



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ**

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**

**ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ  
ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ**

**Μετα-ανάλυση δεδομένων γονιδιακής έκφρασης από  
μικροσυστοιχίες DNA με εφαρμογή στον Διαβήτη**

**Παναγιωτοπούλου Μάρθα**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**Υπεύθυνος**

**Παντελής Μπάγκος**

**Λαμία, 2021**





**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ**  
**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**  
**ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ**  
**ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ**

**Μετα-ανάλυση δεδομένων γονιδιακής έκφρασης  
από μικροσυστοιχίες DNA με εφαρμογή στον  
Διαβήτη**

**Παναγιωτοπούλου Μάρθα**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**Επιβλέπων**

**Παντελής Μπάγκος**

**Λαμία, 2021**

Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις <sup>(1)</sup>, που προβλέπονται από της διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:

1. Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί **χωρίς να τα περικλείω σε εισαγωγικά** και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί παράθεση χωρίς εισαγωγικά χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάστηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.
2. Δέχομαι ότι η αυτολεξεί **παράθεση χωρίς εισαγωγικά**, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.
3. Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια
4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.

Ημερομηνία: ...../...../20.....

Ο – Η Δηλ.

(Υπογραφή)

(1) «Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση του άρθρου 8 παρ. 4 Ν. 1599/1986 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών.

Μετα-ανάλυση δεδομένων γονιδιακής έκφρασης από μικροσυστοιχίες DNA με εφαρμογή στον διαβήτη

Παναγιωτοπούλου Μάρθα

### **Τριμελής Συμβουλευτική Επιτροπή:**

Παντελής Μπάγκος (Επιβλέπων Καθηγητής), Καθηγητής, Τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας

Χατζηγεωργίου Άρτεμις, Καθηγήτρια, Τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας

Μπράλιου Γεωργία, Επίκουρος Καθηγήτρια, Τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας



## Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε στα πλαίσια του προπτυχιακού κύκλου σπουδών του Τμήματος Πληροφορικής με Εφαρμογές στη Βιοϊατρική της Σχολής Θετικών Επιστημών του Πανεπιστημίου Θεσσαλίας κατά τη διάρκεια του ακαδημαϊκού έτους 2020-2021, υπό την επίβλεψη του κυρίου Παντελή Μπάγκου, Καθηγητή του Τμήματος Πληροφορικής με Εφαρμογές στη Βιοϊατρική του Πανεπιστημίου Θεσσαλίας.

Θα ήθελα να ευχαριστήσω θερμά τον Καθηγητή του τμήματος Πληροφορικής με Εφαρμογές στη Βιοϊατρική κ. Παντελή Μπάγκο για την ευκαιρία που μου έδωσε να εκπονήσω το συγκεκριμένο αντικείμενο έρευνας υπό την επίβλεψή του, τη συνεχόμενη καθοδήγηση αλλά και τη αμέριστη βοήθεια του καθ' όλη την διάρκεια εκπόνησης της εργασίας.

Επίσης, θα ήθελα να εκφράσω τις ευχαριστίες μου στην μεταδιδακτορική ερευνήτρια κα. Παναγιώτα Κοντού για την καθοδήγηση αλλά και συνεχή συνεργασία και τον έλεγχο των πειραματικών αποτελεσμάτων σε όλα τα στάδια της εργασίας.

Τέλος θα ήθελα να ευχαριστήσω την μεταδιδακτορική ερευνήτρια κα. Κωνσταντίνα Βέννου για την εξαιρετική συνεργασία μας και για το πνεύμα ομαδικότητας που την διακρίνει, καθώς επίσης για τον χρόνο που προσέφερε κάθε φορά που της ζητούταν.

Η συνεργασία με όλους τους υπεύθυνους για αυτή την εργασία ήταν εξαιρετική και πολύ βοηθητική, καθώς δεν θα τα είχα καταφέρει χωρίς αυτή.





## Περίληψη

Στη παρούσα πτυχιακή εργασία, πραγματοποιήθηκε μετα-ανάλυση γονιδίων, με τη χρήση του στατιστικού πακέτου STATA, για δεδομένα της ασθένειας του σακχαρώδη διαβήτη τύπου II, με στόχο τον εντοπισμό γονιδίων, τα οποία εμφανίζουν διαφορετική έκφραση στη νόσο. Ο μηχανισμός πρόκλησης της διαταραχής είναι μία βλάβη στην παραγωγή και την λειτουργία της ινσουλίνης. Με άλλα λόγια, χαρακτηρίζεται από το συνδυασμό μειωμένης παραγωγής ινσουλίνης και μειωμένης ευαισθησίας των κυττάρων στη δράση της.

Στην συγκεκριμένη εργασία έγινε μια προσπάθεια εντοπισμού γενετικών παραγόντων οι οποίοι οδηγούν στην εμφάνιση διαβήτη τύπου II μέσω της στατιστικής μεθόδου της μετα-ανάλυσης πειραματικών δεδομένων μικροσυστοιχιών. Η έρευνα πραγματοποιήθηκε σε τρεις βάσεις δεδομένων την GEO, την ArrayExpress και την Pubmed και με την χρήση του STATA βρέθηκαν τα γονίδια τα οποία ήταν στατιστικώς σημαντικά διαφορικά εκφραζόμενα.

Κατά τη περάτωση της μετα-ανάλυσης προέκυψαν τα στατιστικώς σημαντικά γονίδια στα οποία εφαρμόστηκαν μέθοδοι διόρθωσης των  $p$ -value προκειμένου να ενισχυθούν τα ευρήματα της έρευνας (διόρθωση FDR, Bonferroni, Holm, Holland, Sidak). Η συμμετοχή σε μεταβολικές δραστηριότητες φανερώνουν την ύπαρξη προδιάθεσης για εμπλοκή στη νόσο. Περαιτέρω εργαλεία, όπως αυτά της PANTHER και της STRING, χρησιμοποιήθηκαν για να μελετηθούν τα μοριακά και βιοχημικά χαρακτηριστικά των επιλεγμένων γονιδίων αλλά και οι τυχόν αλληλεπιδράσεις μεταξύ τους.

Αν και τα αποτελέσματα δίνουν μία «κατεύθυνση», δεν είναι ισχυρά ώστε να ισχυριστούμε την συσχέτιση των γονιδίων αυτών με τη νόσο. Χρειάζεται περαιτέρω ανάλυσή τους ώστε να τεκμηριωθεί και να αξιολογηθεί η συσχέτισή τους τόσο με την εκδήλωση του σακχαρώδη διαβήτη τύπου II όσο και με άλλες μεταβολικές ασθένειες.

Λέξεις κλειδιά: Μικροσυστοιχίες, Σακχαρώδης Διαβήτης, μετα-ανάλυση, διαφορική έκφραση γονιδίων

## Abstract

In this dissertation, gene meta-analysis was performed, with the use of the statistical software package STATA, of type II diabetes mellitus data, with the aim of identifying genes that are differentially expressed in the disease. The cause of the disorder is a defect in the production and function of insulin. In other words, it is characterized by a combination of reduced insulin production and reduced sensitivity of cells to its action.

This study was an attempt to identify genetic factors that lead to the occurrence of type II diabetes mellitus through the statistical method of meta-analysis of gene expression data from microarray experiment. The research was conducted in three databases GEO, ArrayExpress and Pubmed and using STATA genes that were differentially expressed were identified.

From these results statistical correction method applied in the  $p$ -values in order to strengthen the findings of the study (FDR correction, Bonferroni, Holm, Holland, Sidak). Participation in metabolic activities reveals a predisposition to involvement in the disease. Further tools, such as PANTHER and STRING were used to understand the molecular and biochemical characteristics of the differentially expressed genes and any interactions between them.

Although the results give a "direction", they are not enough to claim the association of these genes with the disease. Further analysis is needed to evaluate the association with both the manifestation of diabetes as well as with other metabolic diseases.

Keywords: Microarrays, Diabetes Mellitus, meta-analysis, differentially expressed genes



## Περιεχόμενα

Ευχαριστίες.....	7
Πρόλογος.....	9
Abstract.....	10
<b>Κεφάλαιο 1 : Θεωρητικό υπόβαθρο.....</b>	<b>15</b>
1.1 Σακχαρώδης διαβήτης : Ορισμός και πειραφή.....	16
1.2 Τύποι σακχαρώδη διαβήτη.....	17
1.2.1 Σακχαρώδης διαβήτης τύπου I.....	18
1.2.2 Σακχαρώδης διαβήτης τύπου II.....	18
1.2.3 Σακχαρώδης διαβήτης κήσης.....	19
1.3 Επιπλοκές.....	20
1.3.1 Οξείες επιπλοκές .....	21
1.3.2 Χρόνιες επιπλοκές.....	21
1.4 Μικροσυστοιχίες : Ορισμός και γενικά στοιχεία.....	22
1.4.1 Κατηγορίες μικροσυστοιχιών.....	23
1.4.2 Βασικά βήματα για πείραμα με μικροσυστοιχίες.....	24
1.4.3 Βάσεις δεδομένων μικροσυστοιχιών.....	26
1.5 Στατιστική ανάλυση μικροσυστοιχιών.....	27
1.5.1 Μέθοδοι στατιστικής ανάλυσης.....	27
1.5.2 Διόρθωση.....	28
1.5.2.1 Μέθοδος Bonferroni.....	29
1.5.2.2 Μέθοδος Holm.....	29
1.5.2.3 Μέθοδος Simes.....	30
1.5.2.4 Μέθοδος Sidak.....	30
1.5.2.5 Μέθοδος Holland.....	30
1.6 Μετα-ανάλυση : Ορισμός και χρησιμότητα.....	31
1.6.1 Μονομεταβλητή μετα-ανάλυση.....	32
1.6.2 Προβλήματα κατά την διεξαγωγή μετα-ανάλυσης.....	32
1.6.3 Στατιστικά εργαλεία στη μετα-ανάλυση.....	32
1.6.4 Μετα-ανάλυση και μικροσυστοιχίες.....	33
1.6.4.1 Μοντέλο των σταθερών επιδράσεων (fixed effects model).....	35
1.6.4.2 Μοντέλο των τυχαίων επιδράσεων (random effects model).....	35

1.6.4.3 Έλεγχος ετερογένειας .....	36
1.6.4 IDR και IRR.....	38
1.6.5 Ανάλυση εμπλουτισμού.....	39
1.6.5.1 PANTHER.....	39
1.6.5.2 STRING.....	40
1.7 Σκοπός της παρούσας πτυχιακής εργασίας.....	40
<b>Κεφάλαιο 2 : Υλικά και μέθοδοι.....</b>	<b>41</b>
2.1 Ερευνητικό ερώτημα.....	42
2.2 Συλλογή και καταγραφή δεδομένων.....	42
2.2.1 Συλλογή δεδομένων.....	42
2.2.2 Καταγραφή δεδομένων.....	46
2.2.3 Ονοματολογία δεδομένων.....	46
2.3 Επεξεργασία δεδομένων.....	47
2.3.1 Η εντολή «collapse».....	47
2.3.2 Επιπλοκές .....	49
2.3.3 Συνοπτικά .....	50
2.4 Στατιστική ανάλυση των δεδομένων.....	51
2.4.1 Μετα-ανάλυση.....	51
2.4.2 Μέθοδοι διόρθωσης πολλαπλών μεταβλητών.....	52
<b>Κεφάλαιο 3 : Αποτελέσματα.....</b>	<b>53</b>
3.1 Συνολικά Αποτελέσματα.....	54
3.2 Αποτελέσματα της μετα-ανάλυσης στον λιπώδη ιστό.....	56
3.3 Αποτελέσματα της μετα-ανάλυσης στο ήπαρ.....	58
3.4 Αποτελέσματα της μετα-ανάλυσης στα παγκρεατικά νησίδια.....	59
3.5 Αποτελέσματα της μετα-ανάλυσης στα μονοκύτταρα του περιφερικού αίματος.....	65
3.6 Αποτελέσματα της μετα-ανάλυσης στο σκελετικό μυ.....	70
<b>Κεφάλαιο 4: Συζήτηση.....</b>	<b>76</b>
<b>Βιβλιογραφία.....</b>	<b>78</b>
<b>Παράρτημα 1.....</b>	<b>92</b>
<b>Παράρτημα 2.....</b>	<b>101</b>



## **Κεφάλαιο 1 : Θεωρητικό υπόβαθρο**

## 1.1 Σακχαρώδης διαβήτης : Ορισμός και περιγραφή

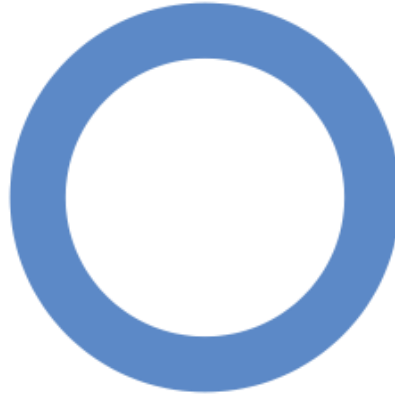
Ο σακχαρώδης διαβήτης είναι μια χρόνια ασθένεια και ορίζεται ως σύνολο μεταβολικών παθήσεων που χαρακτηρίζονται από υπεργλυκαιμία (αύξηση της συγκέντρωσης του σακχάρου στο αίμα), η οποία είναι συνέπεια διαταραχών στην έκκριση ή στην δράση της ινσουλίνης (ορμόνη ελέγχου του σακχάρου στο αίμα, η οποία παράγεται από τα κύτταρα β των νησιδίων του Langerhans του παγκρέατος). Μπορεί να εμφανιστεί είτε όταν το πάγκρεας δεν παράγει αρκετή ινσουλίνη είτε όταν ο ανθρώπινος οργανισμός δεν μπορεί να χρησιμοποιήσει την ινσουλίνη που παράγει [Ο Γιατρός μου, 2002].

Ο διαβήτης αποτελεί ένα σημαντικό πρόβλημα υγείας, το οποίο απασχολεί ένα πολύ μεγάλο αριθμό ατόμων. Τόσο ο αριθμός των περιπτώσεων που έχουν καταγραφεί, όσο και ο επιπολασμός του σακχαρώδη διαβήτη έχουν αυξηθεί κατά τη διάρκεια των τελευταίων δεκαετιών. Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας, σε παγκόσμιο επίπεδο, η υψηλή γλυκόζη αίματος σκοτώνει περίπου 3.4 εκατομμύρια ανθρώπους ετησίως και προβλέπεται ο αριθμός αυτός να διπλασιαστεί μέχρι το 2030 [WHO/Europe, 2016]. Στην Ελλάδα εκτιμάται ότι το ποσοστό του πάσχοντος πληθυσμού ανέρχεται στο 10-12% και έχει αναγνωριστεί ως μία από τις κύριες αιτίες θανάτου παγκοσμίως.

Η χρόνια υπεργλυκαιμία προκαλεί μία σειρά σοβαρών επιπλοκών στον οργανισμό, δυσλειτουργίες και ανεπάρκεια διάφορων οργάνων. Ορισμένες επιπλοκές είναι: καρδιαγγειακή νόσος, διαβητική νεφροπάθεια, βλάβες στα νεύρα, στον αμφιβληστροειδή χιτώνα του ματιού, στυτική δυσλειτουργία κ.α [Υπουργείο Υγείας, 2012]. Ακόμα όμως και αν ένας άνθρωπος δεν παρουσιάζει διαβήτη, δε συμφέρει ότι δεν μπορεί να σχετίζεται άμεσα με την μελλοντική ή όχι εμφάνιση του. Αυτό συμβαίνει διότι μπορεί να έχει ευαισθησία στην ινσουλίνη ή αντίσταση στην ινσουλίνη (insulin sensitivity vs insulin resistance).

Όταν ένας άνθρωπος παρουσιάσει αντίσταση στην ινσουλίνη τότε τα κύτταρα του σταματούν να ανταποκρίνονται στην ορμόνη της ινσουλίνης, με αποτέλεσμα και η ορμόνη αλλά και το σάκχαρο του αίματος να βρίσκονται σε υψηλά επίπεδα στον οργανισμό. Αυτή η κατάσταση μπορεί να οδηγήσει πολύ γρήγορα σε διαβήτη τύπου II. Ακριβώς στην αντίθετη κατεύθυνση βρίσκεται η ευαισθησία στην ορμόνη ινσουλίνη, όπου τα κύτταρα χρησιμοποιούν την ορμόνη με τον καλύτερο δυνατό τρόπο και πιο αποτελεσματικά.

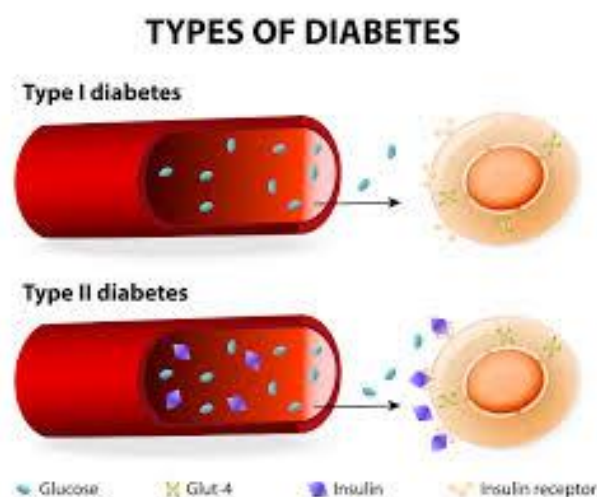




Εικόνα 1: Σύμβολο του διαβήτη, πηγή: <https://www.symbols.com/symbol/blue-circle-for-diabetes>

## 1.2 Τύποι σακχαρώδη διαβήτη

Τα αίτια, η συχνότητα και τα χαρακτηριστικά των ασθενών ποικίλλουν ανάλογα με τον τύπο του σακχαρώδους διαβήτη που εμφανίζεται. Ο σακχαρώδης διαβήτης αποτελείται από τρεις κύριους τύπους και κάποιους άλλους ειδικούς. Οι τρεις κύριοι τύποι είναι: ο διαβήτης τύπου I, ο διαβήτης τύπου II και ο διαβήτης της κήσης. Ενδεικτικά, ορισμένοι από τους ειδικούς τύπους διαβήτη είναι: ο διαβήτης λόγω γενετικών λειτουργικών διαταραχών των β-κυττάρων, ο διαβήτης λόγω γενετικών διαταραχών δράσης της ινσουλίνης, ο διαβήτης λόγω παθήσεων της εξωκρινούς μοίρας του παγκρέατος, ο διαβήτης λόγω ενδοκρινοπαθειών, φαρμακευτικός ή χημικός σακχαρώδης διαβήτης, ο διαβήτης λόγω κάποιων λοιμώξεων, κάποιες σπάνιες μορφές αυτοάνοσου σακχαρώδη διαβήτη και άλλα γενετικά σύνδρομα σχετιζόμενα με σακχαρώδη διαβήτη [Ο Γιατρός μου, 2002].



Εικόνα 2: Κατάσταση των κυττάρων στον σακχαρώδη διαβήτη, πηγή: <https://www.getdoc.com/5-types-of-diabetes/>

### **1.2.1 Σακχαρώδης διαβήτης τύπου I**

Ο σακχαρώδης διαβήτης τύπου I, παλιότερα αποκαλούμενος ως ινσουλινοεξαρτώμενος ή νεανικός-παιδικός διαβήτης, προκαλείται από καταστροφή των β-κυττάρων του παγκρέατος, η οποία οδηγεί σε πλήρη ανεπάρκεια της ινσουλίνης. Διακρίνεται σε αυτοάνοσο και ιδιοπαθή σακχαρώδη διαβήτη [Cooke D., 2008]. Ο διαβήτης τύπου I είναι η κυριότερη αιτία διαβήτη στα παιδιά, όμως μπορεί να προσβάλλει και ενήλικες.

Ο αυτοάνοσος σακχαρώδης διαβήτης αφορά μόνο το 5-10% των ασθενών με διαβήτη και προκαλείται από αυτοάνοση καταστροφή των β-κυττάρων. Η συχνότητα καταστροφής των κυττάρων αυτών μπορεί να ποικίλλει από ταχεία μέχρι βραδεία, ταχεία ιδίως σε νηπιακή και παιδική ηλικία και βραδεία ιδίως στους ενήλικες αντίστοιχα. Η καταστροφή αυτή των κυττάρων αποδίδεται σε πολλούς γενετικούς αλλά και περιβαλλοντικούς παράγοντες. Στον ιδιοπαθή διαβήτη υπάρχει μεγάλος βαθμός κληρονομικότητας και η ανάγκη για εξωγενή χορήγηση ινσουλίνης σε αυτούς τους ασθενείς ποικίλλει.

Οι ασθενείς με διαβήτη τύπου I εξαρτώνται απόλυτα από την εξωγενή χορήγηση ινσουλίνης προκειμένου να μπορέσουν να διατηρηθούν σε φυσιολογικά επίπεδα τα επίπεδα σακχάρου στο αίμα [Ο Γιατρός μου, 2002]. Οι μέθοδοι χορήγησης της ινσουλίνης είναι: η ένεση, η οποία είναι η πιο κοινή μέθοδος, οι αντλίες ινσουλίνης και η εισπνεόμενη ινσουλίνη. Επίσης, μπορεί να πραγματοποιηθεί μεταμόσχευση παγκρέατος ή μεταμόσχευση κυττάρων νησιδίων του παγκρέατος. Στα συμπτώματα του σακχαρώδη διαβήτη τύπου I περιλαμβάνονται η συχνή ούρηση, η αυξημένη δίψα, η ξερροστομία, η αυξημένη όρεξη, η κόπωση και η απώλεια βάρους [Cooke D., 2008].

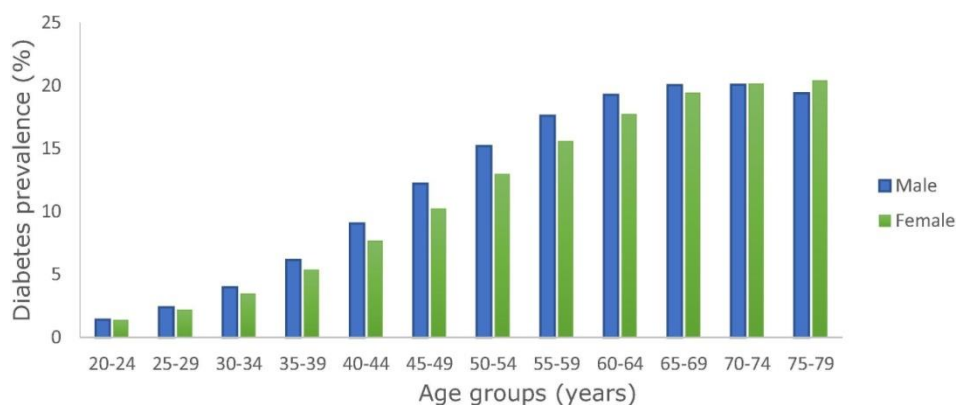
### **1.2.2 Σακχαρώδης διαβήτης τύπου II**

Ο σακχαρώδης διαβήτης τύπου II είναι η κυριότερη αιτία διαβήτη στους ενήλικες. Σε αυτή την περίπτωση διαβήτη, ο μηχανισμός πρόκλησης της διαταραχής είναι μία βλάβη στην παραγωγή και την λειτουργία της ινσουλίνης, γι' αυτό και η ασθένεια εκδηλώνεται σε πιο προχωρημένες ηλικίες (40-60 χρόνων). Με άλλα λόγια, χαρακτηρίζεται από το συνδυασμό μειωμένης παραγωγής ινσουλίνης και μειωμένης ευαισθησίας των κυττάρων στη δράση της. Οι κύριοι παράγοντες προδιάθεσης της εμφάνισης διαβήτη τύπου II είναι η παχυσαρκία, η ηλικία, το οικογενειακό ιστορικό αλλά και ορισμένες φορές το φύλο (κυρίως το γυναικείο) [Ο Γιατρός μου, 2002].

Σε παγκόσμια κλίμακα, το 2010 εκτιμήθηκε ότι το 90% των περιπτώσεων διαβήτη οφειλόταν σε σακχαρώδη διαβήτη τύπου II, ποσοστό που ανέρχεται σε 285 εκατομμύρια ανθρώπους [WHO, 2016]. Ισοδύναμα, αυτό σημαίνει ότι το 6% του παγκόσμιου πληθυσμού πάσχει από διαβήτη τύπου II. Ο διαβήτης εμφανίζεται τόσο στον αναπτυγμένο όσο και στον αναπτυσσόμενο κόσμο. Μεγαλύτερο κίνδυνο φαίνεται να διατρέχουν οι γυναίκες αλλά και

μερικές εθνοτικές ομάδες όπως οι Αμερικάνοι Ιθαγενείς, οι Λατίνοι και οι Νότιο-Ασιάτες κάτοικοι των Νήσων του Ειρηνικού Ωκεανού [Abate N., Chandalia M., 2001]. Κατά παράδοση, ο διαβήτης τύπου II θεωρείται μία ασθένεια που πλήττει του ενήλικες, ωστόσο έχει παρατηρηθεί ότι οι περιπτώσεις παιδιών με σακχαρώδη διαβήτη τύπου II έχουν αυξητική τάση παράλληλα με την αύξηση της παιδικής παχυσαρκίας.

Σε αντίθεση με τον διαβήτη τύπου I, τα συμπτώματα εμφανίζονται σταδιακά, σε διάστημα μηνών ή χρόνων και είναι πιο ελαφριάς μορφής, αυτός είναι και ο λόγος που πολλές φορές μπορεί να περάσουν και απαρατήρητα. Ένα χαρακτηριστικό παράδειγμα είναι η απώλεια βάρους, καθώς, όπως προαναφέρθηκε, αυτό το είδος διαβήτη εμφανίζεται σε παχύσαρκα άτομα. Μερικά κοινά συμπτώματα με τον διαβήτη τύπου I είναι: η πολουρία, η πολυδιψία και η πολυφαγία. Πέραν όμως αυτών, οι ασθενείς με διαβήτη τύπου II μπορεί να παρουσιάσουν ιστορικό θολής όρασης, επαναλαμβανόμενα επεισόδια κοιλιακών μολύνσεων, φαρούρα, περιφερική νευροπάθεια και κόπωση. Στα πρώτα στάδια της νόσου, αλλά συχνά και καθ' όλη την διάρκεια της ζωής τους, οι ασθενείς δεν χρήζουν ινσουλινοθεραπείας για την επιβίωση τους [Ο Γιατρός μου, 2002].



Εικόνα 3: Ποσοστό του σακχαρώδους διαβήτη τύπου II σε άντρες και γυναίκες, πηγή: [https://www.diabetesresearchclinicalpractice.com/article/S0168-8227\(19\)31230-6/fulltext](https://www.diabetesresearchclinicalpractice.com/article/S0168-8227(19)31230-6/fulltext)

### 1.2.3 Σακχαρώδης διαβήτης κύησης

Κατά την διάρκεια της εγκυμοσύνης, προκαλείται πλήθος μεταβολικών αλλαγών, οι οποίες μπορεί να προκαλέσουν κατάσταση υπεργλυκαιμίας, δηλαδή αυξημένης συγκέντρωσης γλυκόζης στο αίμα και έτσι να δωθεί το έναυσμα για πιθανή εμφάνιση διαβήτη. Όταν η διαταραχή αυτή εμφανιστεί κατά την διάρκεια της κύησης, τότε αποκαλείται διαβήτης κύησης. Σύμφωνα με τα στατιστικά, πρόκειται για ένα πολύ συχνό πρόβλημα, το οποίο ανιχνεύεται περίπου στο 10% των εγκύων γυναικών κατά το δεύτερο τρίμηνο της κύησης τους.

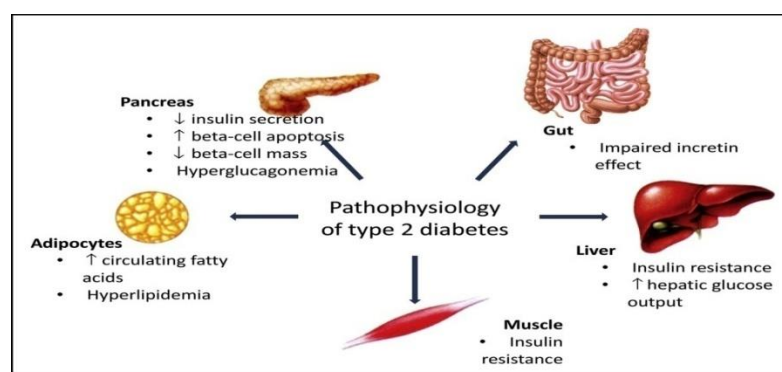
Τα συμπτώματα του διαβήτη της κύησης είναι σχεδόν ίδια με αυτά του σακχαρώδη διαβήτη. Ωστόσο, για στις εγκύους που το εμφανίζουν αυξάνεται η πιθανότητα προεκλαμψίας, κατάθλιψης και μη φυσιολογικού τοκετού (καισαρική). Προκαλείται από

ανεπαρκή παραγωγή ινσουλίνης στο πλαίσιο της αντίστασης στην ινσουλίνη. Ως παράγοντες κινδύνου ορίζονται το βάρος που ξεπερνά το φυσιολογικό, δηλαδή ΔΜΣ (Δείκτης Μάζας Σώματος) > 18.5-25.0, εμφάνιση διαβήτη σε προηγούμενη κύηση, το οικογενειακό ιστορικό και το σύνδρομο των πολυκυστικών ωοθηκών. Η διάγνωση γίνεται με εξετάσεις αίματος. Αν δεν υπάρξει αντιμετώπιση, υπάρχει πιθανότητα να οδηγήσει σε πρόωρο τοκετό ή προσβολή του εμβρύου [Ο Γιατρός μου, 2002].

Όσο αναφορά τα έμβρυα, αν ο διαβήτης δεν αντιμετωπιστεί σωστά, υπάρχει μεγάλος κίνδυνος εμφάνισης δυσμορφιών, παχυσαρκίας, γέννησης με μακροσωμία ή εμφάνιση ίκτερου. Επιπλέον, κατά τον τοκετό ενδέχεται να παρουσιαστούν επιπλοκές, οι οποίες μπορούν να θέσουν σε κίνδυνο την ζωή της μητέρας αλλά και του μωρού. Ευτυχώς, για τις περισσότερες γυναίκες αυτή η κατάσταση είναι προσωρινή και ο διαβήτης υποχωρεί μετά το πέρας της εγκυμοσύνης, όμως παραμένει ο κίνδυνος εμφάνισης διαβήτη τύπου 2 τα επόμενα 10 -20 χρόνια [Ο Γιατρός μου, 2002].

### 1.3 Επιπλοκές

Όταν κάποιος νοσεί με σακχαρώδη διαβήτη μπορεί να εμφανίσει επιπλοκές, των οποίων όμως η πιθανότητα εμφάνισης μπορεί να μειωθεί σημαντικά με σωστή ρύθμιση των τιμών σακχάρου στο αίμα. Οι επιπλοκές που εμφανίζονται κατηγοριοποιούνται ανάλογα με τον ρυθμό εμφάνισης τους σε οξείες και χρόνιες. Οι οξείες εκδηλώνονται απότομα και αποτελούν σοβαρό κίνδυνο θανάτου, ενώ οι χρόνιες αναπτύσσονται μακροπρόθεσμα, μερικές φορές ύστερα από χρόνια εξέλιξης και προκαλούν σοβαρή πτώση στην ποιότητα ζωής και μπορεί να αποβούν και θανατηφόρες. Μερικοί παράγοντες που συμβάλλουν στην ανάπτυξη επιπλοκών είναι τα αυξημένα επίπεδα χοληστερόλης, η αρτηριακή πίεση, η παχυσαρκία, το κάπνισμα και η καθιστική ζωή.



Εικόνα 4: Παθοφυσιολογία του διαβήτη τύπου II, πηγή <https://www.amjmed.com/article/S0002-9343%2813%2900485-3/fulltext>

### 1.3.1 Οξείες επιπλοκές

Οι οξείες επιπλοκές εμφανίζονται ως αποτέλεσμα μίας σοβαρής μεταβολικής ανισορροπίας και αποτελούν επείγοντα περιστατικά, τα οποία πρέπει να κουράρονται άμεσα ώστε να μην αποβαίνουν μοιραία. Οι τρεις κύριες οξείες επιπλοκές είναι η διαβητική κετοξέωση με διαβητικό κόμα, το υπερωσμωτικό μη κετωτικό κόμα και η υπογλυκαιμία.

Η διαβητική κετοξέωση είναι μία πολύ επικίνδυνη επιπλοκή και απαιτεί άμεση μεταφορά του ασθενούς στο νοσοκομείο. Εκδηλώνεται ξαφνικά στο διαβήτη τύπου I και μόνο σε συγκεκριμένες καταστάσεις στο διαβήτη τύπου II. Προκαλείται όταν ο οργανισμός δεν μπορεί να χρησιμοποιήσει την γλυκόζη ως πηγή ενέργειας και προσλαμβάνει ενέργεια μέσω των λιπών, μεταβάλλοντας έτσι σημαντικά το εσωτερικό του περιβάλλον. Ο ασθενής εμφανίζει αφυδάτωση, αναπνοή Kussmaul και απόπνοια ακετόνης. Συχνά εμφανίζεται κοιλιακό άλγος και σταδιακά μπορεί να εμφανιστεί πτώση του επιπέδου συνείδησης, υπνηλία, σύγχυση και τελικά κόμα.

Η δεύτερη οξεία επιπλοκή που αναφέραμε εμφανίζεται κυρίως σε ασθενείς με διαβήτη τύπου II. Τα συμπτώματα του υπερωσμωτικού μη κετωτικού κόματος είναι η αύξηση της ωσμωτικότητας του πλάσματος, τα πολύ υψηλά επίπεδα σακχάρου στο αίμα και απουσία κετοξέωσης. Η επιπλοκή αυτή μπορεί να προκληθεί από πληθώρα αιτιών. Μερικές από αυτές είναι η συχαιμία, το έμφραγμα του μυοκαρδίου, η παγκρεατίτιδα, η αιμορραγία από το γαστρεντερικό και η λήψη φαρμάκων. Οι ασθενείς εμφανίζουν σημάδια αφυδάτωσης, διαταραχές της συνείδησης αλλά και με επαρκείς ηλεκτρολύτες.

Τέλος, η υπογλυκαιμία προκαλείται όταν η συγκέντρωση της γλυκόζης του αίματος είναι χαμηλότερη των 50mg/100ml. Ουσιαστικά, αποτελεί επιπλοκή των ασθενών που κάνουν θεραπεία με ινσουλίνη. Μπορεί να εμφανιστεί ύστερα από λανθασμένα υπερβολική δοσολογία ινσουλίνης, έντονη σωματική άσκηση ή μειωμένη λήψη τροφής και κυρίως υδατανθράκων. Οι ασθενείς που παρουσιάζουν υπογλυκαιμία εμφανίζουν ευερεθιστότητα, κρύο ιδρώτα, θολή όραση, αυξημένους παλμούς και αίσθημα πείνας. Επίσης, μπορεί να προκληθεί απώλεια συνείδησης ή ακόμα και κόμα. Για την αντιμετώπιση αυτής της κατάστασης, απαιτείται άμεση πρόσληψη γλυκόζης από το στόμα και σε περίπτωση που ο ασθενής είναι αναισθητός απαιτείται η εφαρμογή μίας ενδομυϊκής ένεσης γλυκαγόνης, ορμόνης που αυξάνει τη γλυκαιμία [Ο Γιατρός μου, 2002].

### 1.3.2 Χρόνιες επιπλοκές

Σε προχωρημένες φάσεις της νόσου, όταν δεν υπάρχει έλεγχος κατά την διάρκεια αρκετών χρόνων και λόγω διαρκούς υπεργλυκαιμίας, ο σακχαρώδης διαβήτης προκαλεί πολλές επιπλοκές στον οργανισμό που προσβάλλουν πολλούς ιστούς και όργανα. Οι κυριότερες από τις χρόνιες ή αλλιώς καθυστερημένες επιπλοκές είναι η διαβητική μικροαγγειοπάθεια και η

διαβητική μακροαγγειοπάθεια και υπάρχουν και άλλες όπως ευπάθεια σε λοιμώξεις, υπερλιπιδαιμίες και περιοδοντίτιδα.

Η διαβητική μικροαγγειοπάθεια αποτελεί παθολογική κατάσταση που προσβάλλει τα μικρά τριχοειδή και προτριχοειδή αιμοφόρα αγγεία και προκαλεί διαταραχές στην αιμάτωση διάφορων ιστών. Μεταξύ των κύριων επιπτώσεών της περιλαμβάνεται η διαβητική αμφιβληστροειδοπάθεια, κατά την οποία υπάρχει υπερπλασία των τριχοειδών αγγείων στην επιφάνεια του αμφιβληστροειδούς χιτώνα του ματιού και προσβολή της ωχράς κηλίδας, που οδηγούν σε απώλεια όρασης. Επιπλέον, προσβάλλονται τα νεφρά και προκαλούνται βλάβες του σπειράματος, των αγγείων και του διάμεσου ιστού που μπορεί να οδηγήσουν σε νεφρική ανεπάρκεια, η κατάσταση αυτή αποκαλείται διαβητική νεφροπάθεια. Ακόμη μία επίπτωση της διαβητικής μικροαγγειοπάθειας είναι η διαβητική νευροπάθεια. Είναι μία διαταραχή που προκαλεί λειτουργική βλάβη στο νευρικό σύστημα, από την μία προσβάλλονται τα περιφερικά νεύρα των κάτω άκρων και από την άλλη το κεντρικό νευρικό σύστημα, με προβλήματα στην λειτουργία του καρδιαγγειακού και γεννητικού-ουροποιητικού συστήματος.

Άλλη χαρακτηριστική επιπλοκή είναι η διαβητική μακροαγγειοπάθεια, σύμφωνα με την οποία προσβάλλονται αιμοφόρα αγγεία μεσσαίου και μεγάλου διαμετρήματος, προκαλώντας βλάβες παρόμοιες με αυτές της αρτηριοσκλήρωσης, δηλαδή αρτηριακή υπέρταση, στεφανιαία νόσος, εγκεφαλικά επεισόδια και απόφραξη αγγείων από την οποία μπορεί να προκληθεί γάγγραινα[Ο Γιατρός μου, 2002].

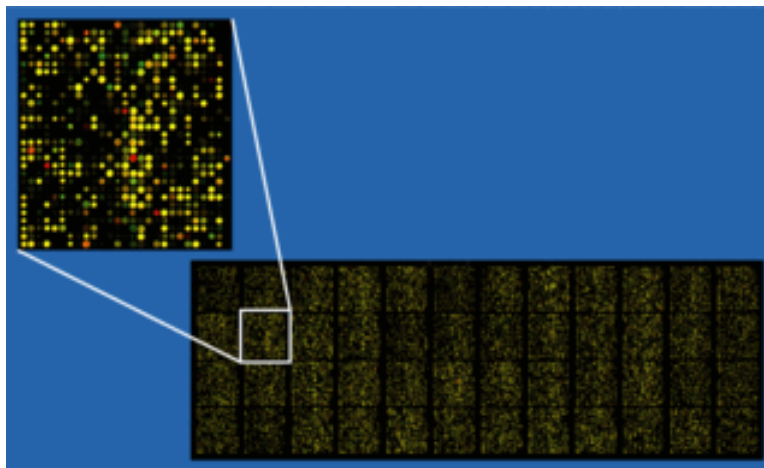
Prevalence of diabetes and related risk factors			
	males	females	total
Diabetes	9.5%	8.8%	9.1%
Overweight	69.6%	60.2%	64.9%
Obesity	23.6%	26.7%	25.1%
Physical inactivity	12.4%	18.2%	15.4%

Εικόνα 5: Επιπολασμός διαβήτη και παράγοντες κινδύνου, πηγή: WHO

#### 1.4 Μικροσυστοιχίες : Ορισμός και γενικά στοιχεία

Η δομή του τσιπ των μικροσυστοιχιών, είναι ένα γυάλινο πλακίδιο που αποτελείται από συγκεκριμένες αλληλουχίες οι οποίες είναι ειδικές για συγκεκριμένα γονίδια, τους ανιχνευτές (probes), οι οποίοι είναι ακινητοποιημένοι σε μία κουκκίδα (spot) της γυάλινης επιφάνειας του πλακιδίου [Lee, 2004]. Πιο συγκεκριμένα, κάθε κουκκίδα του γυάλινου πλακιδίου αποτελείται από νουκλεοτιδικές αλληλουχίες (τους ανιχνευτές) σε μεγάλη συγκέντρωση, οι οποίοι αντιπροσωπεύουν όλα ή σχεδόν όλα τα γονίδια ενός οργανισμού. Το mRNA που βρίσκεται σε κάθε κουκκίδα θα μετατραπεί σε DNA, σύμφωνα με την αρχή συμπληρωματικότητας των βάσεων DNA και RNA και θα σημανθεί με φθορίζουσα ουσία, με

σκοπό να σαρωθεί για να μετρηθεί η ένταση του σήματος σε κάθε σημείο. Η τιμή της μέτρησης θεωρείται αντιπροσωπευτική της ποσότητας του μεταγράφου κάθε γονιδίου στο δείγμα. Αυτό που θα προκύψει από την διαδικασία σάρωσης, θα είναι μία εικόνα με διαβαθμίσεις χρωμάτων.



Εικόνα 6: Παράδειγμα μικροσυστοιχίας γονιδίων, πηγή: <https://el.wikipedia.org/wiki/%CE%9C%CE%B9%CE%BA%CF%81%CE%BF%CF%83%CF%85%CF%83%CF%84%CE%B9%CF%87%CE%AF%CE%B5%CF%82%CE%B3%CE%BF%CE%BD%CE%B9%CE%B4%CE%AF%CF%89%CE%BD>

Οι μικροσυστοιχίες έχουν την δυναμική να απεικονίσουν έκφραση δεκάδων χιλιάδων γονιδίων σε εκατοντάδες δείγματα. Προσφέρουν την δυνατότητα ταυτόχρονης ανάλυσης του τρόπου έκφρασης χιλιάδων γονιδίων από διαφορετικά δείγματα ή από διαφορετικά στάδια ανάπτυξης. Μπορούν να συγκρίνουν έκφραση γονιδίων μεταξύ φυσιολογικής και παθολογικής κατάστασης, καθώς επίσης και απόκριση γονιδίων σε φαρμακευτικές ουσίες, ορμόνες υψηλή θερμοκρασία ή άλλα ερεθίσματα του περιβάλλοντος. Όλες αυτές οι πληροφορίες που μπορούμε να αντλήσουμε από τις μικροσυστοιχίες είναι πολύ χρήσιμες για την καλύτερη κατανόηση της βιολογικής λειτουργίας ενός οργανισμού.

#### 1.4.1 Κατηγορίες μικροσυστοιχιών

Υπάρχουν πολλοί τύποι μικροσυστοιχιών, αν και αρχικά οι πιο ευρέως διαδεδομένες ήταν οι μικροσυστοιχίες DNA. Οι μικροσυστοιχίες αυτές, χωρίζονται σε δύο υποκατηγορίες, τις μονοχρωματικές και τις διχρωματικές. Στις μονοχρωματικές, κάθε δείγμα mRNA σημαίνεται με μία χρωστική ουσία και τοποθετείται για υβριδισμό σε ένα τσιπ. Στην δεύτερη κατηγορία, δύο δείγματα mRNA, ένα δείγμα ελέγχου και ένα δείγμα αναφοράς, σημαίνονται με δύο διαφορετικές χρωστικές και τοποθετούνται για υβριδισμό στο ίδιο τσιπ. Υβριδισμός ορίζεται η διαδικασία κατά την οποία δύο συμπληρωματικές έλικες είτε DNA είτε RNA ενώνονται μεταξύ τους, βασισμένες στον κανόνα συμπληρωματικότητας των βάσεων και κάτω από συγκεκριμένες περιβαλλοντικές συνθήκες, όπως για παράδειγμα η θερμοκρασία υβριδιοποίησης, η σύσταση της αλληλουχίας, το μήκος των ανιχνευτών, η δευτεροταγής

δομή, το pH και η συγκέντρωση των αλάτων και δημιουργούν διπλή έλικα [Brazma A., Vilo J., 2000].

Άλλοι τύποι είναι οι μικροσυστοιχίες πρωτεϊνών, πεπτιδίων, ιστών, χημικών ουσιών, αντισωμάτων, υδατανθράκων, φαινοτύπων, όπως επίσης οι κυτταρικές μικροσυστοιχίες και οι μικροσυστοιχίες επιμόλυνσης. Ακόμα δύο τύποι μικροσυστοιχιών είναι τα MMChips τα οποία χρησιμοποιούνται για τον έλεγχο microRNAs και το IRIS (Interferometric reflectance imaging sensor) [Βέννου, 2020].

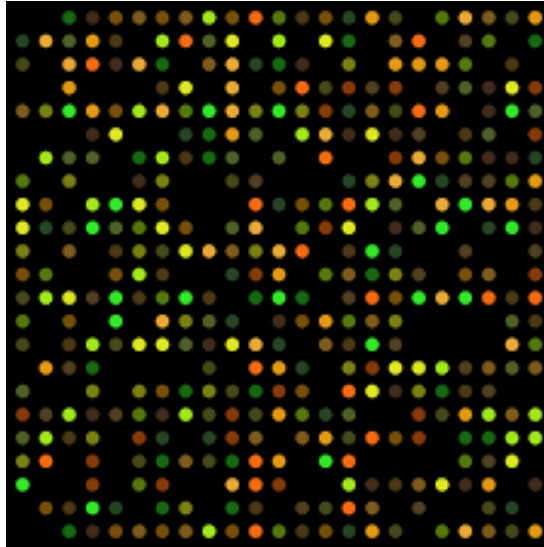
### 1.4.2 Βασικά βήματα για πείραμα με μικροσυστοιχίες

Για την διεξαγωγή ενός πειράματος υπάρχουν κάποια βασικά βήματα που πρέπει να διεξαχθούν. Αυτά είναι τα εξής [Βέννου, 2016]:

1. Διατύπωση του βιολογικού ερωτήματος ► πολύ σημαντικό πρώτο βήμα, καθώς αποτελεί κριτήριο για την επιλογή σωστών μεθόδων και υλικών.
2. Επιλογή του τύπου μικροσυστοιχιών που αρμόζει στην κάθε περίπτωση ► άμεση συνέπεια του πρώτου βήματος, όπως επίσης, και της διαθεσιμότητας των εργαλείων.
3. Απομόνωση RNA από τα εξεταζόμενα δείγματα ► απομόνωση βιολογικού υλικού και χρήση αυτού ως στόχο
4. Σήμανση των δειγμάτων με χρωστικές ουσίες που φθορίζουν ► Το mRNA απομονώνεται από τα κύτταρα ή τους ιστούς και ιχνηθετείται. Χρησιμοποιούνται δύο δείγματα. Σε διχρωματικές μικροσυστοιχίες, συνήθως, το δείγμα αναφοράς (πρόκειται για φυσιολογικό ιστό) ιχνηθετείται με Cy3 (πράσινο) και το δείγμα ελέγχου (πρόκειται για ιστό που νοσεί) ιχνηθετείται με Cy5 (κόκκινο) και μεταγράφονται σε cDNA (μόριο DNA που έχει συντεθεί με βάση το mRNA και δεν περιέχει εσώνια).
5. Υβριδισμός των δειγμάτων στην επιφάνεια του πλακιδίου ► τα δείγματα τοποθετούνται στις μικροσυστοιχίες με σκοπό να υβριδοποιηθούν με τους ανιχνευτές. Η υβριδοποίηση πραγματοποιείται με τα δείγματα να βρίσκονται σε κατάλληλο ρυθμιστικό διάλυμα και επωάζονται μαζί με τους ανιχνευτές για συγκεκριμένη ώρα και θερμοκρασία.
6. Σάρωση των ουσιών για την μέτρηση της έντασης φθορισμού ► μόλις γίνει η υβριδοποίηση, τα δείγματα πλένονται και σαρώνονται από το μηχάνημα. Αφού τα δείγματα ιχνηθετούνται με διαφορετικό χρώμα, η εικόνα που θα δώσουν είναι διαφορετική. Έτσι, ο σαρωτής παράγει δύο εικόνες, μία για κάθε χρώμα φθορισμού.



7. Δημιουργία τελικής εικόνας με την χρήση κατάλληλων προγραμμάτων ανάλυσης ► η ένταση φθορισμού κάθε κουκκίδας αντιστοιχεί αριθμητικά στο επίπεδο έκφρασης του κάθε γονιδίου της μικροσυστοιχίας.



Εικόνα 7: Δεύτερο παράδειγμα μικροσυστοιχίας, πηγή: [https://commons.wikimedia.org/wiki/File:DNA\\_microarray.svg](https://commons.wikimedia.org/wiki/File:DNA_microarray.svg)

Για την καλύτερη κατανόηση του τελικού αποτελέσματος ενός πειράματος μικροσυστοιχιών, παρατίθεται η παραπάνω εικόνα. Όταν η κουκίδα έχει :

- Κόκκινο χρώμα: η ποσότητα δείγματος ελέγχου μεγαλύτερη από την ποσότητα δείγματος αναφοράς
- Πράσινο χρώμα: η ποσότητα δείγματος ελέγχου μικρότερη από την ποσότητα δείγματος αναφοράς
- Κίτρινο χρώμα: η ποσότητα δείγματος αναφοράς ίση με την ποσότητα δείγματος ελέγχου
- Μαύρο χρώμα: κανένα δείγμα δεν έχει υβριδοποιηθεί [Baird, 2012]
- Υπόλοιπες αποχρώσεις: αντίστοιχες ποσότητες των δύο δειγμάτων

Για την υλοποίηση του βήματος 7, που αναφέρθηκε παραπάνω, πραγματοποιούνται διάφορες τεχνικές, οι οποίες αποσκοπούν στην ελαχιστοποίηση σφαλμάτων. Το πρώτο στάδιο που πραγματοποιείται είναι η κανονικοποίηση των αποτελεσμάτων των μικροσυστοιχιών. Κατά την κανονικοποίηση, ελαχιστοποιούνται συστηματικά σφάλματα κατά κύριο λόγο στα επίπεδα έκφρασης των γονιδίων, με σκοπό την καλύτερη αντίληψη βιολογικών διαφορών των δεδομένων. Επιπλέον, η χρήση τέτοιου είδους αποτελεσμάτων, δηλαδή κανονικοποιημένων, είναι πιο αποδοτική όταν γίνεται σύγκριση δεδομένων έκφρασης

από διαφορετικά τσιπς μικροσυστοιχιών. Οι κυριότερες μέθοδοι κανονικοποίησης δεδομένων γονιδιακής έκφρασης είναι οι εξής : στατιστική των λόγων (Chen's ratio statistics), μέση λογαριθμική κεντροποίηση (Mean log centring), κανονικοποίηση ολικής έντασης, γραμμική παλινδρόμηση (linear regression), γραμμική οπισθοδρόμηση τοπικών βαρών (LOcally WEighted Scatterplot Smoothing regression, Lowess) [Βέννου, 2020].

Τα αποτελέσματα που προκύπτουν, είναι ένας πίνακας γονιδιακής έκφρασης με τα γονίδια ως γραμμές και τα δείγματα ως στήλες. Κάθε θέση του πίνακα αναπαριστά την τιμή του εκάστοτε γονιδίου στο συγκεκριμένο δείγμα. Η τιμή αυτή συνήθως προκύπτει από έναν τύπο: τιμή =  $\log_2$  (ένταση του ενός χρωματικού καναλιού / ένταση του άλλου χρωματικού καναλιού).

Τέλος, για την ολοκλήρωση ενός πειράματος με μικροσυστοιχίες απαιτείται στατιστική ανάλυση των αποτελεσμάτων. Λόγω της πολυπλοκότητας των αποτελεσμάτων των μικροσυστοιχιών, οι μέθοδοι που χρησιμοποιούνται είναι διάφορες και περιλαμβάνουν:

- Αναδειγματοληψία, όπως το bootstrap [STINE R., 2016] και το permutation [Dudoit S., Shaffer J, Boldrick J, 2003]
- Αλλαγή διπλώματος [Dalman M. et al, 2012]
- Μπεϋζιανές μεθόδους [Jaynes E., 1986]
- Κανονικοποιημένους ελέγχους  $t$  είτε ελέγχους  $t$  με ποινές
- Έλεγχο βασισμένο στον έλεγχο  $t$  ( $t$  test), ο οποίος πολλές φορές συνδιάζεται και με μεθόδους διόρθωσης της τιμής  $p$  ( $p$ -value)

### 1.4.3 Βάσεις δεδομένων μικροσυστοιχιών

Οι βάσεις δεδομένων μικροσυστοιχιών δημιουργήθηκαν με σκοπό να καλύψουν την ανάγκη για αποθήκευση του, όλο και μεγαλύτερου όγκου δεδομένων πειραμάτων γονιδιακής έκφρασης που πραγματοποιούνταν. Κάθε εγγραφή μέσα στη βάση περιέχει επιπλέον πληροφορίες για το είδος του πειράματος, την πλατφόρμα μικροσυστοιχιών που έχει χρησιμοποιηθεί, τα γονίδια που μελετήθηκαν και άλλα.

Όπως είναι αναμενόμενο η ποικιλομορφία αλλά και ο εκτενής όγκος των πληροφοριών, τα καθιστούν δύσκολα επεξεργάσιμα. Για το λόγο αυτό, αλλά και για να μπορούν τα πειράματα να δημοσιεύονται σε μία δημόσια βάση δεδομένων, έχει εδραιωθεί ένα πρωτόκολλο, το οποίο θα πρέπει να ακολουθούν όλοι οι συγγραφείς και με βάση το οποίο, καταχωρείται η ελάχιστη πληροφορία που περιγράφει ένα πείραμα μικροσυστοιχιών. Το πρωτόκολλο αυτό ονομάζεται MIAME (Minimum Information About Microarray Experiment, Ελάχιστη Πληροφορία σχετικά με Πείραμα Μικροσυστοιχίας) .

Ανάμεσα σε άλλες βάσεις δεδομένων που υπάρχουν, οι πιο διαδεδομένες και ευρέως χρησιμοποιούμενες είναι:

- Gene Expression Omnibus (GEO) <https://www.ncbi.nlm.nih.gov/geo/> (Barrett & Edgar 2006)
- Archive of Functional Genomics Data (ArrayExpress) <https://www.ebi.ac.uk/arrayexpress/> (Athar, Fullgrabe et al. 2019)
- ONCOMINE <https://www.oncomine.org/resource/login.html> (Rhodes, Yu et al. 2004)
- Καρκινικές βάσεις δεδομένων : The Cancer Genome Atlas (TCGA) <http://cancergenome.nih.gov>, Cancer Program Datasets <https://portals.broadinstitute.org/cgi-bin/cancer/datasets.cgi>

## 1.5 Στατιστική ανάλυση μικροσυστοιχιών

Η στατιστική ανάλυση των μικροσυστοιχιών περνάει από κάποια στάδια με σκοπό να μπορέσουμε να αντλήσουμε όσο πιο σωστά και αξιόπιστα αποτελέσματα. Αρχικά, πραγματοποιείται στατιστική ανάλυση για να βρεθούν ποιά γονίδια υπό ή υπερεκφράζονται. Στην συνέχεια, εφαρμόζονται μέθοδοι διόρθωσης. Τελειώνοντας, γίνεται ομαδοποίηση γονιδίων και ταξινόμηση.

### 1.5.1 Μέθοδοι στατιστικής ανάλυσης

Τα δεδομένα ενός πειράματος με μικροσυστοιχίες χρήζουν μίας στατιστικής επιλογής για τον εντοπισμό των διαφορικά εκφραζόμενων γονιδίων, τα οποία όμως είναι στατιστικώς σημαντικά. Για τον σκοπό αυτό υπάρχουν διάφορες μέθοδοι, μερικές από τις οποίες αναφέρθηκαν και παραπάνω, όπως η αλλαγή διπλώματος [Ramasamy, Mondry et al., 2008], τα  $t$  tests, αλλά και άλλες όπως η μέθοδος permute [Dudoit S., Shaffer J, Boldrick J, 2003] ή η μέθοδος bootstrap [STINE R., 2016]. Θα αναφέρουμε περιληπτικά κάποιες πληροφορίες σχετικά με τις μεθόδους αυτές, με σκοπό την καλύτερη κατανόηση του τρόπου λειτουργίας τους.

Για την μέθοδο  $t$  test αξίζει να γνωρίζουμε πως λειτουργεί αξιολογώντας αν τα μέσα των δύο εξεταζόμενων ομάδων παρουσιάζουν στατιστικά σημαντικές διαφορές. Με άλλα λόγια, η μέθοδος αυτή χρησιμοποιείται για την σύγκριση των μέσων τιμών δύο πληθυσμών και έτσι ελέγχεται εάν η διαφορά μέσων τιμών προέκυψε τυχαία ή όχι. Συνήθως, λόγω του μεγέθους των δειγμάτων που έχουν επιλεγεί, η μέθοδος αυτή ακολουθείται από την μέθοδο permute [Dudoit S., Shaffer J, Boldrick J, 2003] ή τη μέθοδο bootstrap [STINE R., 2016], ώστε να παραχθούν ακόμα πιο αξιόπιστα αποτελέσματα.

Και οι δύο τελευταίες μέθοδοι αποτελούν μεθόδους επαναδειγματοληψίας και μπορούν να χρησιμοποιηθούν ακόμα και με άγνωστη κατανομή. Η μέθοδος permutation (μετάθεση) υπολογίζει τη δειγματική κατανομή κάθε στατιστικού ελέγχου. Κάνοντας

μεταθέσεις των ετικετών, υπολογίζει όλες τις πιθανές τιμές του στατιστικού ελέγχου στα δεδομένα. Αν έχουμε ως υπόθεση πως δεν υπάρχει σύνδεση, κάθε μετάθεση των ετικετών για τα δείγματα (case\_control) θεωρείται πιθανή και αντιπροσωπεύει ένα τυχαίο περιστατικό των δεδομένων. Επειδή οι μεταθέσεις αυτές μπορεί να είναι πολλές, επιλέγεται ένα τυχαίο δείγμα μεταθέσεων και δημιουργείται διαδοχικά ένα τυχαίο δείγμα με βάση το προηγούμενο. Στη συνέχεια υπολογίζεται το επίπεδο σημαντικότητας ως η αναλογία των δειγμάτων τυχαιοποίησης με ένα στατιστικό τεστ.

Η μέθοδος bootstrap αποτελεί μία στατιστική μέθοδο βασισμένη στην αναδειγματοληψία. Χρησιμοποιείται για την εκτίμηση της κατανομής δειγματοληψίας ενός εκτιμητή και αποτελείται από τρία στάδια : αναδειγματοληψία, υπολογισμός κατανομής bootstrap, εξαγωγή συμπερασμάτων. Πιο αναλυτικά, μετά από τυχαία αναδειγματοληψία, δημιουργούνται νέα δείγματα ίδιου μεγέθους με το αρχικό σύνολο δεδομένων και υπολογίζεται η στατιστική παράμετρος που επιθυμείται για κάθε ένα δείγμα. Έτσι, δημιουργείται η κατανομή bootstrap, δηλαδή μία εκτίμηση της κατανομής της παραμέτρου. Μέσα από αυτή την διαδικασία αντλούνται πληροφορίες για το τυπικό σφάλμα αλλά και την τυπική απόκλιση της δειγματικής κατανομής αλλά και για το διάστημα εμπιστοσύνης.

Ακόμα μία μέθοδος ανάλυσης αποτελεί η ομαδοποίηση των γονιδίων. Ομαδοποίηση πραγματοποιείται σε γονίδια με βάση το επίπεδο έκφρασης τους. Έπειτα, γίνεται αναπαράσταση των ομάδων αυτών ώστε να βρεθούν πιθανές σχέσεις μεταξύ τους. Οι αλγόριθμοι ομαδοποίησης χωρίζονται σε διάφορες κατηγορίες. Μία από αυτές είναι οι ιεραρχικοί (Single Linkage Clustering, Complete Linkage Clustering, Average Linkage Clustering) και οι μη ιεραρχικοί, ακόμα είναι οι διαχωριστικοί και οι συνενωτικοί και τέλος οι επιβλεπόμενοι και οι μη επιβλεπόμενοι [ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ II].

### 1.5.2 Διόρθωση

Όταν αναφερόμαστε σε πειράματα με μικροσυστοιχίες, θα πρέπει να γνωρίζουμε ότι λόγω του μεγάλου αριθμού γονιδίων, εγκυμονεί ο κίνδυνος τα επίπεδα σημαντικότητας να περιέχουν τυχαία θετικά ή λανθασμένα θετικά στατιστικώς σημαντικά γονίδια (false positive). Τα συνηθέστερα επίπεδα σημαντικότητας για το  $p$ -value είναι  $p < 0.05$  και  $p < 0.01$ . Ανάλογα με το όριο που επιλέγεται, υπάρχει και η ανάλογη εμφάνιση λανθασμένων στατιστικώς σημαντικών γονιδίων. Έτσι, κάθε  $p$ -value κάθε γονιδίου αναπροσαρμόζεται και τα λανθασμένα θετικά γονίδια που παρουσιάζονται, βρίσκονται κάτω από το καθορισμένο όριο. Υπάρχουν διάφορες μέθοδοι διόρθωσης όπως η μέθοδος Bonferroni [Dutoit, Yang et al., 2000], η μέθοδος FDR (Benjamini and Hochberg 1995), η Holm (Holm 1979), η Holland (Holland and Copenhaver 1987) και η Sidak (Sidak 1967).

### 1.5.2.1 Μέθοδος Bonferroni

Η μέθοδος αυτή έχει πάρει το όνομα της από τον δημιουργό της, τον ιταλό μαθηματικό Carlo Emilio Bonferroni (1892-1960). Αναφέρεται γενικότερα ότι η μέθοδος Bonferroni είναι μία συντηρητική μέθοδος, με αποτέλεσμα να μην μπορεί ίσως να δώσει κάποια σημαντικά αποτελέσματα

Η δοκιμή Bonferroni γνωστή ως διόρθωση ή προσαρμογή Bonferroni χρησιμοποιείται στη στατιστική ανάλυση σαν ένα τύπο δοκιμής πολλαπλών συγκρίσεων. Η μέθοδος αυτή προσπαθεί να αποτρέψει την λανθασμένη εμφάνιση στατιστικώς σημαντικών αποτελεσμάτων, κάνοντας μια διόρθωση κατά την διάρκεια της διαδικασίας της σύγκρισης. Στην ουσία, η μέθοδος αυτή υποδηλώνει ότι η  $p$ -value κάθε δοκιμής θα πρέπει να είναι ίση με το  $\alpha$  του, διαιρούμενο από τον αριθμό των δοκιμών που έγιναν (στις μικροσυστοιχίες οι δοκιμές αντιστοιχίζονται στον αριθμό των γονιδίων που ελέγχθηκαν).

Η μέθοδος Bonferroni εφαρμόζεται σε ανεξάρτητες ή εξαρτημένες μελέτες και επαναπροσδιορίζει το επίπεδο σημαντικότητας. Το νέο επίπεδο σημαντικότητας δίνεται από τον τύπο:

$$a_{new} = \frac{\alpha}{n} \quad (1)$$

όπου  $n$ : ο αριθμός των δειγμάτων και οι νέες τιμές δίνονται από τον τύπο

$$p_{cor(i)} = n \times p \quad (2)$$

Αν  $p$ -value <  $a_{new}$  τότε το γονίδιο είναι στατιστικά σημαντικό.

### 1.5.2.2 Μέθοδος Holm

Η μέθοδος Holm ή αλλιώς Bonferroni-Holm, πήρε το όνομά της από τον Sture Holm, ο οποίος την κωδικοποίησε, και τον Carlo Emilio Bonferroni. Προσφέρει μία πιο απλή και πιο ισχυρή διόρθωση από την προηγούμενη και αποσκοπεί στην έλεγχο του ποσοστού οικογενειακών σφαλμάτων (FWER: familywise error rate) [Newson, 2013].

Όταν εξετάζονται πολλές πολλές υποθέσεις, προκύπτει το πρόβλημα της πολλαπλότητας, δηλαδή, όσες περισσότερες υποθέσεις εξετάζονται, τόσο περισσότερο αυξάνεται η πιθανότητα να προκύψουν σφάλματα τύπου I (ψευδώς θετικά). Η μέθοδος αυτή λειτουργεί προσαρμόζοντας τα κριτήρια απόρριψης για κάθε μία από τις μεμονωμένες υποθέσεις. Εδώ γίνεται ταξινόμηση κατά αύξουσα σειρά των  $p$ -value. Το πρώτο στη σειρά πολλαπλασιάζεται με τον αριθμό των γονιδίων που εξετάζονται και τα υπόλοιπα ακολουθούν έναν ενιαίο τύπο (Holm 1979) :

$$p_{cor(i)} = (n - i) \times p_{(i)} \quad (3)$$

### 1.5.2.3 Μέθοδος Simes

Κατά την μέθοδο του Simes, παρουσιάζεται μία τροποποιημένη, λιγότερο συντηρητική, εκδοχή της μεθόδου του Bonferroni για τον έλεγχο των πολλαπλών συγκρίσεων. Στη στατιστική, το ψευδές ποσοστό ανακάλυψης (false discovery rate FDR) είναι μία μέθοδος αναγνώρισης του ποσοστού σφαλμάτων τύπου I κατά την πραγμάτωση πολλαπλών συγκρίσεων. Πιο απλά, το FDR εκφράζει την αναλογία μεταξύ των ψευδών θετικών ταξινομήσεων προς τον συνολικό αριθμό των θετικών ταξινομήσεων. Δηλαδή,  $FDR = FP / (FP + TP)$ .

Η μέθοδος FDR (Benjamini and Hochberg false discovery rate), η οποία χρησιμοποιείται συχνά στις μικροσυστοιχίες, προσφέρει ισορροπία μεταξύ των λανθασμένων και μη στατιστικώς σημαντικών γονιδίων. Η διορθωμένη τιμή για τις  $p$ -values (εκτός από το μεγαλύτερο) δίνεται από τον τύπο (Benjamini and Hochberg 1995):

$$p_{cor(i)} = \frac{n}{n-i} \times p_{(i)} \quad (4)$$

### 1.5.2.4 Μέθοδος Sidak

Στην στατιστική, η μέθοδος Sidak, ή όπως ονομάζεται αλλιώς, η μέθοδος Dunn-Sidak, χρησιμοποιείται και αυτή για να λύσει το πρόβλημα των πολλαπλών συγκρίσεων και κυρίως για έλεγχο σφαλμάτων οικογενειακής φύσεως. Η διορθωση Sidak υποθέτει ότι υπάρχει ανεξαρτησία μεταξύ των μεμονωμένων δοκιμών.

Αν θέσουμε το όριο σημαντικότητας για κάθε δοκιμή να είναι  $\alpha_1$ , τότε (1-η πιθανότητα καμία από τις δοκιμές να είναι σημαντική) είναι η πιθανότητα έστω και μία από τις δοκιμές να είναι σημαντική κάτω από αυτό το όριο. Αφού, λοιπόν, υποτίθεται ότι οι δοκιμές είναι ανεξάρτητες, το προϊόν της εξίσωσης:  $1 - (1 - \alpha_1)^m$  (5) (οι πιθανότητες καμία από τις μεμονωμένες δοκιμές να μην είναι σημαντική), όπου  $m$ : μηδενικές υποθέσεις, είναι η συνολική πιθανότητα όλες οι δοκιμές να μην είναι σημαντικές. Η επιδίωξη μας είναι αυτή η συνολική πιθανότητα να είναι ίση με το  $\alpha$  (επίπεδο σημαντικότητας όλου του σετ δοκιμών ή αλλιώς όσο αφορά τις μικροσυστοιχίες, όλου του σετ γονιδίων). Λύνοντας ως προς  $\alpha_1$ , το αποτέλεσμα είναι:

$$\alpha_1 = 1 - (1 - \alpha)^{\frac{1}{m}} \quad (5)$$

Ο τύπος της για την διορθωση των  $p$ -value είναι :

$$p_{cor(i)} = 1 - (1 - p_{(i)})^{\frac{1}{n}} \quad (6)$$

### 1.5.2.5 Μέθοδος Holland

Η τελευταία μέθοδος, με την οποία εξετάσαμε τα δεδομένα μας ήταν η μέθοδος διορθωσης Holland, η οποία είναι μία παραλλαγή της διορθωσης Holm [Newson, 2013]. Η μέθοδος διορθωσης Holm, η οποία, όπως αναφέρθηκε και προηγουμένως, προέρχεται από την μέθοδο

Bonferroni, και χρησιμοποιεί την ίδια διαδικασία step-down, όπως χρησιμοποιείται για την εξαγωγή της μεθόδου Holland από την μέθοδο Sidak. Η μέθοδος Holland ακολουθεί ακριβώς τα ίδια βήματα με την διόρθωσης Holm, με διαφορά ότι το πρώτο  $p$ -value πολλαπλασιάζεται με  $(n-1)$  συνολικό αριθμό γονιδίων και ο τύπος της διόρθωσης των υπολοίπων  $p$ -values δίνεται από το τύπο (Holland and Copenhaver 1987) :

$$p_{cor(i)} = (n - i + 1) \times p_{(i)} \quad (7)$$

## 1.6 Μετα-ανάλυση : Ορισμός και χρησιμότητα

Είναι μία στατιστική ανάλυση, η οποία συνδυάζει αποτελέσματα πολλών ανεξάρτητων μελετών, οι οποίες θέτουν το ίδιο ερευνητικό ερώτημα και είναι βασισμένη στην συστηματική ανασκόπηση όλων των διαθέσιμων δεδομένων.

Οι λόγοι που καθιστούν την μετα-ανάλυση μία αρκετά χρήσιμη μέθοδο ανάλυσης δεδομένων είναι ποικίλοι. Αρχικά, αποτελεί βασικό εργαλείο σύνδεσης όλων των δεδομένων που σχετίζονται με το πεδίο ενδιαφέροντος κάθε φορά, με σκοπό την αποκόμιση σφαιρικής άποψης επί του θέματος και την διαλεύκανση τυχών διχογνωμιών και αντιπαραθέσεων, σε περίπτωση που τα αποτελέσματα μεμονωμένων μελετών διαφέρουν μεταξύ τους. Επιπλέον, αυξάνεται η στατιστική ισχύς σε σχέση με τις επιμέρους μελέτες και βελτιώνεται η εκτίμηση του μέτρου επιδράσεως κάθε παρέμβασης. Τέλος, δίνεται φως σε ερωτήματα που μπορεί να μην είχαν απαντηθεί ή ακόμα και να μην είχαν τεθεί σε προηγούμενη φάση σε ξεχωριστές μελέτες.

Για να πραγματοποιηθεί μία μετα-ανάλυση, βασικό στάδιο είναι ο σαφής καθορισμός του αντικειμένου της μελέτης και έπειτα η αναζήτηση στη βιβλιογραφία για την εύρεση και καταγραφή όλων των διαθέσιμων μελετών [Normand, 1999]. Απαραίτητη προϋπόθεση, να συγκεντρωθεί ένας απαιτούμενος αριθμός μελετών. Στη συνέχεια, θα πρέπει να καθοριστεί το μέγεθος επίδρασης. Αυτό μπορεί να είναι ένα από τα παρακάτω: ο συντελεστής συσχέτισης, ο λόγος αναλογιών OR(Odds ratio), η διαφορά μέσων τιμών, ο σχετικός κίνδυνος RR(risk ratio) ή η διαφορά κινδύνου RD(risk difference). Για συνεχή δεδομένα επιλέγεται η διαφορά των μέσων τιμών (τυποποιημένη ή μη) μεταξύ των δειγμάτων ελέγχου και αναφοράς, ενώ για διχοτομικά δεδομένα επιλέγεται ο OR ή ο RR ή η RD.

Δύο ακόμα σημαντικά μέτρα για την διεργασία της μετα-ανάλυσης είναι η ισχύς και η δύναμη αυτής. Αυτά τα μέτρα βασίζονται κυρίως σε δύο παράγοντες. Ο πρώτος είναι ο αριθμός των μελετών και ο δεύτερος αφορά την μέθοδο που θα χρησιμοποιηθεί για τον συνδιασμό των μεμονωμένων τιμών του μεγέθους επίδρασης των αρχικών μελετών. Όσο περισσότερες μελέτες μετα-αναλυθούν τόσο ισχυρότερα θα είναι τα αποτελέσματα που θα εξαχθούν. Για τον συνδιασμό των τιμών, όπως αναφέρθηκε, τα μοντέλα που χρησιμοποιούνται είναι δύο: το μοντέλο των τυχαίων επιδράσεων (random effect model) και

το μοντέλο των σταθερών επιδράσεων (fixed effect model) [Normand, 1999]. Όσο αφορά το πρώτο μοντέλο των τυχαίων επιδράσεων, λαμβάνεται υπόψη και η ετερογένεια μεταξύ των πειραμάτων/μελετών.

### **1.6.1 Μονομεταβλητή μετα-ανάλυση**

Όταν αναφερόμαστε σε μία μετα-ανάλυση, συνήθως η κοινή σκέψη πάει σε μία μονομεταβλητή μετα-ανάλυση. Μονομεταβλητή (univariate meta-analysis) ονομάζεται μία μετα-ανάλυση όταν το αντικείμενο της έρευνας μας περιορίζεται ανάμεσα σε δύο καταστάσεις π.χ. ασθενείς και υγιείς. Σε αυτή την περίπτωση, μπορεί να μας ενδιαφέρει η διαφορετική έκφραση γονιδίων ανάμεσα σε αυτές τις δύο ομάδες.

### **1.6.2 Προβλήματα κατά την διεξαγωγή της μετα-ανάλυσης**

Για να μπορέσει να πραγματοποιηθεί μία σωστή μετα-ανάλυση θα πρέπει να γίνει σωστή αναζήτηση και εκτίμηση της διαθέσιμης βιβλιογραφίας και να συμπεριληφθούν όλα τα απαραίτητα δεδομένα. Επιπλέον, θα πρέπει να αποφευχθεί η υπερεκτίμη των αποτελεσμάτων, γι' αυτό και θα πρέπει να ληφθούν υπόψη κάποια προβλήματα που προκύπτουν σε σχέση με την βιβλιογραφία.

Αρχικά, αναφερόμαστε στην γκρίζα και στην ξενόγλωσση βιβλιογραφία. Τα δύο αυτά φαινόμενα αποτελούν το λεγόμενο συστηματικό λάθος δημοσίευσης (publication bias) [Egger, 1997]. Στην αναφερόμενη γκρίζα βιβλιογραφία περιέχονται μελέτες, οι οποίες λόγω των αποτελεσμάτων που έχουν βγάλει (π.χ. μη στατιστικώς σημαντικά) δεν έχουν δημοσιευτεί σε κάποιο επιστημονικό περιοδικό. Στην άλλη κατηγορία, αυτή της ξενόγλωσσης βιβλιογραφίας εντάσσονται μελέτες οι οποίες έχουν διεξαχθεί σε χώρες μη-αγγλόφωνες και έχουν δημοσιευθεί σε τοπικό επίπεδο, με αποτέλεσμα την έλλειψη δεδομένων σχετικών με τις μελέτες αυτές.

Όπως ανέφερα και παραπάνω, κατά την διεξαγωγή μίας μετα-ανάλυσης, λαμβάνεται υπόψη και ο παράγοντας της ετερογένειας μεταξύ των μελετών. Σε περίπτωση ύπαρξης ανομοιογένειας μεταξύ των δειγμάτων μίας μελέτης, μπορεί να υπάρχει διαστρέλωση των αποτελεσμάτων και επομένως η μετα-ανάλυση δεν θα είναι αποδοτική.

### **1.6.3 Στατιστικά εργαλεία στην διεξαγωγή της μετα-ανάλυσης**

Τα δύο ευρέως πιο διαδεδομένα στατιστικά εργαλεία που χρησιμοποιούνται για την διεξαγωγή μιας μετανάλυσης είναι η γλώσσα προγραμματισμού R [Team, 2015] και το στατιστικό πακέτο STATA [StataCorp., 2013].

Το STATA είναι ένα πακέτο λογισμικού γενικής χρήσης, το οποίο προσφέρεται για χειρισμό, επεξεργασία, στατιστική ανάλυση αλλά και οπτικοποίηση δεδομένων. Οι τομείς



όπου εκτείνεται η χρήση του ποικίλλουν από τον τομέα της οικονομίας και της πολιτικής επιστήμης έως τους τομείς της βιοϊατρικής και της επιδημιολογίας.

Γενικά, το STATA περιέχει έτοιμες εντολές και πακέτα για ανάλυση δεδομένων, όμως δεν περιέχει μία έτοιμη εντολή για μετα-ανάλυση σε δεδομένα από μικροσυστοιχίες. Προσφέρεται όμως η δυνατότητα για την δημιουργία κώδικα για την εκτέλεση της. Στην παρούσα εργασία, χρησιμοποιήθηκε ένας έτοιμος κώδικας για την πραγματοποίηση της μετα-ανάλυσης. Ο κώδικας αυτός δημιουργήθηκε από την ομάδα του ερευνητικού εργαστηρίου Υπολογιστικής Γενετικής της σχολής και παραδόθηκε στους φοιτητές για την εκπόνηση των εργασιών με θέμα την μετα-ανάλυση γονιδιακών δεδομένων.

Ακόμα ένα πολύ γνωστό εργαλείο είναι η γλώσσα προγραμματισμού ανοιχτού κώδικα και περιβάλλοντος R. Όπως και το προαναφερθέν εργαλείο, και η R είναι μία γλώσσα που σου δίνει δυνατότητες υπολογιστικής στατιστικής και παραγωγής γραφημάτων.

Στην γλώσσα R υπάρχουν διάφορα έτοιμα πακέτα για την εκτέλεση μετα-ανάλυσης σε δεδομένα μικροσυστοιχιών. Μερικά παραδείγματα είναι το πακέτο metaMA, το πακέτο GeneMeta, το πακέτο MAMA κ.α. Επιπλέον, δίνει την δυνατότητα στον χρήστη να κάνει χρήση μεθόδων διόρθωσης πολλαπλών συγκρίσεων καθώς επίσης και δημιουργία διαγραμμάτων για οπτικοποίηση των αποτελεσμάτων.

#### 1.6.4 Μετα-ανάλυση και μικροσυστοιχίες

Σε γενικό επίπεδο η μετα-ανάλυση χρησιμοποιείται ως εργαλείο για την σύνοψη αποτελεσμάτων μελετών υψηλής απόδοσης, όπως είναι και τα αποτελέσματα που προέρχονται από πειράματα γονιδιακής έκφρασης μικροσυστοιχιών. Στα συγκεκριμένα πειράματα, μετριούνται ταυτόχρονα επιπεδα έκφρασης mRNA από χιλιάδες γονίδια. Έτσι, η εύρεση διαφορετικά εκφραζόμενων γονιδίων ανάγεται σε στατιστικό έλεγχο ισότητας μέσω των τιμών.

Για την διεξαγωγή της μετα-ανάλυσης, δηλαδή για τον συνδιασμό και την ανάλυση των δεδομένων, έχουν αναπτυχθεί τρεις τεχνικές. Η πρώτη τεχνική είναι η μέθοδος των μεγεθών επίδρασης, ή αλλιώς μέθοδος διαφοράς μέσω των τιμών ή αλλιώς μέθοδος *t*-test. Σε αυτή τη μέθοδο ως μέγεθος επίδρασης χρησιμοποιείται η τυποποιημένη μέση διαφορά έκφρασης των δειγμάτων ελέγχου [Choi et al. 2003]. Οι τύποι της μέσης διαφοράς αλλά και της τυπικής απόκλισης είναι:

$$d_i = \frac{\bar{x}_{ti} - \bar{x}_{ci}}{sd_i} \quad (8) \quad sd_i = \sqrt{\frac{(n_{ti}-1)sd_{ti}^2 + (n_{ci}-1)sd_{ci}^2}{n_{ti} + n_{ci} - 2}} \quad (9)$$

όπου  $\bar{x}_{ti}$ : η μέση τιμή του δείγματος αναφοράς και  $\bar{x}_{ci}$ : η μέση τιμή του δείγματος ελέγχου της μελέτης *i*. Ο υπολογισμός των *p*-values και η εκτίμηση της τιμής του FDR (false

discovery rate) βασίζονται στη μέθοδο permute. Με βάση λοιπόν την προϋπόθεση ότι τα μεγέθη επίδρασης των μελετών συνδιάζονται, επιλέγεται το κατάλληλο μοντέλο μετα-ανάλυσης (μοντέλο σταθερών ή τυχαίων επιδράσεων).

Ακόμα μία τεχνική που χρησιμοποιείται είναι η μέθοδος συνδιασμού των  $p$ -values. Η τεχνική αυτή αποτελεί μία εύχρηστη λύση, καθώς μπορεί να εφαρμοστεί σε ανεξάρτητες μεταξύ τους μελέτες, με την προϋπόθεση όμως ότι οι  $p$ -values όλων των γονιδίων είναι γνωστές. Η μέθοδος που χρησιμοποιείται πιο συχνά είναι η μέθοδος αθροίσματος των λογαρίθμων των  $p$ -values. Σε αυτή τη μέθοδο ως μέγεθος επίδρασης χρησιμοποιείται το  $S_i$  και ο τύπος του είναι:

$$S_i = -2 \sum_{k=1}^k \log(p_{ik}) \quad (10),$$

όπου  $S_i$ : το άθροισμα των λογαρίθμων των  $p$ -values για κάθε γονίδιο  $i$  μεταξύ των μελετών  $k$  [Campain and Yang 2010].

Τέλος, άλλη μία είναι η μέθοδος υπολογισμού του γινομένου των βαθμών κατάταξης. Ο τρόπος λειτουργίας αυτής της μεθόδου είναι ο εξής. Αρχικά, γίνεται υπολογισμός του λόγου έκφρασης μεταξύ του δείγματος ελέγχου και αναφοράς και με βάση την τιμή που προκύπτει γίνεται κατάταξη των γονιδίων κατά φθίνουσα σειρά. Ο αριθμός κατάταξης κάθε γονιδίου αποτελεί το βαθμό κατάταξης του. Στη συνέχεια, υπολογίζεται το γινόμενο των βαθμών κατάταξης και η στατιστική σημαντικότητα αυτού μέσω του permutation test. Ως μέγεθος επίδρασης χρησιμοποιείται το γινόμενο των βαθμών κατάταξης, το οποίο δίνεται από τον τύπο :

$$RP_g = (\prod_{i=1}^k r_{gik})^{i/k} \quad (11)$$

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

Ανάλογα με τον τύπο των δεδομένων των μελετών που θα συμπεριληφθούν, επιλέγεται και το κατάλληλο στατιστικό μοντέλο για την μετα-ανάλυση. Τα δεδομένα μπορούν να είναι είτε συνεχή, είτε δυαδικά. Για συνεχή δεδομένα, όπως είναι τα δεδομένα της γονιδιακής έκφρασης, αυτά δηλαδή που εξετάζουμε εμείς, η διαφορά μέσω των τιμών των δειγμάτων αναφοράς και ελέγχου, αποτελεί το μέγεθος επίδρασης  $\theta$  [Normand 1999]. Αυτός (το μέγεθος  $\theta$ ) θα είναι και ο παράγοντας με τον οποίο θα πραγματοποιηθεί η μετα-ανάλυση και η διακύμανση ( $s_i^2$ ) κάθε μελέτης. Η διαφορά των μέσω των τιμών μπορεί να είναι τυποποιημένη ή μη τυποποιημένη. Οι τύποι για την τυποποιημένη και την μη τυποποιημένη διαφορά μέσω των τιμών είναι αυτοί που παραθέτονται παρακάτω αντίστοιχα.

- Τυποποιημένη διαφορά :

$$d_i = \frac{\bar{x}_{ti} - \bar{x}_{ci}}{sd_i} \quad (12)$$

$$sd_i = \sqrt{\frac{(n_{ti}-1)sd_{ti}^2 + (n_{ci}-1)sd_{ci}^2}{n_{ti} + n_{ci} - 2}} \quad (13)$$

- Μη τυποποιημένη διαφορά:

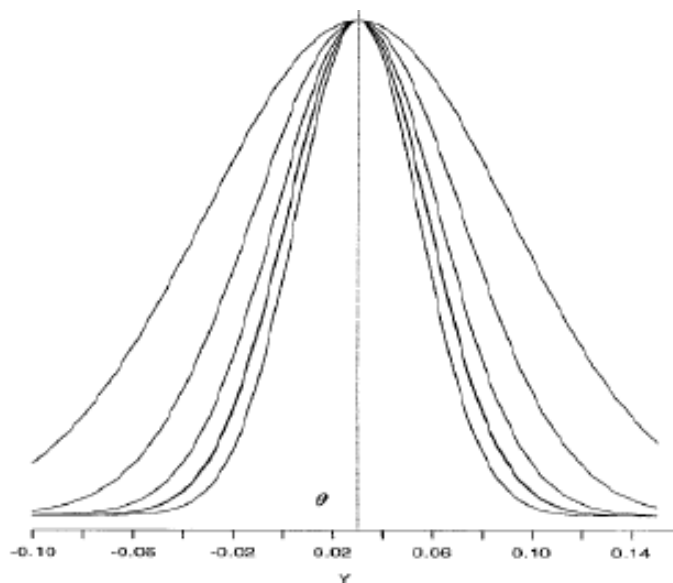
$$d_i = \bar{x}_{1t} - \bar{x}_{2t} \quad (14)$$

όπου  $d_i$ : η τυποποιημένη διαφορά μέσων τιμών και  $sd_i$ : η τυπική απόκλιση

#### 1.6.4.1 Μοντέλο σταθερών επιδράσεων (fixed effects model)

Έστω ότι εξετάζονται  $Y_i$  ανεξάρτητες μελέτες. Στο μοντέλο των σταθερών επιδράσεων (fixed effect model) υποτίθεται ότι αυτές οι μελέτες έχουν ένα μέσο κοινό μέγεθος επίδρασης  $\theta$ , ανεξάρτητο της διακύμανσης και του αριθμού των μελετών. Η κατανομή του μοντέλου αυτού είναι:

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



Εικόνα 8: Μοντέλο σταθερών επιδράσεων. Η κατανομή πέντε υποθετικών επιδράσεων. Η διαφορά ανάμεσα τους αποτελεί την διαφορετική διακύμανση κάθε μελέτης (Normand 1999).

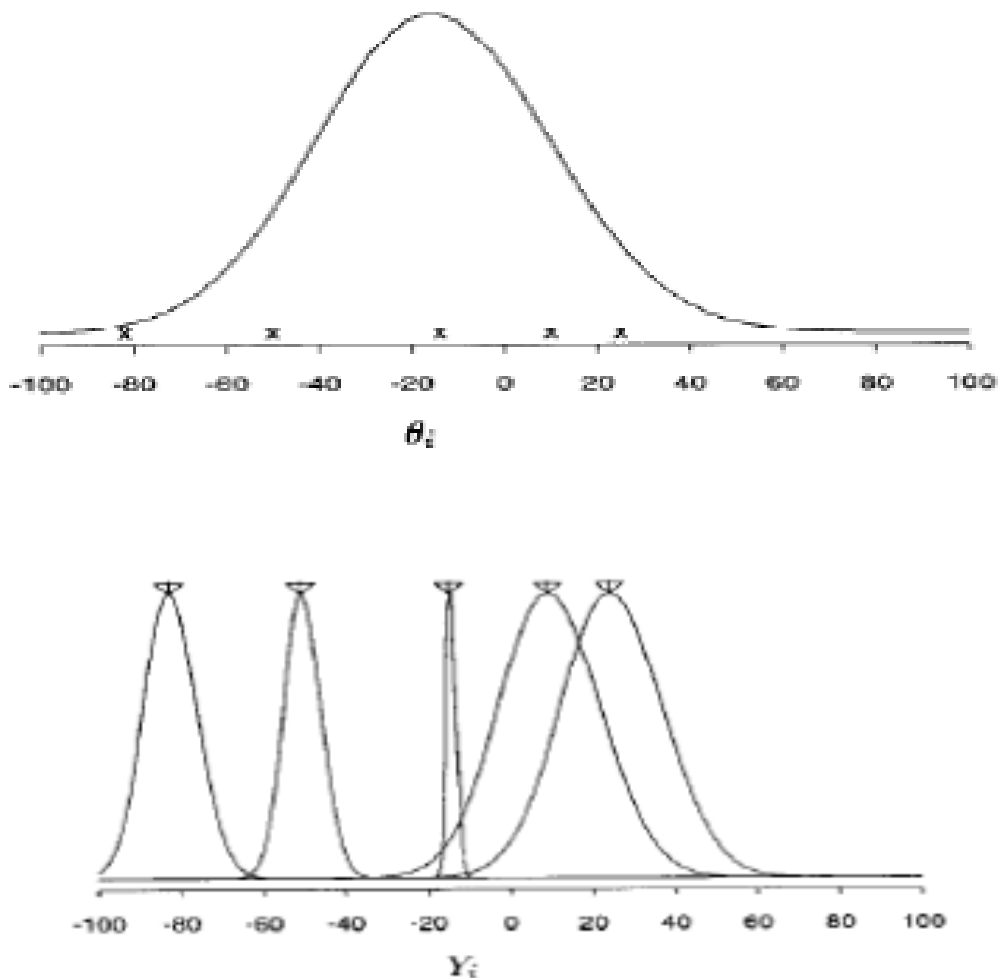
#### 1.6.4.2 Μοντέλο τυχαίων επιδράσεων (random effects model)

Σε αυτό το μοντέλο, θεωρείται ότι η κάθε μία ανεξάρτητη μελέτη  $Y_i$  έχει ένα ξεχωριστό μέγεθος επίδρασης  $\theta_i$  που συνοδεύεται από από μία τιμή διακύμανσης  $s_i^2$ . Αυτό συμβαίνει

διότι θεωρείται ότι τα δείγματα της μελέτης προήλθαν από μία κατανομή πληθυσμών και για αυτό επιτρέπεται η ύπαρξη ετερογένειας.

Εισάγεται μία νέα μεταβλητή, η διακύμανση  $\tau^2$ , η οποία αντιπροσωπεύει, πλέον, την ετερογένεια που παρατηρείται ανάμεσα στις ξεχωριστές μελέτες. Επομένως, το κάθε  $\theta_i$  θεωρείται ότι είναι ένα δείγμα, το οποίο προέρχεται από ένα σύνολο τιμών εκτιμητών με μέση τιμή  $\theta$  και διακύμανση  $\tau^2$ , οι οποίες ακολουθούν αντίστοιχα τους τύπους που ακολουθούν.

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



Εικόνα 9: Μοντέλο τυχαίων επιδράσεων. Η κατανομή πέντε διαφορετικών δειγμάτων. Στο παράδειγμα τα πέντε αποτελέσματα των μοντέλων δημιουργήθηκαν από κάθε ένα από τα μεγέθη επίδρασης που φαίνεται στην πρώτη εικόνα (Normand 1999).

### 1.6.4.3 Έλεγχος ετερογένειας

Ένα κρίσιμο στάδιο κατά τη διεξαγωγή μιας μετα-ανάλυσης είναι η εξέταση του βαθμού ετερογένειας μεταξύ των μεμονωμένων μελετών. Αν υπάρξει ανομοιογένεια μεταξύ των δειγμάτων που εξετάζονται, τότε αυτό μπορεί να επηρεάσει σε μεγάλο βαθμό τα

αποτελέσματα της μετα-ανάλυσης και να οδηγήσει σε υπερεκτίμηση ή υποεκτίμηση του τελικού αποτελέσματος [Thakkinstian et al., 2005].

Η παρουσία ετερογένειας μεταξύ των μελετών προσδιορίζεται από τον τύπο του  $Q$  test του Cochran. Ο τύπος του  $Q$  είναι ο εξής [Villanueva and Zavarsek 2004]:

$$Q = \sum_i^k W_i (d_i - D)^2 \sim \chi_{k-1}^2 \quad (15)$$

$$\text{με } D = \frac{\sum_{i=1}^k w_i d_i}{\sum_{i=1}^k w_i} \quad (16) \quad \text{και } w_i = \frac{1}{\text{var}(d_i)} \quad (17)$$

Από ότι φαίνεται, το  $Q$  ακολουθεί την κατανομή  $\chi^2$  με  $k-1$  βαθμούς ελευθερίας, όπου  $k$  ο αριθμός των μελετών. Ο έλεγχος αυτός μπορεί να εφαρμοστεί για κάθε τύπο δεδομένων, είτε συνεχή, είτε διακριτά. Ο τύπος του Cochran είναι πιο αποδοτικός όταν ο αριθμός των μελετών που εξετάζονται είναι μεγάλος. Ακόμα δύο στατιστικές μέθοδοι για την εκτίμηση της ετερογένειας είναι το  $I^2$  test, το οποίο βασίζεται στο  $Q$  test και ο εκτιμητής  $\tau^2$ .

Το  $I^2$  μπορεί να πάρει τιμές από 0% έως 100%. Τιμές που δεν ξεπερνούν το 25% φανερώνουν μία μικρή ετερογένεια, σε αντίθεση με τιμές πάνω από 50%, οι οποίες υποδηλώνουν ότι η ετερογένεια αποτελεί πρόβλημα. Υπολογίζεται από τον τύπο:

$$I^2 = \max\left(0, \frac{Q - (k-1)}{Q}\right) \quad (18)$$

Όσο αναφορά τον εκτιμητή  $\tau^2$ , υπολογίζει την μεταβλητότητα μεταξύ δύο μελετών και ο τύπος του ορίζεται ως εξής:

$$T_{DL}^2 = \max\left\{0, \frac{Q - (k-1)}{\sum W_i - \frac{\sum w_i^2}{\sum w_i}}\right\} \quad (19)$$

Όταν οι τιμές των δεικτών ετερογένειας είναι ίσοι ή τείνουν στο μηδέν τότε το εξαγόμενο συμπέρασμα είναι ότι δεν υπάρχει ετερογένεια μεταξύ των μελετών [DerSimonian and Laird 1986]. Αντίθετα, όταν οι τιμές αυτές παραμένουν υψηλές υποδηλώνουν την ύπαρξη ετερογένειας.

Στην παρούσα εργασία, ο έλεγχος ετερογένειας γίνεται από το στατιστικό πακέτο το οποίο χρησιμοποιούμε, το STATA13, με την χρήση της εντολής *metan*. Η εντολή αυτή υπολογίζει τους δείκτες ετερογένειας και ξεκινάει κάνοντας δύο υποθέσεις:

- $H_0$  : τα δείγματα είναι ομοιογενή
- $H_1$  : τα δείγματα δεν είναι ομοιογενή

Το αποτέλεσμα της εντολής θα δείξει αν ισχύει η μηδενική υπόθεση  $H_0$ . Αν η υπόθεση αυτή απορριφθεί, τότε θα υπάρχει ετερογένεια ανάμεσα στα δείγματα. Η εντολή *metan* υλοποιεί τους παρακάτω τύπους για να υπολογίσει την ετερογένεια :

$$\theta(\tau)_{MLE} = \frac{\sum_i W_i(\tau) Y_i}{\sum_k W_i(\tau)} \quad (20) \quad W_i(\tau) = \frac{1}{s_i^2 + \tau^2} \quad (21) \quad Y_i = \bar{x}_{1i} - \bar{x}_{2i} \quad (22)$$

### 1.6.5 IDR και IRR

Όταν αναφερόμαστε στην ορθότητα των αποτελεσμάτων μίας μετα-ανάλυσης σε δεδομένα μικροσυστοιχιών δεν θα πρέπει να παραληφθεί η αναφορά στις αναλογίες IDR (Integration-driven discovery, Ανακάλυψη με γνώμονα την ενσωμάτωση) και IRR (Integration-driven revision, Επαναζέταση με γνώμονα την ενσωμάτωση).

Ο όρος Integration-driven discovery τέθηκε το 2003 από τους Choi και συνεργάτες [Choi, Yu et al., 2003], όταν διαφορετικά εκφραζόμενα γονίδια μη στατιστικώς σημαντικά σε καμία μεμονωμένη μελέτη, θεωρούνταν στατιστικώς σημαντικά κατά την μετα-ανάλυση. Ο λόγος IDR υποδεικνύει ότι συλλέγοντας ξεχωριστά σύνολα δεδομένων επιτυγχάνεται αύξηση του μεγέθους του δείγματος αλλά και διαβεβαιώνει ότι μέσω μίας συγκριτικής ανάλυσης, επικυρώνονται τα ξεχωριστά αποτελέσματα κάθε μελέτης. Με άλλα λόγια, η ενσωμάτωση των συνόλων δεδομένων συνέβαλε σε αποτελέσματα πιο αξιόπιστα και με μεγαλύτερη ευασθησία. Ο τύπος του ποσοστού IDR είναι:

$$IDR(x) = \frac{\#γονιδίων[(D_i \geq x) \text{ στην } \mu\epsilon\tau\alpha\text{-}\alpha\lambda\upsilon\sigma\eta] + [(D_i < x) \text{ σε } \tau\omicron\upsilon\lambda\acute{\alpha}\chi\iota\sigma\tau\omicron\nu\text{-}1\text{-}\mu\epsilon\lambda\acute{\epsilon}\tau\eta]}{\#γονιδίων[(D_i \geq x) \text{ στην } \mu\epsilon\tau\alpha\text{-}\alpha\lambda\upsilon\sigma\eta]} \quad (23)$$

Ο όρος Integration-driven revision (IRR) χρησιμοποιήθηκε για πρώτη φορά από τους Stevens και Doerge [Stevens and Doerge 2005] και περιγράφει ακριβώς το αντίθετο από αυτό που περιγράφει ο όρος IDR. Δηλαδή, περιγράφει γονίδια που ενώ θεωρήθηκαν στατιστικώς σημαντικά διαφορετικά εκφραζόμενα στις μεμονωμένες μελέτες, δεν εμφανίστηκαν στατιστικώς σημαντικά στην μετα-ανάλυση. Αυτό συμβαίνει διότι η μετα-ανάλυση κατέληξε στο συμπέρασμα ότι, τα γονίδια εμφανίζονται στατιστικώς σημαντικά λόγω ασυνέπειας στα μεγέθη των αποτελεσμάτων στο σύνολο των μελετών. Ο τύπος του ποσοστού IRR είναι:

$$IRR(x) = \frac{\#γονιδίων[(D_i \geq x) \text{ σε } \tau\omicron\upsilon\lambda\acute{\alpha}\chi\iota\sigma\tau\omicron\nu\text{-}1\text{-}\mu\epsilon\lambda\acute{\epsilon}\tau\eta] + [(D_i < x) \text{ στην } \mu\epsilon\tau\alpha\text{-}\alpha\lambda\upsilon\sigma\eta]}{\#γονιδίων[(D_i \geq x) \text{ σε } \tau\omicron\upsilon\lambda\acute{\alpha}\chi\iota\sigma\tau\omicron\nu\text{-}1\text{-}\mu\epsilon\lambda\acute{\epsilon}\tau\eta]} \quad (24)$$

Στην μετα-ανάλυση μικροσυστοιχιών, συνήθως, εμφανίζονται μεγάλες τιμές IRR, καθώς συνδιάζονται μεγάλες αλλά αντιφατικές εκτιμήσεις μεγέθους αποτελέσματος. Αντίθετα με

τον IDR, ο IRR καταδεικνύει την ανεπάρκεια μίας μετα-ανάλυσης, που συμβαίνει λόγω αυστρότητας ως προς τα αποτελέσματα της.

### **1.6.6 Ανάλυση εμπλουτισμού**

Η ανάλυση εμπλουτισμού ή αλλιώς Enrichment Analysis συνιστά μία ευρέως χρησιμοποιούμενη μέθοδο για την εξαγωγή περαιτέρω πληροφοριών από μία λίστα στατιστικώς σημαντικών γονιδίων/πρωτεϊνών. Η λίστα αυτή μπορεί να προέρχεται από αναλύσεις μεγάλης κλίμακας όπως στην παρούσα εργασία από την υλοποίηση μιας μετα-ανάλυσης.

Πιο συγκεκριμένα, με την ανάλυση εμπλουτισμού επιδιώκεται η ανάκτηση επιπλέον βιολογικών δεδομένων για μία λίστα γονιδίων με βάση τα ήδη γνωστά χαρακτηριστικά τους, προκειμένου να γίνουν περισσότερο κατανοητές οι υποκείμενες βιολογικές διαδικασίες στις οποίες συμμετέχουν και άρα βιοχημικά μονοπάτια στα οποία ενδέχεται να συμμετέχουν και επηρεάζουν.

#### **1.6.6.1 PANTHER**

Στη βιοπληροφορική, το σύστημα ταξινόμησης PANTHER [Thomas, Kejariwal et al., 2003] (Protein Analysis THrough Evolutionary Relationships) (<http://www.pantherdb.org/>) αποτελεί μία επιμελημένη μεγάλη βιολογική βάση δεδομένων των οικογενειών γονιδίων/πρωτεϊνών και των υποοικογενειών τους, τα οποία μπορούν να χρησιμοποιηθούν για την ταξιμόμηση και των προσδιορισμό των γονιδιακών προϊόντων προκειμένου να διευκολυνθεί η ανάλυση υψηλής απόδοσης. Οι πρωτεΐνες ταξινομούνται σύμφωνα με:

- Οικογένεια και υποοικογένεια: οικογένειες είναι ομάδες εξελικτικών συσχετισμένων πρωτεϊνών, οι υποοικογένειες είναι σχετικές πρωτεΐνες που έχουν επίσης την ίδια λειτουργία.
- Μοριακή λειτουργία: η αυτόνομη λειτουργία μίας πρωτεΐνης ή η άμεση αλληλεπίδρασή της με άλλες σε βιοχημικό επίπεδο.
- Βιολογική διαδικασία: η λειτουργία μίας πρωτεΐνης στα πλαίσια ενός μεγαλύτερου δικτύου πρωτεϊνών που αλληλεπιδρούν για να ολοκληρωθεί μια διαδικασία σε επίπεδο του κυττάρου ή οργανισμού.
- Μονοπάτι: παρόμοιο με τη βιολογική διαδικασία, αλλά επιπλέον καθορίζει ρητά τις σχέσεις μεταξύ των αλληλεπιδρώντων μορίων.

### 1.6.6.2 STRING

Η STRING (<http://string-db.org/>) [Szkarczyk, Gable et al., 2019] είναι μία βάση δεδομένων γνωστών και προβλεπόμενων αλληλεπιδράσεων πρωτεϊνών-πρωτεϊνών. Στις αλληλεπιδράσεις περιλαμβάνονται άμεσες και έμμεσες συσχετίσεις, οι οποίες προέρχονται από υπολογιστική πρόβλεψη, από μεταφορά γνώσεων ενός οργανισμού σε άλλον οργανισμό και από αλληλεπιδράσεις που συγκεντρώνονται από άλλες βάσεις. Πιο συγκεκριμένα, οι 5 κυρίες πηγές αλληλεπιδράσεων της STRING είναι: υπολογιστικές προβλέψεις, εργαστηριακά πειράματα (μεγάλης κλίμακας), γνώση που προϋπάρχει σε άλλες βάσεις δεδομένων, αυτοματοποιημένη εξόρυξη κειμένου και συντηρημένη συνέκφραση.

Η περιγραφή των σχέσεων μεταξύ των πρωτεϊνών γίνεται συνήθως με δίκτυο πρωτεϊνικών αλληλεπιδράσεων. Με άλλα λόγια, αναφερόμαστε σε ένα μη κατευθυνόμενο, αβαρή γράφο  $G(V,E)$ , όπου οι πρωτεΐνες αναπαριστούνται σαν σύνολο κόμβων  $V$  και οι αλληλεπιδράσεις μεταξύ τους σαν σύνολο ακμών  $E$ . Έτσι, το σύστημα προσεγγίζεται ολιστικά και δίνεται στο χρήστη τη δυνατότητα να ανακτήσει όλες τις διαθέσιμες πληροφορίες που αφορούν τις σχέσεις που μελετώνται.

## 1.7 Σκοπός της παρούσας πτυχιακής εργασίας

Η συγκεκριμένη πτυχιακή εργασία ασχολήθηκε με την μελέτη και την εφαρμογή τεχνικών μετα-ανάλυσης, με την χρήση του προγραμματιστικού πακέτου STATA13 σε δεδομένα γονιωματικής έκφρασης πειραμάτων μικροσυστοιχιών. Πιο συγκεκριμένα, αντικείμενο της μελέτης ήταν η εύρεση διαφορικής έκφρασης γονιδίων μεταξύ ασθενών και υγιών ατόμων. Το ερώτημα που τέθηκε ήταν το εξής: Ποιες είναι οι διαφορές στις τιμές της γονιωματικής έκφρασης ανάμεσα σε κύτταρα υγιών και ασθενών που πάσχουν από σακχαρώδη διαβήτη τύπου II, σε διαφορετικούς ιστούς. Σκόπος ήταν να εντοπιστούν γονίδια, τα οποία έχουν άμεση σχέση με την εμφάνιση νόσου.



## **Κεφάλαιο 2 : Υλικά και μέθοδοι**

## 2.1 Ερευνητικό ερώτημα

Το ερώτημα που πυροδότησε την αρχή της παρούσας εργασίας ήταν να βρεθεί η διαφορική έκφραση γονιδίων μεταξύ ασθενών με διαβήτη τύπου II και υγιών. Με βάση το ερώτημα αυτό πραγματοποιήθηκε μετα-ανάλυση για την εύρεση των στατιστικά σημαντικών γονιδίων. Τα ευρήματα της ανάλυσης μαζί με περαιτέρω έρευνα θα βοηθήσουν στην εξαγωγή συμπερασμάτων για την καλύτερη κατανόηση πιθανών συσχετίσεων μεταξύ γονιδίων-νόσου.

## 2.2 Συλλογή και καταγραφή δεδομένων

### 2.2.1 Συλλογή δεδομένων

Για την συλλογή όλων των διαθέσιμων δεδομένων για την μελέτη του ερωτήματος που τέθηκε, έγινε αναζήτηση σε διεθνείς βάσεις δεδομένων. Όπως είναι φυσικό, ο όγκος των αποτελεσμάτων που βρέθηκε ήταν πολύ μεγάλος. Για το λόγο αυτό, ο πρώτος περιορισμός που τέθηκε στην αναζήτηση δεδομένων ήταν ότι για την έρευνά μας χρειαζόμασταν μόνο δεδομένα από την σύγκριση ασθενών και υγιών.

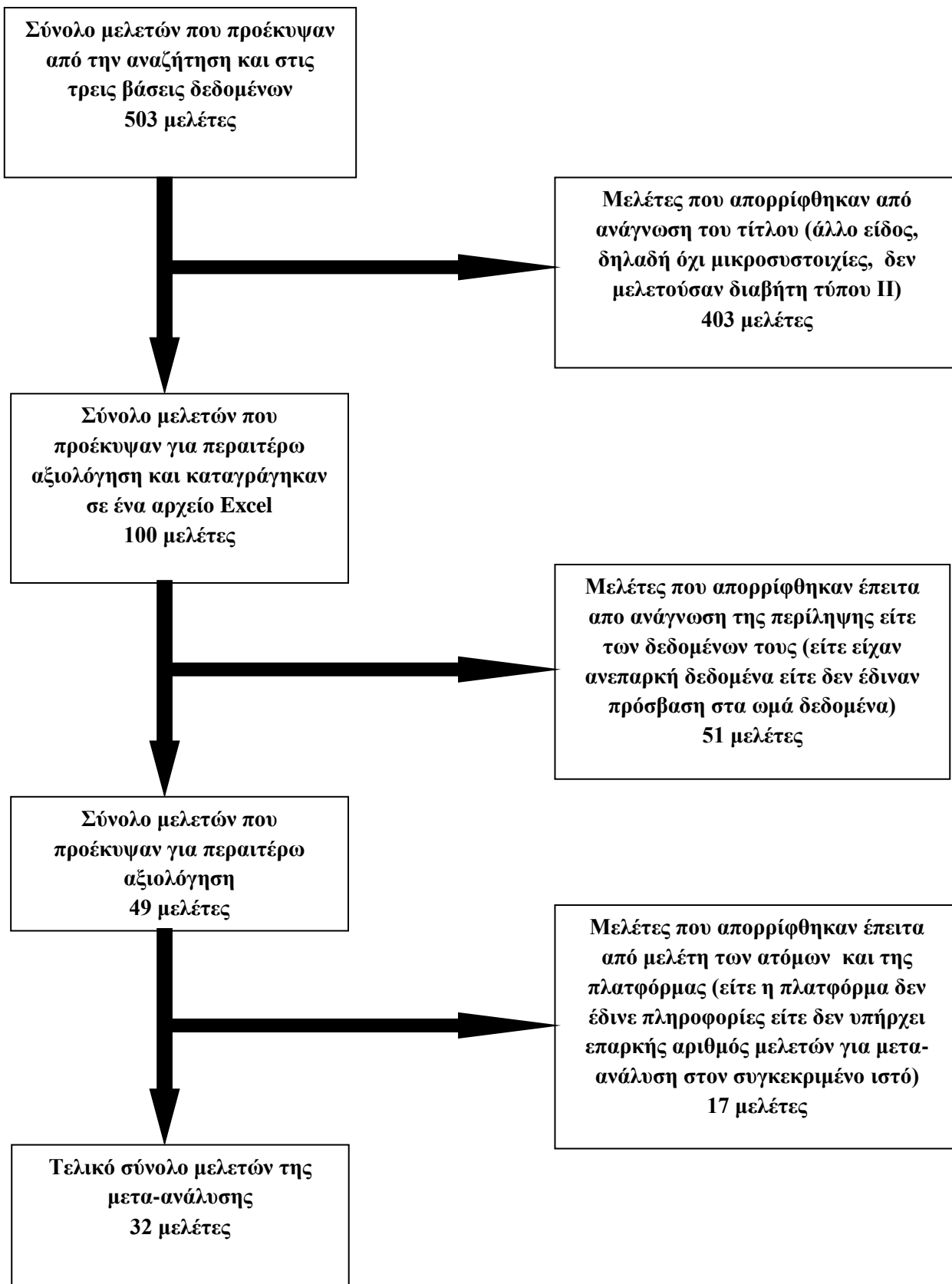
Η πρώτη βάση δεδομένων όπου πραγματοποιήθηκε αναζήτηση ήταν η βάση δεδομένων GEO (Gene Expression Omnibus) του NCBI. Για την ανάκτηση των σετ δεδομένων ενδιαφέροντος, η βάση δίνει την δυνατότητα διατύπωσης ενός query (ερώτημα). Στην προκειμένη περίπτωση το ερώτημα που διατυπώθηκε ήταν το εξής : ("diabetes mellitus, type 2"[MeSH Terms] OR ("diabetes mellitus, type 2"[MeSH Terms] OR type 2 diabetes[All Fields])). Κατά την εισαγωγή αυτού του ερωτήματος, εμφανίστηκαν όλα τα διαθέσιμα σετ δεδομένων. Ωστόσο η αναζήτηση είχε και άλλους περιορισμούς όπως το πεδίο series, το πεδίο organism και το πεδίο expression by array profile. Έτσι, τα κατάλληλα για την αναζήτηση σετ δεδομένων περιορίστηκαν κατά πολύ από τα αρχικά. Παρακάτω παραθέεται μία εικόνα (**Εικόνα 10**) με όλους αυτούς τους περιορισμούς.

Εικόνα 10: Η βάση δεδομένων GEO του NCBI κατά την εφαρμογή του ερωτήματος και των περιορισμών.

Σε κάθε μελέτη εξετάσθηκε η περίληψη της και αν δεν ήταν εμφανές το αντικείμενο της από εκεί, εξεταζόταν εκτενέστερα, με σκοπό την επιλογή της ή όχι. Μετά από όλους αυτούς τους περιορισμούς αλλά και την διεξοδική εξέταση των μελετών, οι χρήσιμες για την εργασία μας μελέτες περιορίστηκαν ακόμα περισσότερο καθώς πολλές απορρίφθηκαν. Οι λόγοι απόρριψης για κάθε μελέτη μπορεί να ήταν διαφορετικοί. Για παράδειγμα, πολλές μελέτες εξέταζαν φάρμακα, εξέταζαν ασθενείς πριν και μετά από δίαιτα ή άσκηση, εξέταζαν μόνο ασθενείς με διαβήτη και καθόλου υγιείς, εξέταζαν παιδιά ή είχαν ελλειπή ή καθόλου προσβάσιμα δεδομένα. Οι μελέτες που εμφανίστηκαν κατά την αναζήτηση ήταν 139, όπως φαίνεται και παραπάνω. Στη συνέχεια, επιλέχθηκαν αυτές που πλήρουν τα κριτήρια που είχαμε θέσει και ομαδοποιήθηκαν με βάση τον ιστό που μελετούσαν. Με βάση αυτή την ομαδοποίηση καταλήξαμε να επιλέξουμε πέντε ιστούς για να κάνουμε μετα-ανάλυση, σε αυτούς που είχαν επαρκή αριθμό μελετών, δηλαδή αριθμό μεγαλύτερο ή ίσο του 3.

Έρευνες έγιναν και σε άλλες βάσεις δεδομένων όπως είναι η Pubmed και η ArrayExpress. Όσο αφορά την βάση Pubmed έγινε έρευνα με το ερώτημα (“type 2 diabetes” AND (microarrays OR microarray OR micro-arrays OR micro-array)). Εκεί οι μελέτες που εμφανίστηκαν ήταν πολλές περισσότερες, πιο συγκεκριμένα 250, όμως καμία δεν χρησιμοποιήθηκε στην τελική έρευνα. Πολλές ήταν οι ίδιες μελέτες με αυτές που χρησιμοποιήθηκαν από την βάση δεδομένων GEO, επομένως η καταγραφή τους ήταν πλεονασμός, και πολλές δεν είχαν διαθέσιμα δεδομένα με αποτέλεσμα να μην υπάρχει πρόσβαση στα δεδομένα της γονιδιακής έκφρασης των συμμετεχόντων στην μελέτη.

Τέλος στην βάση δεδομένων ArrayExpress, αυτό που συναντήθηκε δεν διέφερε πολύ απο την Pubmed. Το ερώτημα που τέθηκε ήταν ("type 2 diabetes" OR "T2D" OR "type II diabetes" AND "Homo sapiens"). Εκεί οι μετέτες που εμφανίστηκαν ήταν 125. Και εκεί πολλές μελέτες βρέθηκαν να είναι ταυτόσημες με αυτές από την GEO, και έτσι δεν καταγράφηκαν για δεύτερη φορά. Επιπλέον, σε κάποιες μελέτες που πληρούσαν τα κριτήρια η καταγραφή των δεδομένων τους για τις τιμές γονιδιακής έκφρασης ήταν αδύνατη, με αποτέλεσμα να μην συμπεριληφθούν ούτε αυτές στην τελική καταγραφή δεδομένων. Τελικά, οι μελέτες που χρησιμοποιήθηκαν για το τελικό αποτέλεσμα ήταν οχτώ.



Εικόνα 11: Απεικόνιση των βημάτων της αξιολόγησης και επιλογής των μελετών οι οποίες προέκυψαν από τις βάσεις δεδομένων GEO, ArrayExpress, Pubmed

### 2.2.2 Καταγραφή δεδομένων

Για κάθε μια βάση δεδομένων που εξετάστηκε, κρατήθηκε ένα υπολογιστικό φύλλο όπου καταγράψαμε τα απαραίτητα δεδομένα για κάθε μελέτη. Αυτά, για κάθε μία μελέτη ήταν: ο αριθμός των ασθενών και των υγιών, ο ανθρώπινος ιστός που εξετάζε, η χρονολογία, ο κωδικός της μελέτης, ο κωδικός της πλατφόρμας που χρησιμοποιεί και λίγα λόγια για αυτή.

Όταν, συγκεντρώθηκαν όλα τα σετ δεδομένων, ανακτήθηκαν για το κάθε ένα σετ το αρχείο του πίνακα τιμών γονιδιακής έκφρασης των ασθενών-υγιών για κάθε γονίδιο και το αρχείο της πλατφόρμας (πίνακας γονιδίων-probes) που χρησιμοποιήθηκε για κάθε μελέτη. Για την αντιστοίχιση των probes (id/id\_ref) με τα γονίδια που αντιπροσωπεύει το καθένα έγινε σύγκριση των δύο πινάκων. Άδεια κελιά, δηλαδή κελιά που δεν είχαν ούτε όνομα γονιδίου (Gene symbol) αλλά ούτε και κωδικό γονιδίου (Gene Id) και κελιά με πολλαπλά γονίδια, δηλαδή ανιχνευτές που εξέταζαν παραπάνω από ένα γονίδια ταυτόχρονα διαγράφηκαν.

### 2.2.3 Ονοματολογία δεδομένων

Για λόγους συμβατότητας αλλά και λόγω του ότι πολλά ονόματα γονιδίων έχουν αλλάξει ή έχουν αποσυρθεί, έγινε έλεγχος των ονομάτων των γονιδίων μέσω της βάσης δεδομένων Gene (<https://www.ncbi.nlm.nih.gov/gene/>), της ιστοσελίδας HUGO HGNC (<https://www.genenames.org/tools/multi-symbol-checker/>) και της ιστοσελίδας bioDBnet (biological DataBase network) (<https://biodbnet-abcc.ncifcrf.gov/db/db2db.php>).

Συνοψίζοντας, οι ιστοί που επιλέχθηκαν να μελετηθούν είναι οι εξής: λιπώδης ιστός (6 μελέτες από την GEO), ήπαρ (3 μελέτες από την GEO), παγκρεατικά νησίδα (5 μελέτες από την GEO και 1 μελέτη από την ArrayExpress), μονοκύτταρα του περιφερικού αίματος ή αλλιώς εν συντομία PBMCs (3 μελέτες από την GEO) και τέλος σκελετικός μυς, όπου υπήρχαν και οι περισσότερες μελέτες (11 μελέτες από την GEO και 3 μελέτες από την ArrayExpress).

**Πίνακας 1:** Συνολικός αριθμός μελετών που ερευνήθηκαν και στην συνέχεια χρησιμοποιήθηκαν για την μετα-ανάλυση

	GEO	ArrayExpress	Pubmed
Συνολικές μελέτες	128	125	250
Ωφέλιμες μελέτες	40	9	0
Λιπώδης ιστός	6	0	0
Ήπαρ	3	0	0
Παγκρεατικά νησίδια	5	1	0
Μονοκύτταρα του περιφερικού αίματος	3	0	0
Σκελετικός μυς	11	3	0
Άλλοι ιστοί	12	1	0

## 2.3 Επεξεργασία δεδομένων

### 2.3.1 Η εντολή «collapse»

Το πρώτο βήμα στην επεξεργασία των δεδομένων ήταν να απαλλάξουμε κάθε αρχείο από εγγραφές γονιδίων που είχαν το ίδιο όνομα αλλά διαφορετικές τιμές έκφρασης. Αυτό συνήθως συμβαίνει γιατί πολλοί ανιχνευτές υβριδοποιούνται από πολλαπλά μετάγραφα του ίδιου γονιδίου. Το βήμα αυτό υλοποιήθηκε με την εντολή *collapse*, η οποία υλοποιείται ως εξής (για λόγους κατανόησης χρησιμοποιούνται τα ονόματα των ασθενών και υγιών που φαίνονται στην παρακάτω εικόνα (**Εικόνα 12**): *collapse(mean) gsm396438-gsm96450, by(gene\_symbol)*).

Το τελικό αρχείο που δημιουργήθηκε είχε την εξής εικόνα και τα εξής χαρακτηριστικά :

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	gene_symbol	gsm39643	gsm39643	gsm39644	gsm39644	gsm39644	gsm39644	gsm39644	gsm39644	gsm39644	gsm39644	gsm39644	gsm39644	gsm39645	gsm39645	gsm39645	gsm39645	gsm39645	gsm39645	gsm39645	gsm39645	gsm39645	gsm39645
2	ALBG	835	1541	1056	1362	797	1845	1173	3825	1203	2867	948	2550	2317	2248	1571					1540	1054	1623
3	AZM	1766	1390	1090	9214	802	4164	3115	12611	3695	622	785	848	12995	6207	7590					1582	937	1719
4	AZML1	71	166	95	140	89	164	102	239	124	163	97	151	99	170	99					168	94	225
5	AGALT	2502	4090	5067	5390	2185	6144	5928	16628	5408	6042	4476	5385	4010	8443	6142	7051	3151	6016	3152	7353	3968	8512
6	AGNT																						
7	AAAS	975	991	690	915	502	1785	1014	1529	1021	1111	927	999	703	1384	1131			1051	619	940	1056	967
8	AACS	1972	1650	1053	1387	797	3302	1747	3499	1856	2699	1878	2287	1974	2803	2177	1028	618	1820	1033	1554	1956	1584
9	AADAC	181	442	367	386	220	364	203	637	393				323	285	599	204				548	245	626
10	AADAT	428	566	413	449	300	790	402	777	471	496	299	451	320	740	387	638	403	518	322	670	576	627
11	AAK1	398	878	498	729	355	1215	536	1313	509	1069	428	877	504	993	494	1421	454	1082	514	1124	405	1324
12	AAMP	1604	1671	1176	1012	533	2452	1627	3480	2118	1880	1508	1762	1136	2702	2284	2094	1588	2003	1175	1334	1715	1387
13	AAAT2	602	699	398	564	341	1061	706	1166	623	886	534	913	662	909	700	1137	734	902	521	648	712	765
14	AAR2	988	1144.5	797.5	799.5	428.5	1769.5	1017.5	2922.5	1292	1965	1204	1753	1047	1610	1254.5	2778	1467	2581	1819	1277	1084	1373
15	AARS1	1047.5	1154.5	1365	1141.5	556	1333.5	958	2357	1474.5	1746	764.5	1179	649.5	2190	970.5	1525	1008	1231.5	901.5	1649.5	1556.5	864.5
16	AARS2	618	745	544	696	383	1100	667	1249	689	824	521	822	668	1047	686	935	613	816	505	815	751	934
17	AARS1D1	1106.5	1212.5	901.5	1180.5	622.5	2496	1396.5	2224.5	1228.5	3672.5	3849	1134.5	719.5	1871.5	1162.5	1540.5	989	1447	959	1374	1207	1460.5
18	AASDH	495	870	435	781	440	963	690	1467	692	959	430	790	535	1053	607					880	569	1204
19	AASDHPTT	2001	2594	3476	3418	1592	1160	652	4921	3908	2933	2913	2823	2404	4440	3322	3912	3762	3585	2976	3311	3243	3898
20	AASS	226.5	402	231	403	219	504	295	549	335.5	390.5	228	356	273	561	288	357	197	431.5	320.5	424.5	337	545
21	AATF	2996	2625	2519	3108	1779	5386	3226	6138	3428	4389	3272	4283	2733	4905	4330	4833	2684	4165	3397	3555	4096	3438
22	AATK	198	370	235	291	172	433	233	491	228	382	221	420	245	443	241	513	244	453	199	433	218	537
23	ABAT	323	446.333	317.333	470.333	305.667	620	402	628	365	405	235	432	276	523.333	371.667	380	188	591	358	510.333	367.333	635
24	ABCA1	441.5	773.5	532.5	794.5	370.5	949	665	1077	672.5	838	462	976	741	1155	634.5	903	426	1032	485	1047.5	736	1240
25	ABCA10	292	515	291.5	533	290.5	877	454	758.5	320	617	313	588.5	350.5	873	377	736.5	275.5	790.5	375.5	880	349.5	894
26	ABCA11P	248	380	289	433	200	489	306	622	337	442	254	393	270	572	310	513	319	480	309	513	335	574
27	ABCA12	395	918	498	720	406	992	501	1028	498	880	455	840	466	932	508	1040	391	999	462	917	448	1258
28	ABCA2	610	937	581	963	491	1289	642	1528	732			964	602	1169	863					1025	671	1235
29	ABCA3	200	403	229	336	184	433	232	451	245	383	232	340	225	350	277	440	242	434	235	395	227	486
30	ABCA4																						
31	ABCA5	475	981	587	657	551	1374	606	1403	597			1269	916	1680	1008							888
32	ABCA6	317	605	366	728	390	704	504	780	598	474	313	426	311	688	357					776	511	918

Εικόνα 12: Απόσπασμα από το αρχείο excel της μελέτης GSE15790 μετά την εκτέλεση της εντολής collapse από το στατιστικό πακέτο STATA13.

Έπειτα, δημιουργήθηκε ένα μεγάλο αρχείο που περιέχει όλες τις μελέτες για κάθε εξεταζόμενο ιστό. Στο σύνολο δημιουργήθηκαν πέντε τέτοια αρχεία, όσοι και οι ιστοί που εξετάζουμε. Το βήμα αυτό πραγματοποιήθηκε με την χρήση της εντολής *append*, ενώνοντας τις μελέτες μεταξύ τους. Η εντολή αυτή υλοποιείται ως εξής: *append using [location of study]* και έχει ως αποτέλεσμα την επόμενη εικόνα (Εικόνα 13).



The screenshot shows the STATA Data Editor interface. The main window displays a dataset with columns for gene symbols (e.g., zswim8, zswim9, zup1, zwi10, zwi1ch, zwi1t, zxda, zxdb, zxdc, zyg11a, zyg11b, zyk, zze1, zzz3, study, case\_control) and rows of data. The 'Variables' panel on the right lists the variables in the dataset, including 'gene\_symbol' and various gene symbols like 'a1bg', 'a1bg\_as1', 'a1cf', 'a2m', 'a2m\_as1', 'a2m1l', 'a2mp1', 'a3galh2', 'a4galt', 'a4gnt', 'aa06', 'aacs', and 'aacs'. The 'Properties' panel shows details for the 'gene\_symbol' variable, such as its name, label, type (string), format (%10s), and value label.

Εικόνα 13: Απόσπασμα από το αρχείο adipose\_genes.dta στο STATA μετά την εντολή append, όπου φαίνεται η ένωση μεταξύ των μελετών.

### 2.3.2 Επιπλοκές

Θα πρέπει επίσης να αναφερθεί ότι λόγω κάποιων επιπλοκών που υπήρξαν κατά τις διαδικασίες της καταγραφής αλλά και εν συνεχεία κατά την εκτέλεση της μετα-ανάλυσης, χρειάστηκε να γίνουν κάποιες επεμβάσεις στην επεξεργασία των δεδομένων, για να έχουμε όσο πιο ορθά αποτελέσματα γινόταν. Πιο συγκεκριμένα, χρειάστηκε τρεις φορές να γίνουν αλλαγές σε κάποια πράγματα. Την πρώτη φορά ήταν όταν το γονίδιο «HBA1» δεν μπορούσε να συμμετέχει στην μετα-ανάλυση του λιπώδη ιστού, καθώς ήταν καταγεγραμμένο μόνο σε μία μελέτη, την GSE15790, με αποτέλεσμα να μην υπάρχουν αρκετά δεδομένα για αυτό και να πρέπει να διαγραφεί. Την δεύτερη φορά, η περίπτωση ήταν διαφορετική. Εκεί υπήρξε μία μελέτη, η GSE44035, η οποία είχε δεδομένα από εννέα υγιή άτομα και ένα ασθενές. Όπως ήταν φυσικό, τα δεδομένα δεν ήταν επαρκή για να μπορεί να σταθεί μόνη της αυτή η μελέτη και έτσι χρειάστηκε να την ενώσουμε με μία άλλη. Μετά από έλεγχο, λοιπόν, υπήρχε μία άλλη μελέτη του ίδιου ιστού, η GSE38642, όπου είχε χρησιμοποιηθεί η ίδια πλατφόρμα και εξετάζαν ακριβώς τα ίδια γονίδια, με αποτέλεσμα να είναι η τέλεια υποψήφια και να μπορέσουμε να τις ενώσουμε.

### 2.3.3 Συνοπτικά

Στον παρακάτω πίνακα παρουσιάζονται συνολικά ο αριθμός των μελετών που χρησιμοποιήθηκε για κάθε ιστό που πραγματοποιήθηκε μετα-ανάλυση, μαζί με τον αριθμό των γονιδίων κάθε μετα-ανάλυσης και τον αριθμό των ατόμων. Επιπλέον, στο Παράρτημα 1 παρουσιάζονται αναλυτικά οι πίνακες για τις μελέτες που χρησιμοποιήθηκαν για κάθε εξεταζόμενο ιστό.

**Πίνακας 2:** Οι μελέτες που χρησιμοποιήθηκαν στην μετα-ανάλυση, ο αριθμός γονιδίων και το πλήθος των ασθενών και υγιών για κάθε ιστό.

Ιστός	Αριθμός μελετών	Συνολικός αριθμός γονιδίων	Σύνολο ατόμων	Ασθενείς	Υγιείς
Λιπώδης ιστός (adipose)	6	23751	243	127	116
Ήπαρ (liver)	3	22105	56	26	30
Παγκρεατικά νησίδα (pancreatic islets)	5	20192	281	81	200
Μονοπύρνα κύτταρα του περιφερικού αίματος (PBMCs)	3	24896	42	21	21
Σκελετικός μυς (skeletal muscle)	14	27582	488	187	301

## 2.4 Στατιστική ανάλυση των δεδομένων

### 2.4.1 Μετα-ανάλυση

Τελειώνοντας και με την επεξεργασία των δεδομένων, η επόμενη κίνηση ήταν η πραγματοποίηση της μετα-ανάλυσης για κάθε έναν ιστό. Όπως αναφέρθηκε και προηγουμένως, το στατιστικό εργαλείο που χρησιμοποιήθηκε ήταν το STATA13 και πιο συγκεκριμένα η έκδοση STATA/SE, η οποία αποτελεί την στάνταρ έκδοση που μπορεί και χειρίζεται σετ δεδομένων πολύ μεγάλου όγκου.

Τα δεδομένα που είχαμε να διαχειριστούμε ήταν δεδομένα γονιδιακής έκφρασης, άρα συνεχή δεδομένα, άρα ως μέσο επίδρασης χρησιμοποιήθηκε η τυποποιημένη διαφορά μέσω των τιμών με έλεγχο τύπου *t*-test. Επίσης, το μοντέλο επίδρασης που επιλέχθηκε, ήταν το μοντέλο τυχαίων επιδράσεων, εφόσον δίνεται διαφορετική βαρύτητα στο μέτρο επίδρασης από κάθε μελέτη. Ο κώδικας που χρησιμοποιήθηκε για την εκτέλεση της μετα-ανάλυσης παρατείνεται στο Παράρτημα 2 μαζί με το αρχείο ado που χρησιμοποιήθηκε.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	gene	g	seg											
2	a1bg	0.066601	0.155178											
3	a1bg_as1	0.347907	0.478312											
4	a1cf	0.644923	0.380935											
5	a2m	-0.16281	0.125851											
6	a2m_as1	-0.44354	0.480208											
7	a2ml1	-0.25722	0.28827											
8	a2mp1	0.576541	0.483587											
9	a3galt2	-0.2879	0.147712											
10	a4galt	0.214352	0.198338											
11	a4gnt	-0.17092	0.253808											
12	aa06	-0.23707	0.476682											
13	aaas	0.057315	0.274118											
14	aacs	0.006687	0.194107											
15	aacsp1	-0.2487	0.140941											
16	aadac	0.109051	0.260475											
17	aadacl2	-0.10025	0.24502											
18	aadacl3	-0.01396	0.305377											
19	aadacl4	-0.11406	0.195594											
20	aadacp1	0.461533	0.480615											
21	aadat	-0.24801	0.1696											
22	aagab	0.098075	0.23112											
23	aak1	-0.04492	0.225977											
24	aamdc	0.3322	0.234571											
25	aamp	-0.28302	0.244241											
26	aanat	0.041711	0.175846											
27	aar2	-0.50165	0.256646											
28	aard	0.523546	0.32111											
29	aars1	0.407507	0.357398											
30	aars2	-0.25366	0.125031											
31	aarsd1	-0.2064	0.192557											
32	aasdh	-0.30425	0.178583											

Εικόνα 13: Αποτελέσματα από μετα-ανάλυση στον λιπώδη ιστό.

Στη συνέχεια, υπολογίστηκαν ακόμα δύο μεταβλητές, η μεταβλητή *z*-score και η μεταβλητή *p*-value. Ουσιαστικά, η μεταβλητή που μας ενδιέφερε κυρίων ήταν η *p*-value,

καθώς από εκεί θα μπορούσαμε να αποφασίσουμε ποιά γονίδια φαίνονται να είναι στατιστικώς σημαντικά και ποιά όχι.

#### 2.4.2 Μέθοδοι διόρθωσης πολλαπλών συγκρίσεων

Σε επόμενη φάση της ανάλυσης των δεδομένων, έπρεπε να τρέξουμε κάποιες μεθόδους διόρθωσης πολλαπλών συγκρίσεων, με βάση τις  $p$ -values που είχαν εξαχθεί προηγουμένως. Οι μέθοδοι αυτοί ελέγχουν την ορθότητα των αποτελεσμάτων που παίρνουμε από την εξαγωγή των  $p$ -values. Με άλλα λόγια, οι μέθοδοι διόρθωσης πολλαπλών συγκρίσεων παίρνουν ως είσοδο μία λίστα  $p$ -values γονιδίων και επιστρέφουν πίσω νέα διορθωμένα  $p$ -values, με σκοπό να βεβαιωθούμε ότι τα διαφορετικά εκφραζόμενα γονίδια που βρέθηκαν να είναι στατιστικώς σημαντικά είναι όντως σημαντικά και ότι δεν απορρίφθηκε κάποιο γονίδιο που μπορεί και αυτό να ήταν στατιστικώς σημαντικό.

Υπάρχουν δύο κατηγορίες ελέγχων, αυτή που ελέγχει το ποσοστό οικογενειακών σφαλμάτων FWER και αυτή που ελέγχει το ποσοστό σφαλμάτων FDR. Η πρώτη κατηγορία ελέγχων εμφανίζεται να είναι πιο αυστηρή στην ανίχνευση σφαλμάτων τύπου I, σε σχέση με την δεύτερη, η οποία παρέχει μία πιο κατανοητή έννοια για τις πολλαπλές συγκρίσεις. Αυτό, όμως, σημαίνει ότι αν και η διαδικασία FDR έχει μεγαλύτερη στατιστική ισχύ, παρέχει αυξημένο κίνδυνο εμφάνισης ποσοστού σφαλμάτων τύπου I.

Ο τρόπος λειτουργίας των ελέγχων FDR είναι ο εξής: αρχικά, κατά την είσοδο των τιμών  $p$ -value των γονιδίων, αυτές ταξινομούνται με αύξουσα σειρά. Συνεχίζοντας, η πρώτη στη σειρά τιμή παραμένει ίδια και οι επόμενες διορθώνονται ακολουθώντας τον τύπο (Benjamini and Hochberg 1995):

$$p_{cor(i)} = \frac{i}{n} \times p_i \quad (25)$$

όπου  $n$ : το πλήθος των εξεταζόμενων γονιδίων.

Ο κώδικας που χρησιμοποιήθηκε για αυτό το βήμα παρατείνεται στο Παράρτημα 2 και οι μέθοδοι όπου χρησιμοποιήθηκαν ήταν: η Simes, η Bonferroni, η Sidak, η Holm και η Holland. Λόγω του ότι το πλήθος των γονιδίων που εξετάζονται είναι πολύ μεγάλο και η πιθανότητα να θεωρηθεί κάτι λανθασμένως σημαντικό, η επιλογή της τιμής του  $p$ -value ήταν σχετικά αυστηρή. Επομένως, στατιστικά σημαντικά θεωρήθηκαν μόνο τα γονίδια που είχαν τιμή  $p$ -value μικρότερη από 0.01.

## **Κεφάλαιο 3 : Αποτελέσματα**

### 3.1 Συνολικά αποτελέσματα

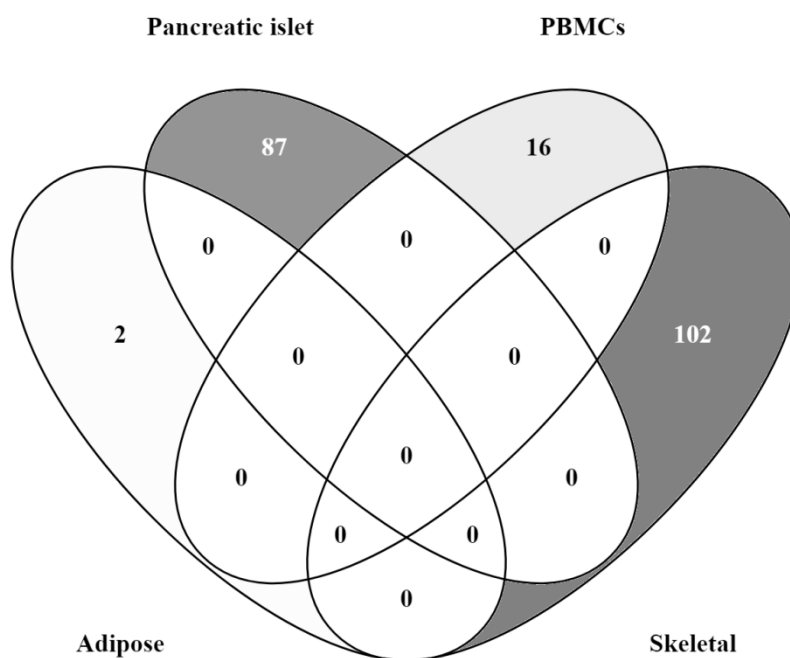
Όπως προαναφέρθηκε, υλοποιήθηκε η μέθοδος της μετα-ανάλυσης για τα δεδομένα που συλλέχθηκαν. Λόγω της ετερογένειας που υπήρχε μεταξύ των μελετών, χρησιμοποιήθηκε το μοντέλο των τυχαίων επιδράσεων και ως μέγεθος απίδρασης επιλέχθηκε η τυποποιημένη διαφορά μέσω τιμών. Τα γονίδια που θεωρήθηκαν στατιστικώς σημαντικά ήταν αυτά που είχαν  $p$ -value μικρότερη από 0.01. Όμως ο μεγάλος αριθμός γονιδίων με  $p$ -value μικρότερη από 0.01 φανερώνει την ύπαρξη false positive αποτελεσμάτων. Έτσι, χρειάστηκε, για αυτές τις  $p$ -values, να γίνει εφαρμογή των μεθόδων διόρθωσης FDR, Bonferroni, Sidak, Holland και Holm. Με την υλοποίηση αυτού του βήματος περιορίστηκαν περισσότερο τα στατιστικώς σημαντικά γονίδια και επιλέχθηκαν αυτά που σε επίπεδο σημαντικότητας 1% (όπως είχε οριστεί) κρίθηκαν σημαντικά από την μέθοδο FDR. Τα συγκεντρωτικά αποτελέσματα φαίνονται παρακάτω στον πίνακα, ξεχωριστά για κάθε ιστό που ερευνήθηκε.

**Πίνακας 3:** Ο αριθμός των γονιδίων κάθε ιστού που θεωρήθηκαν σημαντικά με βάση τις εκάστοτε συνθήκες.

Αριθμός σημαντικών γονιδίων	$p$ -value < 0.05	$p$ -value < 0.01	FDR (0.01)	Bonferroni	Sidak	Holm	Holland
Λιπώδης ιστός (adipose)	1798	479	2	3	3	3	3
Ήπαρ (liver)	1440	373	0	0	0	0	0
Παγκρεατικά νησίδια (pancreatic islets)	2554	1038	87	19	19	19	19
Μονοπύρρηνα κύτταρα του περιφερικού αίματος (PBMCs)	3304	1364	16	7	7	7	7
Σκελετικός μυς (skeletal muscle)	2234	514	102	85	85	85	85

Αξίζει να παρατηρήσουμε ότι τα στατιστικώς σημαντικά διαφορικά εκφραζόμενα γονίδια που βρέθηκαν από τις μετα-αναλύσεις που πραγματοποιήθηκαν ήταν ξεχωριστά για

κάθε ιστό. Αυτό φαίνεται και στο παρακάτω διάγραμμα Venn, όπου κανένα γονίδιο δεν έχει βρεθεί στατιστικά σημαντικό σε παραπάνω από έναν ιστό.



Εικόνα 14: Διάγραμμα Venn μεταξύ των στατιστικά σημαντικών γονιδίων των μετα-αναλύσεων

Για κάθε μελέτη ξεχωριστά χρησιμοποιήθηκε η μέθοδος bootstrap και βρέθηκε ο αριθμός των στατιστικά σημαντικών διαφορικά εκφραζόμενων γονιδίων κάθε μεμονωμένης μελέτης. Στο Παράρτημα 1 βρίσκονται οι πίνακες με τα αναλυτικά αποτελέσματα για την σύγκριση που θέλαμε να κάνουμε, επιλέχθηκαν τα γονίδια με τιμή  $p$  μικρότερη από 0.01 και τα γονίδια που κρίθηκαν σημαντικά από την μέθοδο του FDR. Όπως φαίνεται και παρακάτω οι περιπτώσεις ποικίλλουν. Ορισμένες φορές η μετα-ανάλυση κατάφερε να εντοπίσει περισσότερα σημαντικά αποτελέσματα από τις μεμονωμένες μελέτες και κάποιες τα αποτελέσματα της μετα-ανάλυσης υστερούσαν.

Δύο ακόμα τιμές που έπρεπε να υπολογιστούν για την επικύρωση της μετα-ανάλυσης ήταν τα ποσοστά IDR και IRR. Επίσης, για να ερευνήσουμε τις αλληλεπιδράσεις μεταξύ των γονιδίων που κρίθηκαν στατιστικά σημαντικά από τις μετα-αναλύσεις, όπως βιολογικά μονοπάτια, πρωτεϊνικές αλληλεπιδράσεις κ.α. χρησιμοποιήθηκαν δύο πλατφόρμες, η PANTHER και η STRING. Οι πλατφόρμες αυτές χρησιμοποιούνται για ανάλυση πειραματικών δεδομένων και είναι σχεδιασμένες έτσι ώστε να εμφανίζουν ένα σύνολο βιολογικών πληροφοριών, λαμβάνοντας ως είσοδο μια λίστα γονιδίων/γενετικών τόπων.

Οι πληροφορίες που έχουν ενσωματώσει προέρχονται από πολλές βάσεις βιολογικών δεδομένων και προσφέρουν πλήθος πληροφοριών στον χρήστη (μοριακή λειτουργία, βιολογικά μονοπάτια, αλληλεπιδράσεις πρωτεϊνών με πρωτεΐνες (protein-protein interaction), δίκτυα των γονιδίων της λίστας κ.α.). Δυστυχώς δεν καταφέραμε να αντλήσουμε πληροφορίες για ένα μεγάλο αριθμό γονιδίων που κρίθηκαν στατιστικώς σημαντικά διότι είτε οι βάσεις δεδομένων δεν διέθεταν γνωστές πληροφορίες για αυτά, είτε ήταν γενετικοί τόποι, οι οποίοι παρέμειναν άγνωστοι ακόμα.

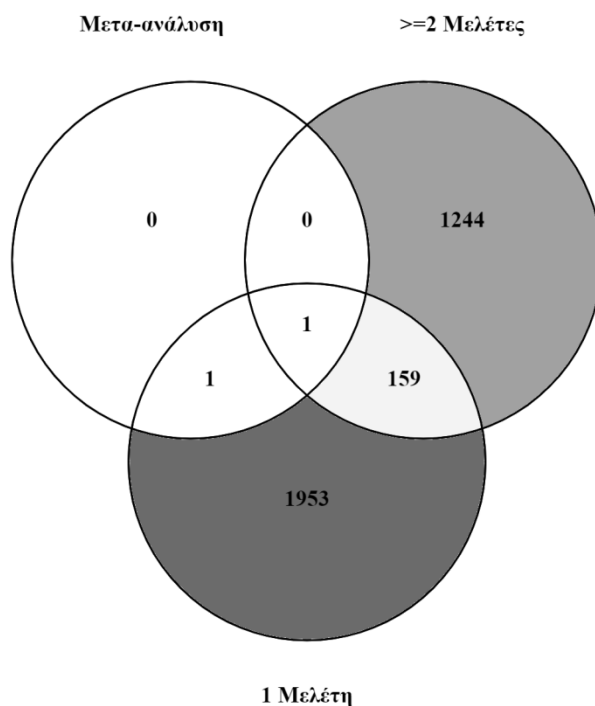
### 3.2 Αποτελέσματα της μετα-ανάλυσης στον λιπώδη ιστό

Στον Πίνακα 4 αναφέρονται αναλυτικά οι αριθμοί των στατιστικών σημαντικών γονιδίων που προέκυψαν από την μέθοδο της μετα-ανάλυσης αλλά και από τις μεμονωμένες μελέτες.

**Πίνακας 4:** Ο αριθμός των στατιστικά σημαντικών γονιδίων με  $p < 0.01$  και της μεθόδου διόρθωσης FDR που προκύπτουν από την μετα-ανάλυση όλων των μελετών συγκεντρωτικά και της κάθε μελέτης ξεχωριστά για τον λιπώδη ιστό.

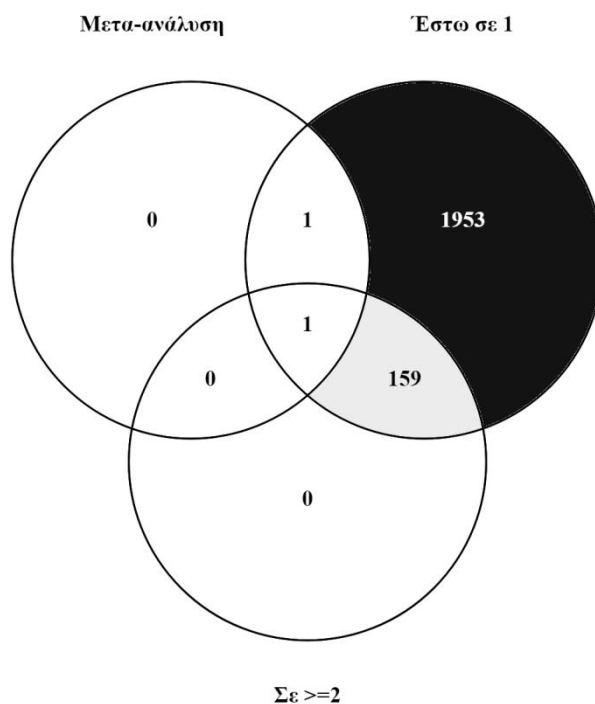
Adipose	Μετα- ανάλυση	GSE1579 0	GSE1641 5	GSE2922 6	GSE2923 1	GSE7141 6	GSE7872 1	
$p < 0.01$		479	20	547	3253	3785	750	503
FDR		2	0	0	1381	2114	23	0





Εικόνα 15: Διάγραμμα Venn σύγκρισης των διαφορετικά εκφραζόμενων γονιδίων που προέκυψαν από την μετα-ανάλυση για τον λιπώδη ιστό και από τις μελέτες GSE29226, GSE29231 και GSE71416 μεμονωμένα (οι οποίες διαθέτουν στατιστικώς σημαντικά γονίδια από τη μέθοδο διόρθωσης FDR σε επίπεδο σημαντικότητας 1%).

Αναφορικά με την **Εικόνα 15** για την μετα-ανάλυση που πραγματοποιήθηκε στον λιπώδη ιστό τα ποσοστά IDR και IRR ήταν 0% και 99.9% αντίστοιχα. Αυτό υποδεικνύει ότι και τα δύο γονίδια που κρίθηκαν στατιστικώς διαφορετικά εκφραζόμενα σύμφωνα με την μετα-ανάλυση δεν βρέθηκαν μόνο από την μετα-ανάλυση. Η τιμή του IRR, από την άλλη μεριά, υποδεικνύει ότι τα γονίδια που βρέθηκαν στατιστικώς σημαντικά στις μεμονωμένες μελέτες ήταν κατά 99.9% περισσότερα, πιο συγκεκριμένα 3516 από τα 3518 βρέθηκαν στατιστικώς σημαντικά μόνο από τις μεμονωμένες μελέτες.



Εικόνα 16: Διάγραμμα Venn σύγκρισης των διαφορικά εκφραζόμενων γονιδίων που προέκυψαν από την μετα-ανάλυση και από τις μεμονωμένες μελέτες για τον λιπώδη ιστό. Συγκρίνονται τα αποτελέσματα που προέκυψαν από την μετα-ανάλυση (2 γονίδια), με τα γονίδια τα οποία προσδιορίζονται σε τουλάχιστον μία μελέτη (2114 γονίδια) και σε τουλάχιστον δύο μελέτες (160 γονίδια).

Για τον λιπώδη ιστό, όπως φαίνεται και στον **Πίνακα 3** δεν βρέθηκαν αρκετά στατιστικά σημαντικά γονίδια και επομένως δεν μπορέσαμε να εξάγουμε κάποιο άλλο συμπέρασμα.

### 3.3 Αποτελέσματα της μετα-ανάλυσης στο ήπαρ

Όσο αναφορά την μετα-ανάλυση που πραγματοποιήθηκε στο ήπαρ δεν βρέθηκε κάποιο στατιστικά σημαντικό αποτέλεσμα. Όπως γαίνεται και στον **Πίνακα 5** ούτε η μετα-ανάλυση αλλά ούτε και κάποια μεμονωμένη μελέτη δεν έδωσε στατιστικά σημαντικά διαφορικά εκφραζόμενα γονίδια που να σχετίζονται με την νόσο που εξετάζουμε.

**Πίνακας 5:** Ο αριθμός των στατιστικά σημαντικών γονιδίων με  $p < 0.01$  και της μεθόδου διόρθωσης FDR που προκύπτουν από την μετα-ανάλυση όλων των μελετών συγκεντρωτικά και της κάθε μελέτης ξεχωριστά για το ήπαρ.

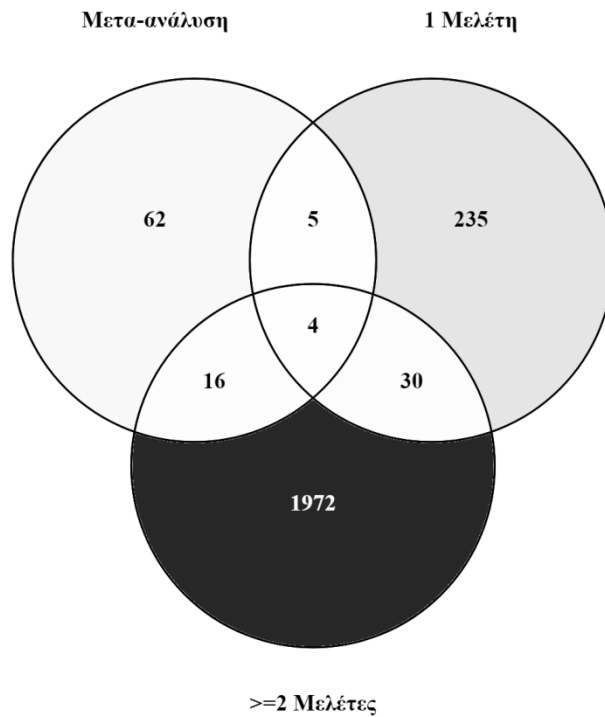
Liver	Μετα-ανάλυση	GSE15653	GSE23343	GSE64998
$p < 0.01$	373	295	581	454
FDR	0	0	0	0

### 3.4 Αποτελέσματα της μετα-ανάλυσης στα παγκρεατικά νησίδια

Η μετα-ανάλυση που πραγματοποιήθηκε για τα παγκρεατικά νησίδια εμφάνισε 87 στατιστικώς σημαντικά διαφορικά εκφραζόμενα γονίδια. Όπως φαίνεται και στον **Πίνακα 6**, στις μισές περιπτώσεις, δηλαδή σε τρεις από τις έξι μεμονωμένες μελέτες, τα στατιστικώς σημαντικά γονίδια που βρέθηκαν από την μετα-ανάλυση είναι παραπάνω, ενώ στις άλλες τρεις είναι κατά πολύ λιγότερα.

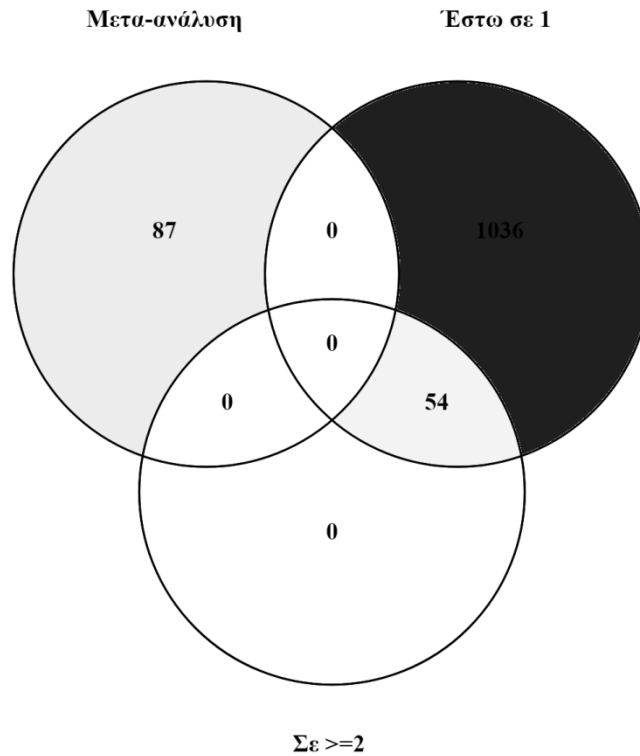
**Πίνακας 6:** Ο αριθμός των στατιστικά σημαντικών γονιδίων με  $p < 0.01$  και της μεθόδου διόρθωσης FDR που προκύπτουν από την μετα-ανάλυση όλων των μελετών συγκεντρωτικά και της κάθε μελέτης ξεχωριστά για τα παγκρεατικά νησίδια.

Pancreatic islets	Μετα-ανάλυση	GSE257	GSE44033_GSE38	GSE768	GSE768	E-CBIL-20_A-AFF-Y-33	E-CBIL-20_A-AFF-Y-33
$p < 0.01$	1038	4049	418	2710	1605	441	251
FDR	87	1090	0	986	274	0	0



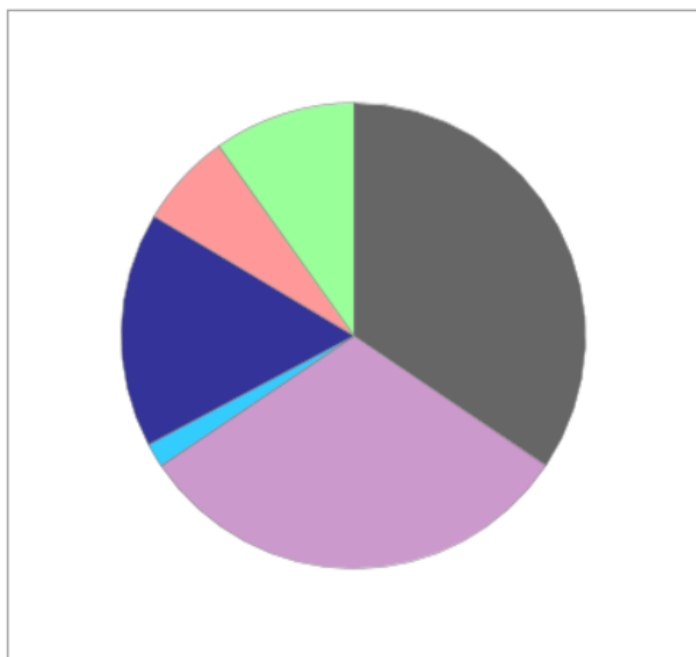
Εικόνα 17: Διάγραμμα Venn σύγκρισης των διαφορετικά εκφραζόμενων γονιδίων που προέκυψαν από την μετα-ανάλυση για τα παγκρεατικά νησίδια και από τις μελέτες GSE25724, GSE76894 και GSE76895 μεμονωμένα (οι οποίες διαθέτουν στατιστικά σημαντικά γονίδια από τη μέθοδο διόρθωσης FDR σε επίπεδο σημαντικότητας 1%).

Επιπλέον, για τα παγκρεατικά νησίδια υπολογίστηκαν τα ποσοστά IDR και IRR. Το IDR βρέθηκε ίσο με 71.2% (62 από τα 87 γονίδια βρέθηκαν μόνο από την μετα-ανάλυση) (Εικόνα 17) και το IRR βρέθηκε ίσο με 97.2% (2188 βρέθηκαν μόνο από τις μεμονωμένες μελέτες από τα 2250 γονίδια σε όλες τις μεμονωμένες μελέτες).



Εικόνα 18: Διάγραμμα Venn σύγκρισης των διαφορεικά εκφραζόμενων γονιδίων που προέκυψαν από την μετα-ανάλυση και από τις μεμονωμένες μελέτες για τα παγκρεατικά νησίδια. Συγκρίνονται τα αποτελέσματα που προέκυψαν την μετα-ανάλυση (87 γονίδια), με τα γονίδια τα οποία προσδιορίζονται σε τουλάχιστον μία μελέτη (1090 γονίδια) και σε τουλάχιστον δύο μελέτες (54 γονίδια).

Συνεχίζοντας, για τα παγκρεατικά νησίδια η PANTHER επέστρεψε πληροφορίες για 84 από τα 87 γονίδια που πήρε ως είσοδο. Πιο αναλυτικά, όσο αναφορά τη μοριακή λειτουργία των 61 γονιδίων (Εικόνα 19), 21 γονίδια συμμετέχουν διεργασίες δεσίματος ή αλλιώς δέσμευσης (34.4%) (όπως δέσμευση πρωτεϊνών, νουκλεϊκών οξέων, κ.α.). 19 γονίδια συμμετέχουν σε καταλυτικές διεργασίες (31.1%) (όπως ενζυμικές δράσεις, κατάλυση μεταλλικών ιόντων, κ.α.). 1 γονίδιο συμμετέχει σε διεργασίες μοριακού προσαρμογέα (1.6%). 10 γονίδια συμμετέχουν σε διεργασίες ρύθμισης της μοριακής λειτουργίας (16.4%). 4 γονίδια συμμετέχουν σε διαδικασίες μοριακού μετατροπέα (6.6%) και 6 γονίδια συμμετέχουν σε διαδικασίες μεταφοράς (9.8%) (όπως διαμεμβρανική μεταφορά αμινοξέων, μεταφορά λιπιδίων και άλλων μορίων, κ.α.).



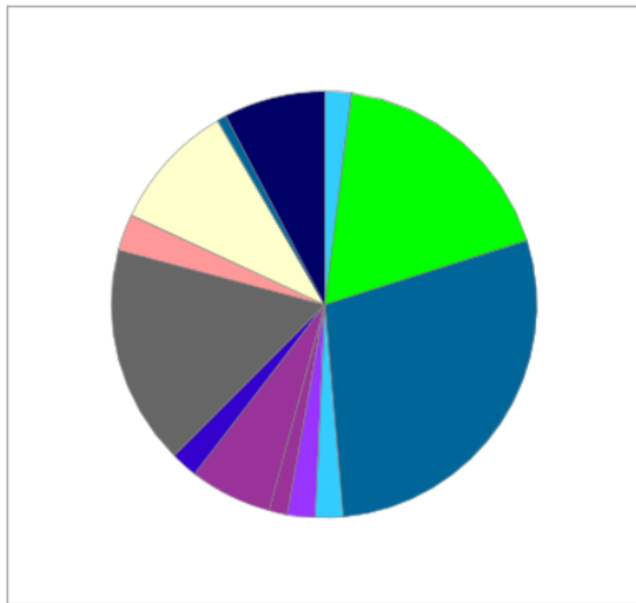
Click to get gene list for a category:

- [binding \(GO:0005488\)](#)
- [catalytic activity \(GO:0003824\)](#)
- [molecular adaptor activity \(GO:0060090\)](#)
- [molecular function regulator \(GO:0098772\)](#)
- [molecular transducer activity \(GO:0060089\)](#)
- [transporter activity \(GO:0005215\)](#)

Color picker powered by 

Εικόνα 19: Κυκλικό διάγραμμα των μοριακών λειτουργιών στις οποίες συμμετέχουν τα στατιστικώς σημαντικά γονίδια που προέκυψαν από την μετα-ανάλυση για τα παγκρεατικά νησίδια. Δεξιά είναι οι μοριακές λειτουργίες στις οποίες συμμετέχουν τα γονίδια.

Όσο αναφορά τις βιολογικές διαδικασίες των 84 γονιδίων τα αποτελέσματα που εμφανίστηκαν ήταν τα εξής. 41 γονίδια (28.5%) συμμετέχουν σε κυτταρικές διεργασίες (όπως κυτταρικός κύκλος, μεταφορά συστατικών κ.α.). 26 γονίδια (18.1%) συμμετέχουν σε βιολογικές ρυθμίσεις (όπως ομοιόσταση). 24 γονίδια (16.7%) συμμετέχουν στις μεταβολικές διαδικασίες (όπως μεταβολισμός φωσφορικών ενώσεων, ασβεστίου κ.α.). 14 γονίδια (9.7%) συμμετέχουν στη διαδικασία απόκρισης σε ερεθίσματα (απόκριση σε εξωτερικά (τσίμπημα) ή εσωτερικά ερεθίσματα). 9 γονίδια (6.3%) συμμετέχουν στον εντοπισμό. 1 γονίδιο (0.7%) συμμετέχει στην ρυθμική διαδικασία (όπως κερκάδικος ρυθμός) και τα υπόλοιπα γονίδια συμμετέχουν στις διεργασίες που φαίνονται παρακάτω στο διάγραμμα (**Εικόνα 20**).



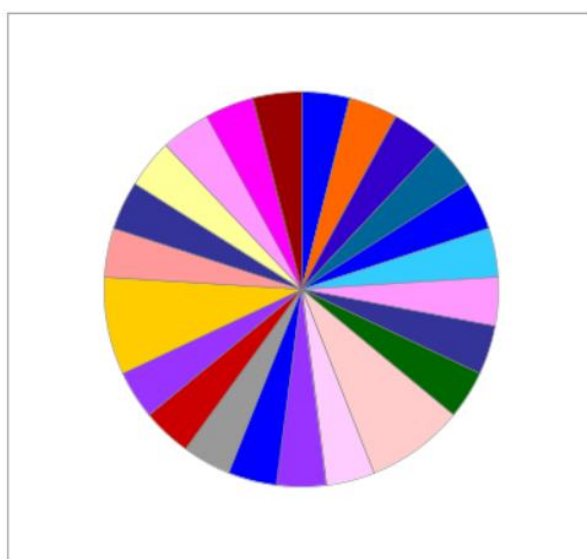
Click to get gene list for a category:

- [biological adhesion \(GO:0022610\)](#)
- [biological regulation \(GO:0065007\)](#)
- [cellular process \(GO:0009987\)](#)
- [developmental process \(GO:0032502\)](#)
- [immune system process \(GO:0002376\)](#)
- [interspecies interaction between organisms \(GO:0044419\)](#)
- [localization \(GO:0051179\)](#)
- [locomotion \(GO:0040011\)](#)
- [metabolic process \(GO:0008152\)](#)
- [multicellular organismal process \(GO:0032501\)](#)
- [response to stimulus \(GO:0050896\)](#)
- [rhythmic process \(GO:0048511\)](#)
- [signaling \(GO:0023052\)](#)

Color picker powered by Web Colors by VisiBone

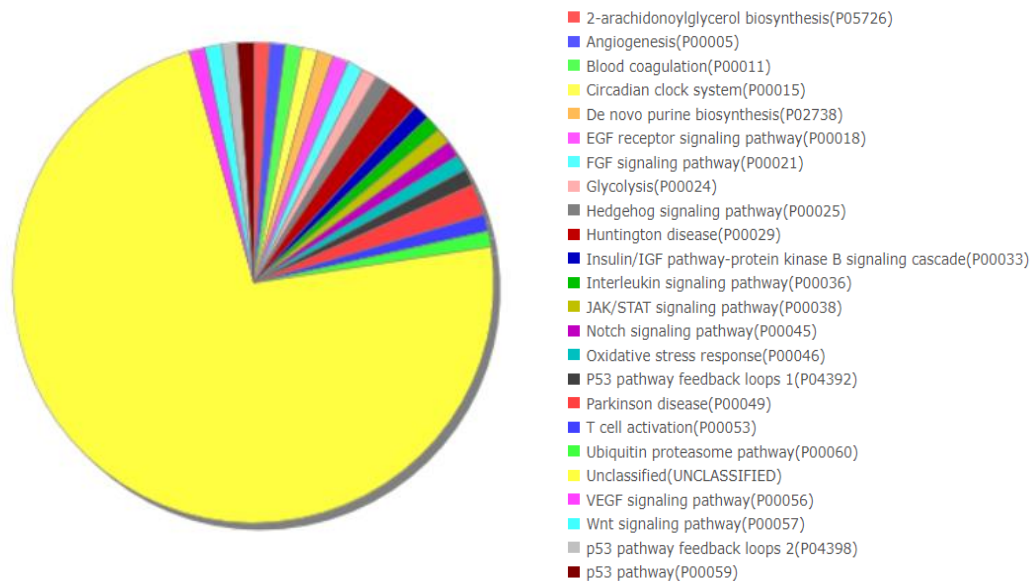
Εικόνα 20: Κυκλικό διάγραμμα των βιολογικών διεργασιών στις οποίες συμμετέχουν τα στατιστικώς σημαντικά γονίδια που προέκυψαν από την μετα-ανάλυση για τα παγκρεατικά νησίδια. Δεξιά είναι οι βιολογικές διαδικασίες στις οποίες συμμετέχουν τα γονίδια.

Ακόμα ένα πεδίο που μελετήθηκε ήταν τα μονοπάτια στα οποία εμπλέκονται τα 84 γονίδια. Οι πληροφορίες που αντλήθηκαν από αυτό το κομμάτι αναφέρονταν μόνο σε 25 από τα 84 γονίδια, τα οποία συμμετείχαν σε 23 μονοπάτια. Μερικά από αυτά είναι η αγγειογέννηση (1 γονίδιο (4.0%)), η γλυκόλυση (1 γονίδιο (4.0%)), η απόκριση οξειδωτικού στρες (1 γονίδιο (4.0%)), η νόσος του Huntington (2 γονίδια (8.0%)), η νόσος του Parkinson (2 γονίδια (8.0%)) και άλλα μονοπάτια που φαίνονται στο διάγραμμα (**Εικόνα 21**). Όπως φαίνεται και στο παρακάτω διάγραμμα (**Εικόνα 22**), μεγάλο είναι το ποσοστό που δεν συμμετέχει σε κανένα μονοπάτι ή τουλάχιστον σε κανένα γνωστό μονοπάτι.



- [2-arachidonoylglycerol biosynthesis \(P05726\)](#)
- [Angiogenesis \(P00005\)](#)
- [Blood coagulation \(P00011\)](#)
- [Circadian clock system \(P00015\)](#)
- [De novo purine biosynthesis \(P02738\)](#)
- [EGF receptor signaling pathway \(P00018\)](#)
- [FGF signaling pathway \(P00021\)](#)
- [Glycolysis \(P00024\)](#)
- [Hedgehog signaling pathway \(P00025\)](#)
- [Huntington disease \(P00029\)](#)
- [Insulin/IGF pathway-protein kinase B signaling cascade \(P00033\)](#)
- [Interleukin signaling pathway \(P00036\)](#)
- [JAK/STAT signaling pathway \(P00038\)](#)
- [Notch signaling pathway \(P00045\)](#)
- [Oxidative stress response \(P00046\)](#)
- [P53 pathway feedback loops 1 \(P04392\)](#)
- [Parkinson disease \(P00049\)](#)
- [T cell activation \(P00053\)](#)
- [Ubiquitin proteasome pathway \(P00060\)](#)
- [VEGF signaling pathway \(P00056\)](#)
- [Wnt signaling pathway \(P00057\)](#)
- [p53 pathway feedback loops 2 \(P04398\)](#)
- [p53 pathway \(P00059\)](#)

Εικόνα 21: Κυκλικό διάγραμμα των μονοπατιών στα οποία συμμετέχουν τα στατιστικώς σημαντικά γονίδια που προέκυψαν από την μετα-ανάλυση για τα παγκρεατικά νησίδια. Δεξιά είναι τα μονοπάτια στα οποία συμμετέχουν τα γονίδια.



Εικόνα 22: Κυκλικό διάγραμμα των μονοπατιών στα οποία συμμετέχουν τα στατιστικώς σημαντικά γονίδια που προέκυψαν από την μετα-ανάλυση για τα παγκρεατικά νησίδια. Δεξιά είναι τα μονοπάτια στα οποία συμμετέχουν τα γονίδια. Σε αυτό το διάγραμμα φαίνεται και το ποσοστό των γονιδίων που ή δεν εμπλέκονται σε κάποιο μονοπάτι ή είναι ακόμα άγνωστη η λειτουργία τους.

Με την βοήθεια της πλατφόρμας STRING πήραμε πληροφορίες σχετικές με τις αλληλεπιδράσεις μεταξύ των πρωτεϊνών που σχετίζονται με τα στατιστικώς σημαντικά γονίδια που βρέθηκαν από την μετα-ανάλυση στα παγκρεατικά νησίδια. Στην **Εικόνα 23** φαίνεται το δίκτυο των 82 γονιδίων που αναγνωρίστηκαν. Όπως φαίνεται, τα γονίδια ενώνονται με γραμμές διαφορετικών χρωμάτων. Αυτό συμβαίνει διότι κάθε χρώμα αντιπροσωπεύει διαφορετικό τύπο ένωσης. Πιο συγκεκριμένα, με την μαύρη γραμμή ενώνονται γονίδια των οποίων οι πρωτεΐνες συνεκφράζονται, με μπλε γονίδια των οποίων οι πρωτεΐνες συνυπάρχουν, με πράσινη γραμμή ενώνονται γειτονικά γονίδια, με κίτρινη γραμμή γονίδια των οποίων οι πρωτεΐνες βρίσκονται σε κοινά κείμενα και με ροζ γραμμή γονίδια των οποίων οι πρωτεΐνες έχουν βρεθεί σε κοινά πειράματα.



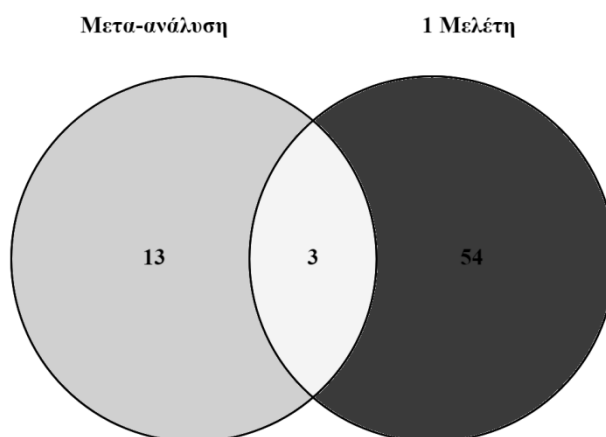


Όπως βλέπουμε, μόνο μία μελέτη έδωσε στατιστικά σημαντικά αποτελέσματα και όπως φαίνεται και στο παρακάτω διάγραμμα (**Εικόνα 24**) μόνο τρία είναι κοινά με αυτά που έδωσε η μετα-ανάλυση.

**Πίνακας 7:** Ο αριθμός των στατιστικά σημαντικών γονιδίων με  $p < 0.01$  και της μεθόδου διόρθωσης FDR που προκύπτουν από την μετα-ανάλυση όλων των μελετών συγκεντρωτικά και της κάθε μελέτης ξεχωριστά για τα μονοπύρηνα κύτταρα του περιφερικού αίματος.

PBMCs	Μετα-ανάλυση	GSE15932	GSE23561	GSE65561
$p < 0.01$	1364	2520	1266	644
FDR	16	57	0	0

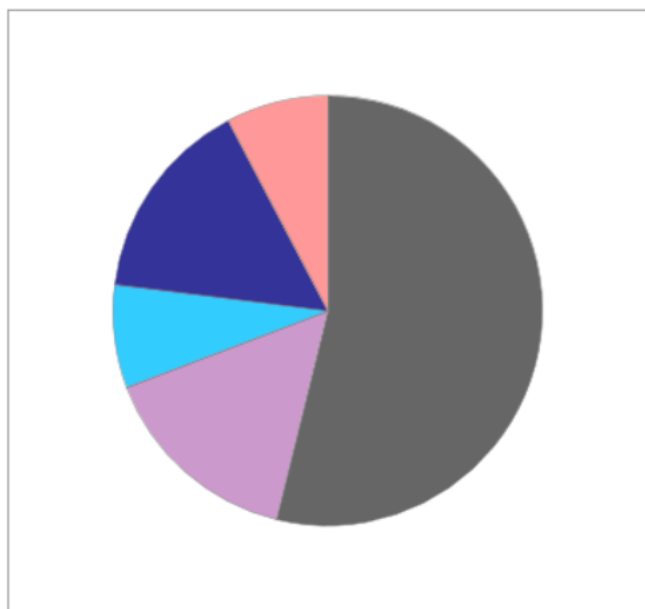
Η τιμή IDR για τα μονοπύρηνα κύτταρα του περιφερικού αίματος υπολογίστηκε ίση με 81.2% (13 από τα 16 γονίδια βρέθηκαν μόνο από την μετα-ανάλυση) και η τιμή του IRR είναι ίση με 94.7% (54 από τα 57 γονίδια μόνο από τις μεμονωμένες μελέτες).



Εικόνα 24: Διάγραμμα Venn σύγκρισης των διαφορεικά εκφραζόμενων γονιδίων που προέκυψαν από την μετα-ανάλυση για τα μονοπύρηνα κύτταρα και από τη μελέτη GSE15932 μεμονωμένα (οι οποίες διαθέτουν στατιστικά σημαντικά γονίδια από τη μέθοδο διόρθωσης FDR σε επίπεδο σημαντικότητας 1%).


Η PANTHER επέστρεψε δεδομένα για 13 από τα 16 γονίδια που κρίθηκαν ως στατιστικά σημαντικά από την μετα-ανάλυση για τα μονοπύρηνα κύτταρα του περιφερικού αίματος (PBMCs). Συνολικά, μελετήθηκαν τα εξής διαγράμματα: το διάγραμμα των μοριακών λειτουργιών, το διάγραμμα των βιολογικών διεργασιών και το διάγραμμα των

μονοπατιών όπου εμπλέκονται αυτά τα γονίδια. Όσο αναφορά τις μοριακές λειτουργίες των γονιδίων, 7 γονίδια εμπλέκονται σε διεργασίες δεσίματος (53.8%), από 2 γονίδια συμμετέχουν σε καταλυτικές διεργασίες (15.4%) και σε διεργασίες μοριακού προσαρμογέα (15.4%) και τέλος από 1 γονίδιο συμμετέχει σε διεργασίες μοριακού προσαρμογέα (7.7%) και σε διαδικασίες μοριακού μετατροπέα (7.7%).



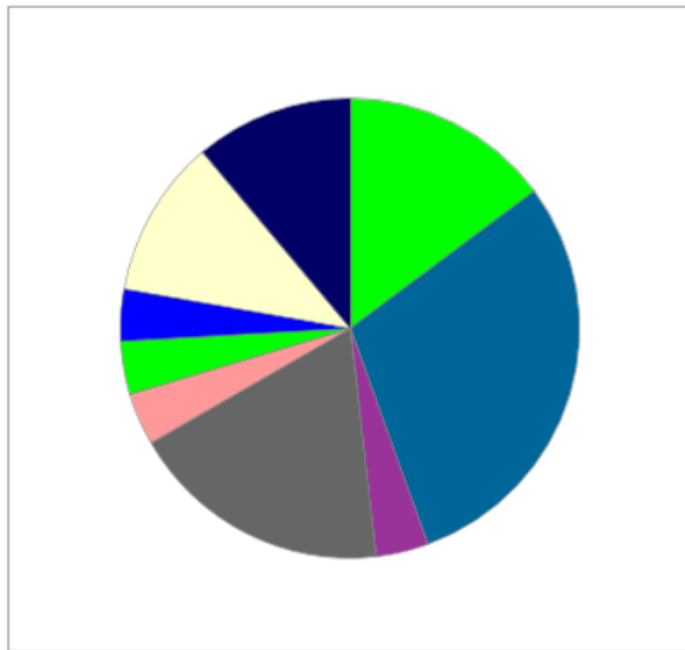
Click to get gene list for a category:

- [binding \(GO:0005488\)](#)
- [catalytic activity \(GO:0003824\)](#)
- [molecular adaptor activity \(GO:0060090\)](#)
- [molecular function regulator \(GO:0098772\)](#)
- [molecular transducer activity \(GO:0060089\)](#)

Color picker powered by 

Εικόνα 25: Κυκλικό διάγραμμα των μοριακών λειτουργιών στις οποίες συμμετέχουν τα στατιστικά σημαντικά γονίδια που προέκυψαν από την μετα-ανάλυση για τα PBMCs. Δεξιά είναι οι μοριακές λειτουργίες στις οποίες συμμετέχουν τα γονίδια.

Από το διάγραμμα των βιολογικών διεργασιών τα δεδομένα που εξάχθηκαν είναι τα παρακάτω. Όπως φαίνεται και στο παρακάτω διάγραμμα (**Εικόνα 26**) τα περισσότερα γονίδια, 8 σε αριθμό, συμμετέχουν σε κυτταρικές διεργασίες (29.6%). 5 γονίδια συμμετέχουν σε μεταβολικές διεργασίες (18.5%). 4 γονίδια συμμετέχουν σε βιολογικές ρυθμίσεις (14.8%). Από 3 γονίδια συμμετέχουν στη διαδικασία απόκρισης σε ερεθίσματα (11.1%) και στη διαδικασία σηματοδότησης (11.1%). Τέλος, από 1 γονίδιο συμμετέχει στη διαδικασία εντοπισμού (3.7%), στην αναπαραγωγή (3.7%), στην αναπαραγωγική διαδικασία (3.7%) και στην πολυκυτταρική οργανική διεργασία (3.7%).



Click to get gene list for a category:

- [biological regulation \(GO:0065007\)](#)
- [cellular process \(GO:0009987\)](#)
- [localization \(GO:0051179\)](#)
- [metabolic process \(GO:0008152\)](#)
- [multicellular organismal process \(GO:0032501\)](#)
- [reproduction \(GO:0000003\)](#)
- [reproductive process \(GO:0022414\)](#)
- [response to stimulus \(GO:0050896\)](#)
- [signaling \(GO:0023052\)](#)

Color picker powered by Web Colors by VisiBone

Εικόνα 26: Κυκλικό διάγραμμα των βιολογικών διεργασιών στις οποίες συμμετέχουν τα στατιστικώς σημαντικά γονίδια που προέκυψαν από την μετα-ανάλυση για τα **PBMCs**. Δεξιά είναι οι βιολογικές διαδικασίες στις οποίες συμμετέχουν τα γονίδια.

Αναφορικά με τα μονοπάτια όπου εμπλέκονται τα γονίδια που μελετήσαμε, οι πληροφορίες που ανακτήθηκαν αναφέρονταν μόνο σε 6 από τα 16 στατιστικώς σημαντικά γονίδια που δώσαμε ως είσοδο. Τα υπόλοιπα 10 γονίδια δεν φαίνεται να έχουν εμπλοκή σε κάποιο μονοπάτι ή δεν υπάρχουν σημαντικές πληροφορίες για αυτά στην πλατφόρμα της PANTHER. Συνολικά, από 1 γονίδιο συμμετέχει σε μία διεργασία. Δηλαδή, 1 γονίδιο συμμετέχει στη διαδικασία ενεργοποίησης των B κυττάρων (16.7%), 1 γονίδιο συμμετέχει στο μονοπάτι του υποδοχέα ορμόνης απελευθέρωσης της γοναδοτροπίνης (16.7%), 1 γονίδιο συμμετέχει στη φλεγμονή που μεσολαβεί από την σηματοδοτική οδό χημειοκίνης και κυτοκίνης (16.7%), 1 γονίδιο συμμετέχει στη μεταβολική τροχολογία ομάδας III υπογοχέα γλουταμινικού (16.7%), 1 γονίδιο συμμετέχει στη διαδικασία ενεργοποίησης των T κυττάρων (16.7%) και 1 γονίδιο συμμετέχει στη WNT διαδικασία σηματοδότησης (16.7%).

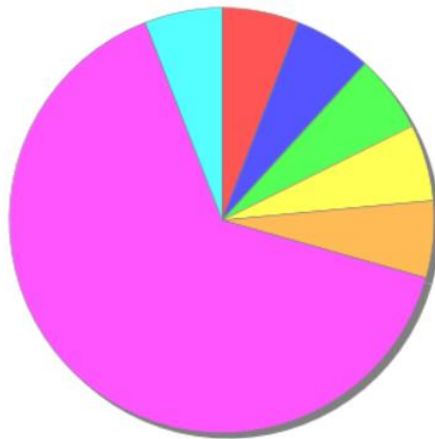


Click to get gene list for a category:

- [B cell activation \(P00010\)](#)
- [Gonadotropin-releasing hormone receptor pathway \(P06664\)](#)
- [Inflammation mediated by chemokine and cytokine signaling pathway \(P00031\)](#)
- [Metabotropic glutamate receptor group III pathway \(P00039\)](#)
- [T cell activation \(P00053\)](#)
- [Wnt signaling pathway \(P00057\)](#)

Color picker powered by Web Colors by VisiBone

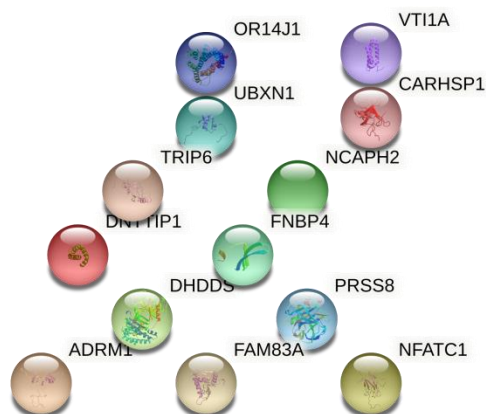
Εικόνα 27: Κυκλικό διάγραμμα των μονοπατιών στα οποία συμμετέχουν τα στατιστικώς σημαντικά γονίδια που προέκυψαν από την μετα-ανάλυση για τα PBMCs. Δεξιά είναι τα μονοπάτια στα οποία συμμετέχουν τα γονίδια.



- B cell activation(P00010)
- Gonadotropin-releasing hormone receptor pathway(P06664)
- Inflammation mediated by chemokine and cytokine signaling pathway(P00031)
- Metabotropic glutamate receptor group III pathway(P00039)
- T cell activation(P00053)
- Unclassified(UNCLASSIFIED)
- Wnt signaling pathway(P00057)

Εικόνα 28: Κυκλικό διάγραμμα των μονοπατιών στα οποία συμμετέχουν τα στατιστικώς σημαντικά γονίδια που προέκυψαν από την μετα-ανάλυση για τα PBMCs. Δεξιά είναι τα μονοπάτια στα οποία συμμετέχουν τα γονίδια. Σε αυτό το διάγραμμα φαίνεται και το ποσοστό των γονιδίων που ή δεν εμπλέκονται σε κάποιο μονοπάτι ή είναι ακόμα άγνωστη η λειτουργία τους.

Όπως και πριν, έγινε έλεγχος και στην βάση δεδομένων STRING, η οποία αναγνώρισε επίσης 13 από τα 16 στατιστικώς σημαντικά γονίδια που πήρε ως είσοδο. Παρόλα ταύτα, δεν φάνηκε να υπάρχει καμία σύνδεση ή αλληλεπίδραση μεταξύ των γονιδίων.



Εικόνα 29: Δίκτυο πρωτεϊνικών αλληλεπιδράσεων των στατιστικά σημαντικών γονιδίων που προέκυψαν από την μετα-ανάλυση στα PBMCs.

### 3.6 Αποτελέσματα της μετα-ανάλυσης στον σκελετικό μυ

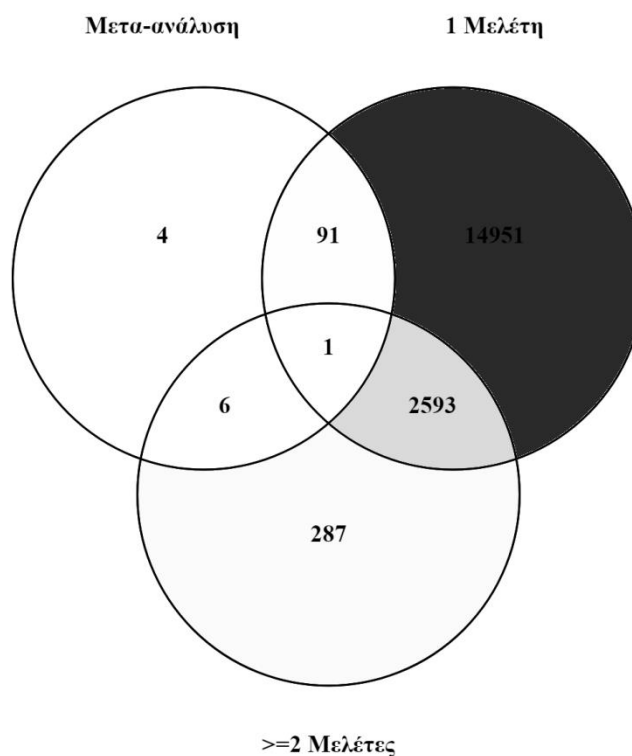
Ο τελευταίος ιστός που μελετήθηκε ήταν ο σκελετικός μυς. Τα αναλυτικά αποτελέσματα που προέκυψαν από την μετα-ανάλυση και από τις μεμονωμένες μελέτες παρουσιάζονται στον **Πίνακα 8**. Τα στατιστικά σημαντικά διαφορικά εκφραζόμενα γονίδια που προέκυψαν από την μετα-ανάλυση είναι 102 και όπως φαίνεται πολλές είναι οι περιπτώσεις που ο αριθμός τους ξεπερνάει το αριθμό των σημαντικών γονιδίων που προήλθαν από κάποια μεμονωμένη μελέτη.

**Πίνακας 8:** Ο αριθμός των στατιστικά σημαντικών γονιδίων με  $p < 0.01$  και της μεθόδου διόρθωσης FDR που προκύπτουν από την μετα-ανάλυση όλων των μελετών συγκεντρωτικά και της κάθε μελέτης ξεχωριστά για το σκελετικό μυ.

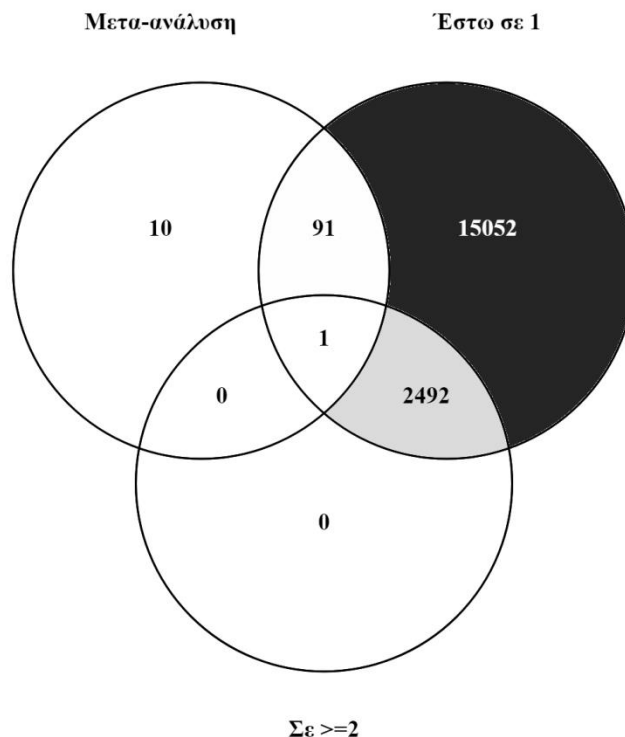
Skeletal muscle	$p < 0.01$	FDR
Μετα-ανάλυση	502	102
GSE12643	69	0
GSE18732	253	0
GSE19420	720	2
GSE21340	25	0
GSE22309	3446	2761
GSE25462	782	8
GSE29221	2315	153
GSE55650	888	0
GSE59363	311	0
GSE73034	823	0
GSE156249	17640	17636

E-CBIL-30	83	0
E-MEXP-1270	752	114
E-MEXP-2559	1572	30

Στο παρακάτω διάγραμμα (Εικόνα 30) φαίνεται ότι τα στατιστικώς σημαντικά διαφορικά εκφραζόμενα γονίδια που προέκυψαν μόνο από την μετα-ανάλυση είναι μόνο 4 από τα συνολικά 102. Χρησιμοποιώντας αυτό το αποτέλεσμα, μπορούμε να προσδιορίσουμε το ποσοστό IDR, το οποίο είναι ίσο με 3.92%. Επιπλέον, υπολογίζουμε και το αντίστοιχο ποσοστό IRR, το οποίο είναι ίσο με 99.5% και υποδικνύει ότι 20602 στατιστικώς σημαντικά γονίδια από τα συνολικά 20704 βρέθηκαν μόνο από τις μεμονωμένες μελέτες και όχι από την μετα-ανάλυση.



Εικόνα 30: Διάγραμμα Venn σύγκρισης των διαφορικά εκφραζόμενων γονιδίων που προέκυψαν από την μετα-ανάλυση για το σκελετικό μυ και από τις μελέτες GSE22309, GSE29221 και GSE156249 μεμονωμένα (οι οποίες διαθέτουν στατιστικώς σημαντικά γονίδια από τη μέθοδο διόρθωσης FDR σε επίπεδο σημαντικότητας 1%).

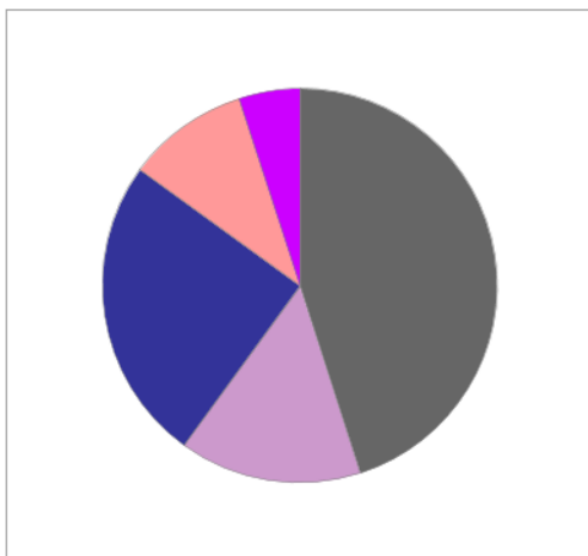


Εικόνα 31: Διάγραμμα Venn σύγκρισης των διαφορεικά εκφραζόμενων γονιδίων που προέκυψαν από την μετα-ανάλυση και από τις μεμονωμένες μελέτες για το σκελετικό ιστό. Συγκρίνονται τα αποτελέσματα που προέκυψαν την μετα-ανάλυση (102 γονίδια), με τα γονίδια τα οποία προσδιορίζονται σε τουλάχιστον μία μελέτη (17636 γονίδια) και σε τουλάχιστον δύο μελέτες (2493 γονίδια).

Εισάγοντας αυτά τα δεδομένα στην PANTHER, οι πληροφορίες που πήραμε αναφέρονταν μόνο για 17 από τα 102 στατιστικώς σημαντικά. Εδώ είναι σημαντικό να αναφερθεί ότι για τον σκελετικό μυ διεκπεραιώθηκαν δύο μετα-αναλύσεις. Αυτή η παρέκλιση αφορά την τελευταία μελέτη που προστέθηκε στο σύνολο, την E-MEXP-2559, η οποία εξέταζε τρεις ομάδες, τους ασθενείς με διαβήτη τύπου II, τους συγγενείς πρώτου βαθμού των προηγούμενων και μεμονωμένους υγείς. Στην πρώτη μετα-ανάλυση περιλήφθηκαν όλοι και τα αποτελέσματα μας έδειξαν 102 στατιστικώς σημαντικά γονίδια. Στην δεύτερη μετα-ανάλυση αφαιρέθηκαν 14 υγιή άτομα (οι συγγενείς πρώτου βαθμού) και τα αποτελέσματα που πήραμε ήταν ακριβώς τα ίδια χωρίς καμία αλλαγή.

Όπως φαίνεται και στην παρακάτω εικόνα (**Εικόνα 32**), το διάγραμμα απεικονίζει τις μοριακές λειτουργίες των 17 γονιδίων. Αναλυτικότερα, 9 γονίδια συμμετέχουν διεργασίες δεσίματος (45.0%). 3 γονίδια συμμετέχουν σε καταλυτικές διεργασίες (15.0%). 5 γονίδια συμμετέχει σε διεργασίες μοριακού προσαρμογέα (25.0%). 2 γονίδια συμμετέχουν σε διαδικασίες μοριακού μετατροπέα (10.0%) και 1 γονίδιο συμμετέχει σε δομικές μοριακές δράσεις (5.0%) (όπως δέσμευση ιόντων ασβεστίου κ.α.).





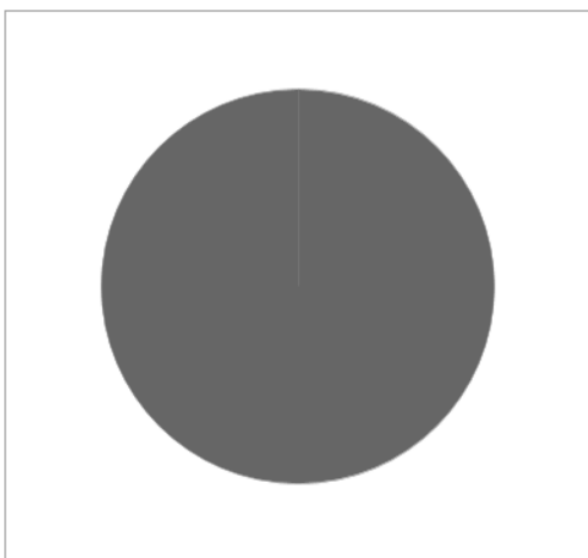
Click to get gene list for a category:

- [binding \(GO:0005488\)](#)
- [catalytic activity \(GO:0003824\)](#)
- [molecular function regulator \(GO:0098772\)](#)
- [molecular transducer activity \(GO:0060089\)](#)
- [structural molecule activity \(GO:0005198\)](#)

Color picker powered by Web Colors by VisiBone

Εικόνα 32: Κυκλικό διάγραμμα των μοριακών λειτουργιών στις οποίες συμμετέχουν τα στατιστικώς σημαντικά γονίδια που προέκυψαν από την μετα-ανάλυση για το σκελετικό μυ. Δεξιά είναι οι μοριακές λειτουργίες στις οποίες συμμετέχουν τα γονίδια.

Επιπλέον, και για αυτόν τον ιστό μελετήθηκε το διάγραμμα των βιολογικών διεργασιών, αλλά και τα μονοπάτια που εμπλέκονται τα 17 γονίδια. Όσο αναφορά τα μονοπάτια είχαμε πληροφορία για ένα μόνο γονίδιο, όπως φαίνεται και στο διάγραμμα παρακάτω (**Εικόνα 33**) και όλα τα υπόλοιπα γονίδια δεν έχουν γνωστή δράση, για αυτό το λόγο ανήκουν στην ομάδα των μη ταξινομημένων (**Εικόνα 34**). Το γονίδιο αυτό εμπλέκεται στο μονοπάτι αποικοδόμησης της 5-υδροξυτρυπταμίνης.

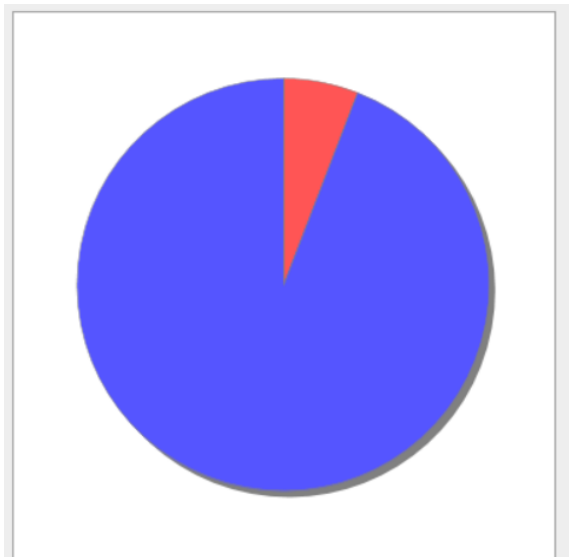


Click to get gene list for a category:

- [5-Hydroxytryptamine degradation \(P04372\)](#)

Color picker powered by Web Colors by VisiBone

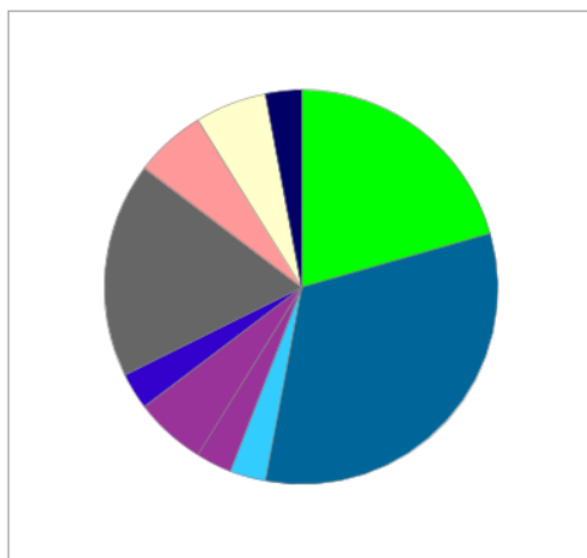
Εικόνα 33: Κυκλικό διάγραμμα του μονοπατιού στο οποίο συμμετέχει το στατιστικώς σημαντικό γονίδιο που προέκυψαν από την μετα-ανάλυση για το σκελετικό μυ. Δεξιά είναι το μονοπάτι στο οποίο συμμετέχει το γονίδιο.



■ 5-Hydroxytryptamine degradation(P04372)  
 ■ Unclassified(UNCLASSIFIED)

Εικόνα 34: Κυκλικό διάγραμμα των μονοπατιών στα οποία συμμετέχουν τα στατιστικά σημαντικά γονίδια που προέκυψαν από την μετα-ανάλυση για το σκελετικό μυ. Δεξιά είναι το αναγνωρισμένο μονοπάτι στο οποίο συμμετέχει το γονίδιο και τα υπόλοιπα που δεν έχουν ταξινομηθεί..

Από αυτό το διάγραμμα των βιολογικών διεργασιών πήραμε ότι 11 γονίδια (32.4%) συμμετέχουν σε κυτταρικές διεργασίες. 7 γονίδια (20.6%) συμμετέχουν σε βιολογικές ρυθμίσεις. 6 γονίδια (17.6%) συμμετέχουν στις μεταβολικές διαδικασίες. 2 γονίδια (5.9%) συμμετέχουν στη διαδικασία απόκρισης σε ερεθίσματα. 2 γονίδια (5.9%) συμμετέχουν στον εντοπισμό. 1 γονίδιο (2.9%) συμμετέχει στην αναπτυξιακή διαδικασία (όπως κυτταρικός θάνατος, εμβρυακή ανάπτυξη, διαφοροποίηση κ.α.) και τα υπόλοιπα συμμετέχουν στις διεργασίες που φαίνονται παρακάτω στο διάγραμμα (Εικόνα 35).



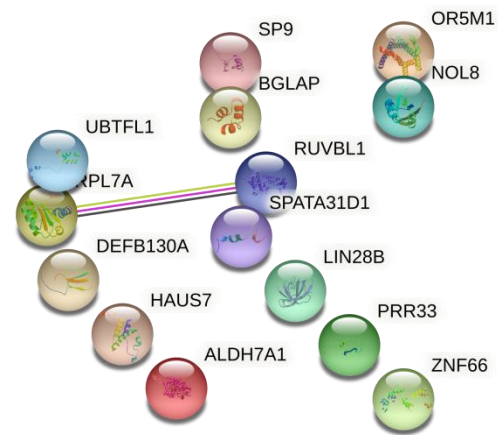
Click to get gene list for a category:

- [biological regulation \(GO:0065007\)](#)
- [cellular process \(GO:0009987\)](#)
- [developmental process \(GO:0032502\)](#)
- [interspecies interaction between organisms \(GO:0044419\)](#)
- [localization \(GO:0051179\)](#)
- [locomotion \(GO:0040011\)](#)
- [metabolic process \(GO:0008152\)](#)
- [multicellular organismal process \(GO:0032501\)](#)
- [response to stimulus \(GO:0050896\)](#)
- [signaling \(GO:0023052\)](#)

Color picker powered by Web Colors by VisiBone

Εικόνα 35: Κυκλικό διάγραμμα των βιολογικών διεργασιών στις οποίες συμμετέχουν τα στατιστικά σημαντικά γονίδια που προέκυψαν από την μετα-ανάλυση για το σκελετικό μυ. Δεξιά είναι οι βιολογικές διαδικασίες στις οποίες συμμετέχουν τα γονίδια.

Προχωρώντας στην STRING, τα γονίδια που αναγνωρίστηκαν και επιστράφηκαν πληροφορίες για αυτά ήταν μόνο 14 από τα 102. Παρακάτω φαίνεται και το πρωτεϊνικό δίκτυο που δημιουργήθηκε, όπως επίσης και ότι υπάρχει ένωση μόνο ανάμεσα σε δύο πρωτεΐνες.



Εικόνα 36: Δίκτυο πρωτεϊνικών αλληλεπιδράσεων των στατιστικά σημαντικών γονιδίων που προέκυψαν από την μετα-ανάλυση στο σκελετικό μυ.

## **Κεφάλαιο 4: Συζήτηση**

Στην παρούσα διπλωματική εργασία, υλοποιήθηκε μετα-ανάλυση μικροσυστοιχιών σε πέντε ιστούς για την εύρεση γονιδίων που εκφράζονται διαφορεικά στο διαβήτη τύπου II. 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.

Τα γονίδια που κρίθηκαν στατιστικά σημαντικά από την μέθοδο διόρθωσης FDR σε επίπεδο σημαντικότητας 1% ήταν για κάθε ιστό τα εξής : λιπώδης ιστός ► 2 γονίδια, ήπαρ ► 0 γονίδια, παγκρεατικά νησίδια ► 87 γονίδια, μονοπύρηννα κύτταρα του περιφερικού αίματος ► 16 γονίδια και τέλος σκελετικός μυς ► 102 γονίδια.

Τα γονίδια εμφανίστηκαν να παίρνουν μέρος σε πληθώρα βιολογικών και μοριακών διεργασιών όπως διεργασίες εντοπισμού και σηματοδότησης, μεταβολικές διεργασίες, δέσμευση, καταλυτικές δραστηριότητες και λειτουργίες προσαρμογέων. Επιπλέον, εμπλέκονται και σε διάφορα μεταβολικά μονοπάτια όπως η αγγειογέννηση, γλυκόλυση, η απόκριση οξειδωτικού στρες, η νόσος του Huntington και η νόσος του Parkinson, η ενεργοποίηση των B και των T κυττάρων κ.α..

Παρόλα τα δικά μας ευρήματα όμως, κρίνεται αναγκαία η περαιτέρω έρευνα για την διερεύνηση της συσχέτισης των γονιδίων που βρέθηκαν στατιστικώς σημαντικά στην συγκεκριμένη μελέτη με την νόσο του διαβήτη τύπου II.

## **Βιβλιογραφία**

Παναγιώτα Κοντού, Διδακτορική Διατριβή, Πανεπιστήμιο Θεσσαλίας, Σχολή Θετικών Επιστημών, Τμήμα Πληροφορικής Με Εφαρμογές Στη Βιοϊατρική, 'Ανάλυση Δεδομένων Γονιδιακής Έκφρασης και Πολύπλοκων Βιολογικών Δικτύων', Λαμία, Δεκέμβριος 2016

Κωνσταντίνα Βέννου, 'Διατριβή: Μεθοδολογίες υπολογιστικής ανάλυσης και μετα-ανάλυσης δεδομένων γονιδιακής έκφρασης - Κωδικός: 47552', Λαμία 2020  
<https://thesis.ekt.gr/thesisBookReader/id/47552#page/1/mode/2up>

Αθανάσιος Σ. Μπαλτζής, ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ 'Ανάλυση του εναλλακτικού ματίσματος από δεδομένα RNA – seq στο μετάλλαγμα Iefkothea του φυτού Arabidopsis.', Εργαστήριο Μοριακής Βιολογίας, Τμήμα Βιοτεχνολογίας, Γεωπονικό Πανεπιστήμιο Αθηνών, Αθήνα 2018,  
[http://dspace.aua.gr/xmlui/bitstream/handle/10329/6735/Baltzis\\_A.pdf?sequence=3](http://dspace.aua.gr/xmlui/bitstream/handle/10329/6735/Baltzis_A.pdf?sequence=3)

Κωνσταντίνα Βέννου, ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ «ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΙΑΤΡΙΚΗ» ΜΕΤΑ-ΑΝΑΛΥΣΗ ΛΕΛΟΜΕΝΩΝ ΓΟΝΙΔΙΑΚΗΣ ΕΚΦΡΑΣΗΣ, ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ, Υπεύθυνος Παντελής Μπάγκος Αναπληρωτής Καθηγητής, 2016

bioDBnet - Biological Database Network, <https://biodbnet-abcc.ncifcrf.gov/db/db2db.php>

Newson 2013, "Bonferroni and Holm approximations for Sidák and Holland-Copenhaver q-values", The Stata Journal, 379-381, 13(2), 2013

eClass Πανεπιστημίου Θεσσαλίας | ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ ΙΙ | Ανάλυση Μικροσυστοιχιών DNA, [https://eclass.uth.gr/modules/units/?course=DIB\\_U\\_191&id=133](https://eclass.uth.gr/modules/units/?course=DIB_U_191&id=133)

Multi-symbol checker | HUGO Gene Nomenclature Committee, <https://www.genenames.org/tools/multi-symbol-checker/>

PANTHER - Gene List Analysis, <http://pantherdb.org/?fbclid=IwAR38m3eu2JIECJ0mKPQxo9LSizpZBpvsnT1Xgwf5JDmA8vKQ7QTz9QtpAgA>

Szklarczyk 2019, Damian Szklarczyk, Annika L. Gable, David Lyon, Alexander Junge, Stefan Wyder, Jaime Huerta-Cepas, Milan Simonovic, Nadezhda T. Doncheva, John H. Morris, Peer Bork, Lars J. Jensen, Christian Von Mering, “STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets”, doi: 10.1093/NAR/GKY1131, D1, Nucleic Acids Research, 1, D607-D613, Oxford University Press, 47, 2019

STRING: functional protein association network, [https://string-db.org/cgi/input?sessionId=bC7YIw7NVL2l&input\\_page\\_show\\_search=on](https://string-db.org/cgi/input?sessionId=bC7YIw7NVL2l&input_page_show_search=on)

H Joseph Newton, Nicholas J Cox, Christopher Baum, Boston College, Rino Bellocco, Karolinska Institutet, David Clayton, Joanne M Garrett, James Hardin, Stephen Jenkins, Jens Lauritsen, J Scott Long, Thomas Lumley, Roger Newson, Marcello Pagano, Sophia Rabe-Hesketh, Philip Ryan, Mark E Schaffer, Jeffrey Wooldridge, “The Stata Journal”, isbn: 9796964600, <http://www.stata-journal.com>

N. Κατσίκη, “Σακχαρώδης διαβήτης : Διάγνωση και ταξινόμηση”, Ελληνικά Διαβητολογικά Χρονικά 23, 1: 78-86, 2010, [https://www.ngda.gr/innet/UsersFiles/ngda/documents/Xronika/2010\\_1o/02\\_Katsiki.pdf](https://www.ngda.gr/innet/UsersFiles/ngda/documents/Xronika/2010_1o/02_Katsiki.pdf)

Kallipos - Chapter07\_geneexpression\_R.pdf, [https://repository.kallipos.gr/pdfviewer/web/viewer.html?file=/bitstream/11419/1585/1/Chapter07\\_geneexpression\\_R.pdf](https://repository.kallipos.gr/pdfviewer/web/viewer.html?file=/bitstream/11419/1585/1/Chapter07_geneexpression_R.pdf)

17b\_goulis\_meta\_analysi.pdf  
[http://promesi.med.auth.gr/mathimata/17b\\_goulis\\_meta\\_analysi.pdf](http://promesi.med.auth.gr/mathimata/17b_goulis_meta_analysi.pdf)

Palsgaard 2009, “Gene Expression in Skeletal Muscle Biopsies from People with Type 2 Diabetes and Relatives: Differential Regulation of Insulin Signaling Pathways”, doi: 10.1371/JOURNAL.PONE.0006575, issn: 1932-6203, 8, PLOS ONE, e6575, Public Library of Science, 2009, <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0006575>

Chibalin AV, Leng Y, Vieira E, Krook A, Björnholm M, Long YC, Kotova O, Zhong Z, Sakane F, Steiler T, Nylén C, Wang J, Laakso M, Topham MK, Gilbert M, Wallberg-Henriksson H, Zierath JR, “Downregulation of diacylglycerol kinase delta contributes to hyperglycemia-induced insulin resistance”, doi: 10.1016/J.CELL.2007.12.035, issn: 1097-



4172, Cell, pmid: 18267070, 2, 375-386, Cell, 132, 2008, <https://pubmed.ncbi.nlm.nih.gov/18267070/>

Supriyo De, Yongqing Zhang, John R. Garner, S. Alex Wang, Kevin G. Becker, “Disease and phenotype gene set analysis of disease-based gene expression in mouse and human”, doi: 10.1152/PHYSIOLGENOMICS.00008.2010, 2, American Physiological Society Bethesda, MD, 162-167, 42 A, 2010, <https://doi.org/10.1152/physiolgenomics.00008.2010>

Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J , Puigserver P, Carlsson E, Ridderstråle M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC, “PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.”, doi: 10.1038/NG1180, issn: 1061-4036, 3, Nature Genetics, 267-273, pmid: 12808457, 34, 2003, <http://europepmc.org/article/MED/12808457>

Gunton JE, Kulkarni RN, Yim S, Okada T, Hawthorne WJ, Tseng YH, Roberson RS, Ricordi C, O'Connell PJ, Gonzalez FJ, Kahn CR, “Loss of ARNT/HIF1beta mediates altered gene expression and pancreatic-islet dysfunction in human type 2 diabetes.”, doi: 10.1016/J.CELL.2005.05.027, issn: 0092-8674, 3, Cell, 337-349, pmid: 16096055, Elsevier B.V., 122, 2005, <http://europepmc.org/article/MED/16096055>

Nascimento EBM, Hangelbroek RWJ, Hooiveld GJEJ, Hoeks J, Van Marken Lichtenbelt WD, Hesselink MHC, Schrauwen P, Kersten S, “Comparative transcriptome analysis of human skeletal muscle in response to cold acclimation and exercise training in human volunteers”, doi: 10.1186/S12920-020-00784-Z, issn: 1755-8794, 1, BMC medical genomics, pmid: 32887608, BMC Med Genomics, 13, 2020, <https://pubmed.ncbi.nlm.nih.gov/32887608/>

Chaudhuri R, Khoo PS, Tonks K, Junutula JR, Kolumam G, Modrusan Z, Samocho-Bonet D, Meoli CC, Hocking S, Fazakerley DJ, Stöckli J, Hoehn KL, Greenfield JR, Yang JYH, James DE, “Cross-species gene expression analysis identifies a novel set of genes implicated in human insulin sensitivity”, doi: 10.1038/NPJSBA.2015.10, issn: 2056-7189, NPJ systems biology and applications, pmid: 28725461, NPJ Syst Biol Appl, 1, 2015. <https://pubmed.ncbi.nlm.nih.gov/28725461/>

Hansen JS, Zhao X, Irmeler M, Liu X, Hoene M, Scheler M, Li Y, Beckers J, Hrabě de Angelis M, Häring HU, Pedersen BK, Lehmann R, Xu G, Plomgaard P, Weigert C, “Type 2

diabetes alters metabolic and transcriptional signatures of glucose and amino acid metabolism during exercise and recovery”, doi: 10.1007/S00125-015-3584-X, issn: 1432-0428, 8, Diabetologia, 1845-1854, pmid: 26067360, Diabetologia, 58, 2015, <https://pubmed.ncbi.nlm.nih.gov/26067360/>

Brown AE, Palsgaard J, Borup R, Avery P, Gunn DA, De Meyts P, Yeaman SJ, Walker M, “p38 MAPK activation upregulates proinflammatory pathways in skeletal muscle cells from insulin-resistant type 2 diabetic patients”, doi: 10.1152/AJPENDO.00115.2014, issn: 1522-1555, 1, American journal of physiology. Endocrinology and metabolism, E63-E70, pmid: 25370850, Am J Physiol Endocrinol Metab, 308, 2015, <https://pubmed.ncbi.nlm.nih.gov/25370850/>

Lerin C, Goldfine AB, Boes T, Liu M, Kasif S, Dreyfuss JM, De Sousa-Coelho AL, Daher G, Manoli I, Sysol JR, Isganaitis E, Jessen N, Goodyear LJ, Beebe K, Gall W, Venditti CP, Patti ME, “Defects in muscle branched-chain amino acid oxidation contribute to impaired lipid metabolism”, doi: 10.1016/J.MOLMET.2016.08.001, issn: 2212-8778, 10, Molecular metabolism, 926-936, pmid: 27689005, Mol Metab, 5, 2016, <https://pubmed.ncbi.nlm.nih.gov/27689005/>

Jin W, Goldfine AB, Boes T, Henry RR, Ciaraldi TP, Kim EY, Emecan M, Fitzpatrick C, Sen A, Shah A, Mun E, Vokes V, Schroeder J, Tatro E, Jimenez-Chillaron J, Patti ME, “Increased SRF transcriptional activity in human and mouse skeletal muscle is a signature of insulin resistance”, doi: 10.1172/JCI41940, issn: 1558-8238, 3, The Journal of clinical investigation, 918-929, pmid: 21393865, J Clin Invest, 121, 2011, <https://pubmed.ncbi.nlm.nih.gov/21393865/>

Wu X, Patki A, Lara-Castro C, Cui X, Zhang K, Walton RG, Osier MV, Gadbury GL, Allison DB, Martin M, Garvey WT, “Genes and biochemical pathways in human skeletal muscle affecting resting energy expenditure and fuel partitioning”, doi: 10.1152/JAPPLPHYSIOL.00293.2010, issn: 1522-1601, 3, Journal of applied physiology (Bethesda, Md. : 1985), 746-755, pmid: 21109598, J Appl Physiol (1985), 110, 2011, <https://pubmed.ncbi.nlm.nih.gov/21109598/>

Wu X, Wang J, Cui X, Maianu L, Rhees B, Rosinski J, So WV, Willi SM, Osier MV, Hill HS, Page GP, Allison DB, Martin M, Garvey WT, “The effect of insulin on expression of genes and biochemical pathways in human skeletal muscle”, doi: 10.1007/S12020-007-0007-

X, issn: 1355-008X, 1, Endocrine, 5-17, pmid: 17709892, Endocrine, 31, 2007, <https://pubmed.ncbi.nlm.nih.gov/17709892/>

Patti ME, Butte AJ, Crunkhorn S, Cusi K, Berria R, Kashyap S, Miyazaki Y, Kohane I, Costello M, Saccone R, Landaker EJ, Goldfine AB, Mun E, DeFronzo R, Finlayson J, Kahn CR, Mandarino LJ, “Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1”, doi: 10.1073/PNAS.1032913100, issn: 0027-8424, 14, Proceedings of the National Academy of Sciences of the United States of America, 8466-8471, pmid: 12832613, Proc Natl Acad Sci U S A, 100, 2003, <https://pubmed.ncbi.nlm.nih.gov/12832613/>

Van Tienen FH, Praet SF, de Feyter HM, van den Broek NM, Lindsey PJ, Schoonderwoerd KG, de Coo IF, Nicolay K, Prompers JJ, Smeets HJ, van Loon LJ, “Physical activity is the key determinant of skeletal muscle mitochondrial function in type 2 diabetes”, doi: 10.1210/JC.2011-3454, issn: 1945-7197, 9, The Journal of clinical endocrinology and metabolism, 3261-3269, pmid: 22802091, J Clin Endocrinol Metab, 97, 2012, <https://pubmed.ncbi.nlm.nih.gov/22802091/>

Gallagher IJ, Scheele C, Keller P, Nielsen AR, Remenyi J, Fischer CP, Roder K, Babraj J, Wahlestedt C, Hutvagner G, Pedersen BK, Timmons JA, “Integration of microRNA changes in vivo identifies novel molecular features of muscle insulin resistance in type 2 diabetes”, doi: 10.1186/GM130, issn: 1756-994X, 2, Genome medicine, pmid: 20353613, Genome Med, 2, 2010, <https://pubmed.ncbi.nlm.nih.gov/20353613/>

Frederiksen CM, Højlund K, Hansen L, Oakeley EJ, Hemmings B, Abdallah BM, Brusgaard K, Beck-Nielsen H, Gaster M, “Transcriptional profiling of myotubes from patients with type 2 diabetes: no evidence for a primary defect in oxidative phosphorylation genes”, doi: 10.1007/S00125-008-1122-9, issn: 0012-186X, 11, Diabetologia, 2068-2077, pmid: 18719883, Diabetologia, 51, 2008, <https://pubmed.ncbi.nlm.nih.gov/18719883/>

“Blood biomarkers of pancreatic cancer associated diabetes identified by peripheral blood-based gene expression profiles”, Second Affiliated Hospital, School of Medicine, Zhejiang University, 2009

Wise AF, Williams TM, Rudd S, Wells CA, Kerr PG, Ricardo SD, “Human mesenchymal stem cells alter the gene profile of monocytes from patients with Type 2 diabetes and end-stage renal disease”, doi: 10.2217/RME.15.74, issn: 1746-076X, 2, Regenerative medicine, 145-158, pmid: 26544198, Regen Med, 11, 2016, <https://pubmed.ncbi.nlm.nih.gov/26544198/>

Grayson BL, Wang L, Aune TM, “Peripheral blood gene expression profiles in metabolic syndrome, coronary artery disease and type 2 diabetes”, doi: 10.1038/GENE.2011.13, issn: 1476-5470, 5, Genes and immunity, 341-351, pmid: 21368773, Genes Immun, 12, 2011, <https://pubmed.ncbi.nlm.nih.gov/21368773/>

Khamis A, Canouil M, Siddiq A, Crouch H, Falchi M, Bulow MV, Eehalt F, Marselli L, Distler M, Richter D, Weitz J, Bokvist K, Xenarios I, Thorens B, Schulte AM, Ibberson M, Bonnefond A, Marchetti P, Solimena M, Froguel P, “Laser capture microdissection of human pancreatic islets reveals novel eQTLs associated with type 2 diabetes”, doi: 10.1016/J.MOLMET.2019.03.004, issn: 2212-8778, Molecular metabolism, 98-107, pmid: 30956117, Mol Metab, 24, 2019, <https://pubmed.ncbi.nlm.nih.gov/30956117/>

Solimena M, Schulte AM, Marselli L, Eehalt F, Richter D, Kleeberg M, Mziaut H, Knoch KP, Parnis J, Bugliani M, Siddiq A, Jörns A, Burdet F, Liechti R, Suleiman M, Margerie D, Syed F, Distler M, Grützmann R, Petretto E, Moreno-Moral A, Wegbrod C, Sönmez A, Pfriem K, Friedrich A, Meinel J, Wollheim CB, Baretton GB, Scharfmann R, Nogoceke E, Bonifacio E, Sturm D, Meyer-Puttlitz B, Boggi U, Saeger HD, Filipponi F, Lesche M, Meda P, Dahl A, Wigger L, Xenarios I, Falchi M, Thorens B, Weitz J, Bokvist K, Lenzen S, Rutter GA, Froguel P, von Bülow M, Ibberson M, Marchetti P, “Systems biology of the IMIDIA biobank from organ donors and pancreatectomised patients defines a novel transcriptomic signature of islets from individuals with type 2 diabetes”, doi: 10.1007/S00125-017-4500-3, issn: 1432-0428, 3, Diabetologia, 641-657, pmid: 29185012, Diabetologia, 61, 2018, <https://pubmed.ncbi.nlm.nih.gov/29185012/>

Hänzelmann S, Wang J, Güney E, Tang Y, Zhang E, Axelsson AS, Nenonen H, Salehi AS, Wollheim CB, Zetterberg E, Berntorp E, Costa IG, Castelo R, Rosengren AH, “Thrombin stimulates insulin secretion via protease-activated receptor-3”, doi: 10.1080/19382014.2015.1118195, issn: 1938-2022, 4, Islets, pmid: 26742564, Islets, 7, 2015, <https://pubmed.ncbi.nlm.nih.gov/26742564/>

Taneera J, Lang S, Sharma A, Fadista J, Zhou Y, Ahlqvist E, Jonsson A, Lyssenko V, Vikman P, Hansson O, Parikh H, Korsgren O, Soni A, Krus U, Zhang E, Jing XJ, Esguerra JL, Wollheim CB, Salehi A, Rosengren A, Renström E, Groop L, “A systems genetics approach identifies genes and pathways for type 2 diabetes in human islets”, doi: 10.1016/J.CMET.2012.06.006, issn: 1932-7420, 1, Cell metabolism, 122-134, pmid: 22768844, Cell Metab, 16, 2012, <https://pubmed.ncbi.nlm.nih.gov/22768844/>

Taneera J, Fadista J, Ahlqvist E, Zhang M, Wierup N, Renström E, Groop L, “Expression profiling of cell cycle genes in human pancreatic islets with and without type 2 diabetes”, doi: 10.1016/J.MCE.2013.05.003, issn: 1872-8057, 1-2, Molecular and cellular endocrinology, 35-42, pmid: 23707792, Mol Cell Endocrinol, 375, 2013, <https://pubmed.ncbi.nlm.nih.gov/23707792/>

Kanatsuna N, Taneera J, Vaziri-Sani F, Wierup N, Larsson HE, Delli A, Skärstrand H, Balhuizen A, Bennet H, Steiner DF, Törn C, Fex M, Lernmark Å, “Autoimmunity against INS-IGF2 protein expressed in human pancreatic islets”, doi: 10.1074/JBC.M113.478222, issn: 1083-351X, 40, The Journal of biological chemistry, 29013-29023, pmid: 23935095, J Biol Chem, 288, 2013, <https://pubmed.ncbi.nlm.nih.gov/23935095/>

Dominguez V, Raimondi C, Somanath S, Bugliani M, Loder MK, Edling CE, Divecha N, da Silva-Xavier G, Marselli L, Persaud SJ, Turner MD, Rutter GA, Marchetti P, Falasca M, Maffucci T, “Class II phosphoinositide 3-kinase regulates exocytosis of insulin granules in pancreatic beta cells”, doi: 10.1074/JBC.M110.200295, issn: 1083-351X, 6, The Journal of biological chemistry, 4216-4225, pmid: 21127054, J Biol Chem, 286, 2011, <https://pubmed.ncbi.nlm.nih.gov/21127054/>

Kirchner H, Sinha I, Gao H, Ruby MA, Schönke M, Lindvall JM, Barrès R, Krook A, Näslund E, Dahlman-Wright K, Zierath JR, “Altered DNA methylation of glycolytic and lipogenic genes in liver from obese and type 2 diabetic patients”, doi: 10.1016/J.MOLMET.2015.12.004, issn: 2212-8778, 3, Molecular metabolism, 171-183, pmid: 26977391, Mol Metab, 5, 2016, <https://pubmed.ncbi.nlm.nih.gov/26977391/>

Misu H, Takamura T, Takayama H, Hayashi H, Matsuzawa-Nagata N, Kurita S, Ishikura K, Ando H, Takeshita Y, Ota T, Sakurai M, Yamashita T, Mizukoshi E, Yamashita T, Honda M, Miyamoto K, Kubota T, Kubota N, Kadowaki T, Kim HJ, Lee IK, Minokoshi Y, Saito Y, Takahashi K, Yamada Y, Takakura N, Kaneko S, “A liver-derived secretory protein, selenoprotein P, causes insulin resistance”, doi: 10.1016/J.CMET.2010.09.015, issn: 1932-7420, 5, Cell metabolism, 483-495, pmid: 21035759, Cell Metab, 12, 2010, <https://pubmed.ncbi.nlm.nih.gov/21035759/>

Pihlajamäki J, Boes T, Kim EY, Dearie F, Kim BW, Schroeder J, Mun E, Nasser I, Park PJ, Bianco AC, Goldfine AB, Patti ME, “Thyroid hormone-related regulation of gene expression in human fatty liver”, doi: 10.1210/JC.2009-0212, issn: 1945-7197, 9, The Journal of clinical

endocrinology and metabolism, 3521-3529, pmid: 19549744, J Clin Endocrinol Metab, 94, 2009, <https://pubmed.ncbi.nlm.nih.gov/19549744/>

Kumar A, Tiwari P, Saxena A, Purwar N, Wahi N, Sharma B, Mathur SK, “The Transcriptomic Evidence on the Role of Abdominal Visceral vs. Subcutaneous Adipose Tissue in the Pathophysiology of Diabetes in Asian Indians Indicates the Involvement of Both”, doi: 10.3390/BIOM10091230, issn: 2218-273X, 9, Biomolecules, 1-23, pmid: 32847136, Biomolecules, 10, 2020, <https://pubmed.ncbi.nlm.nih.gov/32847136/>

Saxena A, Tiwari P, Wahi N, Soni A, Bansiwala RC, Kumar A, Sharma B, Punjabi P, Gupta N, Malik B, Medicherla KM, Suravajhala P, Mathur SK, “Transcriptome profiling reveals association of peripheral adipose tissue pathology with type-2 diabetes in Asian Indians”, doi: 10.1080/21623945.2019.1595269, issn: 2162-397X, 1, Adipocyte, 125-136, pmid: 30894049, Adipocyte, 8, 2019, <https://pubmed.ncbi.nlm.nih.gov/30894049/>

Tiwari P, Saxena A, Gupta N, Medicherla KM, Suravajhala P, Mathur SK, “Systems Genomics of Thigh Adipose Tissue From Asian Indian Type-2 Diabetics Revealed Distinct Protein Interaction Hubs”, doi: 10.3389/FGENE.2018.00679, issn: 1664-8021, JAN, Frontiers in genetics, pmid: 30671081, Front Genet, 9, 2019, <https://pubmed.ncbi.nlm.nih.gov/30671081/>

Doumatey AP, Xu H, Huang H, Trivedi NS, Lei L, Elkahloun A, Adeyemo A, Rotimi CN, “Global Gene Expression Profiling in Omental Adipose Tissue of Morbidly Obese Diabetic African Americans”, doi: 10.14740/JEM286W, issn: 1923-2861, 3, Journal of endocrinology and metabolism, 199-210, pmid: 26504501, J Endocrinol Metab, 5, 2015, <https://pubmed.ncbi.nlm.nih.gov/26504501/>

Jain P, Vig S, Datta M, Jindal D, Mathur AK, Mathur SK, Sharma A, “Systems biology approach reveals genome to phenome correlation in type 2 diabetes”, doi: 10.1371/JOURNAL.PONE.0053522, issn: 1932-6203, 1, PLoS one, pmid: 23308243, PLoS One, 8, 2013, <https://pubmed.ncbi.nlm.nih.gov/23308243/>

Mathur SK, Jindal D, Agarwal A, “Genome wide gene expression profiling of visceral adipose tissue among Asian Indian diabetics”, S.M.S Medical College, 2009

Van Tienen FH, van der Kallen CJ, Lindsey PJ, Wanders RJ, van Greevenbroek MM, Smeets HJ, “Preadipocytes of type 2 diabetes subjects display an intrinsic gene expression profile of

decreased differentiation capacity”, doi: 10.1038/IJO.2010.275, issn: 1476-5497, 9, International journal of obesity (2005), 1154-1164, pmid: 21326205, Int J Obes (Lond), 35, 2011, <https://pubmed.ncbi.nlm.nih.gov/21326205/>

Παπαδημητρίου Κωνσταντίνα, “Μεθοδολογίες Υπολογιστικής Ανάλυσης και Μέτα-ανάλυσης γενετικών δεδομένων και δεδομένων γονιδιακής έκφρασης, εφαρμογή στο Σακχαρώδη Διαβήτη”, isbn: 137.108.70.7, ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ «ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΙΑΤΡΙΚΗ» ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ, Υπεύθυνος ΜΠΑΓΚΟΣ ΠΑΝΤΕΛΗΣ ΑΝΑΠΛΗΡΩΤΗΣ ΚΑΘΗΓΗΤΗΣ, 2016

Παγκόσμια Ημέρα Διαβήτη - Νέα - Υπουργείο Υγείας, <https://www.moh.gov.gr/articles/news/1473-pagkosmia-hmera-diabht>

WHO/Europe | Diabetes - Data and statistics, <https://www.euro.who.int/en/health-topics/noncommunicable-diseases/diabetes/data-and-statistics>

Jung Kyoon Choi and Ungsik Yu and Sangsoo Kim and Ook Joon Yoo, doi : 10.1093/BIOINFORMATICS/BTG1010, issn : 1367-4803, Bioinformatics, i84-i90, Oxford Academic, “Combining multiple microarray studies and modeling interstudy variation”, 19, 2003, [https://academic.oup.com/bioinformatics/article/19/suppl\\_1/i84/227742](https://academic.oup.com/bioinformatics/article/19/suppl_1/i84/227742)

John R Stevens and RW Doerge, doi : 10.1186/1471-2105-6-57, issn : 1471-2105, 1, BMC Bioinformatics 2005 6:1, 1-19, BioMed Central, “Combining Affymetrix microarray results”, 6, 2005, <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-6-57>

Εκδόσεις "ΔΟΜΙΚΗ", Γκούμας - Κωτσιόπουλος, “Ο Γιατρός μου”, ΤΟΜΟΣ Α, Τόμος 3, 156-160, Ελλάδα, 2002

Werner Dubitzky and Martin Granzow and C. Stephen Downes and Daniel Berrar, doi : 10.1007/0-306-47815-3\_1, “A Practical Approach to Microarray Data Analysis”, 1-46, Springer, Boston, MA, Introduction to Microarray Data Analysis, 2003, [https://link.springer.com/chapter/10.1007/0-306-47815-3\\_1](https://link.springer.com/chapter/10.1007/0-306-47815-3_1)

Alvis Brazma and Jaak Vilo, doi : 10.1016/S0014-5793(00)01772-5, issn : 0014-5793, 1, FEBS Letters, 17-24, No longer published by Elsevier, “Gene expression data analysis”, 480, 2000

ROBERT STINE, doi : 10.1177/0049124189018002003, 2-3, <http://dx.doi.org/10.1177/0049124189018002003>, 243-291, SAGE PUBLICATIONS, “An Introduction to Bootstrap Methods: Examples and Ideas”, 18, 2016, <https://journals.sagepub.com/doi/abs/10.1177/0049124189018002003>

Sandrine Dudoit and Juliet Popper Shaffer and Jennifer C. Boldrick, doi : 10.1214/SS/1056397487, issn : 0883-4237, 1, <https://doi.org/10.1214/ss/1056397487>, 71-103, Institute of Mathematical Statistics, “Multiple Hypothesis Testing in Microarray Experiments”, 18, 2003, <https://projecteuclid.org/journals/statistical-science/volume-18/issue-1/Multiple-Hypothesis-Testing-in-Microarray-Experiments/10.1214/ss/1056397487.full>

Mark R Dalman and Anthony Deeter and Gayathri Nimishakavi and Zhong-Hui Duan, doi : 10.1186/1471-2105-13-S2-S11, issn : 1471-2105, 2, BMC Bioinformatics 2012 13:2, 1-4, BioMed Central, “Fold change and p-value cutoffs significantly alter microarray interpretations”, 13, 2012, <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-S2-S11>

E T Jaynes and E T Jaynes, “Bayesian Methods: General Background”, 1986, <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.41.1055>

Normand, Wiley Online Library, 321-359, “Meta-analysis: formulating, evaluating, combining, and reporting”, 18, 1999, [https://onlinelibrary.wiley.com/doi/abs/10.1002/\(SICI\)1097-0258\(19990215\)18:3%3C321::AID-SIM28%3E3.0.CO;2-P](https://onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1097-0258(19990215)18:3%3C321::AID-SIM28%3E3.0.CO;2-P)

Matthias Egger and George Davey Smith and Martin Schneider and Christoph Minder, doi : 10.1136/BMJ.315.7109.629, issn : 0959-8138, 7109, BMJ, 629-634, pmid : 9310563, British Medical Journal Publishing Group, “Bias in meta-analysis detected by a simple, graphical test”, 315, 1997, <https://www.bmj.com/content/315/7109/629>

Anna Campain and Yee Hwa Yang, doi : 10.1186/1471-2105-11-408, issn : 1471-2105, 1, BMC Bioinformatics 2010 11:1, 1-11, BioMed Central, “Comparison study of microarray meta-analysis methods”, 11, 2010, <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-408>



Rainer Breitling and Patrick Armengaud and Anna Amtmann and Pawel Herzyk, doi : 10.1016/J.FEBSLET.2004.07.055, issn : 1873-3468, 1-3, FEBS Letters, 83-92, John Wiley & Sons, Ltd, "Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments", 573, 2004, <https://onlinelibrary.wiley.com/doi/full/10.1016/j.febslet.2004.07.055>

Ammarin Thakkinstian and Patrick McElduff and Catherine D'Este and David Duffy and John Attia, doi : 10.1002/SIM.2010, issn : 1097-0258, 9, Statistics in Medicine, 1291-1306, John Wiley & Sons, Ltd, "A method for meta-analysis of molecular association studies", 24, 2005, <https://onlinelibrary.wiley.com/doi/full/10.1002/sim.2010>

Elmer V Villanueva and Silva Zavarsek, doi : 10.1186/1471-2288-4-18, issn : 1471-2288, 1, BMC Medical Research Methodology 2004 4:1, 1-8, BioMed Central, "Evaluating heterogeneity in cumulative meta-analyses", 4, 2004, <https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-4-18>

Rebecca DerSimonian and Nan Laird, doi : 10.1016/0197-2456(86)90046-2, issn : 0197-2456, 3, Controlled Clinical Trials, 177-188, Elsevier, "Meta-analysis in clinical trials", 7, 1986

Baird, David. 2012. Analysis of Microarray Data. In GenStatfor Windows, edited by GenSta: VSN International, 5 The Waterhouse, Waterhouse Street

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

Holm, S. 1979. "A simple sequentially rejective multiple test procedure." Scandinavian Journal of Statistics 6:65-70

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

Sidak, Zbynek. 1967. "Rectangular Confidence Regions for the Means of Multivariate Normal Distributions." Journal of the American Statistical Association 62 (318):626-633

StataCorp, 2013, Stata 13 Base Reference Manual, College Station, TX: Stata Press

Team, R., 2015, “R Core Team. R: a language and environment for statistical computing

Szklarczyk D and Gable AL and Lyon D and Junge A and Wyder S and Huerta-Cepas J and Simonovic M and Doncheva NT and Morris JH and Bork P and Jensen LJ and Mering CV}, doi : 10.1093/NAR/GKY1131, issn : 1362-4962, D1, Nucleic acids research, D607-D613, pmid : 30476243, Nucleic Acids Res, “STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets”, 47, 2019, <https://pubmed.ncbi.nlm.nih.gov/30476243/>

Paul D. Thomas and Anish Kejariwal and Michael J. Campbell and Huaiyu Mi and Karen Diemer and Nan Guo and Istvan Ladunga and Betty Ulitsky-Lazareva and Anushya Muruganujan and Steven Rabkin and Jody A. Vandergriff and Olivier Doremieux, doi : 10.1093/NAR/GKG115, 1, Nucleic Acids Research, 334, pmid : 12520017, Oxford University Press, “PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification”, 31, 2003, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC165562/>

Burt S. Holland and Margaret DiPonzio Copenhaver, doi : 10.2307/2531823, 2, Biometrics, 417, JSTOR, “An Improved Sequentially Rejective Bonferroni Test Procedure”, 43, 1987

Sandrine Dudoit and Y Ee and Hwa Y Ang and Matthew J Callow and Terence P S P E E D 13, Statistica Sinica, 111-139, “STATISTICAL METHODS FOR IDENTIFYING DIFFERENTIALLY EXPRESSED GENES IN REPLICATED cDNA MICROARRAY EXPERIMENTS”, 12, 2002

Cuilin Zhang and Wei Bao and Ying Rong and Huixia Yang and Katherine Bowers and Edwina Yeung and Michele Kiely, doi : 10.1093/HUMUPD/DMT013, 4, Human Reproduction Update, 376, pmid : 23690305, Oxford University Press, “Genetic variants and the risk of gestational diabetes mellitus: a systematic review”, 19, 2013, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3682671/>

Nicola Abate and Manisha Chandalia, doi : 10.1016/S1056-8727(01)00161-1, issn : 1056-8727, 6, Journal of Diabetes and its Complications, 320-327, Elsevier, “Ethnicity and type 2 diabetes: Focus on Asian Indians”, 15, 2001

David W Cooke, doi : 10.1542/pir.29-11-374, "Type 1 Diabetes Mellitus in Pediatrics", 2008, <http://pedsinreview.aappublications.org>

Pouya Saeedi and Inga Petersohn and Paraskevi Salpea and Belma Malanda and Suvi Karuranga and Nigel Unwin and Stephen Colagiuri and Leonor Guariguata and Ayesha A. Motala and Katherine Ogurtsova and Jonathan E. Shaw and Dominic Bright and Rhys Williams, doi : 10.1016/J.DIABRES.2019.107843, issn : 0168-8227, Diabetes Research and Clinical Practice, pmid : 31518657, Elsevier, "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition", 157, 2019, <http://www.diabetesresearchclinicalpractice.com/article/S0168822719312306/fulltext>

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

## **Παράρτημα 1**

### Λίστα μελετών που χρησιμοποιήθηκαν στις μετα-αναλύσεις

Adipose	Liver	Pancreatic islets	PBMCs	Skeletal muscle
GSE15790	GSE15653	GSE25724	GSE15932	GSE12643
GSE16415	GSE23343	GSE44035	GSE23561	GSE18732
GSE29226	GSE64998	GSE38642	GSE65561	GSE19420
GSE29231		GSE76894		GSE21340
GSE71416		GSE76895		GSE22309
GSE78721		E_CBIL_20		GSE25462
				GSE29221
				GSE55650
				GSE59363
				GSE73034
				GSE156249
				E_CBIL_30
				E_MEXP_1270
				E_MEXP_2559

### Πίνακες μελετών που χρησιμοποιήθηκαν για τις μετα-αναλύσεις

Λιπώδης ιστός	Όνομα πλατφόρμας	Αριθμός probes	Αριθμός γονιδίων	Συνολικός αριθμός ατόμων	Ασθενείς	Υγιείς
GSE15790	GPL8450	25391	15273	39	19	20
GSE16415	GPL2986	32877	16407	9	4	5
GSE29226	GPL6947	48802	20562	23	11	12
GSE29231	GPL6947	48802	20472	23	11	12
GSE71416	GPL570	54674	20250	19	14	5
GSE78721	GPL15207	49394	18739	130	68	62

Ήπαρ	Όνομα πλατφόρμας	Αριθμός probes	Αριθμός γονιδίων	Συνολικός αριθμός ατόμων	Ασθενείς	Υγιείς
GSE15653	GPL96	49392	12499	18	9	9
GSE23343	GPL570	54613	20250	17	10	7
GSE64998	GPL11532	33296	18354	21	7	14

Παγκρεατικά νησίδα	Όνομα πλατφόρμας	Αριθμός probes	Αριθμός γονιδίων	Συνολικός αριθμός ατόμων	Ασθενείς	Υγιείς
GSE25724	GPL96	22282	12501	13	6	7
GSE44035_GSE38642	GPL6244	18807	17246	73	10	63
GSE76894	GPL570	29529	13693	103	19	84
GSE76895	GPL570	29611	14107	68	36	32
e-cbil-20_a-affy-33	A-AFFY-33	22283	12493	12	5	7
e-cbil-20_a-affy-34	A-AFFY-34	22645	9853	12	5	7

Μονοπύρηννα κύτταρα περιφερικού αίματος	Όνομα πλατφόρμας	Αριθμός probes	Αριθμός γονιδίων	Συνολικός αριθμός ατόμων	Ασθενείς	Υγιείς
GSE15932	GPL570	54674	20236	16	8	8
GSE23561	GPL10775	50399	20276	17	8	9
GSE65561	GPL16686	53616	15642	9	5	4

Σκελετικός μυς	Όνομα πλατφόρμας	Αριθμός probes	Αριθμός γονιδίων	Συνολικός αριθμός ατόμων	Ασθενείς	Υγιείς
GSE12643	GPL8300	12557	8589	20	10	10
GSE18732	GPL9486	25769	11496	118	45	73
GSE19420	GPL570	54674	20235	42	18	24
GSE21340	GPL80	7125	5251	15	5	10
GSE22309	GPL91	12625	8587	70	30	40
GSE25462	GPL570	54674	20267	50	10	40
GSE29221	GPL6947	48802	20591	24	12	12
GSE55650	GPL570	54612	20239	23	12	11
GSE59363	GPL6244	33296	21177	14	7	7
GSE73034	GPL6480	36275	18018	13	6	7
GSE156249	GPL11532	33296	18356	25	7	18
e-cbil-30	A-AFFY-33	22283	12493	35	18	17
e-mexp-1270	A-AFFY-32	7129	5251	5	3	2
e-mexp-2559	A-AFFY-44	21058	19331	35	5	30

**Πίνακες με τα αποτελέσματα από τα γονίδια που κρίθηκαν στατιστικώς σημαντικά σε κάθε μεμονωμένη μελέτη για κάθε ιστό**

Adipose	p<0.05	p<0.01	FDR (0.01)	Bonferroni	Sidak	Holm	Holland
GSE15790	161	20	0	0	0	0	0
GSE16415	1679	547	0	0	0	0	0
GSE29226	5665	3253	1381	249	252	251	253
GSE29231	6046	3785	2114	428	434	433	436
GSE71416	2438	750	23	14	14	14	14
GSE78721	2298	503	0	1	1	1	1

Liver	p<0.05	p<0.01	FDR (0.01)	Bonferroni	Sidak	Holm	Holland
GSE15653	1563	295	0	0	0	0	0
GSE23343	2283	581	0	1	1	1	1
GSE64998	1651	454	0	0	0	0	0

Pancreatic islets	p<0.05	p<0.01	FDR (0.01)	Bonferroni	Sidak	Holm	Holland
GSE25724	6800	4049	1090	4	4	4	4
GSE44035_GSE38642	1508	418	0	0	0	0	0
GSE76894	4674	2710	986	82	83	83	83
GSE76895	3311	1605	274	67	68	67	68
e-cbil-20_a-affy-33	1449	441	0	0	0	0	0
e-cbil-20_a-affy-34	852	251	0	1	2	1	2

PBMCs	p<0.05	p<0.01	FDR (0.01)	Bonferroni	Sidak	Holm	Holland
GSE15932	5298	2520	57	9	10	9	10
GSE23561	3677	1266	0	0	0	0	0
GSE65561	2095	644	0	0	0	0	0

Skeletal muscle	p<0.05	p<0.01	FDR (0.01)	Bonferroni	Sidak	Holm	Holland
GSE12643	437	69	0	0	0	0	0
GSE18732	1008	253	0	1	1	1	1
GSE19420	2144	720	2	2	2	2	2
GSE21340	81	25	0	0	0	0	0
GSE22309	4614	3446	2761	693	703	714	719
GSE25462	2481	782	8	6	6	6	6
GSE29221	6130	2315	153	36	36	36	36
GSE55650	2480	888	0	2	2	2	2
GSE59363	1359	311	0	1	1	1	1
GSE73034	2432	823	0	1	1	1	1
GSE156249	17885	17640	17636	15508	15519	16287	16299
e-cbil-30	513	83	0	0	0	0	0
e-mexp-1270	1408	752	114	0	0	0	0
e-mexp-2559	3742	1572	30	5	5	5	5

**Λίστα στατιστικώς σημαντικών γονιδίων για κάθε ιστό**

- **Adipose**

Όνομασία γονιδίου	p-value
CNTNAP3B	3.26E-08
LOC100133284	7.36E-08



- Pancreatic islets

Όνομασία γονιδίου	p-value				
ABRACL	0.0000349	MRPS33	6.31E-06	ZNF233	0.0000226
ACTR3B	0.000018	MSH6	3.07E-06		
APOO	0.0000288	MST1R	0.0000119		
ATP6V1D	0.0000133	NAPA	0.0000102		
C19ORF53	1.66E-06	NDUFV2	0.0000245		
CACHD1	0.0000215	NECTIN3	2.67E-07		
CAPN13	8.10E-08	NEFL	1.98E-06		
CD300LF	0.0000346	NEXMIF	0.0000197		
CD86	3.68E-06	ONECUT2	2.24E-07		
CFAP36	2.38E-07	OPCML	0.0000414		
CFAP65	0.0000165	PDCD2L	8.33E-06		
CLDN12	0.0000247	PDE7B	6.47E-09		
COA5	0.00003	PDPN	0.0000317		
CRLF2	0.0000276	PER3	6.37E-08		
DBI	1.68E-06	PITHD1	7.53E-06		
DHRS7	0.0000127	PLA1A	0.0000422		
DHX40	2.20E-08	PPFIBP1	0.0000132		
DUSP28	5.97E-06	PPP6C	0.0000104		
ERICH5	4.90E-11	PRPF18	0.0000142		
F5	3.97E-06	RCAN2	0.0000115		
FAM102B	1.57E-06	RECK	0.0000319		
FBXW11	0.0000329	RNASE6	6.93E-06		
FFAR4	0.0000125	RNF217	0.0000224		
GAPDH	0.000032	RTL8A	3.02E-07		
GART	0.0000322	RUNX1_IT1	0.0000102		
GOLT1A	3.04E-11	SLAMF7	0.0000261		
GUSBP11	9.94E-06	SLC16A7	0.0000196		
HAVCR2	0.0000338	SLC20A2	0.0000404		
HLA_DPB1	0.0000223	SLC25A34	2.24E-06		
HMCES	0.0000142	SLC4A4	0.0000143		
IL12RB1	0.0000172	SOCS1	2.78E-06		
ITIH4	7.40E-06	SP140L	0.000034		
KBTBD4	9.81E-07	STK16	8.66E-09		
KDELR1	2.99E-06	SWI5	1.70E-06		
LFNG	0.0000188	TFEC	0.0000206		
LIF_AS2	0.0000364	TMEM141	3.18E-06		
LOC100289473	0.000015	TNRC18	2.27E-06		
LSM3	7.63E-06	VAT1L	0.0000351		
LTBP1	0.0000296	VEGFA	0.0000236		
MCOLN1	0.0000108	WBP11	2.93E-06		
MDM2	0.0000149	WFS1	4.32E-06		
MGAT4A	3.01E-06	WIPF1	0.0000348		
MLKL	0.0000133	XBP1P1	0.0000197		

- **PBMCs**

<b>Όνομα γονιδίου</b>	<b>p-value</b>
ADRM1	2.52E-06
CARHSP1	3.17E-06
DHDDS	1.95E-06
DNTTIP1	5.19E-06
DUTP1	6.19E-06
FAM83A	5.61E-07
FNBP4	1.73E-06
GAPDHP73	1.10E-06
NCAPH2	4.30E-06
NFATC1	2.51E-06
OR14J1	2.37E-06
PRSS8	1.07E-06
TFAP2A_AS1	5.05E-07
TRIP6	6.36E-06
UBXN1	1.78E-06
VTI1A	2.91E-06

- Skeletal muscle

Όνομα γονιδίου	p-value		
ALDH7A1	3.43E-05	REXO1L2P	4.07E-06
ATP5MFP4	8.50E-09	RNA18SN5	1.01E-12
BGLAP	7.82E-07	RNU6_285P	2.16E-06
DEFB130A	8.12E-10	RNU6ATAC	4.63E-11
FAM223A	4.16E-11	RNY4P8	5.39E-07
FTH1P12	2.35E-06	RPL23AP71	3.24E-07
FTH1P7	1.08E-05	RUVBL1	1.54E-05
FTH1P8	3.41E-06	SAP30L_AS1	1.70E-07
HAUS7	1.38E-05	SCARNA12	9.15E-12
IGHV3_48	3.90E-10	SCARNA16	1.14E-09
IGKV3D_11	3.68E-07	SLC25A6P2	7.35E-10
LINC02481	2.42E-05	SNORA1	7.87E-10
LOC100128655	3.49E-06	SNORA24	6.29E-10
LOC100289060	2.31E-11	SNORA29	3.96E-07
LOC100289143	2.82E-10	SNORA31	1.31E-11
LOC645877	2.08E-06	SNORA3A	1.48E-10
LOC728554	4.46E-06	SNORA41	3.23E-10
LSP1P3	5.96E-10	SNORA51	1.55E-08
MIR122	1.08E-05	SNORA52	7.84E-12
MIR124_1	1.19E-06	SNORA56	2.01E-08
MIR130B	3.45E-08	SNORA57	5.41E-09
MIR137	3.32E-08	SNORA5A	8.02E-10
MIR144	4.49E-10	SNORA5C	3.60E-11
MIR192	4.82E-09	SNORA6	3.54E-08
MIR197	9.45E-10	SNORA69	1.47E-08
MIR199A2	3.92E-10	SNORD102	3.17E-09
MIR202	1.50E-06	SNORD114_26	2.48E-12
MIR203A	1.25E-05	SNORD115_10	7.66E-07
MIR21	8.59E-07	SNORD116_13	4.09E-11
MIR219A2	1.48E-09	SNORD116_14	6.26E-11
MIR24_2	2.81E-07	SNORD116_22	1.88E-10
MIR29B1	3.11E-07	SNORD116_28	9.19E-08
MIR29B2	3.36E-10	SNORD14C	1.04E-09
MIR29C	3.98E-11	SNORD1A	1.61E-07
MIR330	4.17E-12	SNORD1C	1.23E-06
MIR451A	6.01E-07	SNORD36B	1.35E-10
MIR7_2	2.28E-12	SNORD44	1.78E-11
NDUFB9P2	7.09E-09	SNORD45C	2.23E-12
NOL8	3.80E-06	SNORD49A	5.41E-11
OR5M1	2.53E-08	SNORD50A	1.09E-08
OR7A2P	3.27E-10	SNORD50B	0.000031
PPIAP70	6.13E-06	SNORD58A	1.09E-10
PPP1R2P1	1.30E-11	SNORD6	2.21E-09
PRR33	2.20E-10	SNORD61	1.36E-09
PWWP4	4.71E-10	SNORD74	7.72E-12

SNORD77	5.18E-08
SNORD83B	2.05E-09
SNORD96A	4.37E-09
SP9	2.34E-09
SPATA31D4	1.39E-10
SRSF6P2	7.53E-07
SSU72P3	4.44E-09
TMLHE_AS1	8.09E-12
TPTE2P3	5.75E-11
UBTFL1	1.43E-07
ZNF56	2.93E-11
ZNF66	7.66E-10

## **Παράρτημα 2**

## Εντολές STATA

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

```
Collapse (mean) gsm1 gsm2 gsm3, by (gene_symbol)
```

```
List
```

```
List
```

Εντολή για ένωση πινάκων

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

Αρχείο ADO που χρησιμοποιήθηκε για την εκτέλεση της μετα-ανάλυσης

```
program define myDL5, rclass
```

```
version 10.1
```

```
syntax , x(varlist numeric max=1) Type(varlist numeric max=1) Study(varlist numeric  
max=1) [ Effect(string ) ]
```

```
preserve
```

```
tempvar last
```

```
qui bysort `study': gen `last'=_n==_N
```

```
qui sum `study'
```

```
local max=r(max)
```

```
tempvar x0 x1 sd0 sd1 n0 n1 d sed m gmfm gmfm_1 J g seg
```

```
qui gen double `x0'=.
```

```
qui gen double `x1'=..
```

*qui gen double `sd0'=.*

*qui gen double `sd1'=.*

*qui gen double `n0'=.*

*qui gen double `n1'=.*

*qui gen double `d'=.*

*qui gen double `sed'=.*

*qui gen double `m'=.*

*qui gen double `gmfm'=.*

*qui gen double `gmfm\_1'=.*

*qui gen double `J'=.*

*qui gen double `g'=.*

*qui gen double `seg'=.*

*forvalues i=1(1)`max' {*

*qui sum `x' if `study'==`i' & `type'==0*

*qui replace `x0'=r(mean) if `study'==`i'*

*qui replace `sd0'=r(sd) if `study'==`i'*

*qui replace `n0'=r(N) if `study'==`i'*

*qui sum `x' if `study'==`i' & `type'==1*

*qui replace `x1'=r(mean) if `study'==`i'*

*qui replace `sd1'=r(sd) if `study'==`i'*

*qui replace `n1'=r(N) if `study'==`i'*

```

qui replace `m' = (`n1' + `n0' - 2) if `study' == `i'

qui replace `d' = (`x1' - `x0') / (sqrt( ( (`n1' - 1) * (`sd1')^2 + (`n0' - 1) * (`sd0')^2) / (`n1' + `n0' - 2 ))) if `study' == `i'

qui replace `sed' = sqrt( (`n1' + `n0') / (`n1' * `n0') + (`d')^2 / (2 * (`n1' + `n0'))) if `study' == `i'

qui replace `gmfm' = exp((lngamma(`m'/2))) if `study' == `i'

qui replace `gmfm_1' = exp((lngamma((`m' - 1) / 2))) if `study' == `i'

qui replace `J' = `gmfm' / ((sqrt(`m'/2)) * `gmfm_1) if `study' == `i'

qui replace `g' = `J' * `d' if `study' == `i'

qui replace `seg' = (`J')^2 * `sed' if `study' == `i'

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

}

qui keep if `last'

qui keep if `x' != .

metan `g' `seg', `effect' nograph

**metan `n1' `x1' `sd1' `n0' `x0' `sd0', `effect' nograph

**metan `n1' `x1' `sd1' `n0' `x0' `sd0', fixed nograph hedges

**metan `d' `sed', random nograph

**list

restore

```



```
return scalar g=r(ES)
```

```
return scalar seg=r(seES)
```

```
end
```

Κώδικας για μετα-ανάλυση

```
set more off
```

```
file open meta using resultmyDL5.txt, write append
```

```
file write meta "gene"
```

```
file write meta ", "
```

```
file write meta "g"
```

```
file write meta ", "
```

```
file write meta "seg" _n
```

```
foreach var of varlist a1bg - zzz3 {
```

```
di ""
```

```
di ""
```

```
di in ye "Performing meta-analysis for `var' gene_symbol"
```

```
di ""
```

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

```
file write meta "`var'"
```

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

*file close meta*

Εντολές για δημιουργία z – score & p – value

*generate z= g/ seg*

*generate p=2\*(1-normal(abs( z )))*

Εντολές διόρθωσης p – value

*count if p<0.05*

*count if p<0.01*

*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)*

