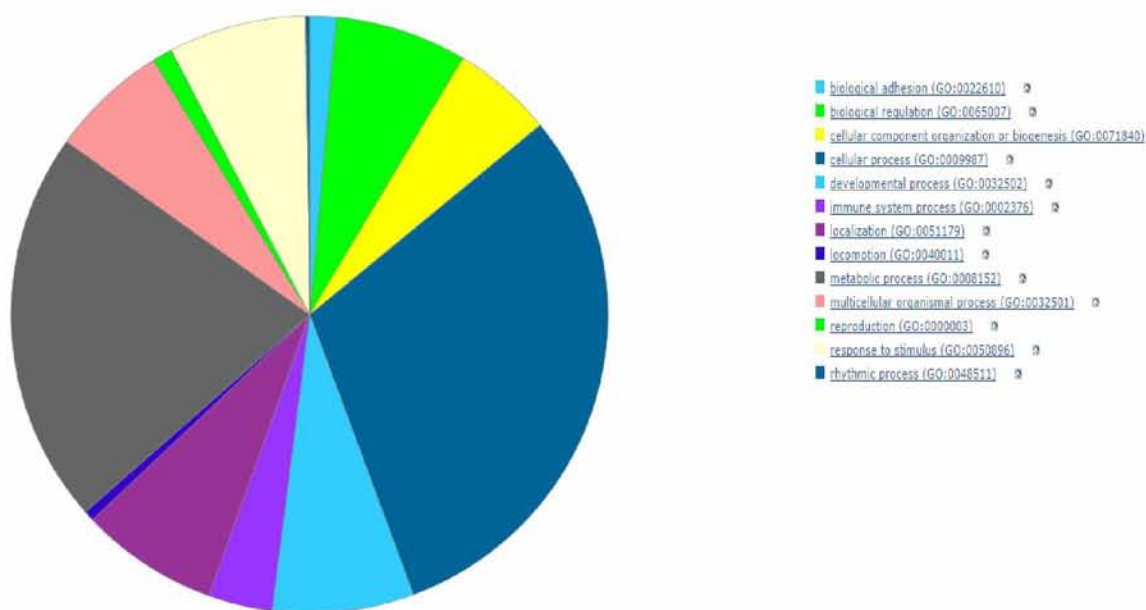
Σύμβολο Γονιδίου	p-value
TBPIP	1.70E-09
MAN1C1	2.00E-09
HK2	2.70E-09
C13orf6	2.90E-09
NEURL	4.10E-09
NYD-SP18	5.00E-09
CL640	5.10E-09
NBS1	6.30E-09
MIF	6.40E-09
SPAG4	7.10E-09
ZNF262	7.10E-09
GALNT2	7.60E-09
KARCA1	9.50E-09
SOX13	1.00E-08
TRIM58	1.20E-08
ARNT2	1.70E-08
PIK3AP1	2.10E-08
RNP	2.30E-08
SFXN3	2.30E-08
OLFML3	2.60E-08
GPR8	2.90E-08
GPR92	2.90E-08
CGB	4.50E-08
BTNL9	4.80E-08
MME	7.30E-08

Όπως είδαμε και στον Πίνακα 5.3 τα στατιστικώς σημαντικά γονίδια που προέκυψαν από την μετα-ανάλυση για την εμφάνιση προεκλαμψίας ήταν 783. Για να ερευνήσουμε τις μεταξύ τους αλληλεπιδράσεις (βιολογικά μονοπάτια, πρωτεϊνική αλληλεπίδραση κ.α.) χρησιμοποιήσαμε δύο πλατφόρμες, η οποίες χρησιμοποιούνται για ανάλυση πειραματικών δεδομένων την PANTHER (<http://www.pantherdb.org/>) και την STRING (<http://string-db.org/>). Και οι δύο είναι σχεδιασμένες έτσι ώστε να λαμβάνουν ως είσοδο μια λίστα γονιδίων/γενετικών τόπων και να εμφανίζουν ένα σύνολο βιολογικών πληροφοριών. Έχουν ενσωματωμένες πληροφορίες από πολλές βάσεις βιολογικών δεδομένων και δίνουν ένα πλήθος πληροφοριών όπως αλληλεπιδράσεις πρωτεϊνών με πρωτεΐνες (protein-protein interaction), της μοριακής λειτουργίας, βιολογικών μονοπατιών, τα δίκτυα των γονιδίων της λίστα και άλλες. Δυστυχώς ένα μεγάλο μέρος των γονιδίων που βρέθηκαν ως στατιστικώς σημαντικά είτε ήταν γενετικοί τόποι άγνωστοι ακόμη (είτε ως προς τις λειτουργίες που ενδέχεται να λαμβάνουν μέρος είτε ως προς την σχέση τους με άλλα γονίδια, είτε της δράσεις και αλληλεπιδράσεις των πρωτεϊνών που παράγουν) είτε οι βάσεις δεδομένων δεν διέθεταν γνωστές πληροφορίες για αυτά. Από τα 783 γονίδια/γενετικοί τόποι που εισάγαμε η PANTHER αναγνώρισε και μας επέστρεψε πληροφορίες για 407 και η STRING 466 αντίστοιχα.

Με την βοήθεια της PANTHER πήραμε τα εξής αποτελέσματα:

Ως προς τη βιολογική διεργασία των 407 γονιδίων βλέπουμε ότι 166 γονίδια (40,8%) συμμετέχουν σε κυτταρικές διεργασίες (όπως κυτταρική επικοινωνία, στον κυτταρικό κύκλο, στην κυκλοφορία κυτταρικών συστατικών κ.α.). 117 γονίδια (28,7%) συμμετέχουν σε μεταβολικές διεργασίες (όπως διεργασίες της κύριας μεταβολικής οδού, μεταβολισμού αζωτούχων ενώσεων, φωσφορικών ενώσεων κ.α.). 42 γονίδια (10,3%) συμμετέχουν σε αναπτυξιακές διεργασίες (όπως ανάπτυξη του συστήματος, κυτταρικό θάνατο, κυτταρική διαφοροποίηση, εμβρυακή ανάπτυξη κ.α.). 41 γονίδια (10,1%) συμμετέχουν στον εντοπισμό (localization) (όπως μεταφορά, πρωτεϊνικός εντοπισμός και εντοπισμός των RNA). 41 γονίδια (10,1%) συμμετέχουν απόκριση σε ερεθίσματα (όπως απόκριση στο στρες, σε εξωτερικά και εσωτερικά ερεθίσματα, ανοσοβιολογική απόκριση κ.α.). 39 γονίδια (9,6%) συμμετέχουν σε βιολογικές ρυθμίσεις (όπως ρυθμίσεις βιολογικών διεργασιών, της κυτταρικής λειτουργίας και της ομοιόστασης). 35 γονίδια (8,6%) συμμετέχουν σε διεργασίες πολυκύτταρων οργανισμών (όπως αναγνώριση κυττάρων, διεργασίες του

συστήματος κ.α.). 19 γονίδια (4,7%) συμμετέχουν σε διεργασίες του ανοσολογικού συστήματος (όπως ανοσολογική ανταπόκριση, ενεργοποίηση μακροφάγων, αντιγονοπαρουσίαση κ.α.). 6 γονίδια (1,5%) στην αναπαραγωγή (όπως γονιμοποίηση και παραγωγή γαμετών) και τα υπόλοιπα γονίδια συμμετέχουν στις ακόλουθες διεργασίες οι οποίες εμφανίζονται στο γράφημα που ακολουθεί (Εικόνα 5.2).

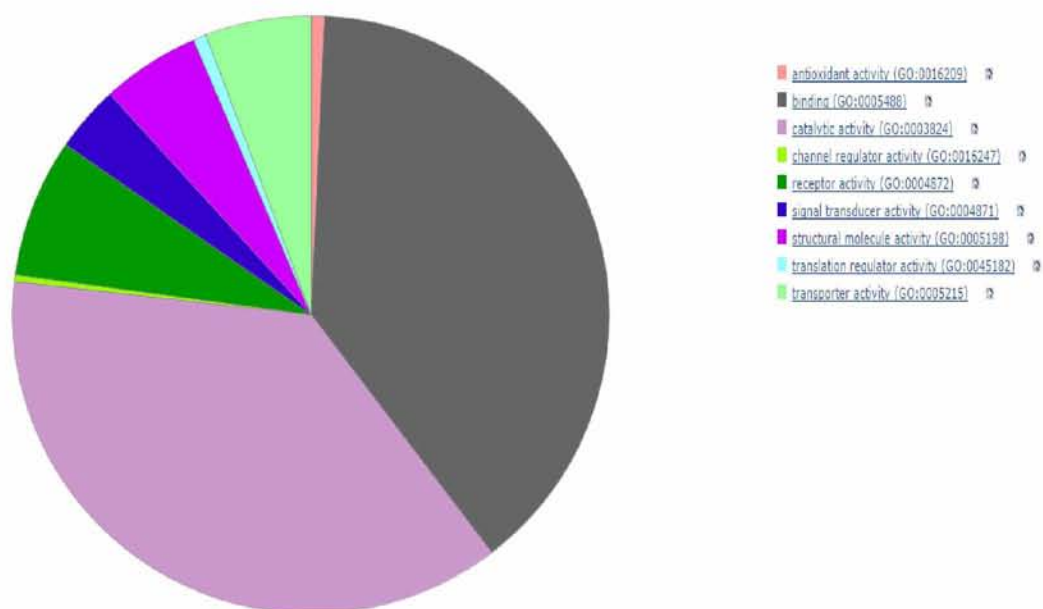


**Εικόνα 5.2** Κυκλικό διάγραμμα των βιολογικών διεργασιών στις οποίες συμμετέχουν τα στατιστικώς σημαντικά γονίδια που προέκυψαν από την μετα-ανάλυση. Δεξιά είναι οι βιολογικές διεργασίες στις οποίες συμμετέχουν τα γονίδια.

Ως προς τη μοριακή λειτουργία των 407 γονιδίων βλέπουμε ότι υπήρχαν πληροφορίες μόνο για 280 γονίδια, από αυτά 109 γονίδια (26,8%) συμμετέχουν στην δέσμευση (όπως δέσμευση ιόντων ασβεστίου, δέσμευση πρωτεϊνών, νουκλεϊκών οξέων, χρωματίνης κ.α.). 104 γονίδια (25,6%) συμμετέχουν σε καταλυτικές δραστηριότητες (όπως δράση της υδρολάσης, της λιγκάσης, ενζυμικές δράσεις κ.α.). 21 γονίδια (5,2%) συμμετέχουν σε δράσεις υποδοχέων. 16 γονίδια (3,9%) συμμετέχουν σε διεργασίες μεταφοράς (όπως διαμεμβρανική μεταφορά, λιπιδιακές μεταφορές, διαμεμβρανική μεταφορά αμινοξέων κ.α.). 15 γονίδια (3,7%) συμμετέχουν σε δομικές μοριακές λειτουργίες. 2 γονίδια (0,5%) συμμετέχουν σε



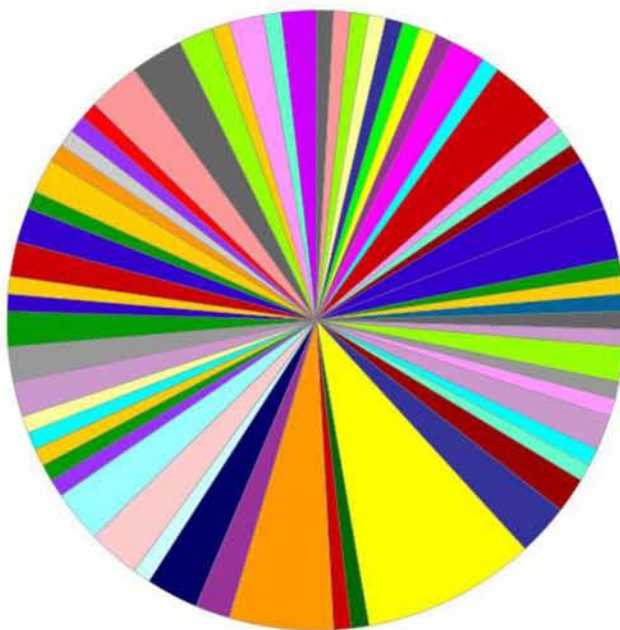
αντιοξειδωτικές λειτουργίες και τα υπόλοιπα γονίδια συμμετέχουν στις ακόλουθες λειτουργίες που εμφανίζονται στο γράφημα (Εικόνα 5.3).



**Εικόνα 5.3** Κυκλικό διάγραμμα των μοριακών λειτουργιών στις οποίες συμμετέχουν τα στατιστικώς σημαντικά γονίδια που προέκυψαν από την μετα-ανάλυση. Δεξιά είναι οι μοριακές λειτουργίες στις οποίες συμμετέχουν τα γονίδια.

Όσο αναφορά τα μονοπάτια στα οποία συμμετέχουν τα 407 στατιστικώς σημαντικά γονίδια υπήρχαν πληροφορίες μόνο για 110 γονίδια και στα οποία βλέπουμε ότι υπάρχει μεγάλο εύρος μονοπατιών (63) όπως φαίνεται και στο διάγραμμα (Εικόνα 5.4). 10 γονίδια (2,5%) εμπλέκονται στο μονοπάτι του ορμονικού υποδοχέα απελευθέρωσης της γονοδοτοπίνης. 6 γονίδια (1,5%) εμπλέκονται στο ετεροτριμερές μονοπάτι-Gi σηματοδότησης G-πρωτεϊνών άλφα και Gs άλφα μεσολάβησης μονοπάτι (Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway). 4 γονίδια (1%) εμπλέκονται στο μονοπάτι της ενεργοποίησης των Β κυττάρων (Β-λεμφοκυττάρων). 3 γονίδια (0.7%) εμπλέκονται στο μονοπάτι σηματοδότησης της Ιντεγκρίνης (ιντεργκίνη α, λαμινίνη και κολλαγόνο). 3 γονίδια (0.7%) εμπλέκονται στο μονοπάτι την νόσου Huntington. 3 γονίδια (0.7%) εμπλέκονται στο μονοπάτι της γλυκόλυσης. 3 γονίδια (0.7%) εμπλέκονται στο μονοπάτι της ενεργοποίησης των Τ κυττάρων (Τ-λεμφοκυττάρων). 3

γονίδια (0.7%) εμπλέκονται στο μονοπάτι της πήξης του αίματος. 2 γονίδια (0.5%) εμπλέκονται στο μονοπάτι της απόπτωσης. 2 γονίδια (0.5%) εμπλέκονται στο μονοπάτι του οξειδωτικού στρες και τα υπόλοιπα γονίδια εμπλέκονται στις ακόλουθες λειτουργίες που εμφανίζονται στο γράφημα (Εικόνα 5.4).



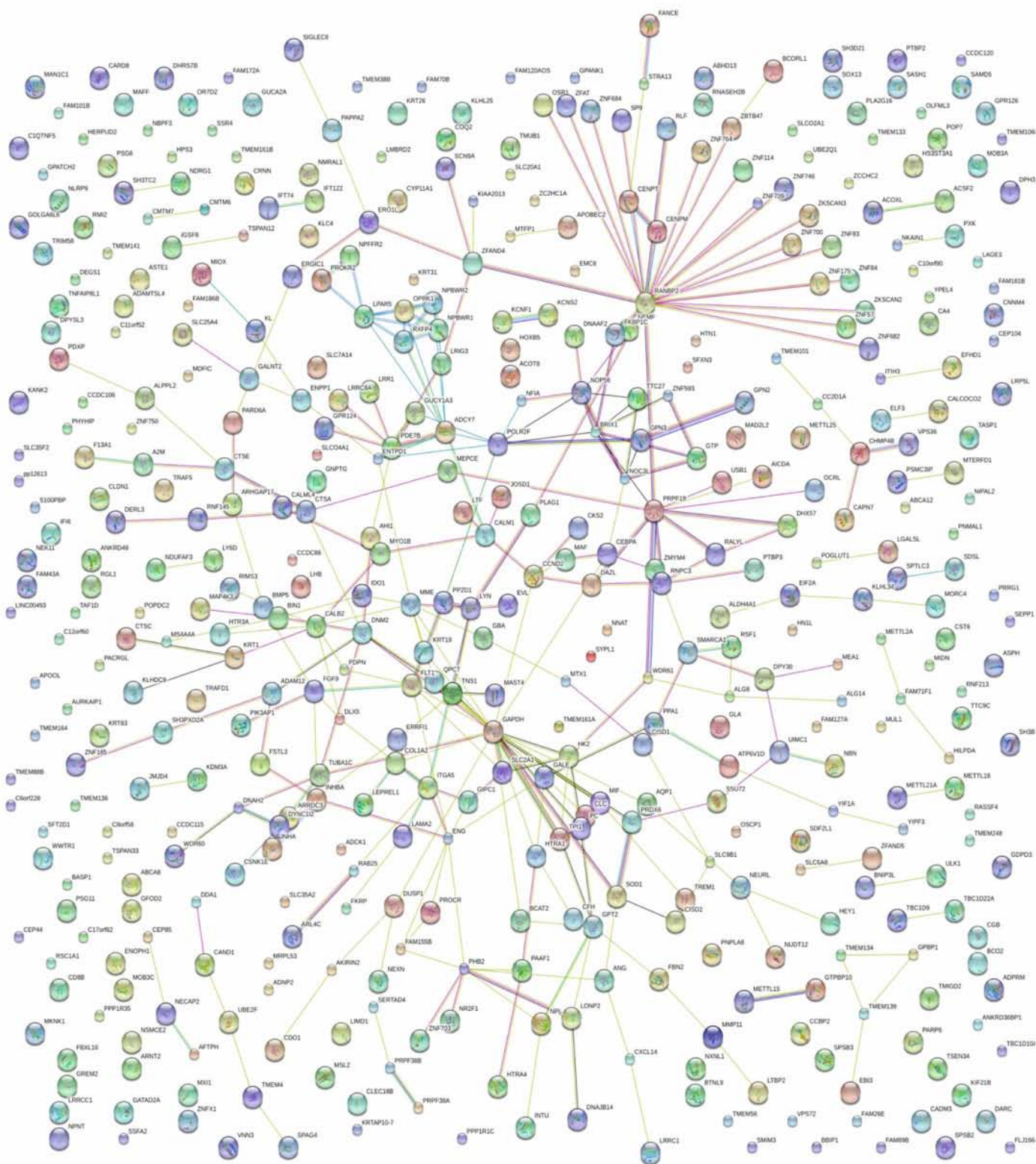
**Εικόνα 5.4** Κυκλικό διάγραμμα των μονοπατιών στις οποίες συμμετέχουν τα στατιστικώς σημαντικά γονίδια που προέκυψαν από την μετα-ανάλυση. Δεξιά είναι τα μονοπάτια στα οποία συμμετέχουν τα γονίδια.

- 5-Hydroxytryptamine degradation (P04372) ⓘ
- 5HT1 type receptor mediated signaling pathway (P04373) ⓘ
- 5HT2 type receptor mediated signaling pathway (P04375) ⓘ
- 5HT4 type receptor mediated signaling pathway (P04376) ⓘ
- Alanine biosynthesis (P02724) ⓘ
- Alzheimer disease-amyloid secretase pathway (P00003) ⓘ
- Alzheimer disease-presenilin pathway (P00004) ⓘ
- Androgen/estrogene/progesterone biosynthesis (P02727) ⓘ
- Apoptosis signaling pathway (P00006) ⓘ
- Axon guidance mediated by semaphorins (P00007) ⓘ
- B cell activation (P00010) ⓘ
- Beta1 adrenergic receptor signaling pathway (P04377) ⓘ
- Beta2 adrenergic receptor signaling pathway (P04378) ⓘ
- Beta3 adrenergic receptor signaling pathway (P04379) ⓘ
- Blood coagulation (P00011) ⓘ
- CCKR signaling map (P06959) ⓘ
- Cadherin signaling pathway (P00012) ⓘ
- Cell cycle (P00013) ⓘ
- Circadian clock system (P00015) ⓘ
- Cytoskeletal regulation by Rho GTPase (P00016) ⓘ
- Dopamine receptor mediated signaling pathway (P05912) ⓘ
- Endothelin signaling pathway (P00019) ⓘ
- Enkephalin release (P05913) ⓘ
- EGF signaling pathway (P00021) ⓘ
- Fructose palactose metabolism (P02744) ⓘ
- GABA-B receptor II signaling (P05731) ⓘ
- General transcription by RNA polymerase I (P00022) ⓘ
- General transcription regulation (P00023) ⓘ
- Glycolysis (P00024) ⓘ
- Gonadotropin-releasing hormone receptor pathway (P06664) ⓘ
- Hedgehog signaling pathway (P00025) ⓘ
- Heme biosynthesis (P02746) ⓘ
- Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway (P00026) ⓘ
- Heterotrimeric G-protein signaling pathway-Go alpha and Go alpha mediated pathway (P00027) ⓘ
- Heterotrimeric G-protein signaling pathway-rod outer segment phototransduction (P00028) ⓘ
- Histamine H2 receptor mediated signaling pathway (P04386) ⓘ
- Huntington disease (P00029) ⓘ
- Integrin signalling pathway (P00034) ⓘ
- Interleukin signaling pathway (P00036) ⓘ
- Isoleucine biosynthesis (P02748) ⓘ
- Leucine biosynthesis (P02749) ⓘ
- Muscarinic acetylcholine receptor 2 and 4 signaling pathway (P00043) ⓘ
- N-acetylglucosamine metabolism (P02756) ⓘ
- Nicotinic acetylcholine receptor signaling pathway (P00044) ⓘ
- Notch signaling pathway (P00045) ⓘ
- Opioid prodynorphin pathway (P05916) ⓘ
- Opioid proenkephalin pathway (P05915) ⓘ
- Opioid proopiomelanocortin pathway (P05917) ⓘ
- Oxidative stress response (P00046) ⓘ
- PDGF signaling pathway (P00047) ⓘ
- PI3 kinase pathway (P00048) ⓘ
- Parkinson disease (P00049) ⓘ
- Pentose phosphate pathway (P02762) ⓘ
- Pyrimidine Metabolism (P02771) ⓘ
- Pyruvate metabolism (P02772) ⓘ
- Ras Pathway (P04393) ⓘ
- T cell activation (P00053) ⓘ
- TGF-beta signaling pathway (P00052) ⓘ
- Transcription regulation by bZIP transcription factor (P00055) ⓘ
- Valine biosynthesis (P02785) ⓘ
- Wnt signaling pathway (P00057) ⓘ
- mRNA splicing (P00058) ⓘ
- p38 MAPK pathway (P05918) ⓘ

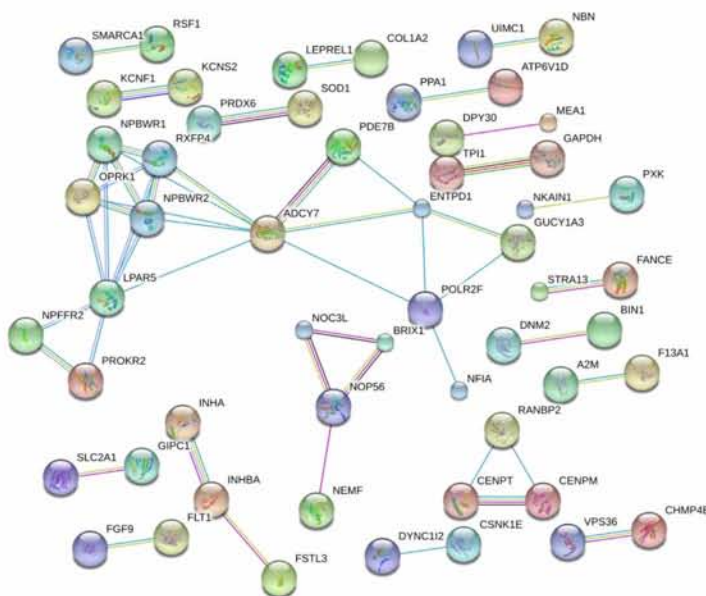
Με την βοήθεια της STRING πήραμε αποτελέσματα σχετικά με τις αλληλεπιδράσεις μεταξύ των πρωτεϊνών. Στην Εικόνα 5.5 βλέπουμε το πρωτεϊνικό

δίκτυο των 466 γονιδίων που αναγνωρίστηκαν από την πλατφόρμα. Με την μαύρη γραμμή ενώνονται γονίδια των οποίων οι πρωτεΐνες συνεκφράζονται, με πράσινη γραμμή γειτονικά γονίδια, με μπλε γραμμή γονίδια των οποίων οι πρωτεΐνες συνυπάρχουν, με κίτρινη γραμμή γονίδια των οποίων οι πρωτεΐνες βρίσκονται σε κοινά κείμενα (text mining), με ροζ γραμμή γονίδια των οποίων οι πρωτεΐνες έχουν βρεθεί σε κοινά πειράματα. Στην Εικόνα 5.6 βλέπουμε τις πρωτεΐνες των στατιστικώς σημαντικών γονιδίων της μετα-ανάλυσης οι οποίες έχουν αυξημένο σκορ αλληλεπιδράσεων (0.9), δυστυχώς δεν είναι ακόμη γνωστό το πως ακριβώς αλληλεπιδρούν ωστόσο έχει βρεθεί ότι κάποιες πρωτεΐνες από αυτές βρίσκονται μαζί σε εξειδικευμένες βάσεις δεδομένων, σε πειράματα, σε διάφορες βιοχημικές έρευνες, αναφέρονται σε διάφορες μελέτες μαζί, οι ομόλογες πρωτεΐνες τους έχουν βρεθεί να συνεκφράζονται είτε να βρίσκονται σε γειτονικούς γενετικούς τόπους σε άλλα είδη και έχουν βρεθεί ότι μπορούν να υποστούν γονιδιακή σύντηξη σε κάποιες ασθένειες.





**Εικόνα 5.5** Δίκτυο πρωτεϊνικών αλληλεπιδράσεων των στατιστικά σημαντικών γονιδίων που προέκυψαν από την μετα-ανάλυση



**Εικόνα 5.6** Δίκτυο πρωτεϊνικών αλληλεπιδράσεων των στατιστικά σημαντικών γονιδίων που προέκυψαν από την μετα-ανάλυση με αυξημένο σκορ αλληλεπιδράσεων.

## 5.2 Συζήτηση

Στην παρούσα διπλωματική εργασία, υλοποιήθηκε μετα-ανάλυση μικροσυστοιχιών για την εύρεση γονιδίων που εκφράζονται διαφορεικά στην προεκλαμψία. Με την πραγματοποίηση της μετα-ανάλυσης ανιχνεύθηκαν γονίδια τα οποία έχουν στατιστικά σημαντική συσχέτιση με τη νόσο και τα οποία δεν θα ήταν δυνατό να ανιχνευθούν από τις επιμέρους μελέτες. Βρέθηκαν πως υπάρχουν 783 στατιστικώς σημαντικά γονίδια από τη μέθοδο διόρθωσης FDR σε επίπεδο σημαντικότητας 1%. Τα γονίδια εμφανίστηκαν να παίρνουν μέρος σε πληθώρα βιολογικών διεργασιών όπως αναπτυξιακές διεργασίες, ανοσολογικές διεργασίες και στην αναπαραγωγή, μοριακών λειτουργιών όπως στην δέσμευση, σε καταλυτικές δραστηριότητες και αντιοξειδωτικές λειτουργίες και μονοπατιών όπως σε μονοπάτια του ορμονικού υποδοχέα απελευθέρωσης της γονοδοτοπίνης, σε μονοπάτια της πήξης του αίματος, σε μονοπάτια της απόπτωσης, σε μονοπάτια του κυττάρων του ανοσοβιολογικού συστήματος και σε μονοπάτια του οξειδωτικού στρες. Αρκετές από τις πρωτεΐνες των γονιδίων που ερευνήσαμε έχουν συμπεριληφθεί σε διάφορες μελέτες, σε κάποιες σχετικές με την προεκλαμψία και το σύνδρομο HELLP (Haram, Mortensen, and Nagy 2014), με την προεκλαμψία (Tejera, Bernardes, and Rebelo 2013), με τον καρκίνο



(Wu 2014), με καρδιαγγειακές νόσους (Toischer et al. 2010) και με τη πίεση του αίματος (Lee et al. 2007). Ακόμη κάποιες από τις πρωτεΐνες των γονιδίων που ερευνήσαμε έχουν βρεθεί οι ομόλογες τους σε άλλα είδη να συνεκφράζονται είτε να προέρχονται από γειτονικούς γενετικούς τόπους. Περαιτέρω έρευνα κρίνεται αναγκαία για τη διερεύνηση της συσχέτισης των γονιδίων που βρέθηκαν στην συγκεκριμένη μελέτη με την εμφάνιση της προεκλαμψίας.



## **Βιβλιογραφία**



- Babu, Madan M. 2004. "An introduction to microarray data analysis " In *Computational Genomics: Theory and Application* edited by Richard P. Grant, 225-249. Cambridge, UK: Horizon Bioscience
- Baird, David. 2012. Analysis of Microarray Data. In *GenStat for Windows*, edited by GenSta: VSN International, 5 The Waterhouse, Waterhouse Street, Hemel Hempstead, Hertfordshire HP1 1ES, UK. Original edition, GenStat for Windows 9th Edition.
- Benjamini, Yoav, and Yosef Hochberg. 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.
- Benson, D. A., K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers. 2015. "GenBank." *Nucleic Acids Res* 43 (Database issue):D30-5. doi: 10.1093/nar/gku1216.
- Bodnar, L. M., J. M. Catov, and J. M. Roberts. 2007. "Racial/ethnic differences in the monthly variation of preeclampsia incidence." *Am J Obstet Gynecol* 196 (4):324.e1-5. doi: 10.1016/j.ajog.2006.11.028.
- Brazma, A., and J. Vilo. 2000. "Gene expression data analysis." *FEBS Lett* 480 (1):17-24.
- Breitling, R., A. Amtmann, and P. Herzyk. 2004. "Iterative Group Analysis (iGA): a simple tool to enhance sensitivity and facilitate interpretation of microarray experiments." *BMC Bioinformatics* 5:34. doi: 10.1186/1471-2105-5-34.
- Breitling, R., P. Armengaud, A. Amtmann, and P. Herzyk. 2004. "Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments." *FEBS Lett* 573 (1-3):83-92. doi: 10.1016/j.febslet.2004.07.055.
- Brown, M. A., W. M. Hague, J. Higgins, S. Lowe, L. McCowan, J. Oats, M. J. Peek, J. A. Rowan, B. N. Walters, and Pregnancy Australasian Society of the Study of Hypertension in. 2000. "The detection, investigation and management of hypertension in pregnancy: full consensus statement." *Aust N Z J Obstet Gynaecol* 40 (2):139-55.
- Buchbinder, Alan, Baha M. Sibai, Steve Caritis, Cora MacPherson, John Hauth, Marshall D. Lindheimer, Mark Klebanoff, Peter VanDorsten, Mark Landon, Richard Paul, Menachem Miodovnik, Paul Meis, and Gary Thurnau. 2002. "Adverse perinatal outcomes are significantly higher in severe gestational hypertension than in mild preeclampsia." *American Journal of Obstetrics & Gynecology* 186 (1): 66–71 doi: <http://dx.doi.org/10.1067/mob.2002.120080>.
- Campaign, Anna, and Yee Hwa Yang. 2010. "Comparison study of microarray meta-analysis methods." *BMC Bioinformatics* 11 (1):408. doi: 10.1186/1471-2105-11-408.
- Carlin, J. B. 2000. "Tutorial in biostatistics. Meta-analysis: formulating, evaluating, combining, and reporting by S-L. T. Normand, *Statistics in Medicine*, 18, 321-359 (1999)." *Stat Med* 19 (5):753-9.
- Carter, M. G., T. Hamatani, A. A. Sharov, C. E. Carmack, Y. Qian, K. Aiba, N. T. Ko, D. B. Dudekula, P. M. Brzoska, S. S. Hwang, and M. S. Ko. 2003. "In situ-synthesized novel microarray optimized for mouse stem cell and early developmental expression profiling." *Genome Res* 13 (5):1011-21. doi: 10.1101/gr.878903.
- Chappell, S., and L. Morgan. 2006. "Searching for genetic clues to the causes of preeclampsia." *Clin Sci (Lond)* 110 (4):443-58. doi: 10.1042/cs20050323.

- Cheung, V. G., M. Morley, F. Aguilar, A. Massimi, R. Kucherlapati, and G. Childs. 1999. "Making and reading microarrays." *Nat Genet* 21 (1 Suppl):15-9. doi: 10.1038/4439.
- Chiaffarino, F., F. Parazzini, D. Paladini, B. Acaia, W. Ossola, L. Marozio, F. Facchinetti, and A. Del Giudice. 2004. "A small randomised trial of low-dose aspirin in women at high risk of pre-eclampsia." *Eur J Obstet Gynecol Reprod Biol* 112 (2):142-4.
- Choi, J. K., U. Yu, S. Kim, and O. J. Yoo. 2003. "Combining multiple microarray studies and modeling interstudy variation." *Bioinformatics* 19 Suppl 1:i84-90.
- Cooper, Desmond W., Shaun P. Brennecke, and Alan N. Wilton. 1993. "Genetics of Pre-Eclampsia." *Hypertension in Pregnancy* 12 (1):1-23. doi: 10.3109/10641959309031050.
- Dalma-Weiszhausz, D. D., J. Warrington, E. Y. Tanimoto, and C. G. Miyada. 2006. "The affymetrix GeneChip platform: an overview." *Methods Enzymol* 410:3-28. doi: 10.1016/S0076-6879(06)10001-4.
- DerSimonian, R., and N. Laird. 1986. "Meta-analysis in clinical trials." *Control Clin Trials* 7 (3):177-88.
- Dudoit, Sandrine, Yee Hwa Yang, Matthew J. Callo, and Terence P. Speed. 2002. "Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments." *Statistica Sinica* (12):111-139.
- Ebrashy, A., M. Ibrahim, A. Marzook, and D. Yousef. 2005. "Usefulness of aspirin therapy in high-risk pregnant women with abnormal uterine artery Doppler ultrasound at 14-16 weeks pregnancy: randomized controlled clinical trial." *Croat Med J* 46 (5):826-31.
- Efron, Bradley, and R. J. Tibshirani. 1994. *An introduction to the bootstrap*: Chapman and Hall/CRC.
- Fay, D. S., and K. Gerow. 2013. "A biologist's guide to statistical thinking and analysis." *WormBook*:1-54. doi: 10.1895/wormbook.1.159.1.
- Freeman, W. M., D. J. Robertson, and K. E. Vrana. 2000. "Fundamentals of DNA hybridization arrays for gene expression analysis." *Biotechniques* 29 (5):1042-6, 1048-55.
- Gifford, RW, PA August, G Cunningham, and et al. 2000. "Report of the National High Blood Pressure Education Program Working Group on High Blood Pressure in Pregnancy." *Am J Obstet Gynecol* 183 (1):S1-S22.
- Guo, R. F., M. Huber-Lang, X. Wang, V. Sarma, V. A. Padgaonkar, R. A. Craig, N. C. Riedemann, S. D. McClintock, T. Hlaing, M. M. Shi, and P. A. Ward. 2000. "Protective effects of anti-C5a in sepsis-induced thymocyte apoptosis." *J Clin Invest* 106 (10):1271-80. doi: 10.1172/jci10793.
- Gursoy, T., D. Aliefendioglu, A. T. Aslan, M. Gunduz, A. Haberal, M. Senes, F. N. Cakmak, and Y. R. Laleli. 2002. "Preeclampsia disrupts the normal physiology of leptin." *Am J Perinatol* 19 (6):303-10. doi: 10.1055/s-2002-34462.
- Halperin, A., A. Buhot, and E. B. Zhulina. 2004. "Sensitivity, specificity, and the hybridization isotherms of DNA chips." *Biophys J* 86 (2):718-30. doi: 10.1016/S0006-3495(04)74150-8.
- Haram, K., J. H. Mortensen, and B. Nagy. 2014. "Genetic aspects of preeclampsia and the HELLP syndrome." *J Pregnancy* 2014:910751. doi: 10.1155/2014/910751.
- Harkin, Denis W., Alex Romaschin, Stephen M. Taylor, Barry B. Rubin, and Thomas F. Lindsay. 2004. "Complement c5a receptor antagonist attenuates multiple



- organ injury in a model of ruptured abdominal aortic aneurysm." *Journal of Vascular Surgery* 39 (1):196-206. doi: <http://dx.doi.org/10.1016/j.jvs.2003.07.001>.
- Hauth, J. C., M. G. Ewell, R. J. Levine, J. R. Esterlitz, B. Sibai, L. B. Curet, P. M. Catalano, and C. D. Morris. 2000. "Pregnancy outcomes in healthy nulliparas who developed hypertension. Calcium for Preeclampsia Prevention Study Group." *Obstet Gynecol* 95 (1):24-8.
- Hoheisel, J. D. 2006. "Microarray technology: beyond transcript profiling and genotype analysis." *Nat Rev Genet* 7 (3):200-10. doi: 10.1038/nrg1809.
- Holland, Burt, and Margaret Copenhaver. 1987. "An Improved Sequentially Rejective Bonferroni Test Procedure." *Biometrics* 43 (2):417.
- Holm, S. 1979. "A simple sequentially rejective multiple test procedure." *Scandinavian Journal of Statistics* 6:65-70.
- Hopewell, S., S. McDonald, M. Clarke, and M. Egger. 2007. "Grey literature in meta-analyses of randomized trials of health care interventions." *Cochrane Database Syst Rev* (2):Mr000010. doi: 10.1002/14651858.MR000010.pub3.
- Hutcheon, J. A., S. Lisonkova, and K. S. Joseph. 2011. "Epidemiology of pre-eclampsia and the other hypertensive disorders of pregnancy." *Best Pract Res Clin Obstet Gynaecol* 25 (4):391-403. doi: 10.1016/j.bpobgyn.2011.01.006.
- Ioannidis, J. P., and T. A. Trikalinos. 2005. "Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials." *J Clin Epidemiol* 58 (6):543-9. doi: 10.1016/j.jclinepi.2004.10.019.
- Kerr, Kathleen F. 2007. "Extended analysis of benchmark datasets for Agilent two-color microarrays." *BMC Bioinformatics* 8:371. doi: 10.1186/1471-2105-8-371.
- Kontou, P.I., A. Pavlopoulou, and P.G. Bagos. 2016 "Methods of Analysis and Meta-Analysis for Identifying Differentially Expressed Genes." *Methods Mol Bio* in press.
- Lachmeijer, A. M., R. Arngrimsson, E. J. Bastiaans, M. L. Frigge, G. Pals, S. Sigurdardottir, H. Stefansson, B. Palsson, D. Nicolae, A. Kong, J. G. Aarnoudse, J. R. Gulcher, G. A. Dekker, L. P. ten Kate, and K. Stefansson. 2001. "A genome-wide scan for preeclampsia in the Netherlands." *Eur J Hum Genet* 9 (10):758-64. doi: 10.1038/sj.ejhg.5200706.
- Lee, H. K., A. K. Hsu, J. Sajdak, J. Qin, and P. Pavlidis. 2004. "Coexpression analysis of human genes across many microarray data sets." *Genome Res* 14 (6):1085-94. doi: 10.1101/gr.1910904.
- Lee, N. H., B. J. Haas, N. E. Letwin, B. C. Frank, T. V. Luu, Q. Sun, C. D. House, S. Yerga-Woolwine, P. Farms, E. Manickavasagam, and B. Joe. 2007. "Cross-talk of expression quantitative trait loci within 2 interacting blood pressure quantitative trait loci." *Hypertension* 50 (6):1126-33. doi: 10.1161/hypertensionaha.107.093138.
- Levine, R. J., L. J. Vatten, G. L. Horowitz, C. Qian, P. R. Romundstad, K. F. Yu, A. N. Hollenberg, A. I. Hellevik, B. O. Asvold, and S. A. Karumanchi. 2009. "Pre-eclampsia, soluble fms-like tyrosine kinase 1, and the risk of reduced thyroid function: nested case-control and population based study." *BMJ* 339:b4336. doi: 10.1136/bmj.b4336.
- Makrides, M., L. Duley, and S. F. Olsen. 2006. "Marine oil, and other prostaglandin precursor, supplementation for pregnancy uncomplicated by pre-eclampsia or

- intrauterine growth restriction." *Cochrane Database Syst Rev* (3):Cd003402. doi: 10.1002/14651858.CD003402.pub2.
- Manten, G. T., M. J. Sikkema, H. A. Voorbij, G. H. Visser, H. W. Bruinse, and A. Franx. 2007. "Risk factors for cardiovascular disease in women with a history of pregnancy complicated by preeclampsia or intrauterine growth restriction." *Hypertens Pregnancy* 26 (1):39-50. doi: 10.1080/10641950601146574.
- Miklos, G. L., and R. Maleszka. 2004. "Microarray reality checks in the context of a complex disease." *Nat Biotechnol* 22 (5):615-21. doi: 10.1038/nbt965.
- Normand, S. L. 1999. "Meta-analysis: formulating, evaluating, combining, and reporting." *Stat Med* 18 (3):321-59.
- Pan, Zhenglun, Thomas A. Trikalinos, Fotini K. Kavvoura, Joseph Lau, and John P. A. Ioannidis. 2005. "Local Literature Bias in Genetic Epidemiology: An Empirical Evaluation of the Chinese Literature." *PLoS Medicine* 2 (12):e334. doi: 10.1371/journal.pmed.0020334.
- Park, J. C., H. S. Kim, and J. J. Kim. 2001. "Bidirectional incremental parsing for automatic pathway identification with combinatory categorial grammar." *Pac Symp Biocomput*:396-407.
- Pease, A. C., D. Solas, E. J. Sullivan, M. T. Cronin, C. P. Holmes, and S. P. Fodor. 1994. "Light-generated oligonucleotide arrays for rapid DNA sequence analysis." *Proc Natl Acad Sci U S A* 91 (11):5022-6.
- Petitti, D. B. 1994. *Meta-analysis, decision analysis, and cost-effectiveness analysis: Methods for quantitative synthesis in medicine*. : New York: Oxford University Press.
- Polyzos, N. P., D. Mauri, M. Tsappi, S. Tzioras, K. Kamposioras, I. Cortinovis, and G. Casazza. 2007. "Combined vitamin C and E supplementation during pregnancy for preeclampsia prevention: a systematic review." *Obstet Gynecol Surv* 62 (3):202-6. doi: 10.1097/01.ogx.0000256787.04807.da.
- Ramasamy, A., A. Mondry, C. C. Holmes, and D. G. Altman. 2008. "Key issues in conducting a meta-analysis of gene expression microarray datasets." *PLoS Med* 5 (9):e184. doi: 10.1371/journal.pmed.0050184.
- Redman, C. W., and I. L. Sargent. 2003. "Pre-eclampsia, the placenta and the maternal systemic inflammatory response--a review." *Placenta* 24 Suppl A:S21-7.
- Roberts, C. J., B. Nelson, M. J. Marton, R. Stoughton, M. R. Meyer, H. A. Bennett, Y. D. He, H. Dai, W. L. Walker, T. R. Hughes, M. Tyers, C. Boone, and S. H. Friend. 2000. "Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles." *Science* 287 (5454):873-80.
- Roberts, James M., Phyllis A. August, George Bakris, John R. Barton, Ira M. Bernstein, Maurice Druzin, Robert R. Gaiser, Joey P. Granger, Arun Jeyabalan, Donna D. Johnson, S. Ananth Karumanchi, Marshall D. Lindheimer, Michelle Y. Owens, George R. Saade, Baha M. Sibai, Catherine Y. Spong, Eleni Tsigas, and James N. Martin Jr. 1996. *Hypertension in Pregnancy*. Vol. 219: American College of Obstetricians and Gynecologists.
- Saftlas, A. F., R. J. Levine, M. A. Klebanoff, K. L. Martz, M. G. Ewell, C. D. Morris, and B. M. Sibai. 2003. "Abortion, changed paternity, and risk of preeclampsia in nulliparous women." *Am J Epidemiol* 157 (12):1108-14.
- Salonen Ros, H., P. Lichtenstein, L. Lipworth, and S. Cnattingius. 2000. "Genetic effects on the liability of developing pre-eclampsia and gestational hypertension." *Am J Med Genet* 91 (4):256-60.



- Seidel, C. 2008. *Introduction to DNA Microarrays, in Analysis of Microarray Data: A Network-Based Approach* (eds F. Emmert-Streib and M. Dehmer). Edited by Weinheim Wiley-VCH Verlag GmbH & Co. KGaA, Germany.
- Sergio, F., D. Maria Clara, F. Gabriella, S. Giorgia, Carolis Sara De, P. Giancarlo, and C. Alessandro. 2006. "Prophylaxis of recurrent preeclampsia: low-molecular-weight heparin plus low-dose aspirin versus low-dose aspirin alone." *Hypertens Pregnancy* 25 (2):115-27. doi: 10.1080/10641950600745517.
- Serrano, N. C. 2006. "Immunology and genetic of preeclampsia." *Clin Dev Immunol* 13 (2-4):197-201. doi: 10.1080/17402520600876903.
- Shoemaker, D. D., E. E. Schadt, C. D. Armour, Y. D. He, P. Garrett-Engele, P. D. McDonagh, P. M. Loerch, A. Leonardson, P. Y. Lum, G. Cavet, L. F. Wu, S. J. Altschuler, S. Edwards, J. King, J. S. Tsang, G. Schimmack, J. M. Schelter, J. Koch, M. Ziman, M. J. Marton, B. Li, P. Cundiff, T. Ward, J. Castle, M. Krolewski, M. R. Meyer, M. Mao, J. Burchard, M. J. Kidd, H. Dai, J. W. Phillips, P. S. Linsley, R. Stoughton, S. Scherer, and M. S. Boguski. 2001. "Experimental annotation of the human genome using microarray technology." *Nature* 409 (6822):922-7. doi: 10.1038/35057141.
- Sibai, B. M. 2003. "Diagnosis and management of gestational hypertension and preeclampsia." *Obstet Gynecol* 102 (1):181-92.
- Sibai, B. M. 2006. "Preeclampsia as a cause of preterm and late preterm (near-term) births." *Semin Perinatol* 30 (1):16-9. doi: 10.1053/j.semperi.2006.01.008.
- Šidák, Zbyněk. 1967. "Rectangular Confidence Regions for the Means of Multivariate Normal Distributions." *Journal of the American Statistical Association* 62 (318):626-633.
- Singh-Gasson, S., R. D. Green, Y. Yue, C. Nelson, F. Blattner, M. R. Sussman, and F. Cerrina. 1999. "Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array." *Nat Biotechnol* 17 (10):974-8. doi: 10.1038/13664.
- Spellman, P. T., G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, and B. Futcher. 1998. "Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization." *Mol Biol Cell* 9 (12):3273-97.
- Stekel, Dov. 2003. *Microarray bioinformatics*: Cambridge University Press.
- Tejera, E., J. Bernardes, and I. Rebelo. 2013. "Co-expression network analysis and genetic algorithms for gene prioritization in preeclampsia." *BMC Med Genomics* 6:51. doi: 10.1186/1755-8794-6-51.
- Thakkinian, A., P. McElduff, C. D'Este, D. Duffy, and J. Attia. 2005. "A method for meta-analysis of molecular association studies." *Stat Med* 24 (9):1291-306. doi: 10.1002/sim.2010.
- Toischer, K., A. G. Rokita, B. Unsold, W. Zhu, G. Kararigas, S. Sossalla, S. P. Reuter, A. Becker, N. Teucher, T. Seidler, C. Grebe, L. Preuss, S. N. Gupta, K. Schmidt, S. E. Lehnart, M. Kruger, W. A. Linke, J. Backs, V. Regitz-Zagrosek, K. Schafer, L. J. Field, L. S. Maier, and G. Hasenfuss. 2010. "Differential cardiac remodeling in preload versus afterload." *Circulation* 122 (10):993-1003. doi: 10.1161/circulationaha.110.943431.
- Tomiuk, S., and K. Hofmann. 2001. "Microarray probe selection strategies." *Brief Bioinform* 2 (4):329-40.
- Trumbo, P. R., and K. C. Ellwood. 2007. "Supplemental calcium and risk reduction of hypertension, pregnancy-induced hypertension, and preeclampsia: an

- evidence-based review by the US Food and Drug Administration." *Nutr Rev* 65 (2):78-87.
- Villanueva, Elmer V., and Silva Zavarsek. 2004. "Evaluating heterogeneity in cumulative meta-analyses." *BMC Medical Research Methodology* 4 (1):18. doi: 10.1186/1471-2288-4-18.
- Walker, J. J. 2000. "Pre-eclampsia." *Lancet* 356 (9237):1260-5. doi: 10.1016/s0140-6736(00)02800-2.
- Wu, Licun. 2014. "Chasing on the way of cancer immunotherapy." *Journal of Thoracic Disease* 6 (6):E137-E138. doi: 10.3978/j.issn.2072-1439.2014.03.14.
- Zhang, J., S. Meikle, and A. Trumble. 2003. "Severe maternal morbidity associated with hypertensive disorders in pregnancy in the United States." *Hypertens Pregnancy* 22 (2):203-12. doi: 10.1081/PRG-120021066.
- Zimmermann, K. C., C. Bonzon, and D. R. Green. 2001. "The machinery of programmed cell death." *Pharmacol Ther* 92 (1):57-70.
- Zintzaras, E., and J. P. Ioannidis. 2008. "Meta-analysis for ranked discovery datasets: theoretical framework and empirical demonstration for microarrays." *Comput Biol Chem* 32 (1):38-46. doi: 10.1016/j.compbiolchem.2007.09.003.
- Μαμόπουλος, Μιχάλης. 1988. *Κόηση υψηλού κινδύνου*.
- Πανάρετος, Ι., and Ε. Ξεκαλάκη. 2000. *Εισαγωγή στη στατιστική σκέψη*. Vol. τόμος II *Εκτιμητική και Έλεγχοι Υποθέσεων*. Αθήνα.

## **Παράρτημα**



## Εντολές Stata

Παρακάτω παρουσιάζονται οι εντολές που χρησιμοποιήθηκαν στο στατιστικό πρόγραμμα Stata

Εντολή που φτιάχνει νέο αρχείο με τις μέσες τιμές απ' όλα τα κελιά

```
collapse(mean) gsm1 gsm2 gsm3 , by(gene_symbol)
list
list
```

Αντιμετάθεση πίνακα

```
xpose, clear
```

Εντολές t-test και bootstrap για κάθε μια μελέτη

```
set more off
file open meta using results.txt, write append
file write meta "gene"
file write meta " , "
file write meta "t"
file write meta " , "
file write meta "r(se) "
    write meta " , "
file write meta "p"
file write meta " , "
file write meta "e(se)"_n

foreach var of varlist alcf-zzz3 {
preserve
qui ttest `var', by(case_control) uneq
file write meta "`var'"
file write meta " , "
file write meta "`r(t)'"
file write meta " , "
file write meta "`r(se)'"
```

```

file write meta " , "
file write meta "`r(p)'"

        bootstrap  t=r(t),  reps(100)  strata(case_control):
        ttest `var',by(case_control)  uneq
mat se=e(se)
local se=se[1,1]
file write meta " , "
file write meta "`se'"_n

restore

}

file close meta

```

### Ενωση πινάκων

```
append using filename [filename ...] [, options]
```

### Εντολές μετα-ανάλυσης

```
encode gene, gen(ngene)
```

```
qui sum ngene
```

```
qui local k=r(max)
```

```

file open metan using finalresults.txt, write append
file write metan "gene"
file write  metan " "
file write metan "ES"
file write  metan " "
file write metan "se(ES)"
file write  metan " "
file write  metan "z"
file write  metan " "

```

```

file write metan "p"
file write metan " "
file write metan "df" _n

forvalues i=1/`k'{
metan t ese if ngene==`i',nograph randomi

    file write metan "`i'"
        file write metan " "
        file write metan "`r(ES)'"
        file write metan " "
        file write metan "`r(seES)'"
        file write metan " "
        file write metan "`r(z)'"
        file write metan " "
        file write metan "`r(p_z)'"
        file write metan " "
        file write metan "`r(df)'" _n

}

file close metan

```

### Εντολές διόρθωσης p-value

```

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)

```

## Λίστα στατιστικώς σημαντικών γονιδίων κατά FDR

(τα οποία χρησιμοποιήθηκαν ως είσοδος και στην PANTHER και στην STRING)

Γονιδιακό Σύμβολο	p-value		
XLOC_008696	7.90E-23	IGSF8	5.60E-10
LOC100509213	8.30E-20	LOC100506303	7.70E-10
PAPPA2	3.40E-16	PPP1R35	9.70E-10
LHB	2.40E-15	HA-1	9.90E-10
PPGB	2.00E-14	FAM164A	1.40E-09
LOC100507473	3.40E-14	AKIP	1.50E-09
FGIF	5.40E-14	XLOC_002602	1.60E-09
FKBP1C	5.40E-14	C13orf9	1.70E-09
HKE2	9.20E-14	SASH1	1.70E-09
LOC100506950	1.20E-13	TBPIP	1.70E-09
LRRC8	1.20E-13	MAN1C1	2.00E-09
LOC100653338	5.10E-13	HK2	2.70E-09
HIG2	1.00E-12	C13orf6	2.90E-09
ZNF651	1.40E-12	LOC100506011	3.90E-09
LOC160313	2.10E-12	NEURL	4.10E-09
SLCO2A1	3.00E-12	XLOC_009474	4.30E-09
DDEF1IT1	3.50E-12	NYD-SP18	5.00E-09
DKFZP564J0123	4.00E-12	CL640	5.10E-09
CKLFSF6	4.50E-12	NBS1	6.30E-09
FLJ12903	6.30E-12	MIF	6.40E-09
LOC100652987	1.00E-11	LOC152015	6.60E-09
GBA	1.90E-11	SPAG4	7.10E-09
MTX1	2.10E-11	ZNF262	7.10E-09
XLOC_L2_000917	3.20E-11	GALNT2	7.60E-09
LOC100652804	4.60E-11	KARCA1	9.50E-09
LOC100506963	8.30E-11	LOC255743	1.00E-08
CLDN1	8.50E-11	SOX13	1.00E-08
LOC100509105	1.50E-10	XLOC_000567	1.00E-08
PTE1	1.90E-10	DKFZP434J0113	1.10E-08
XLOC_002580	2.10E-10	MGC14560	1.20E-08
TMEM88B	2.20E-10	TRIM58	1.20E-08
XLOC_003176	2.40E-10	LOC100652805	1.30E-08
KIAA1764	2.60E-10	DKFZP564D172	1.40E-08
AKIRIN2-AS1	2.90E-10	SNORD82	1.40E-08
P66ALPHA	3.00E-10	ARNT2	1.70E-08
MXI1	3.10E-10	LOC100652876	2.00E-08
C6orf109	3.90E-10	MDS010	2.10E-08
ZCCHC2	4.50E-10	PIK3AP1	2.10E-08
LOC223075	4.60E-10	TIP120A	2.10E-08
FBXL16	5.20E-10	FLJ11712	2.30E-08
		HT008	2.30E-08



LOC90835	2.30E-08
RNP	2.30E-08
SFXN3	2.30E-08
OLFML3	2.60E-08
MGC4368	2.70E-08
XLOC_L2_005465	2.80E-08
GPR8	2.90E-08
GPR92	2.90E-08
XLOC_002211	3.40E-08
XLOC_L2_015397	3.50E-08
CGB	4.50E-08
XLOC_002813	4.50E-08
BTNL9	4.80E-08
MGC17791	5.30E-08
MME	7.30E-08
SLC6A10P	7.70E-08
XLOC_011676	8.10E-08
DPYSL3	8.40E-08
XLOC_006124	8.50E-08
ERGIC1	9.00E-08
XLOC_005426	9.10E-08
XLOC_006416	9.10E-08
AFG3L1	9.20E-08
TMEM255B	9.40E-08
FLJ12610	9.50E-08
ALDH4A1	1.00E-07
HRASLS3	1.00E-07
NPL	1.00E-07
RP4-621O152	1.00E-07
AFTIPHILIN	1.10E-07
FLJ10330	1.10E-07
LOC100507645	1.10E-07
UBE2Q	1.10E-07
ERRFI1	1.20E-07
GLA	1.20E-07
RP11-56A211	1.20E-07
MGC14801	1.30E-07
KIAA1404	1.60E-07
MORC4	1.60E-07
ADAM12	1.70E-07
FLJ13955	1.70E-07
FLJ36131	1.70E-07
MGC12981	1.70E-07
XLOC_012563	1.70E-07

BXDC2	1.90E-07
LOC100287558	2.10E-07
SLC6A8	2.10E-07
LIMD1	2.20E-07
LOC100506338	2.20E-07
FLJ12650	2.30E-07
XLOC_013437	2.30E-07
LOC283377	2.40E-07
MRPL53	2.40E-07
CTSC	2.60E-07
MTP18	2.80E-07
ZCD2	2.80E-07
FLJ26332	2.90E-07
GUCY1A3	3.00E-07
MGC14141	3.00E-07
MTVR1	3.00E-07
XLOC_009452	3.10E-07
LOC100293704	3.20E-07
CEP104	3.50E-07
DKFZP434I1610	3.50E-07
NSE2	3.50E-07
CCDC2	3.60E-07
KRT19	3.60E-07
TPI1	3.60E-07
CFH	3.80E-07
FLJ13841	3.80E-07
XLOC_L2_012210	3.80E-07
METTL18	3.90E-07
FLJ10349	4.00E-07
ANKRD25	4.30E-07
NEMF	4.40E-07
XLOC_001459	4.40E-07
LOC100507663	4.50E-07
XLOC_001083	4.50E-07
FLJ31413	4.60E-07
FKRP	4.80E-07
FLJ11848	4.80E-07
LOC400128	4.80E-07
PRRG1	4.80E-07
SMIM13	4.80E-07
GPR7	5.10E-07
XLOC_L2_001960	5.20E-07
SCN9A	5.30E-07
XLOC_001619	5.30E-07

XLOC_L2_008381	5.30E-07
WWTR1	5.40E-07
LOC391075	5.80E-07
MKNK1	5.80E-07
XLOC_000001	5.80E-07
FLN29	5.90E-07
RLN3R2	6.40E-07
XLOC_006505	6.40E-07
XLOC_L2_005352	6.60E-07
CCND2	6.80E-07
YPEL4	6.80E-07
XLOC_012507	7.00E-07
PSG6	7.40E-07
XLOC_011050	7.40E-07
FLJ14281	7.50E-07
HEY1	7.50E-07
LOC100652757	7.50E-07
XLOC_008033	7.50E-07
FLJ22313	8.00E-07
LENG5	8.00E-07
WDR60	8.10E-07
XLOC_010860	9.40E-07
BCDIN3	9.50E-07
PC	9.70E-07
SLC35A2	9.70E-07
NALP2	1.00E-06
XLOC_008995	1.00E-06
C22orf18	1.10E-06
LOC100505551	1.10E-06
PRDX6	1.10E-06
RAP80	1.10E-06
XLOC_004680	1.10E-06
LOC255104	1.20E-06
METTTL25	1.20E-06
SIGLEC6	1.20E-06
SLC2A1	1.20E-06
TREM1	1.20E-06
LOC100507065	1.30E-06
SH3PXD2A	1.30E-06
XLOC_014177	1.30E-06
ACTR3BP5	1.40E-06
HILPDA	1.40E-06
INHBA	1.40E-06
LOC100506800	1.40E-06

LOC143241	1.40E-06
LOC51249	1.40E-06
NPNT	1.40E-06
TNS	1.40E-06
XLOC_L2_012788	1.40E-06
GREM2	1.50E-06
KNSL8	1.50E-06
LOC340073	1.50E-06
XLOC_003714	1.50E-06
ENTPD1	1.60E-06
LOC91689	1.60E-06
QPCT	1.60E-06
TMEM139	1.60E-06
LRRC1	1.70E-06
XLOC_L2_007271	1.70E-06
EMC8	1.80E-06
FLJ20241	1.80E-06
LOC100506029	1.80E-06
PHYHIP	1.80E-06
XLOC_001406	1.80E-06
SAMD5	1.90E-06
XLOC_008350	1.90E-06
XLOC_011448	1.90E-06
YIF1	1.90E-06
XLOC_001425	2.00E-06
RANBP2	2.10E-06
KLHL25	2.20E-06
HTRA4	2.30E-06
LOC100130430	2.30E-06
MEA	2.30E-06
XLOC_005690	2.30E-06
LRIG3	2.40E-06
XLOC_005863	2.40E-06
GALE	2.50E-06
ITGA5	2.60E-06
LOC100505675	2.60E-06
LOC643940	2.60E-06
SLC35F2	2.60E-06
ATP6V1D	2.70E-06
FLJ20272	2.70E-06
ENG	2.80E-06
LAMA2	2.80E-06
LOC100652993	2.80E-06
POPDC2	2.80E-06

XLOC_L2_009790	2.80E-06
CDO1	2.90E-06
DUSP1	2.90E-06
LOC647179	2.90E-06
BIN1	3.00E-06
LOC100652786	3.00E-06
SUGT1P1	3.00E-06
LOC100506906	3.10E-06
XLOC_007215	3.10E-06
XLOC_011545	3.10E-06
C7orf21	3.20E-06
PCIA1	3.20E-06
JMJD1A	3.40E-06
LOC493861	3.40E-06
LOC494115	3.40E-06
MGC23244	3.40E-06
XLOC_011119	3.50E-06
DNAAF2	3.60E-06
GUCA2A	3.60E-06
MIF-AS1	3.60E-06
XLOC_001148	3.60E-06
C16orf34	3.70E-06
FLJ14936	3.70E-06
LOC100653033	3.70E-06
XLOC_012847	3.70E-06
XLOC_L2_012721	3.70E-06
TMEM133	3.80E-06
MYO1B	3.90E-06
TMEM161B	4.00E-06
TMEM161A	4.20E-06
POP7	4.30E-06
TMEM106B	4.30E-06
XLOC_000824	4.30E-06
LOC100509487	4.40E-06
LOC157567	4.40E-06
CSNK1E	4.60E-06
FLJ34223	4.70E-06
AICDA	4.80E-06
FAM43A	4.90E-06
XLOC_009683	4.90E-06
HOXB5	5.00E-06
XLOC_007207	5.00E-06
NBPF3	5.10E-06
LOC100507430	5.20E-06

SLCO4A1	5.20E-06
TCFL1	5.20E-06
XLOC_003540	5.20E-06
KRTAP10-7	5.30E-06
SEPP1	5.40E-06
XLOC_L2_015893	5.40E-06
TMEM136	5.50E-06
XLOC_008024	5.70E-06
LONP	5.80E-06
FLT1	5.90E-06
REA	5.90E-06
LYN	6.00E-06
ZC2HC1A	6.00E-06
C10orf90	6.10E-06
SH3TC2	6.20E-06
LOC100506035	6.30E-06
NCE2	6.30E-06
SSB3	6.40E-06
XLOC_011755	6.40E-06
XLOC_013968	6.40E-06
LOC100287063	6.60E-06
TBC1D22A	6.70E-06
XLOC_003069	6.70E-06
LOC138046	6.80E-06
RAB25	6.80E-06
ZNF306	6.80E-06
ALG8	7.00E-06
DNHD3	7.00E-06
LTF	7.10E-06
MGC19780	7.10E-06
XLOC_002487	7.10E-06
XLOC_006091	7.10E-06
KCNS2	7.30E-06
DNM2	7.40E-06
XLOC_007544	7.40E-06
LOC100652948	7.70E-06
KIAA1618	7.90E-06
XLOC_003963	7.90E-06
XLOC_L2_009441	8.00E-06
MIDN	8.20E-06
XLOC_003514	8.20E-06
C6orf188	8.30E-06
FLJ14299	8.30E-06
PDPN	8.30E-06

XLOC_L2_011566	8.30E-06
LOC100506732	8.50E-06
LOC284001	8.50E-06
XLOC_003976	8.80E-06
C1QTNF5	8.90E-06
LOC643277	8.90E-06
BASP1	9.10E-06
XLOC_000680	9.10E-06
LOC400662	9.20E-06
CALB2	9.30E-06
CYP11A1	9.50E-06
MIR221	9.50E-06
STRA13	9.50E-06
LOC100652760	9.70E-06
INDO	0.00001
SYPL	0.00001
ALDRL6	0.000011
ENPP1	0.000011
FSTL3	0.000011
NFIA	0.000011
SNORD116-19	0.000011
TMEM56-RWDD3	0.000011
XLOC_001153	0.000011
XLOC_002593	0.000011
XLOC_008760	0.000011
XLOC_L2_008288	0.000011
ZNF175	0.000011
ADCY7	0.000012
BCORL1	0.000012
FER1L5	0.000012
KRTHB3	0.000012
LOC100505951	0.000012
LOC100507458	0.000012
LOC100653003	0.000012
LOC151194	0.000012
XLOC_006731	0.000012
APOBEC2	0.000013
CEBPA	0.000013
GPATC2	0.000013
ITIH3	0.000013
TMEM101	0.000013
XLOC_L2_014518	0.000013
CCBP2	0.000014
ERO1L	0.000014

KCNF1	0.000014
LOC100653120	0.000014
RNF184	0.000014
TSPAN12	0.000014
XLOC_004881	0.000014
XLOC_010456	0.000014
ANG	0.000015
CEP44	0.000015
FBN2	0.000015
OR7D2	0.000015
TBC1D10	0.000015
VNN3	0.000015
XLOC_003014	0.000015
XLOC_004074	0.000015
XLOC_011264	0.000015
XLOC_L2_013506	0.000015
C10orf70	0.000016
FGF9	0.000016
KL	0.000016
RLF	0.000016
XLOC_L2_000864	0.000016
CALM1	0.000017
DKFZP566O084	0.000017
ELF3	0.000017
F13A1	0.000017
LINC00327	0.000017
LOC100506713	0.000017
OPRK1	0.000017
Q6UXP8	0.000017
XLOC_004526	0.000017
XLOC_008111	0.000017
ZCSL2	0.000017
GRCC9	0.000018
LOC126731	0.000018
MGC29649	0.000018
PROCR	0.000018
TSRC1	0.000018
ZNF114	0.000018
ZNF682	0.000018
ASAP1-IT1	0.000019
NEXN	0.000019
PARD6A	0.000019
REC14	0.000019
TMEM164	0.000019

UBE2Q2P2	0.000019
XLOC_001323	0.000019
XLOC_009230	0.000019
XLOC_L2_013868	0.000019
ADPRM	0.00002
SERTAD4	0.00002
XLOC_005290	0.00002
CLEC18B	0.000021
DERL3	0.000021
XLOC_005388	0.000021
CEP85	0.000022
TMEM28	0.000022
XLOC_011682	0.000022
ZA20D2	0.000022
ARL7	0.000023
NDRG1	0.000023
IMAGE3451454	0.000024
XLOC_L2_014931	0.000024
LY6D	0.000025
MIGA2	0.000025
XLOC_013436	0.000025
LEPREL1	0.000026
LOC100129721	0.000026
LOC100507161	0.000026
LOC100652854	0.000026
LOC100653008	0.000026
PPP1R1C	0.000026
RSC1A1	0.000026
SP9	0.000026
XLOC_001747	0.000026
XLOC_003843	0.000026
XLOC_L2_009736	0.000026
HBXAP	0.000027
LINC00493	0.000027
SLC20A1	0.000027
SMARCA1	0.000027
GFOD2	0.000028
RIMS3	0.000028
TMEM248	0.000028
XLOC_000229	0.000028
XLOC_009023	0.000028
XLOC_009709	0.000028
XLOC_L2_008195	0.000028
FLJ43934	0.000029

KIF21B	0.000029
KRT25B	0.000029
XLOC_L2_010118	0.000029
XLOC_L2_010854	0.000029
XLOC_L2_015295	0.000029
ABCA12	0.00003
EIF2A	0.00003
KIAA0063	0.00003
LOC155435	0.00003
PDXP	0.00003
LINC00173	0.000031
OCRL	0.000031
WDR10	0.000031
PRDX6	0.000031
ANKRD36BP1	0.000032
CADM3	0.000032
EBI3	0.000032
FBXO22-AS1	0.000032
PTBP2	0.000032
TMEM141	0.000032
TUBA6	0.000032
C20orf13	0.000033
FLJ25449	0.000033
KIAA0882	0.000033
CD8B1	0.000034
FAM181B	0.000034
SVBP	0.000034
DKFZP434O0213	0.000035
PDE7B	0.000035
PP	0.000035
ROD1	0.000035
XLOC_L2_002281	0.000035
LOC126295	0.000036
LOC285954	0.000036
MAD2L2	0.000036
XLOC_009228	0.000036
FLJ31951	0.000037
GPR126	0.000037
LOC199725	0.000037
FLJ22938	0.000038
HTR3A	0.000038
PLAG1	0.000038
Q46FH2	0.000038
DUSP5P	0.000039

ZFAND4	0.000039
LOC100507175	0.00004
MAFF	0.00004
LGALSL	0.000041
SLC9B1	0.000041
XLOC_011216	0.000041
MGC5306	0.000042
Q6JHZ7	0.000043
ABCA8	0.000044
LOC128977	0.000044
DKFZP761C169	0.000045
LINC00111	0.000045
LOC145474	0.000045
XLOC_009445	0.000045
DKFZP686A10121	0.000046
NR2F1	0.000046
XLOC_013658	0.000046
LOC114926	0.000047
LOC340351	0.000047
GIP3	0.000048
LOC163223	0.000048
Q58MP6	0.000048
SOD1	0.000048
XLOC_006439	0.000048
ASTE1	0.000049
DNCI2	0.000049
INHA	0.000049
FANCE	0.00005
GNPTG	0.00005
ACOXL	0.000051
ASPH	0.000051
CENPT	0.000051
HT001	0.000051
ZNF185	0.000051
LOC100506720	0.000053
LOC389168	0.000053
MOBK12C	0.000053
XLOC_001338	0.000053
XLOC_003508	0.000053
XLOC_004527	0.000053
XLOC_006310	0.000053
XLOC_012253	0.000053
MOB3A	0.000054
XLOC_014396	0.000055

GPANK1	0.000056
LENG4	0.000056
C6orf83	0.000057
FLJ34960	0.000057
ZNF406	0.000057
LOC100505669	0.000059
NEK11	0.000059
PRP19	0.000059
XLOC_000676	0.00006
FLJ42351	0.000061
SDF2L1	0.000061
XLOC_000269	0.000061
DHX57	0.000062
HPF1	0.000062
LOC100289038	0.000062
NPFFR2	0.000063
LOC92912	0.000064
MGC2574	0.000064
XLOC_008349	0.000064
SH3BP2	0.000066
IPLA2GAMMA	0.000067
XLOC_L2_005874	0.000067
C3orf9	0.000068
GPR124	0.000068
LOC387882	0.000068
CXCL14	0.000069
MGC27466	0.000069
CKLFSF7	0.00007
DKFZP434H2226	0.00007
FLJ16641	0.00007
MGC24665	0.00007
CAPN7	0.000072
CLC	0.000072
DARC	0.000072
FAM101B	0.000072
SEC24B-AS1	0.000073
XLOC_000138	0.000073
XLOC_004032	0.000073
ADCK1	0.000074
DDA1	0.000074
HSPC182	0.000074
MS4A4A	0.000074
SSFA2	0.000074
XLOC_002733	0.000074

CCDC106	0.000075
BMP5	0.000076
MGC47869	0.000076
HSCARG	0.000077
OSR1	0.000077
RASSF4	0.000077
NDP52	0.000078
TMEM4	0.000078
TTY9A	0.000078
C6orf166	0.000079
LOC440353	0.000079
C8orf58	0.000081
GPT2	0.000081
SMIM3	0.000081
XLOC_001160	0.000081
COL1A2	0.000082
LOC253982	0.000082
LOC729867	0.000082
XLOC_004420	0.000082
FLJ10781	0.000083
LAGE3	0.000084
LOC389607	0.000084
MGC13138	0.000084
PP12613	0.000084
ALPPL2	0.000085
MGC29898	0.000085
MMP11	0.000085
ZNF694	0.000085
XLOC_000625	0.000086
JM11	0.000087
XLOC_006976	0.000087
EFHD1	0.000089
NNAT	0.000089
PDZK6	0.000089
XLOC_010271	0.000089
PROKR2	0.00009
ULK1	0.00009
DAZL	0.000091
LOC100130442	0.000091
LOC100653149	0.000091
LOC492304	0.000091
SPTLC2L	0.000091
LTBP2	0.000092
SLC22A31	0.000092

XLOC_007286	0.000092
HPS3	0.000093
METTL15	0.000093
BCDO2	0.000094
DLX5	0.000094
NALP9	0.000094
LOC100289627	0.000095
LOC100505856	0.000095
XLOC_007895	0.000095
XLOC_008698	0.000095
XLOC_012868	0.000095
ARHGAP17	0.000096
LOC100508869	0.000096
MAP4K3	0.000096
NUDT12	0.000097
PXK	0.000097
SSR4	0.000097
ENOPH1	0.000098
USB1	0.000098
LOC647911	0.000099
CALML4	0.0001
LOC100652791	0.0001
C1orf10	0.000101
CTSE	0.000101
FLJ10420	0.000101
XLOC_L2_007147	0.000101
ZNF593	0.000101
A2M	0.000102
CNNM4	0.000102
LOC84661	0.000102
BCAT2	0.000103
POLR2F	0.000103
XLOC_004600	0.000103
MOB3C	0.000104
TSPAN33	0.000104
FLJ12760	0.000105
XLOC_L2_000773	0.000105
LOC100506708	0.000106
SLC25A4	0.000106
XLOC_009684	0.000106
LOC100507673	0.000108
MGC33867	0.000108
TMEM38B	0.000108
AHI1	0.000109

MGC14839	0.000109
XLOC_007995	0.000109
XLOC_013274	0.00011
HIC	0.000111
XLOC_011052	0.000112
RGL1	0.000114
FLJ12517	0.000115
JOSD1	0.000115
NOR1	0.000116
SDSL	0.000117
SNORD86	0.000117
AQP1	0.000118
LOC100652915	0.000118
FLJ12875	0.000119
FAM120AOS	0.00012
FAM121A	0.00012
LOC100506073	0.00012
XLOC_005180	0.00012
DEGS1	0.000121
TRAF5	0.000121
KIAA1613	0.000122
XLOC_002682	0.000122
FLJ20920	0.000123
MTERF3	0.000123
XLOC_008152	0.000123
CKS2	0.000124
DOC2GP	0.000124
XLOC_005413	0.000124
XLOC_008610	0.000125
ARRDC3	0.000126
XLOC_005238	0.000126
C20orf178	0.000127
KRT1	0.000127
XLOC_013047	0.000127
XLOC_L2_006013	0.000127
TMEM134	0.000128
XLOC_L2_008689	0.000128
CA4	0.000129
LOC100652849	0.000129
LOC100506183	0.00013
LOC100509541	0.00013
BBIP1	0.000132
HS3ST3A1	0.000133
LOC100133089	0.000133

XLOC_001973	0.000133
LRR1	0.000134
LOC100526771	0.000135
XLOC_000566	0.000135
XLOC_013981	0.000135
TXNL6	0.000136
XLOC_006596	0.000136
HTRA1	0.000137
XLOC_012343	0.000137
LOC284723	0.000138
HIS1	0.000139
ZNF84	0.000139
GAPDH	0.00014
MGC4171	0.00014
XLOC_012831	0.00014
LOC100506307	0.000141
FLJ38281	0.000143
XLOC_001414	0.000143
EVL	0.000144
MAST4	0.000144
XLOC_000569	0.000145
XLOC_L2_003234	0.000145
LOC54499	0.000146
LOC100652878	0.000148
LOC646743	0.000148
XLOC_014071	0.000148
C10orf117	0.000149
XLOC_003286	0.000149
CST6	0.000151
GTP	0.000152
LOC439911	0.000153
PARP6	0.000153
XLOC_000291	0.000154
CXX1	0.000155
KIAA0863	0.000156
LOC100507127	0.000157
RGS19IP1	0.000157
XLOC_013011	0.000157
BNIP3L	0.000158
PSG11	0.000158
LOC100507047	0.000159
MAF	0.000159