



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ
ΒΙΟΙΑΤΡΙΚΗ

ΚΛΗΡΟΝΟΜΗΣΙΜΟΤΗΤΑ ΚΑΙ ΕΠΠΟΛΑΣΜΟΣ
ΣΤΙΣ ΠΟΛΥΠΑΡΑΓΟΝΤΙΚΕΣ ΑΣΘΕΝΕΙΕΣ

ΜΠΟΓΙΑΤΖΟΓΛΟΥ ΑΝΑΣΤΑΣΙΑ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ
Υπεύθυνος
Μπάγκος Παντελεήμων

Λαμία, 2024



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

**ΚΛΗΡΟΝΟΜΗΣΙΜΟΤΗΤΑ ΚΑΙ ΕΠΠΟΛΑΣΜΟΣ ΣΤΙΣ
ΠΟΛΥΠΑΡΑΓΟΝΤΙΚΕΣ ΑΣΘΕΝΕΙΕΣ**

ΜΠΟΓΙΑΤΖΟΓΛΟΥ ΑΝΑΣΤΑΣΙΑ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**Επιβλέπων
Μπάγκος Παντελεήμων**

Λαμία, 2024

Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις ⁽¹⁾, 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: 1/10/2024

Η Δηλούσα.

(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.

**ΚΛΗΡΟΝΟΜΗΣΙΜΟΤΗΤΑ ΚΑΙ ΕΠΠΟΛΑΣΜΟΣ ΣΤΙΣ
ΠΟΛΥΠΑΡΑΓΟΝΤΙΚΕΣ ΑΣΘΕΝΕΙΕΣ**

ΜΠΟΓΙΑΤΖΟΓΛΟΥ ΑΝΑΣΤΑΣΙΑ

Τριμελής Επιτροπή:

ΜΠΑΓΚΟΣ ΠΑΝΤΕΛΕΗΜΩΝ

ΔΕΛΗΜΠΑΣΗΣ ΚΩΝΣΤΑΝΤΙΝΟΣ

ΣΑΧΛΑΣ ΑΘΑΝΑΣΙΟΣ

Περιεχόμενα

ΠΕΡΙΛΗΨΗ	2
Εισαγωγή.....	4
1.1 Κληρονομικότητα και κληρονομησιμότητα.....	4
1.2 Κληρονομησιμότητα στην έρευνα.....	6
1.3 Επιπολασμός και Κληρονομησιμότητα	8
Μέθοδοι	9
2.1 Συλλογή δεδομένων	9
2.2 Διαλογή δεδομένων.....	11
2.3 Συμπλήρωση δεδομένων.....	13
Αποτελέσματα.....	18
3.1 Κληρονομησιμότητα σε γενετικά και μη γενετικά δεδομένα	18
3.2 Κληρονομησιμότητα στα βιολογικά φύλα.....	23
3.3 Κληρονομησιμότητα ανά ομάδα καταγωγής.....	26
3.4 Επιπολασμός ανά ομάδα ασθενειών	28
3.5 Επιπολασμός ανά βιολογικό φύλο.....	29
3.6 Επιπολασμός και κληρονομησιμότητα	31
ΣΥΜΠΕΡΑΣΜΑΤΑ.....	38
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	40

ΠΕΡΙΛΗΨΗ

Η παρούσα εργασία πραγματεύεται θέματα που αφορούν στην κληρονομησιμότητα και στον επιπολασμό πολυπαραγοντικών ασθενειών καθώς και τον τρόπο με τον οποίο τα 2 αυτά χαρακτηριστικά συσχετίζονται μεταξύ τους. Δεδομένου ότι η κληρονομησιμότητα μιας ασθένειας αποτελεί χαρακτηριστικό που επηρεάζεται τόσο από τα δημογραφικά χαρακτηριστικά του δείγματος όσο και από την μέθοδο που ακολουθείται για να υπολογιστεί, σε πολλές περιπτώσεις σημειώνει διακυμάνσεις στην τιμή της ακόμη και αν πρόκειται για μια μόνο ασθένεια. Σκοπός αυτής της προσπάθειας, είναι να εντοπίσει τις διακυμάνσεις στις τιμές κληρονομησιμότητας για τις ασθένειες που προέρχονται από γενετικά και μη δεδομένα, να εξάγει συμπεράσματα σε σχέση με την κληρονομησιμότητα της κάθε ασθένειας και το είδος των δεδομένων, το φύλο και την καταγωγή ενώ επίσης να ελέγξει αν υπάρχει συσχέτιση μεταξύ της κληρονομησιμότητας και του επιπολασμού.

KEYWORDS:

Κληρονομησιμότητα, επιπολασμός, μέθοδος υπολογισμού

SUMMARY

The present work deals with issues concerning heritability and the prevalence of multifactorial diseases as well as the way in which these two characteristics are related to each other. Given that the heritability of a disease is a characteristic that is influenced both by the demographic characteristics of the sample and by the method followed to calculate it, in many cases it shows fluctuations in its value even if it is a single disease. The purpose of this effort is to study these fluctuations and, in common observation with the prevalence, to draw conclusions in relation to the demographic characteristics and to check if there is an influence between heritability and prevalence.

KEYWORDS:

Heritability, prevalence, calculation method

Εισαγωγή

1.1 Κληρονομικότητα και κληρονομησιμότητα

Ως γενικευμένος ορός, η κληρονομικότητα είναι η βασική ιδιότητα των εν ζωή οργανισμών να μεταβιβάζουν τα βασικά χαρακτηριστικά και τις ιδιότητες τους στους απογόνους τους, ενώ στη γενετική, αναφέρεται στα χαρακτηριστικά ή και της παραλλαγές που κωδικοποιούνται στο DNA και μεταβιβάζονται από τον γονέα στον απόγονο κατά την αναπαραγωγή (1). Η πρώτη αναφορά στην ιδιότητα έγινε από τον Αρχαίο Έλληνα ιατρό Ιπποκράτη (460 π.Χ. – 377 π.Χ) (1) ο οποίος μέσα από το έργο «Ιπποκρατική συλλογή» διαφαίνεται ότι υποστήριζε την ύπαρξη της κληρονομικότητας. **Ισχυρίστηκε ότι τόσο οι σωματικές όσο και οι ψυχολογικές ιδιότητες των γονέων μπορούν να κληροδοτηθούν στους απογόνους (1).**

Συγκεκριμένα παρατήρησε πως γονείς που αντιμετώπιζαν πρόβλημα με το αίμα, το φλέγμα την κίτρινη και την μαύρη χολή, αποκτούσαν απογόνους οι οποίοι αντιμετώπιζαν ίδια προβλήματα με τους γονείς τους. Έτσι, και δεδομένης της πεποίθησης ότι το σπέρμα προέρχεται από όλο το σώμα, θεώρησε πως επρόκειτο για κληρονομικότητα της ασθένειας στους απογόνους (2).

Η παρούσα μελέτη ασχολείται με την κληρονομησιμότητα, η οποία αποτελεί μέτρο της επίδρασης των γενετικών παραγόντων στην κληρονομικότητα ενός χαρακτηριστικού. Για την ακρίβεια η κληρονομησιμότητα αποτελεί μια μετρική η οποία μελετά το βαθμό στον οποίο ένα χαρακτηριστικό μπορεί να αποδοθεί σε γενετικούς παράγοντες μέσα σε ένα πληθυσμό (3). Σε αντίθεση με την κληρονομικότητα είναι άγνωστο το πώς προέκυψε ο όρος κληρονομησιμότητα ωστόσο ο τρόπος χρήσης της εξελίχθηκε μέσα σε 3 φάσεις (4):

1^η φάση (1832): χρησιμοποιήθηκε για να δηλώσει την μεταβίβαση κληρονομικών με την ευρύτερη έννοια δηλαδή την μεταβίβαση νομικών και βιολογικών χαρακτηριστικών.

2^η φάση (αρχές 20ου αιώνα): εμφάνιση της **broad sense heritability**.

$$H^2 = \frac{V_G}{V_P}$$

Όπου:

V_G : είναι η γενετική διακύμανση, δηλαδή η διακύμανση ενός φαινοτύπου που αποδίδεται αποκλειστικά σε γενετικούς παράγοντες(5).

V_P : είναι η φαινοτυπική διακύμανση, δηλαδή η διακύμανση ενός φαινοτύπου σε ένα πληθυσμό που αποδίδεται τόσο σε γενετικές όσο και σε περιβαλλοντικές επιδράσεις(5).

Σε αυτή την περίπτωση η κληρονομησιμότητα μετρά το ποσοστό των διαφορών που παρατηρούνται στα χαρακτηριστικά διαχωρίζοντας τις περιβαλλοντικές διακυμάνσεις από τους γονοτυπικές. Υπολογίζει τις πρόσθετες γενετικές επιδράσεις (additive genetic effects), την επικράτηση μεταξύ των αλληλομόρφων γονιδίων καθώς και επιδράσεις από αλληλεπιδράσεις γονιδίων μεταξύ τους (epistasis) (6).

3^η φάση (1936): εμφάνιση της σύγχρονης **narrow sense heritability**.

$$h^2 = \frac{V_A}{V_P}$$

Οπού:

V_A : είναι η προσθετική γενετική διακύμανση (additive effects).

V_P : είναι η φαινοτυπική διακύμανση.

Αυτή αναφέρεται αποκλειστικά στην πρόσθετη γενετική διακύμανση δηλαδή στην διακύμανση που οφείλεται σε γονίδια που μεταβιβάζονται από γονείς σε απογόνους με σταθερό και προβλέψιμο τρόπο. Είναι προφανές ότι η διαφορά μεταξύ των 2 διαφορετικών εννοιών της κληρονομησιμότητα σ έγκειται στο γεγονός ότι η broad sense heritability υπολογίζει όλους τους γενετικούς παράγοντες ενώ οι narrow sense heritability λαμβάνει υπόψιν αποκλειστικά την προσθετική γενετική επίδραση.

1.2 Κληρονομησιμότητα στην έρευνα

Ο υπολογισμός της κληρονομησιμότητας μίας ασθένειας καθώς και ο αλγόριθμος που θα χρησιμοποιηθεί για τον υπολογισμό της, διαφοροποιείται ανάλογα με το είδος των δεδομένων και το από πού προέρχονται αυτά.

Υπάρχουν δυο κατηγορίες δεδομένων: τα γενετικά και τα μη γενετικά δεδομένα. Στην κατηγορία των γενετικών δεδομένων εντάσσονται οι πληροφορίες που προέρχονται από το γενετικό υλικό ενός οργανισμού, δηλαδή τα γονίδια, τις μεταλλάξεις, τους πολυμορφισμούς και γενικότερα πληροφορίες που αφορούν στην αλληλουχία, την επιδραστικότητα και την δραστηριότητα των βάσεων του DNA.

Οι πληροφορίες που δεν σχετίζονται άμεσα με το DNA (7) και προέρχονται από τη μελέτη των περιβαλλοντικών δεδομένων, δηλαδή τον τόπο και τις συνθήκες διαβίωσης του οργανισμού, τη διατροφή και τις συνήθειες του ενώ επιπλέον πληροφορίες όπως το μορφωτικό επίπεδο, το επάγγελμα, το εισόδημα, η συναναστροφή με φίλους και οικογένεια, πληροφορίες επονομαζόμενες και ως κοινωνικό-οικονομικά δεδομένα, θεωρούνται ως μη γενετικά δεδομένα. Στην ουσία ωστόσο, ως μη γενετικά δεδομένα θεωρούνται τα δεδομένα που προκύπτουν, όχι από κάποια γενετική ανάλυση αλλά από προφορική αναφορά και μελέτη των συνθηκών ζωής των οικογενειών καθώς και υποομάδες αυτών, όπως δίδυμα ή τρίδυμα αδέρφια κ.ο.κ. χωρίς να λαμβάνεται υπόψιν η κοινή γενετική σύνδεση. Συνακόλουθα ανάλογα με το αν οι πληροφορίες που έχουμε προέρχονται από γενετικά δεδομένα ή όχι, ακολουθείται διαφορετικός αλγόριθμος για τον υπολογισμό της κληρονομησιμότητας κάποιου χαρακτηριστικού.

Η κληρονόμηση χαρακτηριστικών μελετάται με μαθηματικά μοντέλα και αλγορίθμους πρόβλεψης, αρκετοί εκ των οποίων λαμβάνουν υπόψιν και τους εξωτερικούς παράγοντες (περιβαλλοντικούς και ενίοτε δημογραφικούς) για τον υπολογισμό. Οι αλγόριθμοι που χρησιμοποιούνται για τον υπολογισμό της τιμής της κληρονομησιμότητας δύναται να διακριθούν σε 3 ευρύτερες κατηγορίες:

- **Αλγόριθμοι που περιλαμβάνουν Linear Mixed Models:** Αυτοί χρησιμοποιούνται συχνά για να υπολογίσουν την κληρονομησιμότητα λαμβάνοντας υπόψη την αλληλεπίδραση των γονιδίων και των περιβαλλοντικών παραγόντων (8).
- **Αλγόριθμοι που περιλαμβάνουν Machine Learning:** Εφαρμόζονται σε γενετικά και μη γενετικά δεδομένα και χρησιμοποιούν νευρωνικά δίκτυα ή αλγόριθμους random forest για να υπολογίσουν την πιθανότητα εμφάνισης μιας ασθένειας (9).
- **Αλγόριθμοι που περιλαμβάνουν ανάλυση διακύμανσης:** Μπορούν να χρησιμοποιηθούν για να διαχωρίσουν την επιρροή γενετικών παραγόντων από τους περιβαλλοντικούς με σκοπό να υπολογιστεί το ποσοστό της διακύμανσης που οφείλεται σε κάθε κατηγορία παραγόντων (10).

Ο υπολογισμός της κληρονομησιμότητας μας δείχνει την αναλογία των διακυμάνσεων των φαινοτύπων που επηρεάζονται περισσότερο από γενετικούς από ότι περιβαλλοντικούς παράγοντες, δηλαδή το κατά πόσο ένα χαρακτηριστικό των προγόνων κληροδοτείται στους απόγονους εξαιτίας γενετικών παραγόντων (11). Είναι ένα πολύ ευαίσθητο ζήτημα καθώς επειδή ακριβώς εξαρτάται από πολλές παραμέτρους, στις περισσότερες περιπτώσεις, λαμβάνουμε πολλές διαφορετικές τιμές για το ίδιο χαρακτηριστικό και επομένως απαιτείται περεταίρω μελέτη για να αποφανθεί κανείς για το αν αυτό έχει κληροδοτηθεί γενετικά. Σε γενικές γραμμές η κληρονομησιμότητα όπως υπολογίζεται σήμερα, είναι ο λόγος της γενετικής διακύμανσης προς τη φαινοτυπική διακύμανση και οι τιμές της κληρονομησιμότητας περιορίζονται στο εύρος [0-1] (6) (12).

Όταν η τιμή κληρονομησιμότητας είναι πιο κοντά στο μηδέν ή ίση με μηδέν υποδηλώνει ότι η διαφοροποίηση εκδήλωσης του εκάστοτε χαρακτηριστικού μελέτης οφείλεται κυρίως σε περιβαλλοντικούς παράγοντες και λιγότερο σε γενετικούς παράγοντες, ενώ αντίστοιχα όσο πιο κοντά στο 1 είναι η τιμή της κληρονομησιμότητας για το εκάστοτε χαρακτηριστικό μελέτης τόσο περισσότερο σε οφείλεται η εκδήλωση της γενετικής διαφοροποιήσεως και

αντίστοιχα τόσο λιγότερο σε περιβαλλοντικές επιρροές και τελικά τόσο περισσότερο κληρονομήσιμο είναι το χαρακτηριστικό (12).

1.3 Επιπολασμός και Κληρονομησιμότητα

Ένα εξίσου σημαντικό φαινόμενο που μελετάται παράλληλα με την κληρονομησιμότητα ενός χαρακτηριστικού, και ειδικότερα μίας ασθένειας, είναι ο επιπολασμός (13). Ο επιπολασμός είναι ένα φαινόμενο που υποδηλώνεται με μια αριθμητική τιμή και αποτελεί το ποσοστό εμφάνισης του χαρακτηριστικού μελέτης στον πληθυσμό, μια συγκεκριμένη χρονική στιγμή. Υπολογίζεται ως ο λόγος των ανθρώπων του πληθυσμού που έχουν το χαρακτηριστικό μελέτης προς το σύνολο του πληθυσμού και έχει εύρος τιμών 0-100 ως ποσοστό και αντίστοιχα 0-1 για μη ποσοστιαίες ενδείξεις.

Στην περίπτωση του επιπολασμού όσο πιο κοντά στο 0 είναι η τιμή για ένα χαρακτηριστικό, τόσο λιγότερο συχνά εμφανιζόμενο είναι το χαρακτηριστικό σε συγκεκριμένο πληθυσμό για την συγκεκριμένη χρονική στιγμή, ενώ όσο πιο κοντά στο 1 είναι η τιμή τόσο πιο συχνά εμφανιζόμενο το χαρακτηριστικό στον πληθυσμό.

Η συσχέτιση του επιπολασμού και της κληρονομησιμότητας έχει γίνει αντιληπτή σε ορισμένες περιπτώσεις ασθενειών όπως στην περίπτωση του άρθρου (14), που μελετά την συσχέτιση της υψηλής Κληρονομησιμότητας, του υψηλού επιπολασμού και της μειωμένης αναπαραγωγικής ικανότητας και ακολουθεί το πρότυπο συνδυασμού δυο φαινομένων και αναίρεσης του τρίτου σε διάφορες ασθένειες όπως ο διαβήτης τύπου 2 και η υπέρταση που έχουν υψηλή κληρονομησιμότητα και επιπολασμό και δεν συμβαδίζει με την μειωμένη αναπαραγωγική ικανότητα και τη σχιζοφρένεια που δεν συμβαδίζει με το πρότυπο.

Ωστόσο δεν έχει πραγματοποιηθεί εκτενής μελέτη γύρω από την συσχέτιση των δυο φαινομένων. Σκοπός της συγκεκριμένης μελέτης είναι να εντοπίσει τις διακυμάνσεις στις τιμές κληρονομησιμότητας για τις ασθένειες που προέρχονται από γενετικά και μη δεδομένα, να εξάγει συμπεράσματα σε σχέση με την κληρονομησιμότητα της κάθε ασθένειας και το είδος των δεδομένων, το φύλο και την καταγωγή ενώ επίσης να ελέγξει αν υπάρχει συσχέτιση μεταξύ της κληρονομησιμότητας και του επιπολασμού.

Μέθοδοι

2.1 Συλλογή δεδομένων

Τα δεδομένα που πραγματεύεται η μελέτη και αφορούν στην κληρονομησιμότητα, συγκεντρώθηκαν από 10 μελέτες οι οποίες και επιλέχθηκαν έπειτα από αναζήτηση στην βάση δεδομένων βιοϊατρική βιβλιογραφίας «PUBMED». Οι μελέτες επιλέχθηκαν με κριτήριο της ύπαρξης αριθμητικών δεδομένων που αφορά σε χαρακτηριστικά καθώς και πληροφορίες που αφορούν στον τρόπο υπολογισμού τους. Παρακάτω παρατίθενται οι μελέτες που αξιοποιήθηκαν:

TITLE	YEAR
A global overview of pleiotropy and genetic architecture in complex traits (15)	2019
Phenome-wide heritability analysis of the UK Biobank(16)	2017
Estimate of disease heritability using 4.7 million familial relationships inferred from electronic health records (17)	2016
Estimating heritability and genetic correlations from large health datasets in the absence of genetic data (18)	2019
Assessing the performance of genome-wide association studies for predicting disease risk (19)	2019
Disease heritability inferred from familial relationships reported in medical records (20)	2018
Estimating the proportion of disease heritability mediated by gene expression levels (21)	2017
Estimating SNP-Based Heritability and Genetic Correlation in Case-Control Studies Directly and with Summary Statistics (22)	2018
Finding the missing heritability of complex diseases(23)	2013
Inferring the Nature of Missing Heritability in Human Traits Using Data from the GWAS Catalog(24)	2019

Πίνακας 1 Οι μελέτες κληρονομησιμότητας που αξιοποιήθηκαν στα πλαίσια της εργασίας.

Πιο αναλυτικά, χρησιμοποιήθηκε το εργαλείο Microsoft excel 2019 για να κατασκευαστεί μια βάση δεδομένων. Οι πρώτες 2 στήλες αφορούν στην ταυτότητα της εκάστοτε μελέτης που ερευνήθηκε καθώς αναγράφεται ο τίτλος και το doi.

Σε επόμενη στήλη γίνεται αναζήτηση του τύπου δεδομένων, δηλαδή εδώ αναγράφεται αν τα δεδομένα προήλθαν από μελέτες οικογενειών (Family study)

που συγκρίνουν μέλη οικογενειών μεταξύ τους για να γίνει έλεγχος κοινών χαρακτηριστικών, από μελέτες διδύμων (Twin study) που στοχεύουν σε ότι και οι μελέτες οικογενειών, απλώς μεταξύ δίδυμων αδερφών, από δεδομένα ηλεκτρονικών αρχείων υγείας (EHR data) ή και συνδυασμός αυτών, καθώς και αποτελέσματα αναλύσεων μελετών από γενετικά δεδομένα (GWAS, GWAS summary statistics, Polygenic Risk Score, SNP based data).Για μια περισσότερο ολοκληρωμένη εικόνα για τους τύπους δεδομένων που χρησιμοποίησε η κάθε αναγραφόμενη μελέτη σημειώθηκε και ο τρόπος συλλογής των δεδομένων. Σε αυτή τη στήλη έχει πρακτικά καταγραφεί το αν τα δεδομένα συλλέχθηκαν από έρευνες μεγάλης κλίμακας ή από ηλεκτρονικά αρχεία υγείας από νοσοκομεία και μεγάλα ιατρικά κέντρα , από γενετικές δοκιμές ή είναι αυτοαναφερόμενα (self-reported)(25).

Ακολουθεί η στήλη με την καταγραφή των χαρακτηριστικών όπως αυτά εντοπίστηκαν μέσα στις μελέτες , η καταγραφή των traits με πιο τυποποιημένο και ομαδοποιημένο τρόπο. Η ομαδοποίηση πραγματοποιήθηκε έπειτα από αναζήτηση και καταγραφή των χαρακτηριστικών σύμφωνα με την διεθνή στατιστική ταξινόμηση νόσων και συναφών προβλημάτων υγείας «ICD-10». Αυτό σημαίνει ότι όσα από τα traits της στήλης που καταγράφηκαν στην μορφή όπου εντοπίστηκαν είχαν ίδιο κωδικό ταξινόμησης καταγράφηκαν με το ίδιο όνομα χαρακτηριστικού.

Εκτός από το όνομα και τον κωδικό του χαρακτηριστικού, καταγράφηκε και το όνομα αλλά και ο κωδικός της ευρύτερης ομάδας που ανήκει το εκάστοτε χαρακτηριστικό. Επειδή ωστόσο υπήρχαν και χαρακτηριστικά που δεν είχαν κωδικοποίηση ταξινόμησης στο ICD10, σε εκείνες μόνο τις περιπτώσεις αναγράφηκε στις στήλες κωδικών η έκφραση «no icd10 code», ενώ στην στήλη που αναγράφεται η ευρύτερη κατηγορία των χαρακτηριστικών για τις ανάγκες της μελέτης καταγράφηκαν με βάση την ιδιότητά τους. Περισσότερα για την ομαδοποίηση και την ταξινόμηση των χαρακτηριστικών θα αναφερθούν στην παράγραφο «Διαλογή δεδομένων».

Στα δεδομένα που καταγράφηκαν και αφορούν στην Κληρονομησιμότητα, ανήκουν οι πληροφορίες του πληθυσμού μελέτης που αφορούν στο πλήθος, το φύλο και την καταγωγή καθώς και η τιμή και ο τρόπος υπολογισμού της Κληρονομησιμότητας κάθε χαρακτηριστικού. Σε ότι αφορά τον επιπολασμό των χαρακτηριστικών καταγράφηκαν σε στήλες ο επιπολασμός ανά βιολογικό

φύλο (Male / Female), ανά ηλικιακή ομάδα (adults/ children) και η γενική τιμή επιπολασμού (Prevalence in general) καθώς καταγράφηκε και η πηγή των πληροφοριών για τον επιπολασμό.

2.2 Διαλογή δεδομένων

Τα χαρακτηριστικά που συλλέχθηκαν, αρχικά καταγράφηκαν αυτούσια από τις μελέτες στην στήλη «Phenotypes/Traits» και στην συνέχεια αναζητήθηκε ο κωδικός τους στην κλίμακα ταξινόμησης ICD10. Σύμφωνα με αυτή οι ασθένειες και οι παθήσεις εντάσσονται στις παρακάτω κατηγορίες:

Κωδικός	Κατηγορία
A00-B99	Ορισμένα λοιμώδη και παρασιτικά νοσήματα.
C00-D48	Νεοπλάσματα.
D50-D89	Παθήσεις του αίματος, των αιμοποιητικών οργάνων και ορισμένες διαταραχές του ανοσολογικού μηχανισμού.
E00-E90	Ενδοκρινικές, διατροφικές και μεταβολικές παθήσεις.
F00-F99	Ψυχικές διαταραχές και διαταραχές συμπεριφοράς.
G00-G99	Παθήσεις του νευρικού συστήματος.
H00-H59	Παθήσεις του οφθαλμού και των εξαρτημάτων του.
H60-H95	Παθήσεις του ωτός και της μαστοειδούς απόφυσης.
I00-I99	Παθήσεις του κυκλοφορικού συστήματος.
J00-J99	Παθήσεις του αναπνευστικού συστήματος.
K00-K93	Παθήσεις του πεπτικού συστήματος.
L00-L99	Παθήσεις του δέρματος και του υποδόριου ιστού.
M00-99	Παθήσεις του μυοσκελετικού συστήματος και του συνδετικού ιστού.
N00-N99	Παθήσεις του ουροποιογεννητικού συστήματος.
O00-O99	Κύηση, τοκετός και λοχεία.
P00-P96	Καταστάσεις που έχουν την αρχή τους στην περιγεννητική περίοδο.
Q00-Q99	Συγγενείς ανωμαλίες, διαμαρτίες της διάπλασης και χρωμοσωμικές ανωμαλίες.
R00-R99	Συμπτώματα, σημεία και παθολογικά κλινικά και εργαστηριακά ευρήματα που δεν ταξινομούνται αλλού.
S00-T98	Συνέπειες τραυματισμού, δηλητηρίασης και ορισμένες άλλες εξωγενείς αιτίες.
V01-Y98	Εξωγενή αίτια νοσηρότητας και θνησιμότητας.
Z00-Z99	Παράγοντες που επηρεάζουν την κατάσταση της υγείας και την επικοινωνία με τις υπηρεσίες υγείας.

Πίνακας 2 Κατηγορίες ασθενειών όπως προτείνει η ICD10.

Έπειτα και από την ανεύρεση του κωδικού στην κλίμακα, τα χαρακτηριστικά καταχωρήθηκαν σύμφωνα με την ονομασία του κωδικού ανεύρεσης τους στη στήλη «Phenotypes Without Subcategories» και στην συνέχεια καταγράφηκε και η ευρύτερη κατηγορία που ανήκει το κάθε χαρακτηριστικό, δηλαδή το Chapter level. Επειδή ωστόσο, υπήρχαν χαρακτηριστικά τα οποία δεν είχαν κωδικό καταχώρησης στην κλίμακα ταξινόμησης και συνακόλουθα δεν ανήκαν σε ευρύτερη κατηγορία, για την εξυπηρέτηση της ταξινόμησης, τα χαρακτηριστικά αυτά ταξινομήθηκαν με βάση την ιδιότητα τους.

Έτσι όσα από τα «αταξινόμητα» χαρακτηριστικά δεν ήταν ασθένειες ή παθήσεις και αφορούσαν σε ανατομικές δομές του σώματος ή ακόμη και χαρακτηριστικά του φαινοτύπου όπως για παράδειγμα το χρώμα μαλλιών καταχωρήθηκαν στην κατηγορία «**Structure**» ενώ τα χαρακτηριστικά που αναφέρονταν σε λειτουργίες ή μετρήσεις επιπέδων διεργασιών του σώματος αποτέλεσαν μέλη της κατηγορίας «**Functions**». Στην κατηγορία «**Activities**» καταχωρήθηκαν τα χαρακτηριστικά που αφορούν σε διάφορες δραστηριότητες του ατόμου στην κοινωνική του ζωή, ενώ αντίστοιχα, στην κατηγορία «**Questionnaire**» εντάχθηκαν τα χαρακτηριστικά που αντλήθηκαν υπό μορφή ερωτηματολογίου και αφορούν σε διάφορες πτυχές της ζωής του ατόμου. Τα χαρακτηριστικά που σχετίζονταν με την χρήση προϊόντων ιατροφαρμακευτικής περίθαλψης και τη λήψη φαρμάκων αποτέλεσαν την κατηγορία «**Healthcare products and medication**».

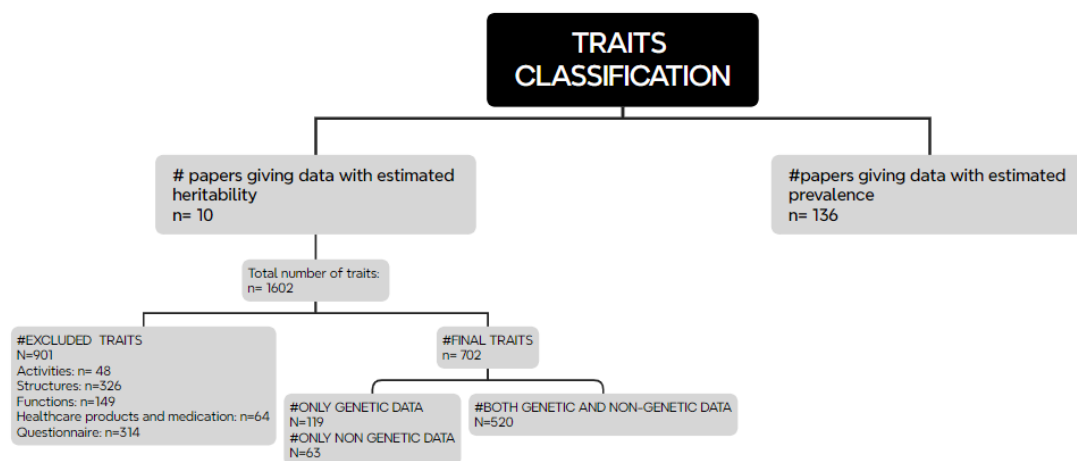


Figure 1 Διάγραμμα διαχείρισης των δεδομένων της μελέτης.

Μετά και από την κατανομή των χαρακτηριστικών σε ομάδες, η βάση δεδομένων διαθέτει **1602** ομαδοποιημένα χαρακτηριστικά, εκ των οποίων τα **48** χαρακτηριστικά εντάσσονται στην κατηγορία «**Activities**», **149** στην κατηγορία «**Functions**», **326** στην κατηγορία «**Structures**», **313** στην κατηγορία «**Questionnaire**», και **64** καταχωρήθηκαν στην κατηγορία «**Healthcare products and medication**». Η διαδικασία διαλογής και καταχώρησης δεδομένων σε ομάδες, περιγράφεται αναλυτικά στο παρακάτω διάγραμμα

Δεδομένου ότι τα χαρακτηριστικά ενδιαφέροντος της μελέτης είναι τα χαρακτηριστικά που αφορούν σε ασθένειες και παθήσεις, απομακρύνθηκαν σε πρώτη φάση όλα τα δεδομένα τα οποία δεν ήταν σχετικά. Σε δεύτερη φάση πραγματοποιήθηκε απομάκρυνση **182** χαρακτηριστικών που εμφανίζονταν μόνο μια φορά στην βάση εκ των οποίων τα **119** αφορούσαν σε γενετικά δεδομένα και **63** σε μη γενετικά δεδομένα. Συμπερασματικά πραγματοποιήθηκε ανάλυση σε **520** χαρακτηριστικά.

Για τον επιπολασμό πραγματοποιήθηκε στοχευμένη αναζήτηση για το εκάστοτε χαρακτηριστικό και εντοπίστηκαν τιμές για 186 από τον συνολικό αριθμό των χαρακτηριστικών. Στο εξής κάθε φορά που θα αναφέρεται ο όρος «χαρακτηριστικό» θα εννοείται ασθένεια ή πάθηση, μιας και η μελέτη πραγματεύεται αποκλειστικά τέτοια χαρακτηριστικά.

2.3 Συμπλήρωση δεδομένων

Βασικό κριτήριο επιλογής των δεδομένων ήταν η ύπαρξη της τιμής της Κληρονομησιμότητας καθώς και ο τρόπος υπολογισμού της για κάθε χαρακτηριστικό. Από τα δεδομένα προέκυψαν Ωστόσο, όπως είναι φυσικό από διαφορετικούς τρόπους υπολογισμού, προκύπτουν διαφορετικές τιμές Κληρονομησιμότητας για το ίδιο χαρακτηριστικό. Συγκεκριμένα κάποιοι από τις μεθόδους υπολογισμού σε πολλές περιπτώσεις έδιναν πολύ αρνητικές τιμές ή τιμές πολύ μεγαλύτερες από το 1 κάτι που δεν συμβαδίζει με το πρωτόκολλο υπολογισμού της Κληρονομησιμότητας. Αυτό μπορεί να συμβαίνει για 3 λόγους (26):

- **Χρήση μοντέλου:** Πολλά μαθηματικά μοντέλα που χρησιμοποιούνται ως πανάκεια για τον υπολογισμό τιμών γενικότερα, όταν πρόκειται για τον υπολογισμό τιμών Κληρονομησιμότητας, δεν λαμβάνουν υπόψη ειδικές παραμέτρους, όπως για παράδειγμα την συνεπίδραση γονιδίων για την έκφραση ενός χαρακτηριστικού (gene - gene interaction).
- **Θόρυβος:** Σε περιπτώσεις όπου το δείγμα έρευνας είναι μικρό και άρα οι παρατηρήσεις είναι περιορισμένες ή σημειώνεται μεγάλη διακύμανση τιμών στα δεδομένα.
- **Λαθεμένη θεώρηση φαινοτύπου:** Υπάρχουν περιπτώσεις χαρακτηριστικών όπου παρότι το γενετικό υπόβαθρο καθορίζει συγκεκριμένο φαινότυπο, εξαιτίας περιβαλλοντικών παραγόντων παρουσιάζουν διαφορετικό φαινότυπο από το αναμενόμενο. Αυτή η διαφορά προκαλεί πρόβλημα στα μοντέλα που δεν συμπεριλαμβάνουν πρόσθετες γενετικές επιδράσεις καθώς υπολογίζουν τις φαινοτυπικές ομοιότητες δεδομένες ως απόρροια των γενετικών ομοιοτήτων.

Για τους λόγους αυτούς, στα δεδομένα όπου σημειώθηκε αρνητική τιμή ή τιμή μεγαλύτερη της μονάδας, παρότι οι τιμές αυτές δεν πρέπει να απορριφθούν ως ανούσιες ή στατιστικά λάθος καθώς μπορεί να περιέχουν πληροφορία (26) τις διαχωρίσαμε από τις υπόλοιπες μελέτες της ανάλυσης.

Μια βασική παράμετρος για την εξαγωγή συμπερασμάτων αποτελεί και η καταγωγή. Επειδή, υπήρχαν πολλές μελέτες που πραγματοποιήθηκαν και καταχωρήθηκαν σε συγκεκριμένη χώρα ή περιπτώσεις δεδομένων που πραγματοποιήθηκαν σε συγκεκριμένες κοινοτικές ομάδες (λόγου χάρη Ashkenazi) οι πληθυσμοί της μελέτης καταχωρήθηκαν σύμφωνα με κωδικό καταγωγής. Ποιο συγκεκριμένα χρησιμοποιήθηκαν οι κωδικές ονομασίες(27):

EUR: Για πληθυσμούς με καταγωγή από χώρες της Ευρώπης, το Ηνωμένο Βασίλειο και την Αυστραλία.

EAS: Για πληθυσμούς με καταγωγή από Ανατολική Ασία όπως η Κίνα, Ιαπωνία, Κορέα.

SAS: Για πληθυσμούς με καταγωγή από την νότια Ασία όπως η Ινδία, Πακιστάν, Νεπάλ.

AFR: Για πληθυσμούς με καταγωγή από χώρες της Αφρικής.

AMR: Για πληθυσμούς που αποτελούν μίγμα ημεδαπών Αμερικανών, Ευρωπαίων Αμερικανών και Αμερικανών της Αφρικής.

MIDDLE EAST: Για πληθυσμούς που αποτελούν περιπτώσεις συγκεκριμένων κοινοτικών ομάδων. Επειδή αυτή η κατηγορία εμφανίστηκε μόλις 6 φορές, εν τέλει δεν χρησιμοποιήθηκε στην διεξαγωγή αποτελεσμάτων που σχετίζονται με την καταγωγή.

Σε πολλές περιπτώσεις, όπως είναι φυσικό, ο πληθυσμός αποτέλεσε συνδυασμό των παραπάνω κωδικών.

Οι τιμές της κληρονομησιμότητας κατηγοριοποιήθηκαν σύμφωνα και με τον τρόπο υπολογισμού τους, καθώς αυτό αποτελεί μια ακόμη εξήγηση για την όποια διακύμανση μεταξύ υπολογισμών Κληρονομησιμότητας για το ίδιο χαρακτηριστικό. Οι μέθοδοι υπολογισμού που συνάντησε η μελέτη ήταν ονομαστικά οι: GREML, LDSC, LDAK, PCGC, ACE, SOLAR, PRS, Linear mixed models, Genetic risk και Sibling recurrence risk. Η κάθε μια από τις παραπάνω μεθόδους λειτουργεί με διαφορετικές παραμέτρους και μπορούν να κατηγοριοποιηθούν σε ομάδες ανάλογα με τον τύπο δεδομένων που διαχειρίζονται.

Η μέθοδος GREML (28) χρησιμοποιεί γενετικά δεδομένα (SNP-based data) από άτομα που δεν συνδέονται με οικογενειακούς δεσμούς μεταξύ τους για να υπολογίσει SNP Κληρονομησιμότητα, ενώ ο αλγόριθμος LDAK (29) λειτουργεί όπως ο GRML με την διαφορά ότι επειδή υποθέτει ότι η διακύμανση που εξηγείται από σπάνιες μεταλλάξεις είναι πιο μεγάλη από αυτή που εξηγείται από κοινές μεταλλάξεις και επομένως καλείται να χρησιμοποιήσει κάποιες τεχνικές όπως ομοιόμορφη κλιμάκωση στον πίνακα συγγένειας για να αντιμετωπίσει το bias όταν προκύπτει αρνητική ή πολύ μεγάλη τιμή στο LD score. Η μέθοδος LDSC (30) υπολογίζει και αυτή τιμή κληρονομησιμότητας μέσω του LD score με την διαφορά ότι χρησιμοποιεί genetic correlation μεταξύ των χαρακτηριστικών χρησιμοποιώντας SNP-based data. Τα Linear Mixed Models, όπως αναφέρθηκε και παραπάνω, είναι μαθηματικά μοντέλα που για τον υπολογισμό της κληρονομησιμότητας αξιοποιούν γενετικά δεδομένα ενώ και οι μέθοδοι Polygenic risk score

(PRS)(31) και GWAS ROCS (19) χρησιμοποιούν αποκλειστικά γενετικά δεδομένα για τον υπολογισμό της Κληρονομησιμότητας (29).

Από την άλλη, οι μέθοδοι ACE (32), SOLAR (33), Sibling recurrence risk και Genetic risk χρησιμοποιούν μη γενετικά δεδομένα και αξιοποιούν την φαινοτυπική διακύμανση από δεδομένα οικογενειών και αδελφών ως παράμετρο για να εξάγουν την τιμή της κληρονομησιμότητας.

Ο επιπολασμός από την άλλη έχει μελετηθεί σε συγκριτικά πολύ λιγότερες ασθένειες από την κληρονομησιμότητα και αυτός είναι και ο λόγος που η αναζήτηση του έγινε σε δεύτερο χρόνο με αναζήτηση όρων της εκάστοτε ασθένειας και της λέξης «prevalence» στην βιβλιογραφική βάση «PUBMED».

Σε περιπτώσεις όπου ο γενικός επιπολασμός δινόταν αποκλειστικά κατά κριτήριο, υπολογίστηκε ο μέσος όρος και η τιμή που προέκυψε θεωρήθηκε ο επιπολασμός του γενικού πληθυσμού. Τέλος επειδή σε αρκετές μελέτες, η τιμή του επιπολασμού δινόταν περιφραστικά, ως αναλογία των 100,000 ατόμων ή ως ποσοστό, μετατρέψαμε όλες τις τιμές των δεδομένων σε δεκαδικού αριθμού 4^ο ψηφίου σύμφωνα με τον παρακάτω τύπο (34):

$$P() = TC/TP$$

P():prevalence rate

TC: the total number of cases

TP: the total population size

Για ευνόητους λόγους επιλέχθηκαν οι τιμές επιπολασμού που πρότειναν οι μελέτες που πραγματοποιήθηκαν πιο πρόσφατα και αποτελούσαν review. Σε αρκετές περιπτώσεις, ο επιπολασμός δινόταν με κριτήριο την ηλικιακή ομάδα δηλαδή ενήλικες (πληθυσμός ηλικίας >18) και παιδιά (πληθυσμός ηλικίας <18) ή ακόμη και με κριτήριο το βιολογικό φύλο (άντρας ή γυναίκα).

Στα πλαίσια της έρευνας χρησιμοποιήθηκε το data analysis button του excel για την υλοποίηση t-tests paired two samples for means που εφαρμόστηκαν στους μέσους όρους των 2 μεταβλητών, της κληρονομησιμότητας από γενετικά και της κληρονομησιμότητας από μη γενετικά δεδομένα ανά κατηγορία του ICD10, με σκοπό να βρεθεί αν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των κατηγοριών δεδομένων ή όχι. Στον παρακάτω πίνακα παρατίθενται τα βασικά μέρη του πίνακα αποτελεσμάτων του t-tests paired two samples for means:

Mean: δίνει την μέση τιμή κάθε μεταβλητής.

Variance: πληροφορεί για την διακύμανση των τιμών της κάθε μεταβλητής.

Observations: δηλώνει το πλήθος τιμών που πραγματεύεται η κάθε μεταβλητή.

Pearsons correlation: υποδηλώνει αν υπάρχει συσχέτιση μεταξύ των 2 μεταβλητών μελέτης. Οι τιμές ανήκουν στο φάσμα $[-1,1]$ (35).

- **Τιμή= -1:** υποδεικνύει τέλεια φθίνουσα γραμμική συσχέτιση μεταξύ των μεταβλητών.
- **Τιμή= 1:** υποδεικνύει τέλεια αύξουσα γραμμική συσχέτιση μεταξύ των μεταβλητών.
- **Τιμή που ανήκει στο διάστημα (-1,1):** όσο πιο κοντά βρίσκεται η τιμή στο -1 ή το 1 τόσο μεγαλύτερη η συσχέτιση των 2 μεταβλητών.

t-value: αποτελεί την τιμή του t-test και μπορεί να είναι θετική ή αρνητική.

- **Αρνητική τιμή t-value:** η πρώτη μεταβλητή έχει χαμηλότερο μέσο όρο από τη δεύτερη.
- **Θετική τιμή t-value:** η πρώτη μεταβλητή έχει υψηλότερο μέσο όρο από τη δεύτερη.

P(T≤t): p-value του t-test. Αν η τιμή σε αυτό το πεδίο είναι <0.05 τότε υπάρχει στατιστικά σημαντική διαφορά μεταξύ των μεταβλητών μελέτης (36).

Επιπλέον πραγματοποιήθηκε μια One-way ANOVA για την μελέτη της κληρονομησιμότητας στις φυλές της έρευνας. Η τιμή p-value και σε αυτή την περίπτωση αποτελεί την σημαντική ένδειξη στατιστικής σημαντικότητας μεταξύ των φυλών. Για την πραγματοποίηση των υπολογισμών που πραγματοποιήθηκαν καθώς και την οπτικοποίηση των αποτελεσμάτων χρησιμοποιήθηκε το εργαλείο ανάλυσης και οπτικοποίησης «Tableau» (<https://www.tableau.com/>).

Σε πρώτη φάση, για την πλήρη αξιοποίηση των δεδομένων, υπολογίστηκε ο μέσος όρος της κληρονομησιμότητας για κάθε ένα από τα χαρακτηριστικά της μελέτης. Στην συνέχεια, δεδομένου ότι όλα τα χαρακτηριστικά της μελέτης που δεν απορρίφθηκαν είχαν τουλάχιστον 2 παρατηρήσεις από γενετικά δεδομένα

και τουλάχιστον 2 παρατηρήσεις από μη γενετικά δεδομένα , υπολογίστηκε ο μέσος όρος κληρονομησιμότητας από γενετικά δεδομένα και ο μέσος όρος κληρονομησιμότητας από μη γενετικά δεδομένα.

Αποτελέσματα

Σκοπός της μελέτης είναι η διεξαγωγή συμπερασμάτων της συσχέτισης της κληρονομησιμότητας των χαρακτηριστικών μελέτης με το φύλο, την εθνικότητα αλλά και τον επιπολασμό ενώ εξίσου σημαντική είναι η παρατήρηση της διαφοράς των αποτελεσμάτων αυτών από γενετικά και μη γενετικά δεδομένα.

3.1 Κληρονομησιμότητα σε γενετικά και μη γενετικά δεδομένα

Επειδή στο σύνολο δεδομένων υπάρχουν και χαρακτηριστικά τα οποία έχουν γενετικά και μη γενετικά δεδομένα αλλά και χαρακτηριστικά που έχουν μόνο γενετικά δεδομένα ή μόνο μη γενετικά δεδομένα, πραγματοποιήσαμε 2 t-test για να παρατηρήσουμε εάν σημειώνονται σημαντικές διαφορές ως προς το εύρος τιμών της κληρονομησιμότητας. Στο Διάγραμμα 1 απεικονίζεται με κόκκινο χρώμα το ιστόγραμμα της κληρονομικότητας των χαρακτηριστικών εκείνων που έχουν γενετικά δεδομένα ενώ με μπλε χρώμα το ιστόγραμμα της κληρονομησιμότητας των χαρακτηριστικών εκείνων που έχουν μη γενετικά δεδομένα. Θεωρώντας ως κατώφλι τα 150 counts, παρατηρούμε ότι για την κληρονομησιμότητα που προέρχεται από γενετικά δεδομένα το εύρος τιμών κυμαίνεται στο διάστημα $[0,0- 0,3]$ ενώ για την κληρονομησιμότητα που προέρχεται από μη γενετικά δεδομένα το εύρος τιμών κυμαίνεται στο διάστημα $[0,2- 0,66]$.

Heritability histogram of traits with genetic and non genetic data

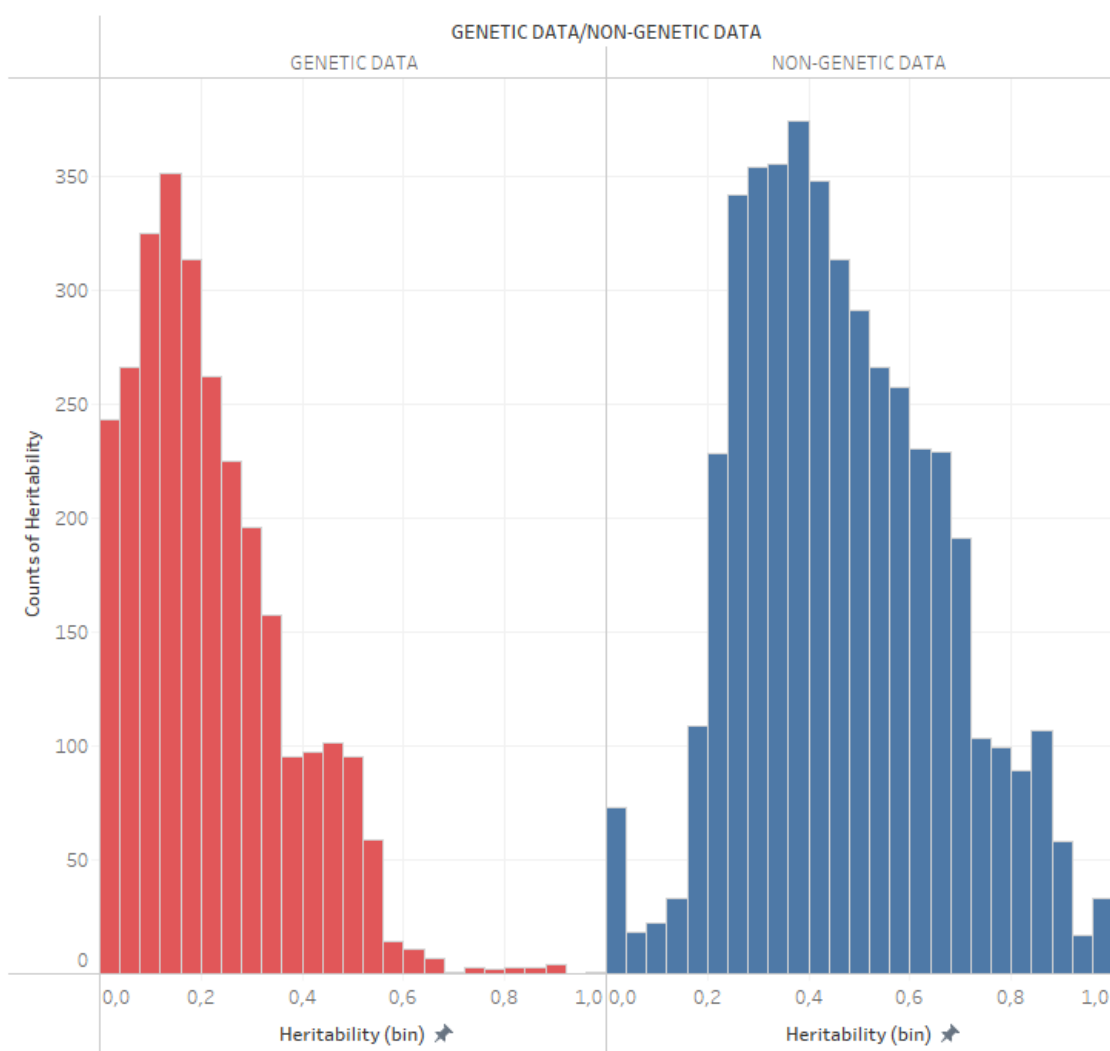


Figure 2 Ιστογράμματα κληρονομησιμότητας όσον αφορά σε γενετικά δεδομένα (κόκκινο χρώμα) και σε μη γενετικά δεδομένα (μπλε χρώμα). Τα ιστογράμματα προέκυψαν από χαρακτηριστικά τα οποία διατηρούν στην βάση δεδομένων τόσο γενετικά όσο και μη γενετικά δεδομένα.

Ο πίνακας 1 παρουσιάζει τα αποτελέσματα του t-test που πραγματοποιήθηκε στα δεδομένα του Διαγράμματος 1. Παρατηρούμε ότι ο μέσος όρος κληρονομησιμότητας των γενετικών δεδομένων είναι σημαντικά χαμηλότερος από τον αντίστοιχο μέσο όρο των μη γενετικών δεδομένων κάτι που φυσικά επιβεβαιώνεται από την αρνητική τιμή του t και τη σημαντικά χαμηλότερη από το 0,05 p-value.

Έλεγχος t του μέσου δύο δειγμάτων συσχετισμένων ζευγών	Genetic data	Non-Genetic data
Μέσος	0,188293387	0,439139875
Διακύμανση	0,013819997	0,019194574
Μέγεθος δείγματος	517	517
Συσχέτιση Pearson	0,825507507	
Υποτιθέμενη διαφορά μέσων	0	
βαθμοί ελευθερίας	516	
t	-72,88245569	
P(T<=t) μονόπλευρη	4,2349E-274	
t κρίσιμο, μονόπλευρο	1,647812009	
P(T<=t) δίπλευρη	8,4698E-274	
t κρίσιμο, δίπλευρο	1,964572029	

Πίνακας 3 Αποτελέσματα t-test

Στο Διάγραμμα 2 παρουσιάζονται τα οπτικοποιημένα αποτελέσματα των τιμών κληρονομησιμότητας από χαρακτηριστικά που παρουσιάζονται στο σύνολο δεδομένων τόσο με γενετικά όσο και με μη γενετικά δεδομένα. Σύμφωνα με τα αποτελέσματα της ανάλυσης, οι τιμές κληρονομησιμότητας των γενετικών δεδομένων παρουσιάζουν και πάλι διαφορά με τις αντίστοιχες τιμές των μη γενετικών δεδομένων. Μάλιστα διατηρώντας το ίδιο κατώφλι των 150 counts, γίνεται εμφανές ότι οι τιμές κληρονομησιμότητας κυμαίνονται στα ίδια επίπεδα με τα προηγούμενα ιστογράμματα. Για τον λόγο αυτό και δεδομένου ότι δεν δύναται να πραγματοποιηθεί t-test σε μεταβλητές οι οποίες διαθέτουν κενά κελιά, λόγου χάρη τα χαρακτηριστικά που έχουν μόνο γενετικά δεδομένα ή μόνο μη γενετικά δεδομένα, στο εξής θα χρησιμοποιηθούν τα χαρακτηριστικά που διαθέτουν ζεύγη γενετικών και μη γενετικών δεδομένων.

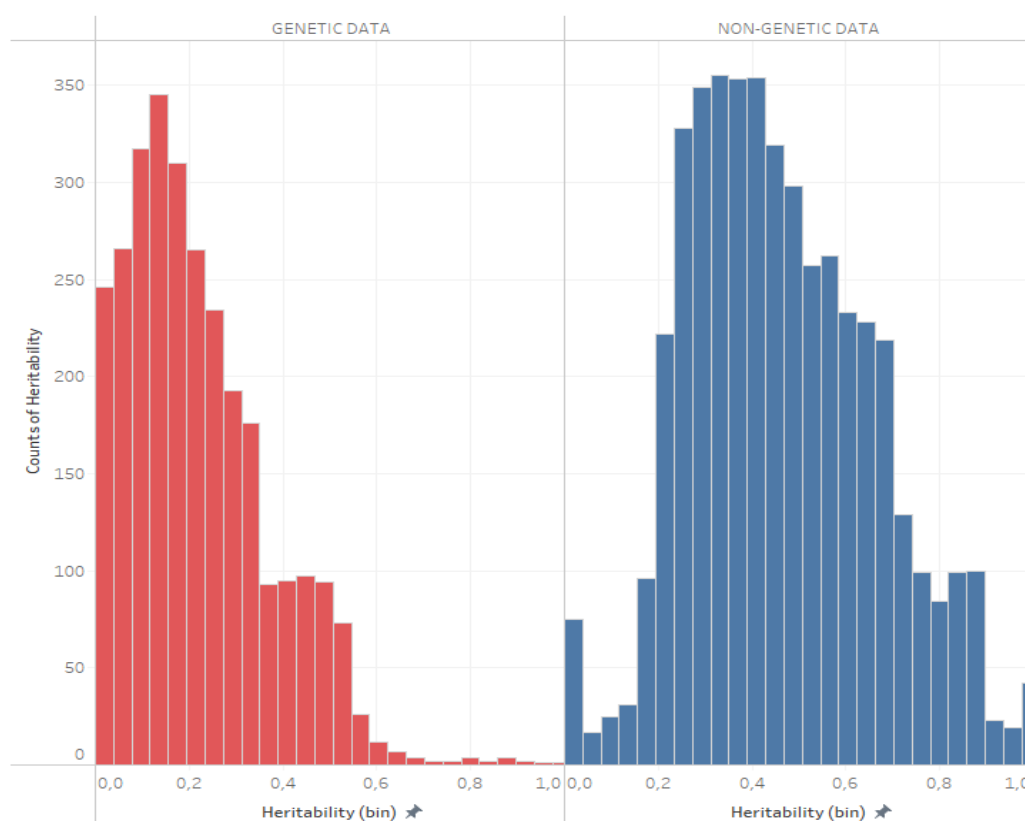


Figure 3 Ιστογράμματα κληρονομησιμότητας που αφορούν σε γενετικά δεδομένα (κόκκινο χρώμα) και σε μη γενετικά δεδομένα (μπλέ χρώμα)

Η διαφορά αυτή μεταξύ «γενετικής» και «μη γενετικής κληρονομησιμότητας» έχει μεγάλη σημασία, καθώς οι τιμές που προκύπτουν από γενετικά δεδομένα έχουν τιμή που είναι πλησιέστερα στο 0 και επομένως η πλειοψηφία των χαρακτηριστικών πολύ λιγότερη επιρροή από γενετικούς από ότι περιβαλλοντικούς παράγοντες. Με το ίδιο επιχείρημα τα μη γενετικά δεδομένα προτείνουν τιμές που πλησιάζουν περισσότερο το 1 και δηλώνουν μια γενετική επιρροή μεγαλύτερη από ότι τα γενετικά δεδομένα.

Αυτή η αυξημένη τιμή κληρονομησιμότητας που προκύπτει από μη γενετικά δεδομένα μπορεί να αιτιολογηθεί από 3 παράγοντες:

- **Μεγαλύτερη επίδραση περιβαλλοντικών από ότι γενετικών παραγόντων:** στα μη γενετικά δεδομένα όπως οι μελέτες οικογενειών και οι μελέτες διδύμων τα μέλη που συμμετέχουν και έχουν μεταξύ τους σχέση, εκτός από κοινά γονίδια μοιράζονται και κοινό περιβάλλον όπως για παράδειγμα τη διατροφή

την έκθεση σε κοινούς παράγοντες όπως είναι ο καπνός και γενικότερα τις ίδιες συνήθειες και ο τρόπος ζωής (37).

- **Επιγενετική:** η επιγενετική είναι ένας ακόμη παράγοντας ενδεχομένως που επηρεάζει τις διαφορές των γενετικών και των μη γενετικών δεδομένων. Εξ ορισμού επιγενετικά θεωρούνται τα χαρακτηριστικά που κληροδοτούν φαινότυπους οι οποίοι προκύπτουν από αλλαγές σε κάποιο χρωμόσωμα χωρίς αλλοιώσεις στην αλληλουχία τους DNA (38) . Τέτοιου είδους επιδράσεις μπορεί να μην λαμβάνονται υπόψη από γενετικά δεδομένα και τους αλγόριθμους υπολογισμού της κληρονομησιμότητας αλλά αντ'αυτού να καταγράφονται και να συνυπολογίζονται στις μετρήσεις μη γενετικών δεδομένων.
- **Πολυπαραγοντικές ασθένειες:** σε περιπτώσεις όπως οι πολυπαραγοντικές ασθένειες οι τιμές της Κληρονομησιμότητας είναι ούτως ή άλλως μικρότερες από ότι θα έπρεπε να είναι καθώς αυτές εξαρτώνται από πολλούς γενετικούς παράγοντες και συνακόλουθα ενδεχομένως εξαρτώνται από πολλά γονίδια που δεν λαμβάνονται υπόψη κατά τον υπολογισμό της Κληρονομησιμότητας αυτών είτε λόγω έλλειψης γνώσης από τον ερευνητή είτε λόγω έλλειψης αντίστοιχων παραμέτρων από την εκάστοτε μέθοδο υπολογισμού(39).

Η διαφορά αυτή των αποτελεσμάτων έγινε σαφής και στο διάγραμμα σύγκρισης της μέσης κληρονομησιμότητας ανά κατηγορία ασθενειών (Chapter level) καθώς για μία ακόμη φορά, η Κληρονομησιμότητα προερχόμενη από μη γενετικά δεδομένα έτεινε να είναι μεγαλύτερη από τις αντίστοιχες τιμές γενετικών σε κάθε κατηγορία.

Genetic and Non-Genetic Heritability

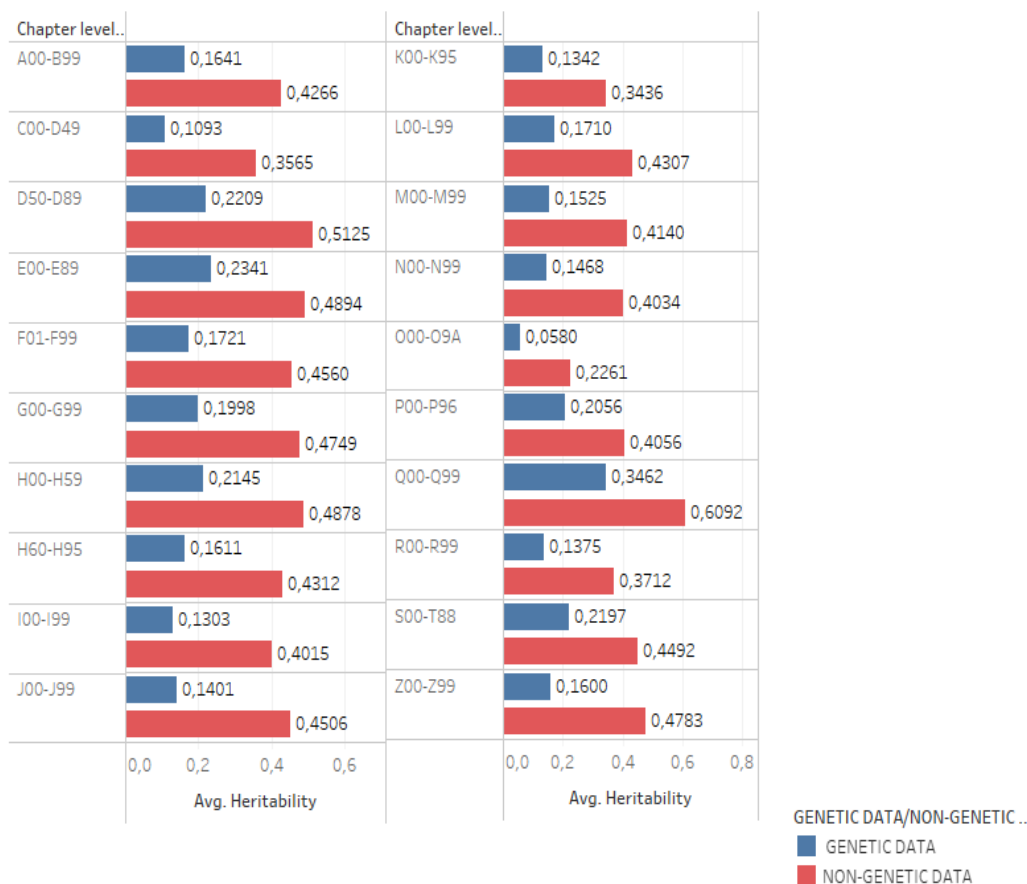


Figure 4 Κληρονομησιμότητα από γενετικά και μη γενετικά δεδομένα ανά κατηγορία ασθενειών του ICD10.

Παρατηρούμε πως η κατηγορία των χρωμοσωμικών ανωμαλιών σημείωσε την μεγαλύτερη τιμή κληρονομησιμότητας τόσο από γενετικά όσο και από μη γενετικά δεδομένα ενώ η αντίστοιχη χαμηλότερη τιμή σημειώθηκε στην κατηγορία που αφορά στην εγκυμοσύνη.

3.2 Κληρονομησιμότητα στα βιολογικά φύλα

Η «απόσταση» στις τιμές της κληρονομησιμότητας μεταξύ γενετικών και μη γενετικών δεδομένων δεν φαίνεται να επηρεάζεται από το φύλο του πληθυσμού μελέτης. Με εξαίρεση της δυο κατηγορίες ασθενειών που σχετίζονται με την προγεννητική περίοδο και τις ασθένειες που σχετίζονται με την εγκυμοσύνη, που αφορά αποκλειστικά σε γυναικείο πληθυσμό, οι

υπόλοιπες κατηγορίες εμφανίζουν διαφορά $\sim 0,01$ ως προς την τιμή μεταξύ των 2 βιολογικών φύλων.

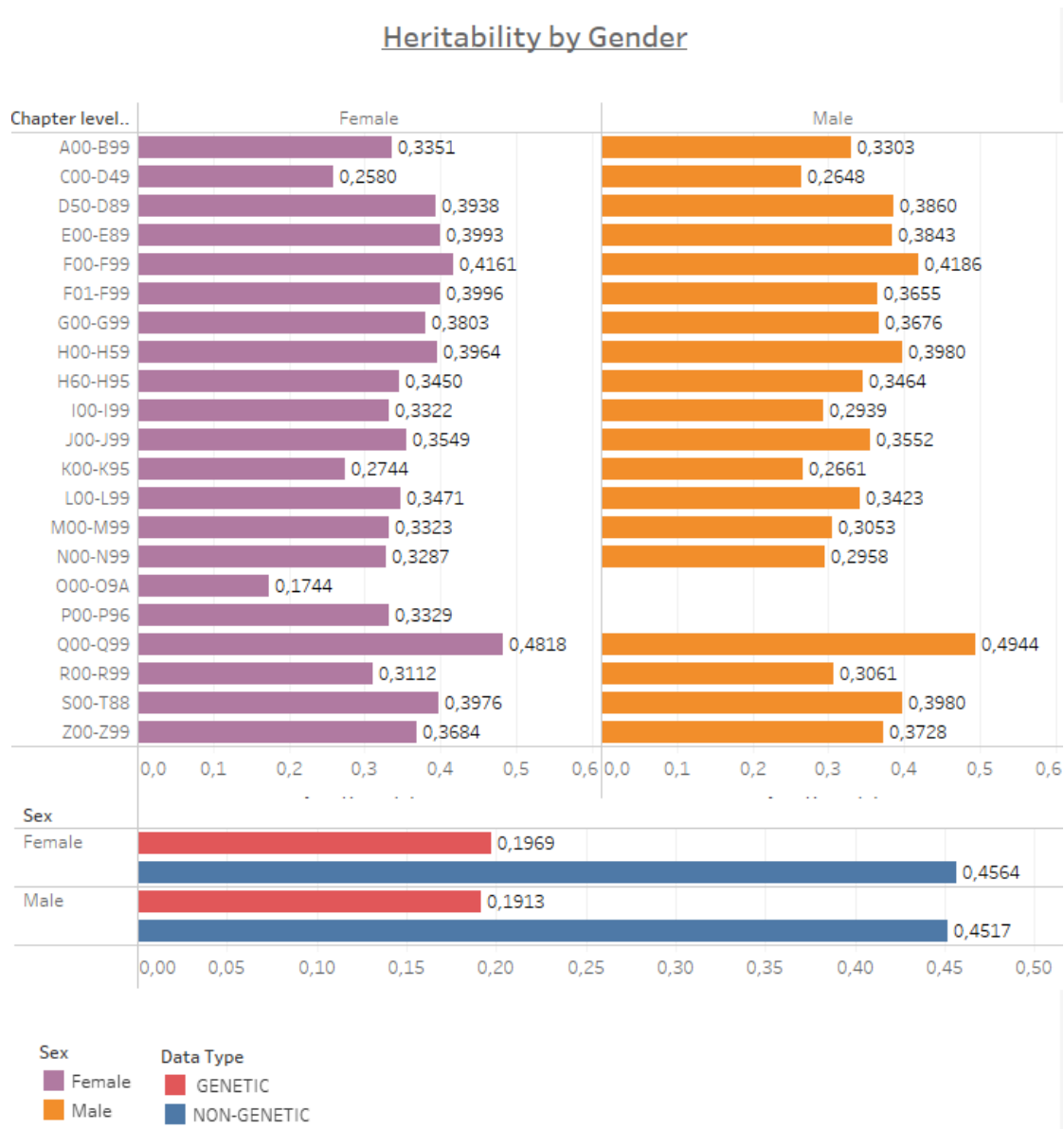


Figure 5 Κληρονομησιμότητα ανά βιολογικό φύλο. Με μοβ χρώματα σημειώνονται οι τιμές κληρονομησιμότητας για το γυναικείο φύλο ενώ με πορτοκαλί για το ανδρικό φύλο ανά κατηγορία ασθένειας. Με κόκκινο χρώμα σημειώνεται η κληρονομησιμότητα από γενετικά δεδομένα και με μπλε χρώμα από μη γενετικά δεδομένα.

Για την μελέτη ύπαρξης στατιστικά σημαντικής διαφοράς κληρονομησιμότητας μεταξύ γενετικών και μη γενετικών δεδομένων ανά φύλο, πραγματοποιήθηκαν 4 t-tests, όπου και αποδείχθηκε ότι υπάρχει στατιστικά σημαντική διαφορά μεταξύ γενετικών και μη γενετικών δεδομένων ανά φύλο αλλά και προς τύπο δεδομένων.

Έλεγχος t του μέσου δύο δειγμάτων συσχετισμένων ζευγών					
	Female Genetic data	Female Non Genetic data		Female Genetic data	Male Genetic data
Μέσος	0,184460183	0,440311443	Μέσος	0,190315481	0,182287846
Διακύμανση	0,003208775	0,005174064	Διακύμανση	0,002582724	0,003181577
Μέγεθος δείγματος	20	20	Μέγεθος δείγματος	18	18
Συσχέτιση Pearson	0,895802567		Συσχέτιση Pearson	0,961102732	
Υποτιθέμενη διαφορά μέσων	0		Υποτιθέμενη διαφορά μέσων	0	
βαθμοί ελευθερίας	19		βαθμοί ελευθερίας	17	
t	-34,77261284		t	2,136196797	
P(T<=t) μονόπλευρη	5,74388E-19		P(T<=t) μονόπλευρη	0,023750751	
t κρίσιμο, μονόπλευρο	1,729132812		t κρίσιμο, μονόπλευρο	1,739606726	
P(T<=t) δίπλευρη	1,14878E-18		P(T<=t) δίπλευρη	0,047501501	
t κρίσιμο, δίπλευρο	2,093024054		t κρίσιμο, δίπλευρο	2,109815578	
	Male Genetic data	Male Non Genetic data		Female Non-Genetic data	Male Non-Genetic data
Μέσος	0,182287846	0,446242531	Μέσος	0,454136233	0,446242531
Διακύμανση	0,003181577	0,003340417	Διακύμανση	0,002811655	0,003340417
Μέγεθος δείγματος	18	18	Μέγεθος δείγματος	18	18
Συσχέτιση Pearson	0,900972958		Συσχέτιση Pearson	0,976230341	
Υποτιθέμενη διαφορά μέσων	0		Υποτιθέμενη διαφορά μέσων	0	
βαθμοί ελευθερίας	17		βαθμοί ελευθερίας	17	
t	-44,00614495		t	2,580314854	
P(T<=t) μονόπλευρη	2,93646E-19		P(T<=t) μονόπλευρη	0,009727756	
t κρίσιμο, μονόπλευρο	1,739606726		t κρίσιμο, μονόπλευρο	1,739606726	
P(T<=t) δίπλευρη	5,87293E-19		P(T<=t) δίπλευρη	0,019455513	
t κρίσιμο, δίπλευρο	2,109815578		t κρίσιμο, δίπλευρο	2,109815578	

Πίνακας 4 t-test ανα τύπο δεδομένων (στήλες 1-3) και ανά φύλο(στήλες 4-6).

Στα t-tests που μελετούν αν υπάρχει στατιστική διαφορά μεταξύ κληρονομησιμότητας γενετικών και μη γενετικών δεδομένων στο ίδιο φύλο παρατηρήθηκε ότι τόσο στους άντρες όσο και στις γυναίκες ο μέσος όρος των μη γενετικών δεδομένων (0,4403 και 0,4462 αντίστοιχα) έτεινε να είναι σημαντικά μεγαλύτερος από τον αντίστοιχο μέσο όρο των γενετικών δεδομένων (0,1844 και 0,1822 αντίστοιχα) καθώς οι τιμές p-value ήταν σημαντικά χαμηλότερες από 0,05.

Στα t test που πραγματοποιήθηκαν ανά φύλο, παρότι οι μέσες τιμές ήταν αρκετά κοντά τόσο στην κληρονομησιμότητα των γενετικών δεδομένων (0,1903 για τις γυναίκες και 0,1822 για τους άντρες) όσο και στην κληρονομησιμότητα των μη γενετικών δεδομένων (0,4541 για τις γυναίκες και 0,4462 για τους άντρες), υπάρχει στατιστική διαφορά μεταξύ των 2 φύλων που επιβεβαιώνεται και από τις τιμές p-value στα 2 tests. Αυτό ενδεχομένως, να αποτελεί ένδειξη συσχέτισης γενετικών και μη γενετικών δεδομένων με το φύλο. Ωστόσο για να επιβεβαιωθεί η συσχέτιση του τύπου δεδομένων με το φύλο είναι απαραίτητες πληροφορίες όπως η γονιδιακή έκφραση και οι παράγοντες κινδύνου κάθε μίας από τις ασθένειες που μελετούμε, πληροφορίες ωστόσο που δεν έχουν συλλεχθεί στην παρούσα βάση δεδομένων.

3.3 Κληρονομησιμότητα ανά ομάδα καταγωγής

Η κληρονομησιμότητα μελετήθηκε και ως προς την καταγωγή με σύγκριση των αποτελεσμάτων κληρονομησιμότητας της κάθε ομάδας καταγωγής ανά ομάδα ασθενειών (Chapter level). Η ομάδα ασθενειών που εμφάνισε την μεγαλύτερη τιμή Κληρονομησιμότητας στην πλειοψηφία των ομάδων καταγωγής (EAS,AMR,EUR,SAS,AFR) είναι η κατηγορία που αφορά σε χρωμοσωμικές ανωμαλίες (Q00-Q99) ενώ η αντίστοιχη χαμηλότερη ήταν η κατηγορία με ασθένειες που αφορούν στην εγκυμοσύνη.

Chapter level..	EAS	AMR	EUR	SAS	AFR
A00-B99	0,3407	0,3443	0,3274	0,3402	0,3406
C00-D49	0,2753	0,2805	0,2553	0,2825	0,2799
D50-D89	0,3970	0,4166	0,3857	0,3966	0,3966
E00-E89	0,3894	0,4064	0,3818	0,3903	0,3918
F01-F99	0,3595	0,3688	0,3372	0,3595	0,3576
G00-G99	0,3633	0,3764	0,3534	0,3678	0,3663
H00-H59	0,3970	0,4044	0,3805	0,3993	0,4001
H60-H95	0,3312	0,3385	0,3339	0,3312	0,3312
I00-I99	0,2983	0,3086	0,2818	0,3021	0,3021
J00-J99	0,3382	0,3462	0,3345	0,3427	0,3405
K00-K95	0,2705	0,2782	0,2540	0,2719	0,2747
L00-L99	0,3301	0,3351	0,3396	0,3322	0,3348
M00-M99	0,3227	0,3289	0,3096	0,3281	0,3291
N00-N99	0,3263	0,3304	0,2939	0,3282	0,3263
O00-O9A	0,2203	0,1990	0,1873	0,2203	0,2203
P00-P96	0,3358	0,3329	0,3358	0,3358	0,3358
Q00-Q99	0,4882	0,5144	0,4893	0,4882	0,4882
R00-R99	0,3110	0,3142	0,2870	0,3110	0,3090
S00-T88	0,3863	0,3779	0,3720	0,3863	0,3863
Z00-Z99	0,3769	0,3841	0,3555	0,3769	0,3697

Πίνακας 5 Κληρονομησιμότητα ανά ομάδα καταγωγής. Στον πίνακα σημειώνονται οι μέσες τιμές κληρονομησιμότητας ανά κατηγορία ασθενειών

Σε γενικές γραμμές οι ομάδα καταγωγής AMR σημείωσε της μεγαλύτερες τιμές Κληρονομησιμότητας στην πλειοψηφία των κατηγοριών ασθένειας. Εξαίρεση αποτελούν οι κατηγορίες που αφορούν τα Νεοπλάσματα

(C00-D49) όπου υψηλότερες τιμές σημείωσε η ομάδα SAS, η κατηγορία παθήσεων του δέρματος και του υποδόριου ιστού (L00-L99) όπου υψηλότερη τιμή σημείωσε η EUR και τέλος οι κατηγορίες που αφορούν την εγκυμοσύνη (O00-O99), την προγεννητική περίοδο (P00-P99) αλλά και την κατηγορία των τραυματισμών και των δηλητηριάσεων (S00-T98) όπου υψηλότερες τιμές συγκέντρωσαν οι ομάδες EAS, EUR, SAS, AFR.

Στη συνέχεια πραγματοποιήθηκε ένας έλεγχος ONE WAY ANOVA για να μελετηθεί αν υπάρχει στατιστικά σημαντική διαφορά με της κληρονομησιμότητας μεταξύ των ομάδων καταγωγής.

Ανάλυση διακύμανσης κατά ένα παράγοντα				
ΣΥΜΠΕΡΑΣΜΑ				
Ομάδες	Πλήθος	Άθροισμα	Μέσος όρος	Διακύμανση
EAS	20	6,183325313	0,309166266	0,003336824
AFR	20	6,200413176	0,310020659	0,003311798
EUR	20	6,196752111	0,309837606	0,003306257
AMR	20	6,20087482	0,310043741	0,003314349
SAS	20	6,203278646	0,310163932	0,003303515

ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ						
Προέλευση διακύμανσης SS	βαθμοί ελευθερίας	MS	F	τιμή-P	κριτήριο F	
Μεταξύ ομάδων	1,26559E-05	4	3,16398E-06	0,000954574	0,9999981	2,4674936
Μέσα στις ομάδες	0,314882106	95	0,003314548			
Σύνολο	0,314894762	99				

Πίνακας 6 ONE WAY ANOVA στις ομάδες καταγωγής.

Τα αποτελέσματα της ANOVA όπως αυτά δίνονται στον πίνακα 3, φανερώνουν ότι οι μέσες τιμές των ομάδων καταγωγής είναι πολύ κοντά μεταξύ τους ενώ το ίδιο συμβαίνει και με τις διακυμάνσεις τιμών σε κάθε ομάδα. Επιπλέον η τιμή p-value είναι σημαντικά μεγαλύτερη από το 0,05 κάτι που υποδηλώνει ότι δεν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των μέσων τιμών των 5 ομάδων.

3.4 Επιπολασμός ανά ομάδα ασθενειών

Στην πλειοψηφία των ασθενειών της μελέτης, οι τιμές του επιπολασμού δεν ξεπέρασαν το 0,02 ελάχιστες ασθένειες σημείωσαν τιμή μέχρι και 0,25 και μόλις 2 ασθένειες ξεπέρασαν το 0,5 όπως άλλωστε υποδηλώνει και το ιστόγραμμα του διαγράμματος 7.

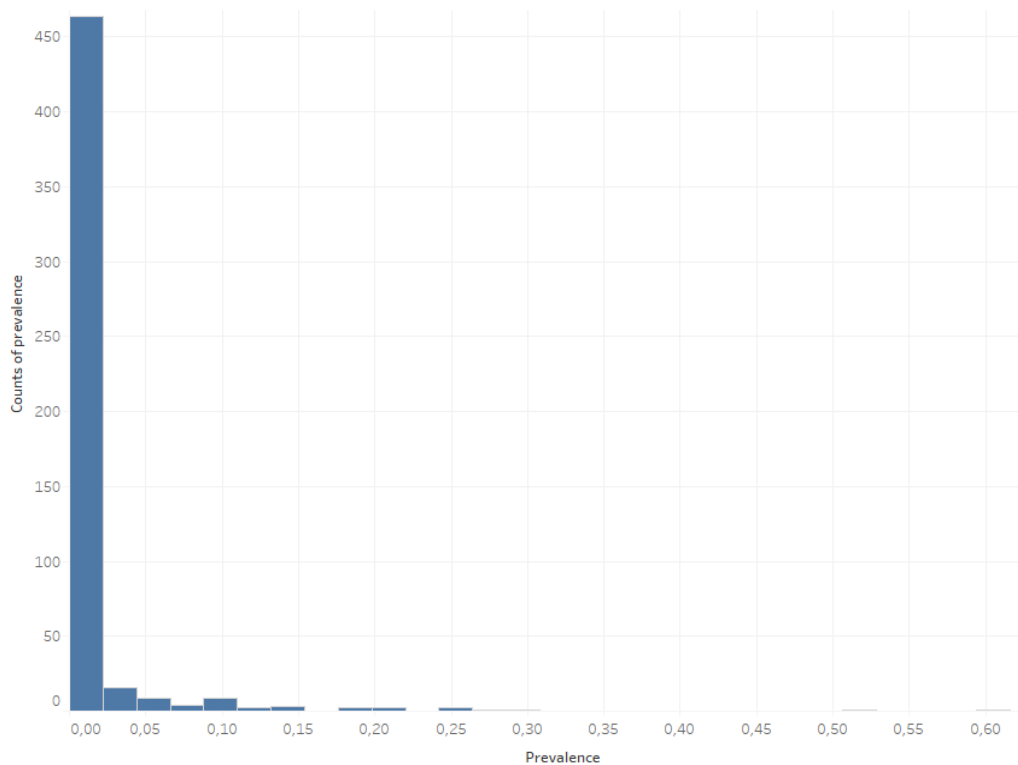


Figure 6 Ιστόγραμμα του επιπολασμού.

Έπειτα από την σύγκριση των τιμών της κληρονομησιμότητας ανά κατηγορία ασθενειών σύμφωνα με την ICD10, δημιουργήθηκε η ανάγκη για ανεύρεση των τιμών του επιπολασμού ανά κατηγορία ασθενειών και σύγκριση με τις αντίστοιχες τιμές κληρονομησιμότητας. Τα αποτελέσματα εμφανίζονται στο παρακάτω διάγραμμα.

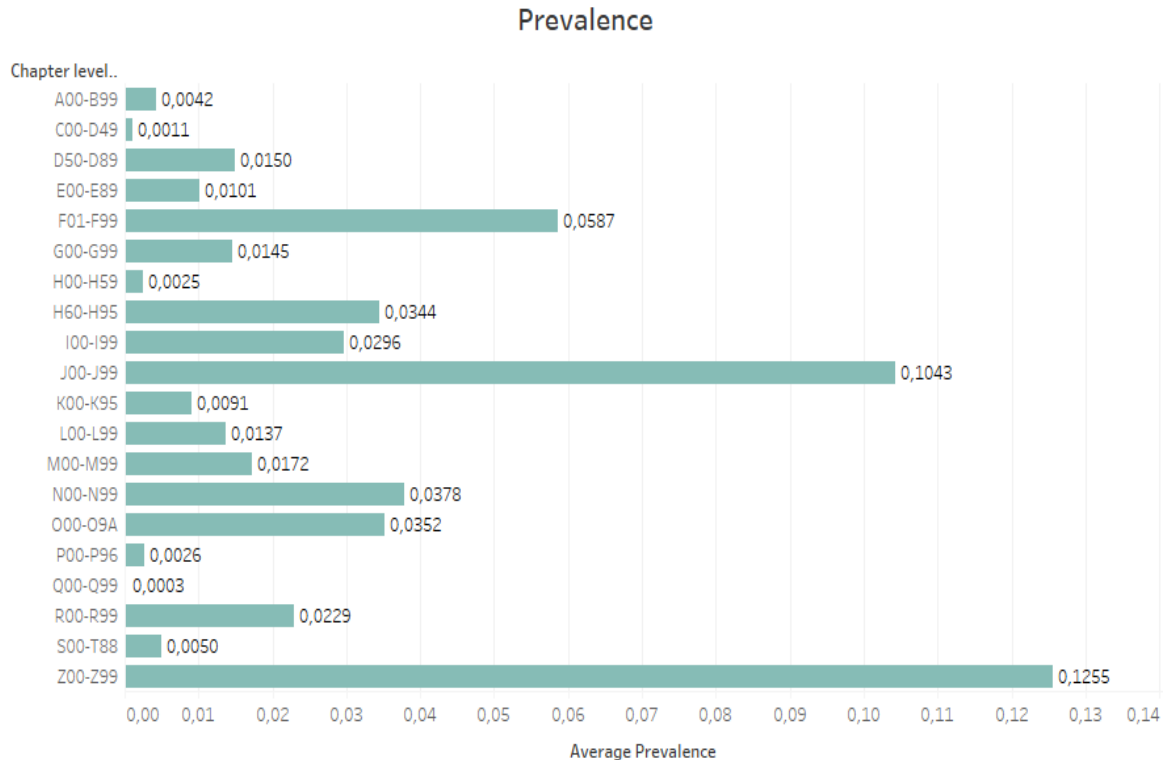


Figure 7 Επιπολασμός ανα κατηγορία ασθένειας

Οι τιμές επιπολασμού των ασθενειών που πραγματεύεται η μελέτη δεν ξεπέρασαν την τιμή 0,13. Η κατηγορία που συγκέντρωσε την μικρότερη τιμή επιπολασμού ήταν η κατηγορία ασθενειών που αφορά σε χρωμοσωμικές ανωμαλίες (Q00-Q99) με τιμή 0,0003 ενώ η μεγαλύτερη τιμή ήταν 0,1255 και σημειώθηκε στην κατηγορία των παραγόντων που επηρεάζουν την κατάσταση της υγείας και την επικοινωνία με τις υπηρεσίες υγείας (Z00-Z99). Οι κατηγορίες παθήσεων του αναπνευστικού συστήματος και οι κατηγορία ψυχικών διαταραχών ξεχώρισαν επίσης με τιμές επιπολασμού 0,1043 και 0,0587 ενώ οι υπόλοιπες κατηγορίες δεν ξεπέρασαν το 0,04.

3.5 Επιπολασμός ανά βιολογικό φύλο

Στην βάση δεδομένων δεν ήταν εφικτό να γίνει συλλογή πληροφορίες επιπολασμού για τις κατηγορίες ψυχικής διαταραχής και διαταραχής συμπεριφοράς (F00-F99), παθήσεων κατά την εγκυμοσύνη (O00-O99) και την προγεννητική περίοδο (P00-P99) και στην κατηγορία συμπτωμάτων και

ευρημάτων που δεν μπορούν να ταξινομηθούν αλλού (R00-R99) ανά βιολογικό φύλο και επομένως δεν υπάρχει εικόνα επιπολασμού ανά φύλο για αυτές τις κατηγορίες.

Αν και μικρές, σημειώθηκαν κάποιες διαφορές στις τιμές επιπολασμού μεταξύ των δύο βιολογικών φύλων. Οι μεγαλύτερες τιμές επιπολασμού τόσο σε άνδρες όσο και σε γυναίκες συγκεντρώθηκαν στις κατηγορίες παραγόντων που επηρεάζουν την κατάσταση της υγείας και την επικοινωνία με τις υπηρεσίες υγείας (Z00-Z99) και ακολουθούν οι παθήσεις του ωτός και της μαστοειδούς απόφυσης (H60-H95) και παθήσεις του αίματος και των αιμοποιητικών οργάνων (D50-D89) με τους άντρες να σημειώνουν μεγαλύτερες τιμές επιπολασμού από τις γυναίκες.

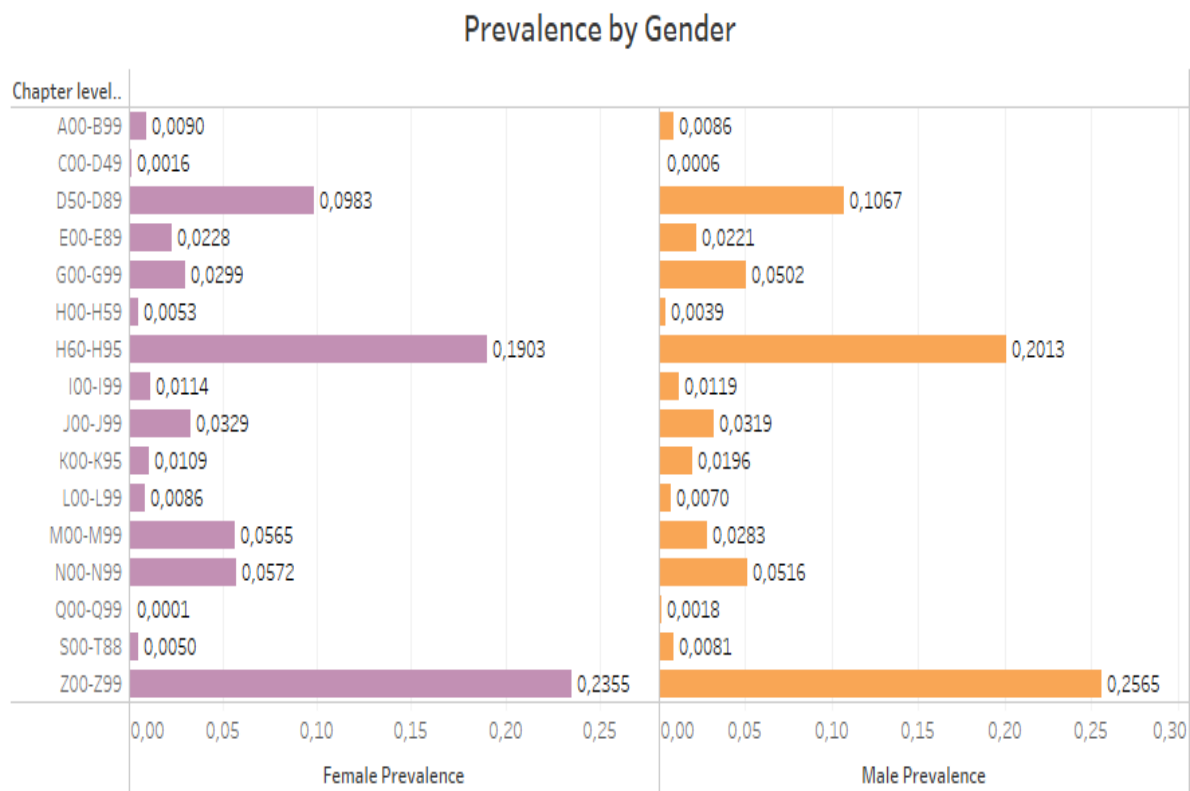


Figure 8 Επιπολασμός ανά φύλο.

Οι παθήσεις του νευρικού συστήματος (G00-G99) εμφανίζουν διαφορά μεταξύ των δύο φύλων καθώς η τιμή του επιπολασμού είναι κατά 0,02 μεγαλύτερη στους άντρες από ότι στις γυναίκες ενώ στην περίπτωση των

παθήσεων του μυοσκελετικού συστήματος και του συνδετικού ιστού (M00-M99) οι γυναίκες παρουσιάζουν μεγαλύτερη τιμή επιπολασμού κατά 0,03. Στις υπόλοιπες κατηγορίες οι διαφορές του επιπολασμού υπήρχαν αλλά ήταν τάξης 0,001.

Οι γυναίκες παρουσίασαν ελάχιστα μεγαλύτερες τιμές από τους άντρες στις κατηγορίες που αφορούν λοιμώδη και παρασιτικά νοσήματα (A00-B99), Νεοπλάσματα (C00-D48), παθήσεις που σχετίζονται με το ενδοκρινικό και τον μεταβολισμό (E00-E90), παθήσεις του οφθαλμού (H00-H59), του αναπνευστικού (J00-J99), του δέρματος (L00-L99) και του ουροποιογεννητικού συστήματος (N00-N99). Ωστόσο, όπως είναι προφανές, αυτές οι διαφορές στις τιμές του επιπολασμού μεταξύ των φύλων είναι στατιστικά μη σημαντικές, κάτι που επιβεβαιώνει και το p-value του t-test που πραγματοποιήθηκε και τα αποτελέσματά του φαίνονται στον πίνακα 7.

Έλεγχος t του μέσου δύο δειγμάτων συσχετισμένων ζευγών	Female prevalence	Male prevalence
Μέσος	0,043176246	0,047851057
Διακύμανση	0,00507934	0,005992899
Μέγεθος δείγματος	15	15
Συσχέτιση Pearson	0,976617075	
Υποτιθέμενη διαφορά μέσων	0	
βαθμοί ελευθερίας	14	
t	-1,052765455	
P(T<=t) μονόπλευρη	0,155140232	
t κρίσιμο, μονόπλευρο	1,761310136	
P(T<=t) δίπλευρη	0,310280464	
t κρίσιμο, δίπλευρο	2,144786688	

Πίνακας 7 t-test του επιπολασμού ανά βιολογικό φύλο.

3.6 Επιπολασμός και κληρονομησιμότητα

Η συσχέτιση της κληρονομησιμότητας με τον επιπολασμό αποτέλεσε βασικό ερώτημα για την μελέτη. Για να μελετηθεί η σχέση των 2 φαινομένων, κατασκευάστηκαν 2 scatter plots, όπου το ένα έχει ως άξονα x την κληρονομησιμότητα από γενετικά δεδομένα και άξονα y τον επιπολασμό, ενώ το άλλο έχει ως άξονα x την κληρονομησιμότητα από μη γενετικά δεδομένα και άξονα y τον επιπολασμό. Και στα 2 διαγράμματα επισημάνθηκαν οι κατηγορίες των ασθενειών με τα ίδια χρώματα.

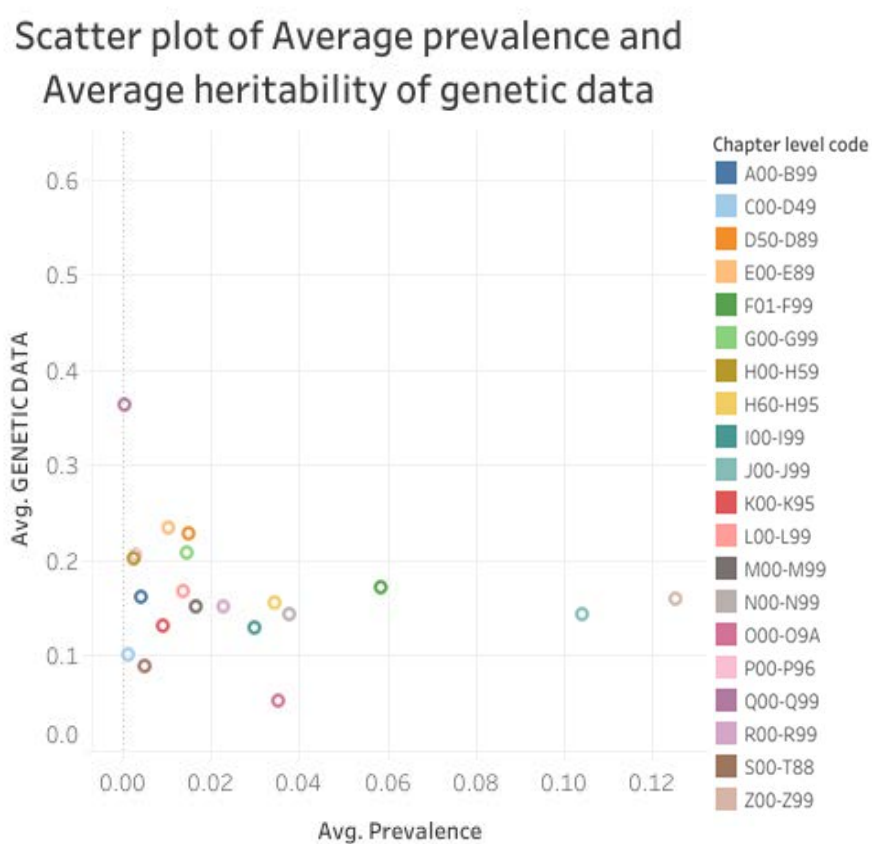
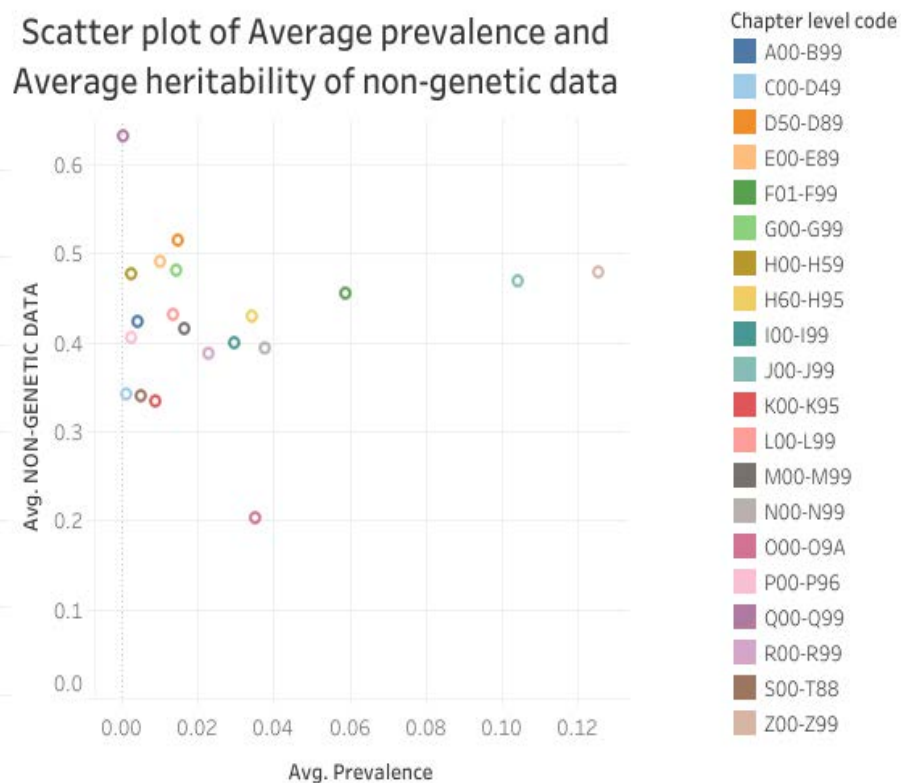


Figure 9 Scatterplots του επιπολασμού με την κληρονομησιμότητα από μη γενετικά δεδομένα (πάνω σχήμα) και με την κληρονομησιμότητα από γενετικά δεδομένα (κάτω σχήμα).

Από το διάγραμμα 4 φαίνεται πώς οι ασθένειες των περισσότερων κατηγοριών έχουν επιπολασμό που κυμένεται μεταξύ 0,0-0,02. Εξαίρεση αποτελούν ασθένειες των κατηγοριών F01-F99, J00-J99 και Z00-Z99 που έχουν τιμές 0,06 , 0,103 και 0,123 αντίστοιχα. Για να γίνει έλεγχος αν υπάρχει στατιστικά σημαντική διαφορά του επιπολασμού με την κληρονομησιμότητα πραγματοποιήθηκαν t-tests.

Έλεγχος t του μέσου δύο δειγμάτων συσχετισμένων ζευγών		
	<i>Heritability from Genetic data</i>	<i>Prevalence</i>
Μέσος	0,173889468	0,028878128
Διακύμανση	0,003482494	0,00121097
Μέγεθος δείγματος	20	20
Συσχέτιση Pearson	-0,314930234	
Υποτιθέμενη διαφορά μέσων	0	
βαθμοί ελευθερίας	19	
t	8,381365136	
P(T<=t) μονόπλευρη	4,16745E-08	
t κρίσιμο, μονόπλευρο	1,729132812	
P(T<=t) δίπλευρη	8,33491E-08	
t κρίσιμο, δίπλευρο	2,093024054	
	<i>Heritability from Non-Genetic data</i>	<i>Prevalence</i>
Μέσος	0,430917682	0,026088814
Διακύμανση	0,005887239	0,001133433
Μέγεθος δείγματος	20	20
Συσχέτιση Pearson	0,018038366	
Υποτιθέμενη διαφορά μέσων	0	
βαθμοί ελευθερίας	19	
t	21,75198314	
P(T<=t) μονόπλευρη	3,44181E-15	
t κρίσιμο, μονόπλευρο	1,729132812	
P(T<=t) δίπλευρη	6,88363E-15	
t κρίσιμο, δίπλευρο	2,093024054	

Πίνακας 8 t-tests για τον έλεγχο ύπαρξης στατιστικά σημαντικής διαφοράς μεταξύ γενετικών και μη γενετικών δεδομένων με τον επιπολασμό.

Όπως διαφαίνεται στον πίνακα 5, το p-value στο t-test με μεταβλητές ενδιαφέροντος τον επιπολασμό και την κληρονομησιμότητα από γενετικά δεδομένα έχει τιμή $< 0,05$ και συνεπώς υποδηλώνεται ότι υπάρχει στατιστικά σημαντική διαφορά μεταξύ των μεταβλητών ενδιαφέροντος. Το ίδιο συμβαίνει και στην περίπτωση της κληρονομησιμότητας από μη γενετικά δεδομένα.

Αξίζει να παρατηρήσουμε, ότι η συσχέτιση Pearson παρουσιάζει διαφορά μεταξύ των 2 t-test, καθώς στην περίπτωση της κληρονομησιμότητας από γενετικά δεδομένα υπάρχει μεγαλύτερη γραμμική συσχέτιση με τον επιπολασμό (0,314) από ότι η κληρονομικότητα από μη γενετικά δεδομένα (0,018).

Σε ότι αφορά την σύγκριση μεταξύ των τιμών της κληρονομησιμότητας και του επιπολασμού, αποδείχθηκε για μια ακόμη φορά ότι τα δύο φαινόμενα είναι ανεξάρτητα. Στον παρακάτω πίνακα, γίνονται εμφανείς οι διαφορές μεταξύ των τιμών των 2 δεδομένων ανά κατηγορία. Η ένταση του χρώματος είναι ανάλογη της τιμής, καθώς όσο πιο μεγάλη είναι η τιμή τόσο πιο έντονο είναι το χρώμα.

Chapter level c..	Prevalence	Heritability	Chapter level c..	Prevalence	Heritability
A00-B99	0,0042	0,3332	K00-K95	0,0091	0,2569
C00-D49	0,0011	0,2501	L00-L99	0,0137	0,3376
D50-D89	0,0150	0,4060	M00-M99	0,0172	0,3065
E00-E89	0,0101	0,3935	N00-N99	0,0378	0,3004
F01-F99	0,0587	0,3467	O00-O9A	0,0352	0,1744
G00-G99	0,0145	0,3618	P00-P96	0,0026	0,3329
H00-H59	0,0025	0,3844	Q00-Q99	0,0003	0,5154
H60-H95	0,0344	0,3405	R00-R99	0,0229	0,2903
I00-I99	0,0296	0,2830	S00-T88	0,0050	0,3647
J00-J99	0,1043	0,3392	Z00-Z99	0,1255	0,3567

Figure 10 Σύγκριση τιμών κληρονομησιμότητας και επιπολασμού ανά κατηγορία ασθενειών.

Εκτός από την συγκριτικά μεγάλη διαφορά μεταξύ των τιμών των 2 φαινομένων σε όλες τις κατηγορίες, παρατηρούμε ότι δεν διατηρήθηκε κάποια αναλογία, όπως για παράδειγμα η ίδια ένταση τιμών στις ίδιες κατηγορίες από κάθε φαινόμενο.

Στο figure 11 παρατίθενται για λόγους σύγκρισης τα διαγράμματα του επιπολασμού με την κληρονομικότητα από γενετικά δεδομένα και από μη γενετικά δεδομένα ανά ασθένεια. Κάθε ασθένεια είναι χρωματισμένη σύμφωνα

με το χρώμα της κατηγορίας ασθένειας που ανήκει σύμφωνα με την κλίμακα ICD10.

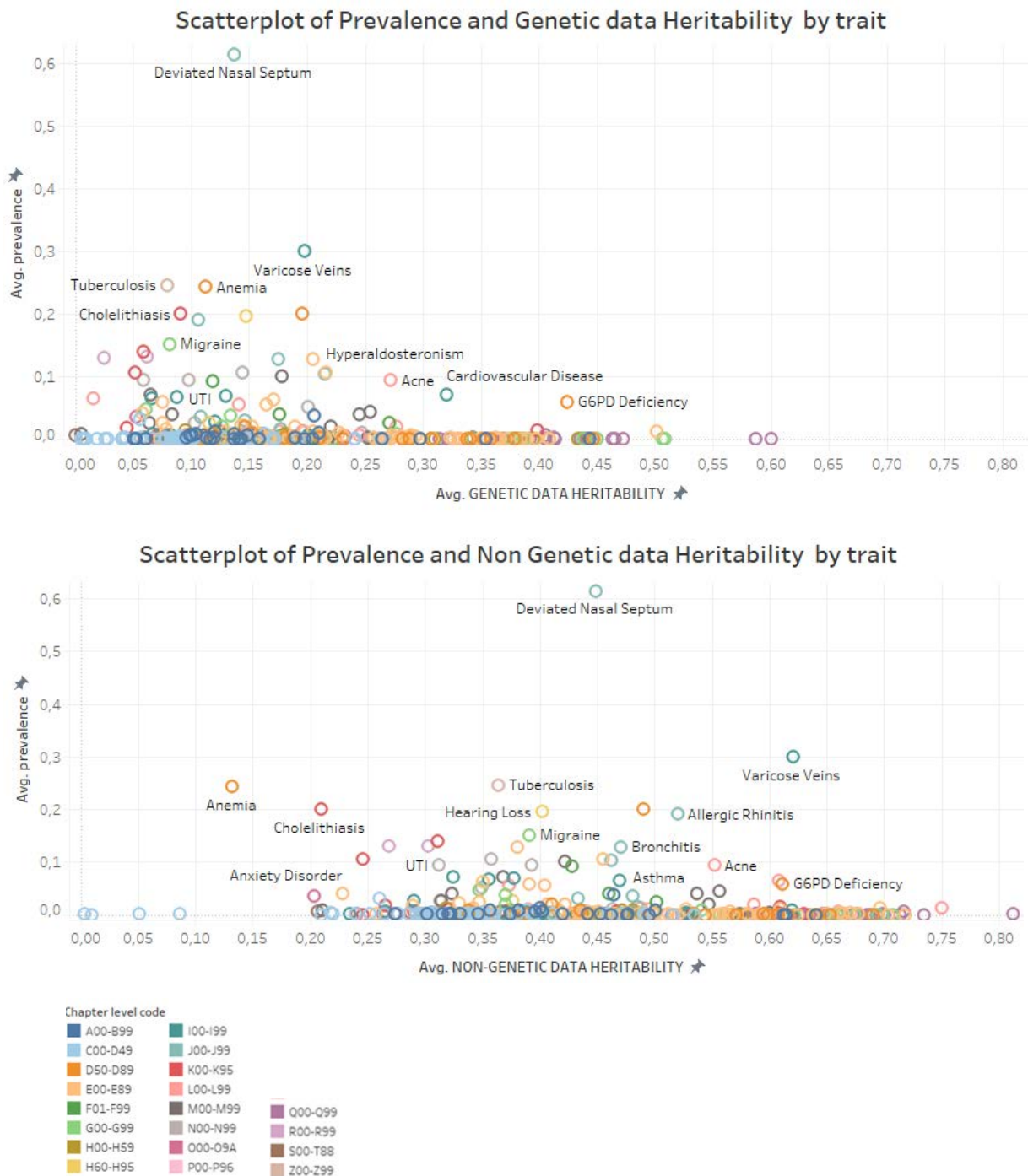


Figure 11 Διάγραμμα επιπολασμού με την κληρονομησιμότητα από γενετικά δεδομένα (πάνω) και από μη γενετικά δεδομένα (κάτω).

Από αυτά τα διαγράμματα είναι ευκολό να σημειωθούν κάποιες παρατηρήσεις σε σχέση με μεμονομένες περιπτώσεις ασθενειών και παθήσεων. Αρχικά παρατηρούμε πως από το σύνολο των χαρακτηριστικών η εκτροπή

ρινικού διαφράγματος (deviated nasal septum) σημειώσε την υψηλότερη τιμή επιπολασμού και ακολουθούν οι κυρσοί (vericose veins), η φυματίωση (tuberculosis) και η αναιμία (anemia).

Σε ότι αφορά την κληρονομησιμότητα τα αποτελέσματα στις περισσότερες περιπτώσεις, είναι διαφορετικά ανάλογα με τον τύπο δεδομένων. Για την κληρονομησιμότητα από μη γενετικά δεδομένα μεγαλύτερη τιμή σημείωσε η συγγενής ανωμαλία των ανδρικών γεννητικών οργάνων (male genital congenital anomaly) με τιμή 0,8119 και ακολουθούν λεύκη (vitiligo), hirschsprung's disease και φαινυλκετονουρία με τιμές 0,7500 , 0,7342 και 0,7182 αντίστοιχα . Για την κληρονομικότητα από γενετικά δεδομένα οι ασθένειες /παθήσεις με την μεγαλύτερη τιμή ήταν κατά σειρά: συγγενής ανωμαλία των ανδρικών γεννητικών οργάνων, neural tube defect, η δυστροφία των χόνδρων και η οικογενής δυσαντονομία με τιμές 0,6003 , 0,5874 , 0,5088 και 0.5064 αντίστοιχα .

Σε όλες τις ασθένειες, η διαφορά των τιμών κληρονομησιμότητας από γενετικά δεδομένα με την τιμή κληρονομησιμότητας από μη γενετικά δεδομένα ήταν εμφανής. Στο figure12 απεικονίζεται η διαφορά αυτή σε σχέση με τον επιπολασμό. Τα χαρακτηριστικά έχουν χρωματιστεί σύμφωνα με την κατηγορία ασθενειών / παθήσεων που ανήκουν. Η απόκλιση υπολογίστηκε σύμφωνα με τον τύπο:

$$\frac{NON_GENETIC\ DATA - GENETIC\ DATA}{NON_GENETIC\ DATA}$$

Όπου:

NON-GENETIC DATA: η κληρονομησιμότητα από μη γενετικά δεδομένα

GENETIC DATA: η κληρονομησιμότητα από γενετικά δεδομένα

Θετική τιμή αποτελέσματος σημαίνει πως η τιμή της κληρονομησιμότητας από μη γενετικά δεδομένα είναι μεγαλύτερη από ότι προτείνουν τα γενετικά δεδομένα, ενώ το ανάποδο συμβαίνει όταν το αποτέλεσμα έχει τιμή μεγαλύτερη από το 0. Αποτέλεσμα ίσο με το 0, σημαίνει πως η τιμή κληρονομησιμότητας είναι ίδια σε γενετικά και μη γενετικά δεδομένα.

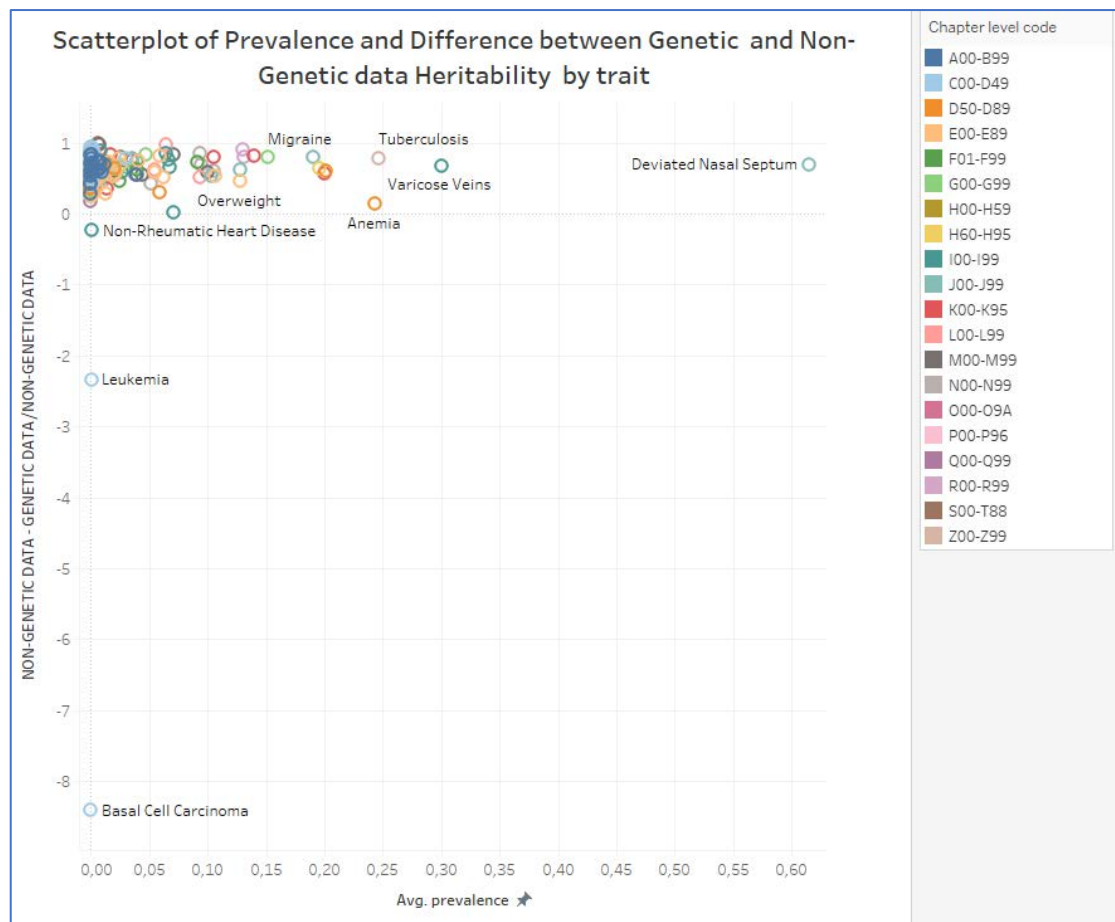


Figure 12 Scatterplot επιπολασμού και διαφοράς μεταξύ κληρονομικότητας μη γενετικών δεδομένων και κληρονομικότητας γενετικών δεδομένων ανά χαρακτηριστικό.

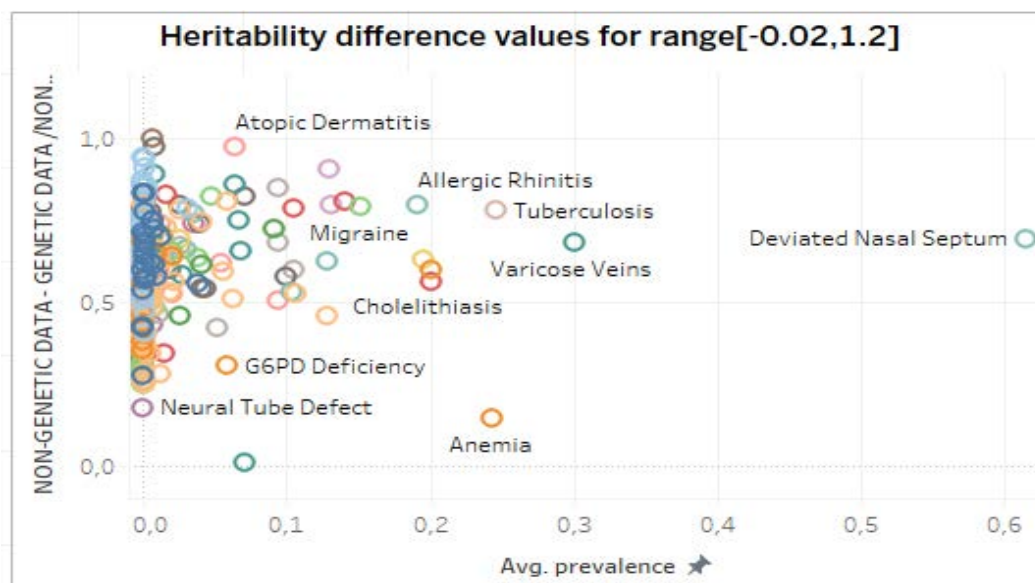


Figure 13 Scatterplot επιπολασμού και διαφοράς μεταξύ κληρονομικότητας μη γενετικών δεδομένων και κληρονομικότητας γενετικών δεδομένων για το εύρος τιμών διαφοράς -0,02 έως 1,2.

Παρατηρούμε ότι το βασικοκυτταρικό καρκίνωμα (basal cell carcinoma) και η λευχαιμία αποτέλεσαν οι περιπτώσεις ασθένειας που η τιμή κληρονομησιμότητας από γενετικά δεδομένα είναι μεγαλύτερη από ότι οι αντίστοιχες τιμές από μη γενετικά δεδομένα. Η πλειοψηφία ωστόσο των χαρακτηριστικών όπως φαίνεται και στο figure 13 σημείωσε μεγαλύτερη τιμή κληρονομησιμότητας προερχόμενη από μη γενετικά δεδομένα, με διαφορά που αποκτά θετικές τιμές και μάλιστα, με μεγαλύτερη τιμή την 1,02 που σημειώνεται σε τραυματισμούς στο κεφάλι.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Ως απόρροια των παραπάνω αποτελεσμάτων, έχει γίνει εμφανές το γεγονός ότι παρόλο που το φαινόμενο της κληρονομησιμότητας και το φαινόμενο του επιπολασμού είναι εξίσου αναγκαία και αναζητούνται μαζί στα επιδημιολογικά δεδομένα των ασθενειών, δεν έχουν κάποια άμεση αλληλεπίδραση μεταξύ τους.

Όσον αφορά στα χαρακτηριστικά ενδιαφέροντος της μελέτης, τόσο σε μεμονωμένες περιπτώσεις ασθενειών όσο και στις κατηγορίες αυτών σημειώθηκαν διαφορές στην τιμή κληρονομησιμότητας μεταξύ γενετικών και μη γενετικών δεδομένων, επιβεβαιώνοντας τον σημαντικό ρόλο του τύπου δεδομένων και τον αλγόριθμο υπολογισμού που διαχειριζόμαστε για να αντλήσουμε κληρονομικά δεδομένα. Σε όλες σχεδόν τις περιπτώσεις, η κληρονομησιμότητα που προκύπτει από μη γενετικά δεδομένα, έτεινε να έχει τιμές πιο κοντά στο 1 από ότι η κληρονομησιμότητα από γενετικά δεδομένα, κάτι που υποδηλώνει ότι οι ασθένειες της μελέτης δέχονται μεγάλη επίδραση από περιβαλλοντικούς παράγοντες.

Σε ότι αφορά τα χαρακτηριστικά της καταγωγής και πάλι σε μεμονωμένες ασθένειες και παθήσεις διαφαίνεται ότι υπάρχει συσχέτιση, σε γενικότερο όμως επίπεδο οι αποστάσεις των τιμών κληρονομησιμότητας δεν είναι σημαντικές.

Στην περίπτωση της εξέτασης ύπαρξης συσχέτισης μεταξύ του φύλου και της κληρονομησιμότητας οι έλεγχοι υπέδειξαν ότι υπάρχει στατιστικά σημαντικό διαφορά μεταξύ των 2 φύλων τόσο στην κληρονομησιμότητα από

γενετικά όσο και από μη γενετικά δεδομένα, κάτι που αποτελεί ένδειξη συσχέτισης του φύλου με την κληρονομησιμότητα. Ωστόσο αυτό για να επιβεβαιωθεί απαιτεί περαιτέρω ανάλυση και έλεγχος επιρροής από επιπλέον περιβαλλοντικούς παράγοντες.

Σημαντική παρατήρηση αποτελεί το γεγονός ότι στο σύνολο των χαρακτηριστικών η μέση τιμή κληρονομησιμότητας, φάνηκε να μην ξεπερνά την τιμή 0,5 και άρα να διαφαίνεται πώς η επιρροή από μη γενετικούς παράγοντες όπως το περιβάλλον για παράδειγμα είναι μεγαλύτερη. Αυτό το εύρημα αποτελεί σπουδαίο εφαλτήριο για έναν νέο κύκλο διερεύνησης της συσχέτισης κληρονομησιμότητας με παράγοντες όπως την έκθεση στον καπνό και την διατροφή και το κατά πόσο αυτοί οι παράγοντες την επηρεάζουν.

Τα δεδομένα που συγκεντρώθηκαν για τον επιπολασμό, δυστυχώς αφορούσαν λιγότερο από το 50% των δεδομένων χαρακτηριστικών της μελέτης. Η ανεύρεση δεδομένων επιπολασμού αποτέλεσε μια ιδιαίτερα επίπονη διαδικασία καθώς οι μελέτες που αφορούσαν στα ποσοτικά στατιστικά δεδομένα επιπολασμού και μάλιστα παγκόσμιου επιπέδου ήταν πραγματικά ολίγιστες. Αυτό, αυτόματα θα μπορούσε να προκαλέσει εύλογη αμφισβήτηση των αποτελεσμάτων που σχετίζονται με τον επιπολασμό, κυρίως σε ότι αφορά στη σύγκρισή του με την κληρονομησιμότητα.

Κατέστη λοιπόν σαφές πως η ανάγκη για ολοκλήρωση περισσότερων μελετών που αφορούν στον επιπολασμό και ιδιαίτερα τον παγκόσμιο είναι μεγάλη καθώς αποτελεί συνθήκη για την εξέλιξη ερευνών, αφού όσο αυξάνονται τα δεδομένα και οι πληροφορίες για τον επιπολασμό αυξάνονται και οι πιθανότητες να προκύψει μια νέα πτυχή που θα συσχετίζει άμεσα τον επιπολασμό με την κληρονομησιμότητα.

BIBΛΙΟΓΡΑΦΙΑ

1. Inherited [Internet]. 2024. Available from: <https://www.genome.gov/genetics-glossary/Inherited>.
2. Connell SM. Aristotle on Animal Generation and Hereditary Resemblance. *The Cambridge Companion to Aristotle's Biology*. 2021:142.
3. Steup M, Ram Neta E. *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), Edward N. Zalta.
4. Bell AE. Heritability in retrospect. *Journal of Heredity*. 1977;68(5):297-300.
5. Falconer D. *Introduction to quantitative genetics*: Longman; 1996.
6. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era—concepts and misconceptions. *Nature reviews genetics*. 2008;9(4):255-66.
7. Cabej NR. On the origin and nature of nongenetic information in eumetazoans. *Annals of the New York Academy of Sciences*. 2023;1525(1):104-27.
8. Heckerman D, Gurdasani D, Kadie C, Pomilla C, Carstensen T, Martin H, et al. Linear mixed model for heritability estimation that explicitly addresses environmental variation. *Proceedings of the National Academy of Sciences*. 2016;113(27):7377-82.
9. Hai Y, Wen Y. A Bayesian linear mixed model for prediction of complex traits. *Bioinformatics*. 2020;36(22-23):5415-23.
10. Frouin A, Dandine-Roulland C, Pierre-Jean M, Deleuze J-F, Ambroise C, Le Floch E. Exploring the link between additive heritability and prediction accuracy from a ridge regression perspective. *Frontiers in Genetics*. 2020;11:581594.
11. Barry C-JS, Walker VM, Cheesman R, Davey Smith G, Morris TT, Davies NM. How to estimate heritability: A guide for genetic epidemiologists. *International Journal of Epidemiology*. 2023;52(2):624-32.
12. Moore DS, Shenk D. The heritability fallacy. *Wiley Interdisciplinary Reviews: Cognitive Science*. 2017;8(1-2):e1400.
13. Tenny S, Hoffman MR. *Prevalence*. 2017.
14. Doi N, Hoshi Y, Itokawa M, Usui C, Yoshikawa T, Tachikawa H. Persistence criteria for susceptibility genes for schizophrenia: a discussion from an evolutionary viewpoint. *PLoS one*. 2009;4(11):e7799.
15. Watanabe K, Stringer S, Frei O, Umićević Mirkov M, de Leeuw C, Polderman TJC, et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet*. 2019;51(9):1339-48.
16. Ge T, Chen CY, Neale BM, Sabuncu MR, Smoller JW. Phenome-wide heritability analysis of the UK Biobank. *PLoS Genet*. 2017;13(4):e1006711.
17. Polubriaginof F, Vanguri R, Quinnes K, Belbin GM, Yahi A, Salmasian H, et al. Estimate of disease heritability using 7.4 million familial relationships inferred from electronic health records. *bioRxiv*. 2017:066068.
18. Jia G, Li Y, Zhang H, Chattopadhyay I, Boeck Jensen A, Blair DR, et al. Estimating heritability and genetic correlations from large health datasets in the absence of genetic data. *Nat Commun*. 2019;10(1):5508.
19. Patron J, Serra-Cayuela A, Han B, Li C, Wishart DS. Assessing the performance of genome-wide association studies for predicting disease risk. *PLoS One*. 2019;14(12):e0220215.

20. Polubriaginof FCG, Vanguri R, Quinnes K, Belbin GM, Yahi A, Salmasian H, et al. Disease Heritability Inferred from Familial Relationships Reported in Medical Records. *Cell*. 2018;173(7):1692-704.e11.
21. O'Connor LJ, Gusev A, Liu X, Loh P-R, Finucane HK, Price AL. Estimating the proportion of disease heritability mediated by gene expression levels. *bioRxiv*. 2017:118018.
22. Weissbrod O, Flint J, Rosset S. Estimating SNP-Based Heritability and Genetic Correlation in Case-Control Studies Directly and with Summary Statistics. *Am J Hum Genet*. 2018;103(1):89-99.
23. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747-53.
24. López-Cortegano E, Caballero A. Inferring the Nature of Missing Heritability in Human Traits Using Data from the GWAS Catalog. *Genetics*. 2019;212(3):891-904.
25. Ormondroyd E, Border P, Hayward J, Papanikitas A. Genomic health data generation in the UK: a 360 view. *European Journal of Human Genetics*. 2022;30(7):782-9.
26. Steinsaltz D, Dahl A, Wachter KW. On negative heritability and negative estimates of heritability. *Genetics*. 2020;215(2):343-57.
27. Gaspar HA, Breen G. Probabilistic ancestry maps: a method to assess and visualize population substructures in genetics. *BMC bioinformatics*. 2019;20:1-11.
28. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*. 2011;88(1):76-82.
29. Srivastava AK, Williams SM, Zhang G. Heritability estimation approaches utilizing genome-wide data. *Current protocols*. 2023;3(4):e734.
30. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Consortium SWGotPG, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics*. 2015;47(3):291-5.
31. Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nature protocols*. 2020;15(9):2759-72.
32. Benckek PH, Morris NJ. How meaningful are heritability estimates of liability? *Human genetics*. 2013;132:1351-60.
33. Kochunov P, Patel B, Ganjgahi H, Donohue B, Ryan M, Hong EL, et al. Homogenizing estimates of heritability among SOLAR-Eclipse, OpenMx, APACE, and FPHI software packages in neuroimaging data. *Frontiers in Neuroinformatics*. 2019;13:16.
34. Academy C. Prevalence Calculator 2023.
35. Pripp AH. Pearson's or Spearman's correlation coefficients. *Tidsskrift for den Norske lægeforening: tidsskrift for praktisk medicin, ny række*. 2018;138(8).
36. Kwak S. Are Only p-values less than 0.05 significant? A p-value greater than 0.05 is also significant! *Journal of Lipid and Atherosclerosis*. 2023;12(2):89.
37. Plomin R, DeFries JC, Knopik VS, Neiderhiser JM. Top 10 replicated findings from behavioral genetics. *Perspectives on psychological science*. 2016;11(1):3-23.
38. Berger SL, Kouzarides T, Shiekhata R, Shilatifard A. An operational definition of epigenetics. *Genes & development*. 2009;23(7):781-3.
39. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747-53.

