



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΥΓΕΙΑΣ
ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ**

**ΤΟΜΕΑΣ ΚΟΙΝΩΝΙΚΗΣ ΙΑΤΡΙΚΗΣ ΚΑΙ ΨΥΧΙΚΗΣ ΥΓΕΙΑΣ
ΕΡΓΑΣΤΗΡΙΟ ΥΓΙΕΙΝΗΣ ΚΑΙ ΕΠΙΔΗΜΙΟΛΟΓΙΑΣ**

**ΓΕΝΕΤΙΚΕΣ ΣΥΣΧΕΤΙΣΕΙΣ ΓΙΑ ΓΕΝΕΤΙΚΟΥΣ ΠΟΛΥΜΟΡΦΙΣΜΟΥΣ
ΧΑΜΗΛΗΣ ΣΥΧΝΟΤΗΤΑΣ ΣΕ ΜΕΛΕΤΕΣ ΓΟΝΙΔΙΑΚΗΣ ΣΑΡΩΣΗΣ**

ΟΡΕΣΤΗΣ ΠΑΝΑΓΙΩΤΟΥ

ΙΑΤΡΟΣ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΙΩΑΝΝΙΝΑ 2014



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΥΓΕΙΑΣ
ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ**

**ΤΟΜΕΑΣ ΚΟΙΝΩΝΙΚΗΣ ΙΑΤΡΙΚΗΣ ΚΑΙ ΨΥΧΙΚΗΣ ΥΓΕΙΑΣ
ΕΡΓΑΣΤΗΡΙΟ ΥΓΙΕΙΝΗΣ ΚΑΙ ΕΠΙΔΗΜΙΟΛΟΓΙΑΣ**

**ΓΕΝΕΤΙΚΕΣ ΣΥΣΧΕΤΙΣΕΙΣ ΓΙΑ ΓΕΝΕΤΙΚΟΥΣ ΠΟΛΥΜΟΡΦΙΣΜΟΥΣ
ΧΑΜΗΛΗΣ ΣΥΧΝΟΤΗΤΑΣ ΣΕ ΜΕΛΕΤΕΣ ΓΟΝΙΔΙΑΚΗΣ ΣΑΡΩΣΗΣ**

ΟΡΕΣΤΗΣ ΠΑΝΑΓΙΩΤΟΥ

ΙΑΤΡΟΣ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΙΩΑΝΝΙΝΑ 2014

Η έγκριση της διδακτορικής διατριβής από το Τμήμα Ιατρικής του Πανεπιστημίου
Ιωαννίνων δεν υποδηλώνει αποδοχή των γνώμων του συγγραφέα Ν. 5343/32, άρθρο
202, παράγραφος 2 (νομική κατοχύρωση του Ιατρικού Τμήματος).

Ημερομηνία αίτησης του κ. Παναγιώτου Ορέστη: 28-1-2010

Ημερομηνία ορισμού Τριμελούς Συμβουλευτικής Επιτροπής: 677^α/9-2-2010

Μέλη Τριμελούς Συμβουλευτικής Επιτροπής:

Επιβλέπων

Ιωαννίδης Ιωάννης Καθηγητής Υγιεινής

Μέλη

Δημολιάτης Ιωάννης Επίκουρος Καθηγητής Υγιεινής

Ντζάνη Ευαγγελία Επίκουρη Καθηγήτρια Υγιεινής με ιδιαίτερη έμφαση στην Επιδημιολογία

Ανασύσταση Τριμελούς Συμβουλευτικής Επιτροπής: Αριθμ Συνεδρ. 752^α/16-5-2013

Επιβλέπων

Ντζάνη Ευαγγελία, Επίκουρη Καθηγήτρια Υγιεινής με έμφαση στην Επιδημιολογία

Μέλη

Ιωαννίδης Ιωάννης Καθηγητής Υγιεινής

Δημολιάτης Ιωάννης Επίκουρος Καθηγητής Υγιεινής

Ημερομηνία ορισμού θέματος: 1-3-2010

«Γενετικές συσχετίσεις για γενετικούς πολυμορφισμούς χαμηλής συχνότητας σε μελέτες γονιδιακής σάρωσης».

ΔΙΟΡΙΣΜΟΣ ΕΠΤΑΜΕΛΟΥΣ ΕΞΕΤΑΣΤΙΚΗΣ ΕΠΙΤΡΟΠΗΣ : 763^α/10-4-2014

1. Ιωαννίδης Ιωάννης, C.F. Rehnborg Professor in Disease Prevention in the School of Medicine Stanford University, USA
2. Μαυρέας Βενετσάνος Καθηγητής Ψυχιατρικής του Τμήματος Ιατρικής του Παν/μίου Ιωαννίνων
3. Παυλίδης Νικόλαος Καθηγητής Παθολογίας-Ογκολογίας του Τμήματος Ιατρικής του Παν/μίου Ιωαννίνων
4. Δημολιάτης Ιωάννης Επίκουρος Καθηγητής Υγιεινής του Τμήματος Ιατρικής του Παν/μίου Ιωαννίνων
5. Ντζάνη Ευαγγελία Επίκουρη Καθηγήτρια Υγιεινής με ιδιαίτερη έμφαση στην Επιδημιολογία του Τμήματος Ιατρικής του Παν/μίου Ιωαννίνων
6. Ευαγγέλου Ευάγγελος Λέκτορας Υγιεινής με έμφαση στην Κλινική και Μοριακή Επιδημιολογία του Τμήματος Ιατρικής του Παν/μίου Ιωαννίνων
7. Τσιλίδης Κωνσταντίνος Λέκτορας Υγιεινής με έμφαση στην Επιδημιολογία του Τμήματος Ιατρικής του Παν/μίου Ιωαννίνων

Έγκριση Διδακτορικής Διατριβής με βαθμό «ΑΡΙΣΤΑ» στις 10-6-2014

ΠΡΟΕΔΡΟΣ ΤΟΥ ΤΜΗΜΑΤΟΣ ΙΑΤΡΙΚΗΣ

Φωτόπουλος Ανδρέας

Καθηγητής Πυρηνικής Ιατρικής

Η Γραμματέας του Τμήματος



ΚΑΠΗΤΟΠΟΥΛΟΥ ΜΑΡΙΑ

ΕΥΧΑΡΙΣΤΙΕΣ

Η διατριβή αυτή ξεκίνησε το 2010 και εκπονήθηκε υπό την επίβλεψη και το mentoring του Καθηγητή κ. Ιωαννίδη, ο οποίος από νωρίς με μύησε στην έρευνα και μου δίδαξε τη συνεχή αναζήτηση της επιστημονικής αλήθειας. Τον ευχαριστώ που μου έδωσε τη δυνατότητα να συνεργαστώ μαζί του, καθώς και για τη διαρκή καθοδήγηση και υποστήριξή του από όταν έκανα τα πρώτα δειλά ερευνητικά μου βήματα, όντας ακόμα τριτοετής φοιτητής της Ιατρικής, έως και σήμερα.

Θα ήθελα, επίσης, να ευχαριστήσω θερμά τους συνεπιβλέποντες κ.κ. Ντζάνη και Δημολιάτη για την υποστήριξη και την καθοδήγησή τους όλα αυτά τα χρόνια.

Από τα μέλη ΔΕΠ του Εργαστηρίου Επιδημιολογίας, θα ήθελα να ξεχωρίσω τους κ.κ. Ευαγγέλου, Τσιλίδη και Τζουλάκη. Ιδιαίτερες ευχαριστίες οφείλω στον κ. Ευαγγέλου για την αμέριστη βοήθειά του και την καθοριστικής σημασίας συμβολή στην εκπόνηση αυτής της διατριβής. Επίσης θα ήθελα να ευχαριστήσω θερμά τους κ.κ. Τσιλίδη και Τζουλάκη για τη συνεισφορά τους και την υποστήριξή τους όλα αυτά τα χρόνια, καθώς και την κ. Κουτσοτόλη για την παροχή κάθε είδους διοικητικής-γραμματειακής βοήθειας οποτεδήποτε αυτή χρειάστηκε. Τα προαναφερθέντα μέλη του Εργαστηρίου δημιούργησαν για μένα ένα ερευνητικά εποικοδομητικό περιβάλλον όλα αυτά τα χρόνια.

Τέλος, ιδιαίτερες ευχαριστίες οφείλω στην οικογένειά μου για τη στήριξή της όλα αυτά τα χρόνια.

Ο.Α.Π.

North Bethesda, MD / Ιωάννινα, Ελλάδα

Φεβρουάριος 2014

Αντί προλόγου ...

*... παρατίθενται τα τέσσερα πρώτα άρθρα από την
«Καθολική Διακήρυξη για το Ανθρώπινο Γονιδίωμα και τα Ανθρώπινα Δικαιώματα»
της UNESCO.*

Ο.Α.Π.

North Bethesda, MD / Ιωάννινα, Ελλάδα

Φεβρουάριος 2014

UNESCO
Paris, 11 November 1997

Universal Declaration on the Human Genome and Human Rights

A. Human dignity and the human genome

Article 1

The human genome underlies the fundamental unity of all members of the human family, as well as the recognition of their inherent dignity and diversity. In a symbolic sense, it is the heritage of humanity.

Article 2

- (a) Everyone has a right to respect for their dignity and for their rights regardless of their genetic characteristics.
- (b) That dignity makes it imperative not to reduce individuals to their genetic characteristics and to respect their uniqueness and diversity.

Article 3

The human genome, which by its nature evolves, is subject to mutations. It contains potentialities that are expressed differently according to each individual's natural and social environment, including the individual's state of health, living conditions, nutrition and education.

Article 4

The human genome in its natural state shall not give rise to financial gains.

[...]

ΠΕΡΙΕΧΟΜΕΝΑ

ΑΝΤΙ ΠΡΟΛΟΓΟΥ	1
ΠΕΡΙΕΧΟΜΕΝΑ	5
ΕΥΡΕΤΗΡΙΟ ΕΙΚΟΝΩΝ	7
ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ	9
ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ	11
ΚΕΦΑΛΑΙΟ 1: ΓΕΝΙΚΟ ΜΕΡΟΣ – ΕΙΣΑΓΩΓΗ	13
1.1. ΓΕΝΙΚΟΤΗΤΕΣ.....	13
1.1.1. Το Ανθρώπινο Γονιδίωμα – Μία Σύντομη Εισαγωγή.....	13
1.1.2. Η Ποικιλομορφία του Ανθρώπινου Γονιδιώματος.....	15
1.1.3. Η Γενετική Επιδημιολογία των Πολυσύνθετων Νοσημάτων	18
1.1.4. Οι Μελέτες Γενετικής Συσχέτισης	21
1.1.5. Οι Πολυμορφισμοί Χαμηλής Συχνότητας.....	36
1.2. ΣΚΟΠΟΣ ΤΗΣ ΔΙΑΤΡΙΒΗΣ	39
ΚΕΦΑΛΑΙΟ 2: ΕΙΔΙΚΟ ΜΕΡΟΣ	41
2.1. ΕΜΠΕΙΡΙΚΗ ΑΠΟΤΙΜΗΣΗ ΤΟΥ ΔΥΝΑΜΙΚΟΥ ΤΩΝ ΜΕΤΑ-ΑΝΑΛΥΣΕΩΝ ΜΕΛΕΤΩΝ ΓΟΝΙΔΙΑΚΗΣ ΣΑΡΩΣΗΣ.....	41
2.1.1. Σκεπτικό και Σκοπός.....	41
2.1.2. Μέθοδοι.....	45
2.1.3. Αποτελέσματα	48
2.1.4. Συζήτηση.....	93
2.2. ΑΝΑΣΚΟΠΗΣΗ ΤΩΝ ΠΟΛΥΜΟΡΦΙΣΜΩΝ ΧΑΜΗΛΗΣ ΣΥΧΝΟΤΗΤΑΣ ΣΕ ΜΕΛΕΤΕΣ ΓΟΝΙΔΙΑΚΗΣ ΣΑΡΩΣΗΣ.....	97
2.2.1. Σκεπτικό και Σκοπός.....	97
2.2.2. Μέθοδοι.....	100
2.2.3. Αποτελέσματα	104
2.2.4. Συζήτηση.....	131

2.3. Το ΠΑΡΑΔΕΙΓΜΑ ΤΟΥ ΜΕΛΑΝΩΜΑΤΟΣ	137
2.3.1. Σκεπτικό και Σκοπός	137
2.3.2. Μέθοδοι.....	139
2.3.3. Αποτελέσματα.....	143
2.3.4. Συζήτηση	157
ΠΑΡΑΡΤΗΜΑ 1.....	161
ΠΑΡΑΡΤΗΜΑ 2	169
ΣΥΜΠΕΡΑΣΜΑΤΑ ΤΗΣ ΔΙΑΤΡΙΒΗΣ	171
ΠΕΡΙΛΗΨΗ.....	173
ABSTRACT	175
ΒΙΒΛΙΟΓΡΑΦΙΑ	177

ΕΥΡΕΤΗΡΙΟ ΕΙΚΟΝΩΝ

Εικόνα 1. Μεταβολή του κόστους γονοτύπησης.	25
Εικόνα 2. Συσχέτιση των μεγεθών αποτελέσματος στον Ελληνικό πληθυσμό και στις αρχικές μελέτες.....	148
Εικόνα 3. Συσχέτιση ανάμεσα στις συχνότητες των αλληλίων κινδύνου στον Ελληνικό πληθυσμό και τις συχνότητες των αντίστοιχων αλληλίων στην Ευρωπαϊκή ομάδα του HarMap.....	154
Εικόνα 4. Καμπύλες ROC όπου φαίνεται η διακριτική ικανότητα των 3 μοντέλων που μελετήθηκαν.	156

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

Πίνακας 1. Περιγραφή των 139 μετα-αναλύσεων μελετών γονιδιακής σάρωσης με μέγεθος δείγματος ≥ 10.000	49
Πίνακας 2. Πλατφόρμες γονοτύπησης, αριθμός μελετών, μέθοδοι μετα-ανάλυσης και μέθοδοι αναπλήρωσης ελλειπόντων γονοτύπων	57
Πίνακας 3. Κριτήρια ελέγχου ποιότητας για τους πολυμορφισμούς στα 88 άρθρα	64
Πίνακας 4. Όρια στατιστικής σημαντικότητας, μέθοδοι διερεύνησης της ετερογένειας και διαδικασίες επαλήθευσης	70
Πίνακας 5. Φυλετικές καταγωγές συμμετεχόντων, ποικίλοι φαινότυποι και διαφανοτυπικοί έλεγχοι	79
Πίνακας 6. Λειτουργικές αναλύσεις που συνόδευσαν τις μετα-αναλύσεις	88
Πίνακας 7. Γενετική διακύμανση που εξηγείται από τον συνολικό αριθμό των ανακαλυφθέντων γενετικών τόπων στις μετα-αναλύσεις μελετών γονιδιακής σάρωσης	91
Πίνακας 8. Γενετικές σχέσεις με συχνότητα ελάσσονος αλληλίου έως 5% και $P \leq 10^{-7}$	105
Πίνακας 9. Αριθμός των ανακαλυφθεισών (Α) και προσδοκώμενων (Π) γενετικών σχέσεων με πολυμορφισμούς χαμηλής συχνότητας με βάση τα τεταρτημόρια του λόγου αναλογιών και τις κατηγορίες της συχνότητας ελάσσονος αλληλίου.	109
Πίνακας 10. Πολυμορφισμοί στους ίδιους γενετικούς τόπους με τους πολυμορφισμούς χαμηλής συχνότητας, όπως περιγράφονται σε άλλες GWAS για τον ίδιο φαινότυπο	113
Πίνακας 11. Αριθμός μελετών γενετικής συσχέτισης υποψήφιων γονιδίων για τους γενετικούς τόπους που περιείχαν πολυμορφισμούς χαμηλής συχνότητας, καθώς και τα σχόλια των συγγραφέων των μελετών GWAS για την πρότερη τεκμηρίωσή τους.	120
Πίνακας 12. Μεταλλάξεις στους ίδιους γενετικούς τόπους με τους πολυμορφισμούς χαμηλής συχνότητας που προκαλούν τα ίδια ή παρόμοια φαινοτυπικά αποτελέσματα σύμφωνα με τη βάση OMIM	124
Πίνακας 13. Συχνότητες ελασσόνων αλληλίων των μη-κοινών πολυμορφισμών που αναλύθηκαν στη διατριβή στα 4 δείγματα πληθυσμών του HarMap	129
Πίνακας 14. Δημογραφικά χαρακτηριστικά και κατανομή των φαινοτύπων χρώματος στους ασθενείς με μελάνωμα και στους υγιείς μάρτυρες	144
Πίνακας 15. Πολυμορφισμοί που επιλέχθηκαν για επικύρωση στον Ελληνικό πληθυσμό..	146
Πίνακας 16. Αποτελέσματα μονοπαραγοντικής και πολυπαραγοντικής ανάλυσης για τη σχέση μεταξύ πολυμορφισμών και μελανώματος	151
Πίνακας 17. Γονοτυπηθέντα SNPs, αλληλία κινδύνου και συχνότητές τους στον Ελληνικό πληθυσμό και στην Ευρωπαϊκή ομάδα του HarMap (HarMap CEU)	153

ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ

ΔΕ: διάστημα εμπιστοσύνης

ΛΑ: λόγος αναλογιών

ΠΑΝ: περιφερική αρτηριακή νόσος

ΣΕΑ: συχνότητα ελάσσονος αλληλίου

AUC: area under the curve

γ-GT: γ-glutamyl transpeptidase

FEV₁: forced expiratory volume in 1 second

FVC: forced vital capacity

GWAS: genome-wide association study

HapMap: Haplotype Map

HDL: high-density lipoprotein

HOMA-B: homeostatic model assessment beta cell function

HOMA-IR: homeostatic model assessment insulin resistance

HuGE: Human Genome Epidemiology

IgE: immunoglobulin E

LDL: low-density lipoprotein

ROC: receiver operating characteristic

SNP: single-nucleotide polymorphism

ΚΕΦΑΛΑΙΟ 1: ΓΕΝΙΚΟ ΜΕΡΟΣ – ΕΙΣΑΓΩΓΗ

1.1. ΓΕΝΙΚΟΤΗΤΕΣ

1.1.1. Το ΑΝΘΡΩΠΙΝΟ ΓΟΝΙΔΙΩΜΑ – ΜΙΑ ΣΥΝΤΟΜΗ ΕΙΣΑΓΩΓΗ

Οι τελευταίες δεκαετίες χαρακτηρίζονται από μία αλματώδη ανάπτυξη των βιοϊατρικών επιστημών γενικά, αλλά και της Γενετικής ειδικότερα. Οι εξελίξεις στον κλάδο της Γενετικής, κυρίως τα τελευταία 15 χρόνια, θα μπορούσαν να χαρακτηριστούν ως παραπάνω από ραγδαίες. Η πρόοδος που έχει συντελεστεί από το 1865 – όταν ο Gregor Mendel διατύπωνε τους θεμελιώδεις νόμους της Γενετικής – μέχρι σήμερα, έχουν καταδείξει εμφανώς το ρόλο της γενετικής πληροφορίας τόσο στη φυσιολογία όσο και στην παθολογία του ανθρώπινου οργανισμού. Η πληροφορία αυτή κωδικοποιείται με τη μορφή των βάσεων του DNA (1, 2) το οποίο αποτελεί και το φορέα της κληρονομικότητας (3).

Το σύνολο της γενετικής πληροφορίας του ανθρώπου συνιστά το ανθρώπινο γονιδίωμα (*human genome*). Υπολογίζεται ότι το απλοειδές ανθρώπινο γονιδίωμα (δηλαδή αυτό που απαντάται στους γαμέτες) αποτελείται από περίπου 3,3 δισεκατομμύρια ζεύγη βάσεων (*base pairs, bp*) – τα οποία από βιοχημικής άποψης είναι νουκλεοτίδια (2)– και σύμφωνα με τις πιο πρόσφατες εκτιμήσεις περιέχει περίπου 20,000 γονίδια (*genes*) που κωδικοποιούν πρωτεΐνες (4, 5). Ο αριθμός αυτός είναι μικρότερος από τα περίπου 30,000 γονίδια που υπολογίζονταν κατά την ολοκλήρωση του Προγράμματος του Ανθρώπινου Γονιδιώματος (*Human Genome Project*) το 2003 (6, 7). Υπολογίζεται ότι λιγότερο από το 2% του ανθρώπινου γονιδιώματος «χρησιμεύει» για την κωδικοποίηση πρωτεϊνών.

Οι επιτυχίες που συντελέστηκαν κατά τη λεγόμενη «Γονιδιωματική Εποχή» (*genomic era*) (8) έχουν δημιουργήσει υψηλές προσδοκίες ότι η κατάλληλη χρήση της γενετικής πληροφορίας θα αλλάξει εκ βάθρων το τρίπτυχο «πρόληψη – διάγνωση – θεραπεία», που αποτελεί το θεμέλιο της Ιατρικής επιστήμης (9). Ήδη έχουν πραγματοποιηθεί κάποια πρώτα βήματα προς αυτή την κατεύθυνση, τελικός προορισμός της οποίας δεν είναι άλλος από την «εξατομικευμένη ιατρική» (*personalized medicine; individualized medicine*) (10-12).

1.1.2. Η ΠΟΙΚΙΛΟΜΟΡΦΙΑ ΤΟΥ ΑΝΘΡΩΠΙΝΟΥ ΓΟΝΙΔΙΩΜΑΤΟΣ

*«Variation Is The Spice Of Life»
Kruglyak & Nickerson (13)*

Το γονιδίωμα δύο ατόμων χωρίς μεταξύ τους συγγένεια είναι πανομοιότυπο κατά 99.9%, ενώ το υπόλοιπο 0.1% ποικίλλει με διάφορους τρόπους (14). Αυτό σημαίνει ότι στο ανθρώπινο γονιδίωμα υπάρχουν περίπου 3 εκατομμύρια πολυμορφικές θέσεις, ή αλλιώς το γονιδίωμα κάθε ατόμου περιέχει περισσότερες από 3 εκατομμύρια «διαφορές» από το γονιδίωμα αναφοράς (*reference genome*). Οι εναλλακτικές αλληλουχίες που βρίσκονται σε μία πολυμορφική θέση καλούνται αλλήλια (*alleles*) (15). Για μία δεδομένη θέση ένα άτομο μπορεί να είναι ομόζυγο ή ετερόζυγο, δηλαδή να έχει αντιστοίχως τα ίδια ή διαφορετικά αλλήλια στα δύο χρωμοσώματα. Έχει υπολογιστεί ότι περίπου 1 βάση σε κάθε 250-1000 βάσεις DNA είναι πολυμορφική.

Η ποικιλομορφία του ανθρώπινου γονιδιώματος κυμαίνεται από μεγάλες δομικές αλλαγές ολόκληρων χρωμοσωμικών περιοχών μέχρι πολυμορφισμούς λίγων ή και ενός μόνο νουκλεοτιδίου (13, 15-18). Πολύ αδρά, θα μπορούσε να γίνει ο διαχωρισμός σε ποικιλομορφία αλληλουχίας (*sequence variation*) που αφορά μικρής έκτασης αλληλουχία DNA, και σε δομική ποικιλομορφία (*structural variation*) που αφορά σε κυτταρογενετικά ορατές ή υπομικροσκοπικές (παρ)αλλαγές του DNA (18).

Οι πολυμορφισμοί του ενός νουκλεοτιδίου ή μονήρεις πολυμορφισμοί (*single-nucleotide polymorphisms, SNPs*) αποτελούν την πιο συχνά απαντώμενη μορφή ποικιλομορφίας στο ανθρώπινο DNA (19-21) (χαρακτηριστικά μέχρι τα τέλη του 2011 η βάση δεδομένων dbSNP [<http://www.ncbi.nlm.nih.gov/projects/SNP/>]

περιείχε συνολικά περίπου 42 εκατομμύρια επικυρωμένα διακριτά/μοναδικά SNPs). Οι πολυμορφισμοί συνίστανται στην αντικατάσταση μιας βάσης από κάποια άλλη (π.χ. μία αδενίνη A αντικαθίσταται από κυτοσίνη C, A>C) και έχουν συνήθως μικρή έως μέτρια διεισδυτικότητα (*penetrance*). Τα SNPs απαντώνται με συχνότητα $\geq 1\%$ στον πληθυσμό, ενώ για πολυμορφισμούς με μικρότερες συχνότητες στον πληθυσμό και μεγάλη διεισδυτικότητα χρησιμοποιείται ο όρος μετάλλαξη (*mutation*) (17). Ακριβώς λόγω της αφθονίας τους στο ανθρώπινο γονιδίωμα, τα SNPs αποτελούν το επίκεντρο των μελετών που αναζητούν τη γενετική αρχιτεκτονική των πολυσύνθετων νοσημάτων και φαινοτύπων (15).

Η πλειοψηφία των πολυμορφικών θέσεων του γονιδιώματος όπου εντοπίζονται τα SNPs είναι κατά κύριο λόγο διαλληλικές, δηλαδή σε επίπεδο πληθυσμού απαντώνται δύο διαφορετικές μορφές, π.χ. A ή G. Δεδομένου ότι σε έναν συγκεκριμένο μη-νοσούντα πληθυσμό τα δύο αυτά αλληλία θα πρέπει να βρίσκονται σε ισορροπία Hardy-Weinberg, το ένα από τα δύο αλληλία μιας πολυμορφικής θέσης έχει συχνότητα $< 50\%$ – και ονομάζεται έλασσον (*minor allele*) – και το έτερο έχει συχνότητα $> 50\%$ – και ονομάζεται μείζον (*major allele*). Επιπλέον, με βάση τη συχνότητα του ελάσσονος αλληλίου οι πολυμορφισμοί αυτοί διακρίνονται σε κοινούς (με συχνότητα $> 5\%$) και σε σπάνιους ή μη-κοινούς (με συχνότητα 5% ή μικρότερη). Είναι αξιοσημείωτο ότι ο αριθμός των SNPs με μεγάλες συχνότητες ελάσσονος αλληλίου είναι αισθητά μικρότερος από αυτόν των SNPs με μικρότερες συχνότητες (22).

Τέλος, άλλοι τύποι γενετικής ποικιλομορφίας λιγότερο συχνά απαντώμενοι είναι οι προσθέσεις (*insertions*), οι αφαιρέσεις (*deletions*), οι αναστροφές (*inversions*), οι διπλασιασμοί (*duplications*) και οι πολυμορφισμοί αριθμού αντιγράφων (*copy*

number variations) (23, 24). Αυτές οι αλλαγές αφορούν μεγαλύτερες γενετικές περιοχές που κυμαίνονται από μερικά ζεύγη βάσεων μέχρι ολόκληρα χρωμοσωμικά τμήματα.

1.1.3. Η ΓΕΝΕΤΙΚΗ ΕΠΙΔΗΜΙΟΛΟΓΙΑ ΤΩΝ ΠΟΛΥΣΥΝΘΕΤΩΝ ΝΟΣΗΜΑΤΩΝ

Σύμφωνα με τον ορισμό της Παγκόσμιας Οργάνωσης Υγείας (<http://www.who.int/topics/epidemiology/en/>), Επιδημιολογία είναι η επιστήμη που μελετά την κατανομή της συχνότητας καταστάσεων που σχετίζονται με την ανθρώπινη υγεία, καθώς και τους παράγοντες που τις επηρεάζουν, αλλά και τη συνακόλουθη εφαρμογή αυτής της γνώσης στον έλεγχο της ανθρώπινης νοσηρότητας (25). Κατ' επέκταση, η γενετική επιδημιολογία – ως κλάδος της επιδημιολογίας – μελετά το ρόλο των γενετικών χαρακτηριστικών των ατόμων και των περιβαλλοντικών τους (αλληλ)επιδράσεων στην κατανομή και εξέλιξη διαφόρων φυσιολογικών φαινοτύπων και νόσων στο επίπεδο του πληθυσμού (26, 27).

Τα πρώτα επιτεύγματα της γενετικής επιδημιολογίας αφορούσαν μονογονιδιακά νοσήματα (28), η οικογενής εμφάνιση των οποίων φαίνεται ότι είναι απόλυτα συμβατή με τους νόμους της μενδελιανής κληρονομικότητας. Τέτοια νοσήματα, όπως η φαινυλκετονουρία ή η κυστική ίνωση, προκαλούνται από μεταλλάξεις χαμηλής συχνότητας και υψηλής διεισδυτικότητας (*penetrance*) σε ένα μόνο γονίδιο (29). Για παράδειγμα, η κυστική ίνωση προκαλείται από μεταλλάξεις στο γονίδιο *CFTR*. Όσον αφορά αυτά τα ομοιογενή από άποψης γενετικής αρχιτεκτονικής νοσήματα, οι μελέτες γενετικής διασύνδεσης (*genetic linkage studies*) (30) είχαν πολύ σημαντικές επιτυχίες στη χαρτογράφησή τους (28, 31), με κυριότερη την ανακάλυψη του γονιδίου της κυστικής ίνωσης (32, 33).

Αντίθετα, για τα περισσότερα σύνθετα (*complex*) ή πολυπαραγοντικά (*multifactorial*) νοσήματα, όπως ο καρκίνος, το έμφραγμα του μυοκαρδίου, ο διαβήτης ή η σχιζοφρένεια, οι μελέτες γενετικής διασύνδεσης απέτυχαν να

εξακριβώσουν τη γενετική αρχιτεκτονική τους (34). Αυτή η αποτυχία οφείλεται κυρίως στη χαμηλή ισχύ των μελετών αυτών για τα πολυπαραγοντικά νοσήματα που επηρεάζονται από περισσότερα του ενός γονίδια, καθώς και στις μεγάλες χρωμοσωμικές περιοχές που μοιράζονται τα μέλη μιας οικογένειας, στις οποίες είναι δύσκολο να περιοριστεί το σήμα της διασύνδεσης ώστε να εντοπιστεί το αιτιολογικό γονίδιο (35). Τα νοσήματα αυτά αποκαλούνται επίσης και κοινά (*common*) γιατί, σε αντίθεση με τα σπάνια Μενδελιανά νοσήματα, συνιστούν ένα μεγάλο μέρος του ολικού φορτίου νοσηρότητας (36), σαφώς μεγαλύτερο σε σχέση με τις μονογονιδιακές διαταραχές, και ως εκ τούτου η γενετική διερεύνηση της αιτιολογίας και της παθοφυσιολογίας τους αποκτά μεγάλη σημασία για την κλινική πρακτική και συγκεκριμένα την πρόληψη, την πρόγνωση, τη διάγνωση (προγεννητική ή μη) και τη θεραπεία. Σε αντίθεση με τις μονογονιδιακές διαταραχές, τα πολυπαραγοντικά νοσήματα καθορίζονται από έναν αριθμό γενετικών και περιβαλλοντικών (επίκτητων) παραγόντων, καθώς και από τις μεταξύ τους αλληλεπιδράσεις (37, 38).

1.1.3.1. Μοντέλα γενετικής αρχιτεκτονικής

Για τη γενετική αρχιτεκτονική των κοινών νοσημάτων είχε διατυπωθεί αρχικά το μοντέλο του «κοινού νοσήματος – κοινού πολυμορφισμού» (*common disease – common variant model, CDCV*) (39-41). Σύμφωνα με το μοντέλο αυτό, το γενετικό υπόβαθρο για τα κοινά νοσήματα προέρχεται από προδιαθεσικά αλληλία με σχετικά υψηλές συχνότητες (κοινά αλληλία) και μικρά μεγέθη αποτελέσματος, τα οποία βρίσκονται σε μικρό αριθμό στους υποκείμενους γενετικούς τόπους (42, 43). Ευρήματα για την εφαρμογή του μοντέλου «κοινού νοσήματος – κοινού πολυμορφισμού» σε διάφορα νοσήματα προέρχονταν τόσο από θεωρητικές (44) όσο

και από εμπειρικές μελέτες (45, 46). Ωστόσο, για τα περισσότερα νοσήματα δεν υπήρχαν επαρκείς ενδείξεις που να στηρίζουν την καθολικότητα αυτού του μοντέλου (39).

Έτσι, διατυπώθηκε το μοντέλο του «κοινού νοσήματος – σπάνιου πολυμορφισμού» (*common disease – rare variant model, CDRV*) (47-49). Σύμφωνα με αυτό το μοντέλο, ένα σημαντικό μέρος της γενετικής μεταβλητότητας στα πολυσύνθετα νοσήματα μπορεί να αποδοθεί και να εξηγηθεί από την πολλαπλασιαστική δράση πολλών σπάνιων πολυμορφισμών με μέσα και μεγάλα μεγέθη αποτελέσματος (42, 43, 50). Το μοντέλο αυτό επιβεβαιώθηκε τόσο από εμπειρικές μελέτες αλληλούχισης (51-53) όσο και από θεωρητικές μελέτες (47).

Σήμερα η επικρατούσα άποψη είναι ότι η γενετική αρχιτεκτονική των σύνθετων νοσημάτων καθορίζεται από την αλληλεπίδραση τόσο των κοινών όσο και των σπάνιων πολυμορφισμών, οι οποίοι συνεισφέρουν ανεξάρτητα στο αντίστοιχο νόσημα/φαινότυπο (54). Μάλιστα, για πολλά νοσήματα φαίνεται ότι συνυπάρχουν στον ίδιο γενετικό τόπο κοινοί πολυμορφισμοί με μικρά μεγέθη αποτελέσματος και σπάνιοι πολυμορφισμοί με μεγαλύτερα μεγέθη αποτελέσματος ή ακόμα και μεταλλάξεις (55, 56).

1.1.4. ΟΙ ΜΕΛΕΤΕΣ ΓΕΝΕΤΙΚΗΣ ΣΥΣΧΕΤΙΣΗΣ

Οι μελέτες γενετικής συσχέτισης (*genetic association studies*) (57-59) άρχισαν να πραγματοποιούνται κυρίως από το 1990 και μετέπειτα, όταν και κατέστη σαφής η περιορισμένη δυναμική των μελετών γενετικής διασύνδεσης (34). Σε αντίθεση με τις τελευταίες, που πραγματοποιούνταν σε επιλεγμένες οικογένειες (*family-based studies*), οι μελέτες γενετικής συσχέτισης σε πληθυσμιακό επίπεδο (*population-based association studies*) τεκμηριώθηκαν ως η λύση για τη γενετική χαρτογράφηση των σύνθετων νοσημάτων (60).

Στόχος των μελετών γενετικής συσχέτισης είναι η αναζήτηση της ύπαρξης συσχέτισης ανάμεσα σε μία νοσολογική οντότητα/φαινότυπο και σε στοιχεία γενετικής ποικιλομορφίας. Στην πλειοψηφία των περιπτώσεων εξετάζεται αν η συχνότητα συγκεκριμένων γενετικών δεικτών είναι στατιστικά σημαντικά υψηλότερη σε άτομα τα οποία έχουν τη νόσο που μας ενδιαφέρει να μελετήσουμε. Από μεθοδολογικής πλευράς, ο πιο απλός επιδημιολογικός σχεδιασμός για τον έλεγχο της παραπάνω υπόθεσης είναι μία μελέτη ασθενών-μαρτύρων (*case-control study*). Οι πιο ευρέως χρησιμοποιούμενοι γενετικοί δείκτες είναι τα SNPs λόγω του μεγάλου αριθμού τους στο γονιδίωμα, ενώ άλλοι δείκτες που χρησιμοποιούνται πρόσφατα είναι οι μικροδορυφόροι (*microsatellites*), οι προσθέσεις ή αφαιρέσεις, τα CNVs κ.α. (61).

1.1.4.1. Μελέτες υποψηφίων γονιδίων

Οι μελέτες των υποψηφίων γονιδίων ή γονιδίων-στόχων (*candidate-gene association studies*) αναζητούν συσχετίσεις μεταξύ φαινοτύπων και γενετικών πολυμορφισμών σε συγκεκριμένα γονίδια ή γονιδιακούς τόπους (*loci*) (62). Η

επιλογή αυτών των γονιδίων και των αντίστοιχων SNPs προς μελέτη στηρίζεται σε πρότερη πληροφορία που αφορά τη λειτουργία και το βιολογικό ρόλο των γονιδίων αυτών (63). Τα λειτουργικά SNPs είναι συνήθως SNPs που προκαλούν μη-συνώνυμες αλλαγές στην αλληλουχία των αμινοξέων μιας πρωτεΐνης (64, 65).

Οι μελέτες αυτές είχαν αρκετά μειονεκτήματα. Ήταν κατά κύριο λόγο μικρές και με μικρή ισχύ για να ανιχνεύσουν τα μεγέθη αποτελέσματος των κοινών πολυμορφισμών (57), χρησιμοποιούσαν ως επίπεδο στατιστικής σημαντικότητας το συμβατικό 0.05 με βάση ένα 5% σφάλμα τύπου I ή άλλα όρια τα οποία ήταν πολύ χαμηλά δεδομένης χαμηλής πρότερης πιθανότητας ακόμα και για τα «καλύτερα» γονίδια (66), και μελετούσαν έναν ή λίγους πολυμορφισμούς κάθε φορά. Επίσης, συχνά δεν υπήρχε καμία μέριμνα για συγχυτικούς παράγοντες, όπως η πληθυσμιακή διαστρωμάτωση (*population stratification*) (67, 68). Τέλος, η πλειοψηφία των μελετών παρουσίαζε σε μεγάλη κλίμακα συστηματικά σφάλματα, όπως το συστηματικό σφάλμα αναφοράς (*reporting bias*) (69). Συνέπεια των ανωτέρω ήταν ο μεγάλος αριθμός ψευδώς θετικών ευρημάτων (70), τα περισσότερα από τα οποία καταρρίφθηκαν σε μεταγενέστερες μελέτες ή μετα-αναλύσεις αυτών, καθώς επίσης και η αδυναμία επικύρωσης πολλών συσχετίσεων (71, 72), που σε κάποιο βαθμό έχει να κάνει με το γεγονός ότι στα πρώτα στάδια οι ερευνητές ίσως υπερεκτίμησαν την ικανότητά τους να επιλέξουν τα καλύτερα υποψήφια γονίδια (73). Το φαινόμενο αυτό έχει περιγραφεί γενικότερα για τους βιολογικούς δείκτες (*biomarkers*) πέραν των γενετικών δεικτών, πολλοί από τους οποίους έχουν αποτελέσει το επίκεντρο της βιοϊατρικής έρευνας κατά το παρελθόν (74).

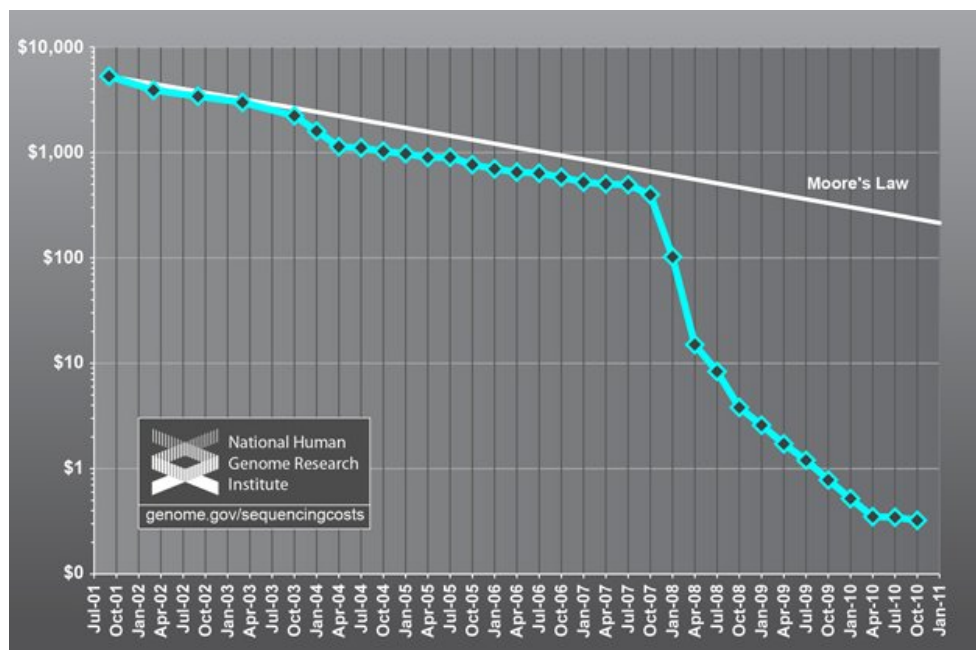
1.1.4.2. Μελέτες σάρωσης του γονιδιώματος

Οι μελέτες σάρωσης του γονιδιώματος (*genome-wide association studies*, *GWAS*) είναι μελέτες που χρησιμοποιούν υψηλής απόδοσης τεχνολογία γονοτύπησης DNA (75), με την οποία γονοτυπούνται συγχρόνως εκατοντάδες χιλιάδες έως εκατομμύρια SNPs σε όλη την έκταση του γονιδιώματος. Παρόλο που η ισχύς των μελετών γονιδιακής σάρωσης για την ανακάλυψη γενετικών σχέσεων σε κοινά νοσήματα τεκμηριώθηκε σε θεωρητικό επίπεδο ήδη από τα μέσα της δεκαετίας του '90 (60), η πρακτική εφαρμογή τους σε μεγάλη κλίμακα κατέστη εφικτή την τελευταία δεκαετία, χάρη σε δύο κυρίως παράγοντες.

Ο πρώτος – και ίσως σημαντικότερος – παράγοντας είναι ο καθορισμός των προτύπων ανισορροπίας διασύνδεσης (*linkage disequilibrium*, *LD*) στο ανθρώπινο γονιδίωμα μέσω του προγράμματος HapMap (76-78). Το φαινόμενο της ανισορροπίας διασύνδεσης αφορά στην ύπαρξη ορισμένων συνδυασμών αλληλίων οι οποίοι απαντώνται μαζί σε έναν πληθυσμό πιο συχνά από ό,τι θα αναμέναμε από τύχη και μόνο (15, 17). Με άλλα λόγια, πρόκειται για τη μη-τυχαία συσχέτιση αλληλίων σε δύο ή περισσότερους γενετικούς τόπους (17, 79, 80) λόγω του γεγονότος ότι αυτά τα SNPs δεν κληρονομούνται ανεξάρτητα αλλά σε συστάδες (*blocks*) (81, 82). Υπολογίζεται ότι σε όλο το γονιδίωμα περίπου 10.000.000 SNPs μεταβιβάζονται από τη μία γενιά στην επόμενη σε τέτοιες συστάδες. Κατά συνέπεια, είναι δυνατόν να γονοτυπηθούν μερικά μόνο SNPs (που ονομάζονται tag-SNPs) – τα οποία λειτουργούν ως «εξουσιοδοτημένοι» δείκτες (*proxies*) για άλλα SNPs που βρίσκονται σε ανισορροπία διασύνδεσης – για να «αντικατοπτρίσουν» την ποικιλομορφία σε κάθε συστάδα αντί να γονοτυπηθούν όλοι οι πολυμορφισμοί στην περιοχή αυτή (83-85). Στη γενετική πληθυσμών (*population genetics*),

χρησιμοποιούνται δύο μέτρα για την ποσοτικοποίηση της ανισορροπίας διασύνδεσης. Το πρώτο είναι το r^2 (κατ' ουσίαν ένας συντελεστής συσχέτισης [*correlation coefficient*] ανάμεσα σε δύο πολυμορφισμούς) που κυμαίνεται από 0 έως 1 με τιμές >0.80 να υποδηλώνουν πολύ ισχυρή συσχέτιση – και κατ' επέκταση και ανισορροπία διασύνδεσης – μεταξύ των πολυμορφισμών. Το δεύτερο μέτρο είναι ο δείκτης D' που μπορεί να ερμηνευθεί ως η διαφορά των αναμενόμενων από τις παρατηρούμενες συχνότητες των απλοτύπων· τιμές D' κοντά στο 0 είναι ενδεικτικές ανεξάρτητων πολυμορφισμών, ενώ τιμές κοντά στο 1 δηλώνουν ισχυρή διασύνδεση. Έτσι, σύμφωνα με ορισμένες εκτιμήσεις περίπου 300.000 έως 1.000.000 SNPs μπορούν να καλύψουν πάνω από το 80% της ποικιλομορφίας του ανθρώπινου γονιδιώματος που αφορά κοινούς πολυμορφισμούς (δηλαδή με συχνότητα $>5\%$ στο γενικό πληθυσμό) (85, 86). Το μειονέκτημα αυτής της έμμεσης προσέγγισης έγκειται στο γεγονός ότι οι πολυμορφισμοί που ανιχνεύονται είναι στις περισσότερες περιπτώσεις κάποιοι δείκτες και όχι κατ' ανάγκη οι αιτιολογικοί πολυμορφισμοί για το εξεταζόμενο νόσημα, ο προσδιορισμός των οποίων απαιτεί λεπτομερή χαρτογράφηση (*fine mapping*) (85, 87).

Ο δεύτερος παράγοντας που συντέλεσε στην ευρεία πραγματοποίηση των μελετών γονιδιακής σάρωσης είναι η ραγδαία μείωση του κόστους γονοτύπησης κατά την τελευταία δεκαετία (85), όπως φαίνεται και στην **Εικόνα 1**. Το κόστος αυτό έχει μειωθεί σε λιγότερο από 0.01 US\$ ανά γονότυπο, με αποτέλεσμα μία μελέτη σάρωσης του γονιδιώματος για 500,000 πολυμορφισμούς σε 2,000 άτομα να κοστίζει σήμερα πολύ λιγότερο από 1 εκατομμύριο δολάρια. Συγκριτικά, πριν από 10 χρόνια η ίδια μελέτη θα απαιτούσε τη γονοτύπηση περίπου 10,000,000 SNPs με κόστος 1 US\$ ανά πολυμορφισμό και συνολικό κόστος 20 εκατομμύρια δολάρια!



Εικόνα 1. Μεταβολή του κόστους γονοτύπησης.

Πηγή: National Human Genome Research Institute, National Institutes of Health, Bethesda, MD

Οι πρώτες εμπορικές πλατφόρμες που χρησιμοποιήθηκαν για τη γονοτύπηση περιλάμβαναν περίπου 500,000 SNPs, ενώ οι πιο πρόσφατες περιλαμβάνουν πάνω από 2,000,000. Αν και δεν υπάρχει ένας κοινώς αποδεκτός αριθμός SNPs που να καθορίζει πότε μια μελέτη χαρακτηρίζεται ως GWAS, εντούτοις ο ορισμός των Εθνικών Ιδρυμάτων Υγείας (*National Institutes of Health, NIH*) των ΗΠΑ χρησιμοποιεί ως όριο τους 100,000 πολυμορφισμούς. Με βάση αυτόν τον ορισμό, η βάση «A Catalog of Published Genome-Wide Association Studies» του Εθνικού Ινστιτούτου για την Έρευνα του Ανθρώπινου Γονιδιώματος (*National Human Genome Research*

Institute) των ΗΠΑ καταλογογραφεί τις μελέτες που γονοτυπούν τουλάχιστον 100,000 πολυμορφισμούς (88, 89).

Συμπερασματικά, η ευρεία πραγματοποίηση των μελετών σάρωσης του γονιδιώματος, ωθούμενη από όλα τα παραπάνω, αύξησε σημαντικά τον αριθμό των γενετικών συσχετίσεων με πολύ υψηλά επίπεδα στατιστικής σημαντικότητας για πολλά σύνθετα νοσήματα (87, 90-92).

1.1.4.2.1. Βασικός σχεδιασμός των μελετών σάρωσης του γονιδιώματος

Τα SNPs που γονοτυπούνται σε μία μελέτη σάρωσης του γονιδιώματος επιλέγονται κατά τέτοιον τρόπο, ώστε να καλύψουν ένα μεγάλο μέρος της ποικιλομορφίας της αλληλουχίας σε όλο το εύρος του γονιδιώματος. Έτσι, έχοντας διαθέσιμο τον γονότυπο σε 500,000 έως 2,000,000 θέσεις (ανάλογα με την πλατφόρμα) για κάθε ασθενή στη μελέτη μας, μπορούμε να συσχετίσουμε αυτούς τους πολυμορφισμούς με τους διάφορους φαινοτύπους ή νοσήματα που μας ενδιαφέρουν (91, 93, 94). Με άλλα λόγια οι μελέτες σάρωσης του γονιδιώματος καλύπτουν ταυτόχρονα ένα πολύ μεγάλο μέρος του γονιδιώματος χωρίς να κάνουν κάποια εκ των προτέρων υπόθεση για τη θέση ή τη λειτουργία των υπεύθυνων πολυμορφισμών (85), εφαρμόζοντας έτσι μία αγνωστικιστική (*agnostic*) προσέγγιση (95, 96).

Οι πλατφόρμες γονοτύπησης στις μελέτες σάρωσης του γονιδιώματος προσφέρουν πολύ καλή κάλυψη σε όλο το γονιδίωμα για πολυμορφισμούς με συχνότητα ελάσσονος αλληλίου μεγαλύτερη του 5% (97, 98). Η επιλογή αυτού του εύρους των συχνοτήτων οφείλεται κυρίως σε τεχνολογικούς (π.χ. αδυναμία κάλυψης σπάνιων πολυμορφισμών λόγω απουσίας τους στο πρόγραμμα HarMap) και

οικονομικούς περιορισμούς. Έτσι, πολυμορφισμοί με χαμηλότερες συχνότητες ελάσσονος αλληλίου είτε αποκλείονται συστηματικά από τις εμπορικές πλατφόρμες είτε τυγχάνουν ανεπαρκούς κάλυψης (97, 99). Για το λόγο αυτό, οι μελέτες σάρωσης του γονιδιώματος που έχουν πραγματοποιηθεί και δημοσιευθεί μέχρι σήμερα έχουν χρησιμοποιήσει ως επί το πλείστον κοινούς γενετικούς δείκτες και έχουν ανακαλύψει χαμηλής διεισδυτικότητας (*low-penetrance*) πολυμορφισμούς με μικρά μεγέθη αποτελέσματος (50, 97).

Ένα βασικό στάδιο σε κάθε μελέτη γονιδιακής σάρωσης αφορά στον έλεγχο ποιότητας (*quality control*) της μελέτης (100-102). Το στάδιο αυτό διασφαλίζει ότι αναλύονται στη μελέτη άτομα αλλά και γονότυποι οι οποίοι έχουν κριθεί κατάλληλοι, ώστε να μην προκαλούν συστηματικά σφάλματα τα οποία αυξάνουν τον αριθμό των ψευδώς θετικών και ψευδώς αρνητικών αποτελεσμάτων. Τα κριτήρια που χρησιμοποιούνται στον έλεγχο ποιότητας ενός πολυμορφισμού περιλαμβάνουν συνοπτικά: (i) τη συχνότητα ελάσσονος αλληλίου του πολυμορφισμού, με τις περισσότερες πλατφόρμες γονοτύπησης να είναι σχεδιασμένες ώστε να περιλαμβάνουν κοινούς πολυμορφισμούς (συχνότητες >1% ή >5%), (ii) την ισορροπία κατά Hardy-Weinberg σύμφωνα με την οποία οι συχνότητες των γονοτύπων σε έναν πληθυσμό θα πρέπει να παραμένουν σταθερές από γενιά σε γενιά (πολυμορφισμοί που αποκλίνουν από την ισορροπία αυτή αποκλείονται από την ανάλυση), (iii) τον ρυθμό ανάκλησης (*call rate*) των γονοτύπων, δηλαδή το ποσοστό των γονοτύπων σε κάθε πολυμορφισμό που δεν έχουν ελλιπείς τιμές κατά τη γονοτύπηση (κατά κανόνα ρυθμοί >95% θεωρούνται αποδεκτοί) και (iv) τη σημαντική διαφορά ελλειπουσών τιμών γονοτύπων μεταξύ των συγκρινόμενων ομάδων ατόμων (π.χ. ασθενών και μαρτύρων). Παρόμοια κριτήρια εφαρμόζονται και

όσον αφορά τον αριθμό των γονοτυπηθέντων ατόμων που τελικά θα συμπεριληφθούν στην ανάλυση.

1.1.4.2.2. Τα πρώτα μαθήματα από τις μελέτες σάρωσης του γονιδιώματος

Οι ευρείας κλίμακας συστηματικές αποτιμήσεις των ευρημάτων των μελετών σάρωσης του γονιδιώματος ανέδειξαν ορισμένες εντυπωσιακές παρατηρήσεις. Η πρώτη ήταν ότι πολλά από τα νέα SNPs που ανακαλύφθηκαν δεν βρίσκονταν σε γονίδια που θεωρούνταν ότι σχετίζονταν με κάποιες ασθένειες με βάση την πρότερη γνώση που είχαμε (89, 103). Επιπλέον, πολλά SNPs δεν εντοπίζονταν ουδόλως σε γονιδιακές περιοχές, αλλά σε περιοχές που δεν υπήρχαν γονίδια και που χαρακτηρίζονται ως «γονιδιακές έρημοι» (*gene deserts*) (104-106). Έτσι, νέοι παθοφυσιολογικοί μηχανισμοί και δυνητικοί θεραπευτικοί στόχοι προέκυψαν από αυτές τις μελέτες, όπως π.χ. το μονοπάτι της φλεγμονής στην ηλικιακή εκφύλιση της ωχράς (107, 108), το μονοπάτι του υποδοχέα της νικοτίνης στον καρκίνο του πνεύμονα (109) αλλά και διάφορα άλλα. Η δεύτερη εντυπωσιακή παρατήρηση ήταν το γεγονός ότι οι πολυμορφισμοί που ανακαλύφθηκαν επέφεραν μόνο μια μικρή αύξηση του κινδύνου για τις αντίστοιχες νόσους (103), με τον τυπικό λόγο αναλογιών να είναι της τάξης του 1.20-1.30 (89, 91). Τέλος, ακόμα και μετά από αναλύσεις γενετικών δεδομένων χιλιάδων ατόμων μέσω διεθνών συνασπισμών (*consortia*), οι δεκάδες ή και εκατοντάδες ανακαλυφθέντες πολυμορφισμοί εξηγούν ένα μικρό μόνο μέρος της κληρονομικότητας των αντίστοιχων φαινοτύπων (55, 110-112).

Έχουν διατυπωθεί πολλές υποθέσεις για τους λόγους στους οποίους οφείλεται αυτή η αδυναμία των μελετών σάρωσης του γονιδιώματος να βρουν

γενετικές σχέσεις με μεγάλα μεγέθη. Πρώτον, η διαδικασία της εξέλιξης και η φυσική επιλογή αναμένονται να επιδράσουν προς την κατεύθυνση της απάλειψης τέτοιων πολυμορφισμών από τον πληθυσμό (113). Δεύτερον, πολλά από τα SNPs που χρησιμοποιούνται στις πλατφόρμες γονοτύπησης βρίσκονται σε ανισορροπία διασύνδεσης σε ποικίλο βαθμό με τα αιτιολογικά SNPs, γεγονός που πιθανόν οδηγεί σε υποεκτίμηση του μεγέθους της γενετικής σχέσης. Τρίτον, σημαντική συνεισφορά έχουν οι διάφοροι δομικοί πολυμορφισμοί και οι πολυμορφισμοί χαμηλής συχνότητας (114, 115), οι οποίοι, όπως προαναφέρθηκε, δεν περιλαμβάνονται στις περισσότερες μελέτες σάρωσης του γονιδιώματος, οι οποίες εστιάζονται κατά κύριο λόγο – αν και όχι αποκλειστικά – σε πολυμορφισμούς με συχνότητες >5% (55, 97). Τέταρτον, οι περισσότερες μελέτες μέχρι τώρα δεν περιλάμβαναν στο σχεδιασμό και τις αναλύσεις τους τις διάφορες αλληλεπιδράσεις τόσο μεταξύ γονιδίων (116) όσο και μεταξύ γονιδίων και περιβαλλοντικών παραγόντων (117). Τέλος, τα μικρά μεγέθη δείγματος που χρησιμοποιήθηκαν στις περισσότερες μελέτες σάρωσης του γονιδιώματος πριν θεμελιωθούν οι διεθνείς συνασπισμοί, δεν ήταν επαρκή για την ανίχνευση μικρών μεγεθών αποτελέσματος που προκύπτουν από τους κοινούς πολυμορφισμούς (55). Αυτές οι παράμετροι αποτελούν το αντικείμενο εντατικής διερεύνησης στο πεδίο της γενετικής επιδημιολογίας, καθώς υπάρχουν ισχυρά τεκμήρια ότι μπορούν να εξηγήσουν ένα σημαντικό μέρος της άγνωστης γενετικής αρχιτεκτονικής – γνωστής επίσης και ως «γενετικής σκοτεινής ύλης» (*genetic dark matter*) (54, 55).

1.1.4.2.3. Οι πολλαπλές συγκρίσεις στις μελέτες γονιδιακής σάρωσης

Στις μελέτες σάρωσης του γονιδιώματος πραγματοποιούνται συγχρόνως εκατοντάδες χιλιάδες έως εκατομμύρια έλεγχοι υποθέσεων (όσοι δηλαδή και οι πολυμορφισμοί που αναλύονται). Έτσι, στην περίπτωση που εξετάζουμε π.χ. 1,000,000 πολυμορφισμούς, η εφαρμογή του κλασικού ορίου στατιστικής σημαντικότητας $P=0.05$ (που αντιστοιχεί σε σφάλμα τύπου I $\alpha=0.05$) έχει ως συνέπεια $0.05 \times 1,000,000 = 50,000$ πολυμορφισμοί να «συσχετίζονται» με τον εξεταζόμενο φαινότυπο από τύχη και μόνο, δηλαδή αναμένουμε έναν μεγάλο αριθμό ψευδώς θετικών γενετικών συσχετίσεων. Γίνεται, λοιπόν, αντιληπτό ότι όλες αυτές οι πολλαπλές συγκρίσεις αποτελούν ένα μείζον θέμα στο σχεδιασμό και την ανάλυση των μελετών σάρωσης του γονιδιώματος (118). Θα πρέπει επομένως το όριο στατιστικής σημαντικότητας σε αυτές τις μελέτες να είναι τέτοιο ώστε να λαμβάνεται υπόψη και να «διευθετείται» το θέμα των πολλαπλών συγκρίσεων που πραγματοποιούνται (119-121). Ωστόσο, σε μεγάλο βαθμό υπάρχει αβεβαιότητα στη βιβλιογραφία ως προς το ποιο είναι το κατάλληλο όριο στατιστικής σημαντικότητας που πρέπει να εφαρμόζεται, με τις διάφορες μελέτες να έχουν χρησιμοποιήσει ποικίλες μεθόδους αλλά και τιμές P προκειμένου να θέσουν το όριο στατιστικής σημαντικότητας, χωρίς κανένα από αυτά τα όρια να είναι απόλυτο (122, 123).

Πολλές στατιστικές μέθοδοι έχουν αναπτυχθεί και χρησιμοποιηθεί προκειμένου να αντιμετωπισθεί αυτό το θέμα. Οι Martin et al. παρουσιάζουν μία διεξοδική ανασκόπηση όλων αυτών των προσεγγίσεων (124). Συνοπτικά, αυτές περιλαμβάνουν μεθόδους στατιστικής ανεξαρτησίας (125), μεθόδους προσαρμογής του ποσοστού σφάλματος (όπως η διόρθωση κατά Bonferroni (126), η διόρθωση κατά Sidak (127), το ποσοστό ψευδών ανακαλύψεων (*false discovery rate*), δηλαδή

το ποσοστό των στατιστικά σημαντικών σχέσεων που είναι στην πραγματικότητα ψευδώς θετικά ευρήματα (121)), και τέλος, μεθόδους αναγωγής δεδομένων (*data reduction methods*) (128). Από αυτές τις μεθόδους η πιο συχνά χρησιμοποιούμενη είναι η διόρθωση κατά Bonferroni (103), όπου η τιμή P προκύπτει διαιρώντας τη συμβατική τιμή $P=0.05$ με τον αριθμό των συγκρίσεων που πραγματοποιούνται. Έτσι, για την πλειονότητα των περιπτώσεων όπου ελέγχονται περίπου 1,000,000 ($=10^6$) πολυμορφισμοί, το όριο στατιστικής σημαντικότητας είναι $0.05/10^6=5 \times 10^{-8}$, ενώ οι πρώτες μελέτες που γονοτυπούσαν 500,000 ($=5 \times 10^5$) πολυμορφισμούς χρησιμοποιούσαν ως όριο το $0.05/5 \times 10^5=1 \times 10^{-7}$.

Εναλλακτική προσέγγιση του ορίου στατιστικής σημαντικότητας για τον έλεγχο της αξιοπιστίας μιας γενετικής σχέσης αποτελούν διάφορες Μπεϋζιανές (*Bayesian*) προσεγγίσεις. Οι μέθοδοι αυτές ενσωματώνουν την πρότερη πιθανότητα μιας σχέσης με βάση εξωτερικά τεκμήρια, όπως π.χ. τα χαρακτηριστικά του νοσήματος ή του πολυμορφισμού, το μέγεθος αποτελέσματος κ.ά. (66, 95, 129, 130). Μία τέτοια προσέγγιση είναι ο υπολογισμός του συντελεστή Bayes (*Bayes factor, BF*), ο οποίος ορίζεται ως ο λόγος της πιθανότητας των δεδομένων υπό τη μηδενική υπόθεση προς την πιθανότητα των δεδομένων υπό την εναλλακτική υπόθεση (131):

$$BF = \frac{\Pr(data | H_0)}{\Pr(data | H_a)}$$

Από την εξίσωση αυτή συνάγεται ότι, όταν $BF > 1$, υπάρχει τεκμηρίωση υπέρ της μηδενικής υπόθεσης, ενώ όταν $BF < 1$ υπάρχει τεκμηρίωση υπέρ της εναλλακτικής υπόθεσης. Η προσέγγιση αυτή προσφέρει πλεονεκτήματα έναντι των τιμών P (129, 132) και μπορεί να χρησιμοποιηθεί για τη βαθμονόμηση (*calibration*) της αξιοπιστίας των ευρημάτων των μελετών σάρωσης του γονιδιώματος (95). Τέτοιες προσεγγίσεις

μπορούν να ενσωματώσουν στην εκτίμηση της μετα-τεκμηριωματικής πιθανότητας για την ύπαρξη μιας γενετικής σχέσης, τις πρότερες πεποιθήσεις των ερευνητών σχετικά με την ύπαρξη αυτής της σχέσης ή όχι (133).

Τέλος, προκειμένου να μειώσουν τον αριθμό των ψευδώς θετικών ευρημάτων, πολλές μελέτες επιλέγουν ένα σχεδιασμό πολλαπλών σταδίων, όπου τα ισχυρότερα SNPs που ανακαλύπτονται στο στάδιο I μεταφέρονται προς επαλήθευση σε άλλα δείγματα στα στάδια II ή/και III. Ακόμα, από τους πλέον καθιερωμένους σχεδιασμούς είναι τα στάδια αυτά να ακολουθούνται από μετα-ανάλυση (134) των ευρημάτων κάθε σταδίου. Με τον τρόπο αυτό μειώνεται ο αριθμός των ψευδώς θετικών ευρημάτων, ενώ συγχρόνως μειώνεται το κόστος και διατηρείται ικανοποιητική η στατιστική ισχύς (135, 136).

1.1.4.2.4. Περιορισμοί των μελετών σάρωσης του γονιδιώματος

Παρόλα τα θεαματικά ευρήματα και τις επιτυχίες των μελετών σάρωσης του γονιδιώματος, θα πρέπει να έχει κανείς υπόψη ορισμένες παραμέτρους που αφορούν την ερμηνεία των αποτελεσμάτων αυτών των μελετών. Κατ'αρχήν, η ανίχνευση μιας συσχέτισης – ακόμα και καλά επικυρωμένης – ανάμεσα σε κάποιο γενετικό πολυμορφισμό και μία νόσο, δεν σημαίνει απαραίτητα ότι αυτός είναι και ο υπεύθυνος αιτιολογικός πολυμορφισμός, καθώς είναι πολύ πιθανό αυτός να είναι ένας πολυμορφισμός σε υψηλή ανισορροπία διασύνδεσης που γονοτυπήθηκε στη συγκεκριμένη εμπορική πλατφόρμα (91, 92). Κατά κανόνα οι μελέτες συσχέτισης καταδεικνύουν έναν γενετικό τόπο που σχετίζεται με μία νόσο, αλλά παρέχουν λίγα δεδομένα όσον αφορά τη λειτουργικότητα. Για το λόγο αυτό συνήθως απαιτούνται επιπλέον λειτουργικές μελέτες, π.χ. μελέτες γονιδιακής έκφρασης ή έκφρασης RNA,

μελέτες σε ζώα, ιστικές μελέτες κ.ά., ή ακόμα και αλληλούχιση ορισμένων περιοχών προκειμένου να βρεθεί ο υπεύθυνος αιτιολογικός πολυμορφισμός (92). Πράγματι, πολλές μελέτες σάρωσης του γονιδιώματος συνοδεύονται συχνά από τέτοιες λειτουργικές μελέτες (137).

Επίσης, οι διάφορες μεθοδολογικές προσεγγίσεις που στοχεύουν στη μείωση των ψευδώς θετικών ευρημάτων διατηρώντας παράλληλα επαρκή ισχύ, είναι πιθανό ορισμένες φορές να «παραβλέψουν» μία αληθινή σχέση. Αυτό είναι ιδιαίτερα συχνό φαινόμενο όταν για λόγους κόστους μόνο ένας μικρός αριθμός πολυμορφισμών – κατά βάση αυτοί με τη μεγαλύτερη στατιστική σημαντικότητα – μεταφέρονται στα επόμενα στάδια μιας μελέτης. Στο πλαίσιο αυτό, έχουν προταθεί λιγότερο αυστηρά όρια στατιστικής σημαντικότητας (122).

Τέλος, η πλειοψηφία των μελετών σάρωσης του γονιδιώματος (κυρίως οι πρώτες μελέτες που πραγματοποιήθηκαν) δεν είχε σε ικανοποιητικό βαθμό τη δυνατότητα ανίχνευσης πολυμορφισμών χαμηλής συχνότητας ή δομικών πολυμορφισμών ή ακόμα τη δυνατότητα ταυτόχρονης αξιολόγησης γενετικών και περιβαλλοντικών παραγόντων.

1.1.4.2.5. Σύνοψη των μελετών σάρωσης του γονιδιώματος

Οι μελέτες σάρωσης του γονιδιώματος οδήγησαν στην ανακάλυψη ενός μεγάλου αριθμού γενετικών σχέσεων με πολύ υψηλή στατιστική σημαντικότητα και μεγάλα ποσοστά επαλήθευσης και επικύρωσης. Αυτή η επιτυχία οφείλεται στους ακόλουθους λόγους. Πρώτον, το φορτίο των πολλαπλών συγκρίσεων και ο μεγάλος αριθμός ψευδώς θετικών ευρημάτων που αυτό συνεπάγεται, έχει οδηγήσει στην υιοθέτηση πολύ αυστηρών κριτηρίων που χρησιμοποιούνται για την ανακάλυψη των

γενετικών σχέσεων, όπου το σφάλμα τύπου I είναι συνήθως 5×10^{-8} αντί του 0.05 (122, 138, 139).

Δεύτερον, όλο και περισσότερες ερευνητικές ομάδες συνεργάζονται δημιουργώντας συνασπισμούς (*consortia*) (140, 141) που μοιράζονται τα δεδομένα, γεγονός που οδήγησε σε μεγάλη αύξηση των δειγμάτων που χρησιμοποιούνται για την ανακάλυψη και επαλήθευση των γενετικών σχέσεων μέσω προοπτικών μετα-αναλύσεων (134, 137, 142). Μάλιστα, το Εργαστήριο Υγιεινής & Επιδημιολογίας της Ιατρικής Σχολής του Πανεπιστημίου Ιωαννίνων αποτέλεσε το μεθοδολογικό κέντρο των συνασπισμών GEFOS (*GEnetic Factors in OSteoporosis*) και TREAT-OA (*Translational Research in Europe Applied Technologies in OsteoArthritis*) που οδήγησαν στην ανακάλυψη γενετικών τόπων για την οστική πυκνότητα (143) και την οστεοαρθρίτιδα (144), αντίστοιχα.

Τέλος, στις μελέτες γονιδιακής σάρωσης χρησιμοποιούνται αγνωστικιστικές πλατφόρμες γονοτύπησης για εκατομμύρια πολυμορφισμούς που επιτρέπουν τον ταυτόχρονο έλεγχο όλων των γονοτύπων, περιορίζονται δραστικά μέσω κατάλληλης μεθοδολογίας οι συγχυτικοί παράγοντες λόγω πληθυσμιακής διαστρωμάτωσης, και παρουσιάζονται όλα τα αποτελέσματα χωρίς συστηματικά σφάλματα αναφοράς (145).

1.1.4.3. Αλληλούχιση

Μέχρι πριν από μερικά χρόνια η ευρείας κλίμακας αλληλούχιση περιοχών του γονιδιώματος ήταν μία χρονοβόρα και ακριβή διαδικασία που απαιτούσε τεράστια υπολογιστική ισχύ. Ωστόσο, το 2008 κυκλοφόρησαν οι πρώτες συσκευές αλληλούχισης επόμενης γενιάς (*next-generation sequencing, NGS*) οι οποίες μείωσαν

δραστικά το σχετικό κόστος, ενώ έχουν έως και 50,000 φορές μεγαλύτερη ταχύτητα αλληλούχισης σε σχέση με την τεχνολογία που χρησιμοποιήθηκε στο Πρόγραμμα του Ανθρώπινου Γονιδιώματος (146). Η νέα αυτή τεχνολογία επεξεργάζεται και αναλύει εκατομμύρια τμηματικές περιοχές (*reads*) του γονιδιώματος κάθε φορά (αντί για μόνο 96 περιοχές που ήταν το μέγιστο όριο των παλαιότερων τεχνικών αλληλούχισης), με ελάχιστα σφάλματα και με ακρίβεια που φτάνει έως και το 99.9%, ενώ απαιτούνται μόνο μερικά μg αίματος (147, 148).

Η επόμενη γενιάς αλληλούχιση μόλις τα τελευταία χρόνια έχει αρχίσει να εφαρμόζεται σε πληθυσμιακές μελέτες γενετικής συσχέτισης για διάφορα πολυσύνθετα νοσήματα, όπως διάφορες νεοπλασίες (149, 150), το σύνδρομο νοσούντος φλεβόκομβου (151) ή τα επίπεδα λιπιδίων στο αίμα (152). Οι τεχνικές αυτές μπορούν να εφαρμοστούν είτε στο σύνολο των εξονίων του γονιδιώματος (*exome sequencing*), είτε σε ολόκληρο το γονιδίωμα (*whole-genome sequencing*), είτε σε ορισμένες γονιδιωματικές περιοχές ενδιαφέροντος (*targeted sequencing*) ανάλογα με το στόχο της μελέτης (ανακάλυψη νέων πολυμορφισμών, επιβεβαίωση προηγούμενων ευρημάτων ή ενδεδειγμένη χαρτογράφηση του γονιδιώματος [*fine mapping*]) (153). Οι μελέτες αυτές ήδη έχουν ανακαλύψει έναν σημαντικό αριθμό πολυμορφισμών με συχνότητα ελάσσονος αλληλίου <5%, οι οποίοι δεν έχουν επαρκή κάλυψη στις μελέτες GWAS, ενώ αναμένεται να αποσαφηνίσουν σε ακόμα μεγαλύτερο βαθμό το ρόλο της ποικιλομορφίας χαμηλής συχνότητας στη γενετική αρχιτεκτονική των πολυσύνθετων νοσημάτων, καθώς προσφέρουν ανάλυση του γονιδιώματος σε πολύ μεγαλύτερο βάθος (153, 154).

1.1.5. ΟΙ ΠΟΛΥΜΟΡΦΙΣΜΟΙ ΧΑΜΗΛΗΣ ΣΥΧΝΟΤΗΤΑΣ

Όπως αναφέρθηκε παραπάνω, οι πολυμορφισμοί που συνεισφέρουν στον κίνδυνο των πολυσύνθετων νοσημάτων καλύπτουν ένα μεγάλο εύρος όσον αφορά τις συχνότητές τους στον πληθυσμό (39, 49, 154). Στο ένα άκρο βρίσκονται οι πολυμορφισμοί χαμηλής συχνότητας, δηλαδή αυτοί που απαντώνται σε συχνότητα μικρότερη από 5% στο γενικό πληθυσμό (56). Διακρίνονται σε μη-κοινούς (*uncommon*), με συχνότητα ελάσσονος αλληλίου $\leq 5\%$, και σε σπάνιους (*rare*), με συχνότητα ελάσσονος αλληλίου $< 0.5\%$ (55, 97), – αν και η διάκριση αυτή δεν θα πρέπει να θεωρείται απόλυτη (π.χ. παραμένει ασαφές ποια είναι η σχετική συνεισφορά αυτών των διαφορετικών τάξεων στον κίνδυνο της νόσου (55)).

Τέτοιας χαμηλής συχνότητας πολυμορφισμοί με ποικίλη λειτουργικότητα υπάρχουν σε αφθονία στο ανθρώπινο γονιδίωμα (περίπου 1 πολυμορφισμός σε κάθε 17 βάσεις) και είναι πιθανό να έχουν σημαντικές γενετικές επιδράσεις για διάφορα νοσήματα, όπως προκύπτει από λειτουργικές μελέτες (*functional studies*) αλληλούχισης του γονιδιώματος (155, 156). Τα συμπεράσματα αυτά ενισχύονται από ευρήματα που τεκμηριώνουν ότι η διαδικασία της φυσικής επιλογής περιορίζει τα περισσότερα τέτοια επιβλαβή αλληλία σε χαμηλές συχνότητες $< 1\%$ στον πληθυσμό μέσω διαφόρων μοντέλων της εξέλιξης (113, 157).

Επιπλέον των ανωτέρω πειραματικών ή/και θεωρητικών μελετών, ο ρόλος των πολυμορφισμών χαμηλής συχνότητας επιβεβαιώνεται και από μελέτες γενετικής συσχέτισης που έχουν βρει συγκεκριμένες γενετικές σχέσεις μεταξύ τέτοιων πολυμορφισμών και ορισμένων νοσημάτων. Τέτοια παραδείγματα είναι ο σακχαρώδης διαβήτης τύπου I (53), η υπέρταση (158), καθώς και η υψηλής

πυκνότητας λιποπρωτεΐνης (HDL) χοληστερόλη (159) και η χαμηλής πυκνότητας λιποπρωτεΐνης (LDL) χοληστερόλη (52).

Δυστυχώς οι πολυμορφισμοί αυτοί ήταν δύσκολο έως σήμερα να ανιχνευθούν στις μελέτες σάρωσης του γονιδιώματος κυρίως λόγω της ανεπαρκούς κάλυψης, των απαιτήσεων για μεγάλα μεγέθη δείγματος και των ανεπαρκών μεθοδολογιών ανάλυσης (97, 160). Έτσι, τα τελευταία χρόνια αναπτύχθηκαν νεότερες πλατφόρμες και τεχνικές γονοτύπησης, όπως η αλληλούχιση του συνόλου των εξονίων (*exome sequencing*) (161, 162) και η αλληλούχιση του πλήρους γονιδιώματος (*whole-genome sequencing*) (154, 163), καθώς και νέες μεθοδολογικές προσεγγίσεις (160, 164, 165), οι οποίες εφαρμόζονται ήδη για την αναζήτηση γενετικών συσχετίσεων που αφορούν μη-κοινούς και σπάνιους πολυμορφισμούς (166-169).

Παρόλα αυτά, ακόμα και «παραδοσιακές» μελέτες σάρωσης του γονιδιώματος έχουν ανακαλύψει περιστασιακά γενετικές συσχετίσεις που αφορούν τέτοιους χαμηλής συχνότητας πολυμορφισμούς (170, 171).

1.2. ΣΚΟΠΟΣ ΤΗΣ ΔΙΑΤΡΙΒΗΣ

Σκοπός της παρούσας διατριβής είναι η εμπειρική αξιολόγηση των μελετών σάρωσης του γονιδιώματος όσον αφορά την επιδημιολογική τους εφαρμογή στη διερεύνηση της γενετικής αρχιτεκτονικής των πολυσύνθετων νοσημάτων και φαινοτύπων γενικά, αλλά και ειδικότερα στην ανακάλυψη πολυμορφισμών χαμηλής συχνότητας. Για το σκοπό αυτό γίνεται αρχικά μία εμπειρική αποτίμηση δημοσιευμένων μετα-αναλύσεων μελετών γονιδιακής σάρωσης (137) και ακολούθως μία συστηματική ανασκόπηση των πολυμορφισμών χαμηλής συχνότητας που έχουν βρεθεί σε μελέτες γονιδιακής σάρωσης (56). Τέλος, παρατίθεται μία επιδημιολογική μελέτη δείκτου-ελέγχου ή αλλιώς ασθενών-μαρτύρων (*case-control study*) για τη γενετική συσχέτιση ανάμεσα σε κοινούς πολυμορφισμούς και πολυμορφισμούς χαμηλής συχνότητας και στο μελάνωμα (172).

ΚΕΦΑΛΑΙΟ 2: ΕΙΔΙΚΟ ΜΕΡΟΣ

2.1. ΕΜΠΕΙΡΙΚΗ ΑΠΟΤΙΜΗΣΗ ΤΟΥ ΔΥΝΑΜΙΚΟΥ ΤΩΝ ΜΕΤΑ-ΑΝΑΛΥΣΕΩΝ ΜΕΛΕΤΩΝ ΓΟΝΙΔΙΑΚΗΣ ΣΑΡΩΣΗΣ

2.1.1. ΣΚΕΠΤΙΚΟ ΚΑΙ ΣΚΟΠΟΣ

Οι μελέτες γονιδιακής σάρωσης έχουν οδηγήσει στην ανακάλυψη ενός ιδιαίτερα μεγάλου αριθμού γενετικών σχέσεων με ισχυρή στατιστική σημαντικότητα (92, 103). Για παράδειγμα, μέχρι τον Ιανουάριο του 2013, ο Κατάλογος Δημοσιευμένων Μελετών Γονιδιακής Σάρωσης που διατηρείται από το Εθνικό Ινστιτούτο για την Έρευνα του Ανθρώπινου Γονιδιώματος (National Human Genome Research Institute) των ΗΠΑ (<http://www.genome.gov/gwastudies>) (89) περιλάμβανε περισσότερες από 8000 συσχετίσεις ανάμεσα σε πολυμορφισμούς του ενός νουκλεοτιδίου και σύνθετους φαινοτύπους με $P < 10^{-5}$, που προέρχονταν από σχεδόν 1500 δημοσιευμένες μελέτες γονιδιακής σάρωσης. Η πραγματοποίηση των μελετών αυτών διευκολύνθηκε σε μεγάλο βαθμό από τις τεχνολογικές εξελίξεις στην ευρείας κλίμακας γονοτύπηση, που κατέστησαν εφικτή την ακριβή και επαναλήψιμη γονοτύπηση, σε συνδυασμό με τη σταδιακή μείωση του κόστους γονοτύπησης. Ακόμη, τα αποτελέσματα του Διεθνούς Προγράμματος HarMap (*International HarMap Project*) (76-78) και του Προγράμματος 1000 Γονιδιωμάτων (*1000 Genomes Project*) (173, 174), προσέφεραν επιπλέον πληροφορίες για τη γενετική ποικιλομορφία μέσω της συστηματικής καταγραφής των κοινών αλλά και των χαμηλής συχνότητας πολυμορφισμών και μέσω του χαρακτηρισμού των προτύπων

ανισορροπίας διασύνδεσης του ανθρώπινου γονιδιώματος, πληροφορία η οποία αξιοποιείται κατά κόρον στις μελέτες γονιδιακής σάρωσης.

Η συνεισφορά των μελετών γονιδιακής σάρωσης όσον αφορά τη χαρτογράφηση γονιδίων υπήρξε εξαιρετικά σημαντική (87, 92) με την ανακάλυψη δεκάδων γενετικών σχέσεων, πολλές από τις οποίες επαληθεύθηκαν επιτυχώς σε μεταγενέστερες μελέτες (122). Ωστόσο, πρώιμες μελέτες χρησιμοποιούσαν μικρά μεγέθη δείγματος και είχαν κατά συνέπεια μικρή ισχύ για να ανακαλύψουν τα μικρά μεγέθη αποτελέσματος που αναμένονταν σύμφωνα με το μοντέλο «κοινού νοσήματος-κοινού πολυμορφισμού» (39). Τέτοιοι πολυμορφισμοί απαιτούν μεγάλα μεγέθη δείγματος, ειδικά όταν η συχνότητά τους είναι μικρή (56). Κατά συνέπεια, ορισμένα ευρήματα από αρχικές μελέτες γονιδιακής σάρωσης καταρρίφθηκαν στη συνέχεια από μεταγενέστερες μελέτες (175-177). Αν και όχι ιδιαίτερα συχνή στις μελέτες γονιδιακής σάρωσης, αυτή η έλλειψη επαλήθευσης ήταν αρκετά συχνή στις μελέτες υποψηφίων-γονιδίων στόχων. Τα αποτελέσματα πολλών τέτοιων μελετών είχαν εκπληκτικά χαμηλό βαθμό επαναληψιμότητας σε μεταγενέστερες μελέτες και μετα-αναλύσεις (72, 74, 178, 179) αποκαλύπτοντας έναν μεγάλο αριθμό ψευδώς θετικών ευρημάτων (70, 180). Έτσι η μετα-ανάλυση διαθέσιμων δεδομένων από μελέτες γονιδιακής σάρωσης αναγνωρίστηκε γρήγορα ως η κατάλληλη μέθοδος για την επίτευξη επαρκών μεγεθών δείγματος και ιδανικής στατιστικής ισχύος για την ανακάλυψη γενετικών σχέσεων με μέτρια μεγέθη αποτελέσματος (136, 181-183).

Ως στατιστική μέθοδος που επιτρέπει την ποσοτική σύνθεση των αποτελεσμάτων από διαφορετικές μελέτες, προκειμένου να υπολογιστεί ένα κοινό συνοπτικό μέγεθος αποτελέσματος, η μετα-ανάλυση έχει το μείζον πλεονέκτημα της αύξησης της στατιστικής ισχύος, δηλαδή της μείωσης των ψευδώς αρνητικών (182,

184). Επιπλέον πλεονεκτήματα αποτελούν η αυξημένη ακρίβεια των εκτιμητών του αποτελέσματος των ανιχνευόμενων γενετικών σχέσεων, η μείωση της έκτασης του φαινομένου της «κατάρας του νικητή» (*winner's curse*) (δηλαδή του φαινομένου υπερεκτίμησης των πραγματικών μεγεθών των γενετικών επιδράσεων, ειδικά όταν οι αντίστοιχες μελέτες δημοσιεύονται σε υψηλής απήχησης περιοδικά (185)), και τέλος, η δυνατότητα μελέτης και μέτρησης του βαθμού συνέπειας ή ετερογένειας των γενετικών επιδράσεων στις διαφορετικές μελέτες που συντίθενται (186-188).

Η μετα-ανάλυση των μελετών γονιδιακής σάρωσης έχει αναδειχθεί σε μία αποτελεσματική στρατηγική για την αξιολόγηση των ευρημάτων των μελετών αυτών ως προς την περαιτέρω διερεύνησή τους (189) μέσω λειτουργικών αναλύσεων ή προσπαθειών επαλήθευσης, διαφαινοτυπικών ελέγχων ή μελετών Μενδελιανής τυχαιοποίησης (190, 191). Έτσι, η μετα-ανάλυση πολλαπλών μελετών γονιδιακής σάρωσης είναι πλέον μία καλά τεκμηριωμένη και επικυρωμένη τεχνική για την ανακάλυψη και επαλήθευση γενετικών συσχετίσεων (145, 189, 192).

Η μετα-ανάλυση έχει καταστεί ιδιαίτερα επιτυχής χάρη στην εξέλιξη των μεθόδων αναπλήρωσης μη-γονοτυπημένων πολυμορφισμών (*genotype imputation*) (193-195), αλλά και στη δημιουργία ερευνητικών συνασπισμών που διευκόλυναν την παροχή δεδομένων μεταξύ διαφορετικών μελετών. Ωστόσο, παρά τις πολυάριθμες επιτυχίες των μελετών γονιδιακής σάρωσης και των μετα-αναλύσεών τους, οι γενετικοί τόποι που έχουν ανακαλυφθεί έως τώρα εξηγούν μόνο ένα μικρό μέρος από την εκτιμώμενη συνολική κληρονομικότητα των περισσότερων κοινών νοσημάτων (112).

Εδώ, αποτιμήθηκε εμπειρικά το δυναμικό της μετα-ανάλυσης μελετών γονιδιακής σάρωσης μέσω μιας συστηματικής ανασκόπησης των χαρακτηριστικών

τους και μέσω του υπολογισμού της συνεισφοράς των ανακαλυφθέντων γενετικών τόπων στη διαμόρφωση του γενετικού υποστρώματος των αντίστοιχων κοινών φαινοτύπων ως ποσοστού της γενετικής διακύμανσης που εξηγείται από τους πολυμορφισμούς αυτούς. Τέλος, γίνεται μία περιγραφή των χαρακτηριστικών των ανακαλυφθέντων πολυμορφισμών.

2.1.2. ΜΕΘΟΔΟΙ

Για την αναζήτηση των μετα-αναλύσεων μελετών γονιδιακής σάρωσης χρησιμοποιήθηκε η ηλεκτρονική βάση «Κατάλογος Δημοσιευμένων Μελετών Γονιδιακής Σάρωσης» (<http://www.genome.gov/gwastudies/>) του Εθνικού Ινστιτούτου για την Έρευνα του Ανθρώπινου Γονιδιώματος (National Human Genome Research Institute) των ΗΠΑ (Bethesda, MD). Η τελευταία αναζήτηση έγινε στις 15 Ιουνίου 2012.

Βρέθηκαν αρχικά μελέτες γονιδιακής σάρωσης με μέγεθος δείγματος ≥ 10.000 άτομα στο στάδιο της ανακάλυψης (*discovery stage*). Το συγκεκριμένο κριτήριο χρησιμοποιήθηκε έτσι, ώστε να εντοπιστούν μελέτες γονιδιακής σάρωσης όπου οι προσπάθειες ανακάλυψης είχαν ικανοποιητική ισχύ, για να βρουν κοινούς πολυμορφισμούς με μέτρια μεγέθη αποτελέσματος (λόγος αναλογιών ανά αλληλίο > 1.20 και συχνότητα ελάσσονος αλληλίου $> 5\%$) (196). Με μικρότερα μεγέθη δείγματος πολλοί πολυμορφισμοί με υπολογίσιμα αποτελέσματα θα παραβλέπονταν και έτσι υπάρχει ακόμα αρκετή αβεβαιότητα σχετικά με την αρχιτεκτονική των γενετικών κινδύνων και σχετικά με το τι μπορεί να επιτύχει η μετα-ανάλυση επαρκών δεδομένων για τους φαινοτύπους υπό διερεύνηση. Μετά από ανασκόπηση των αντίστοιχων άρθρων επιλέχθηκαν δημοσιεύσεις όπου είχαν πραγματοποιηθεί μετα-αναλύσεις 2 ή περισσότερων μελετών γονιδιακής σάρωσης στο στάδιο ανακάλυψης, ανεξαρτήτως από το εάν ακολουθήθηκαν από στάδιο επαλήθευσης (*replication stage*). Μετα-αναλύσεις σε πληθυσμούς διαφορετικής εθνικής καταγωγής συμπεριλήφθηκαν ως διακριτές περιπτώσεις. Όταν δύο οι περισσότερες δημοσιεύσεις ήταν διαθέσιμες για τον ίδιο φαινότυπο και τον ίδιο πληθυσμό, συμπεριλήφθηκε μόνο εκείνη με το μεγαλύτερο μέγεθος δείγματος.

Από κάθε δημοσίευση και αντίστοιχη μετα-ανάλυση έγινε εξαγωγή των εξής δεδομένων: πρώτος συγγραφέας, έτος και περιοδικό δημοσίευσης, νόσος/φαινότυπος υπό μελέτη, μεγέθη δείγματος στα στάδια ανακάλυψης ή/και επαλήθευσης, εθνική καταγωγή των αντίστοιχων πληθυσμών, πλατφόρμα γονοτύπησης και αριθμός πολυμορφισμών που πληρούσαν τα κριτήρια ποιότητας, αριθμός μελετών που συντέθηκαν με μετα-ανάλυση, μέθοδος σύνθεσης που χρησιμοποιήθηκε (μοντέλο σταθερών επιδράσεων [*fixed-effects*], μοντέλο τυχαίων επιδράσεων [*random-effects*], άλλο), έλεγχος για ετερογένεια, όριο στατιστικής σημαντικότητας, μέθοδοι αναπλήρωσης γονοτύπων και αριθμός αναπληρούμενων γονοτύπων, κριτήρια ποιότητας, κριτήρια για μεταφορά πολυμορφισμών στο στάδιο επαλήθευσης, γενίκευση σε πληθυσμούς άλλης εθνικής καταγωγής, επέκταση μελέτης των πολυμορφισμών σε άλλους φαινοτύπους, διαφαινοτυπικοί έλεγχοι (*phenotype cross-checks*), και τυχόν λειτουργικές αναλύσεις.

Επιπλέον, από κάθε μετα-ανάλυση έγινε καταγραφή των ανεξάρτητων πολυμορφισμών που ανακαλύφθηκαν και των αντίστοιχων γενετικών τόπων, χρωμοσωμικών περιοχών, ελάσσονων αλληλίων και αλληλίων αναφοράς, καθώς και των σχετικών συχνοτήτων, των μεγεθών αποτελέσματος, και των μέτρων αβεβαιότητας (τυπικά σφάλματα ή 95% διαστήματα εμπιστοσύνης). Ακολουθώς, αυτά τα μεγέθη αποτελέσματος εκφράστηκαν ως συνεισφορά των αντίστοιχων γενετικών τόπων στην ολική γενετική διακύμανση του φαινοτύπου χρησιμοποιώντας την προσέγγιση των Park et al. (197). Ο όρος γενετική διακύμανση (σ_G^2) αφορά το κομμάτι της διακύμανσης ενός φαινοτύπου (σ_P^2) σε έναν πληθυσμό που μπορεί να εξηγηθεί από γενετικούς παράγοντες (το υπόλοιπο κομμάτι της φαινοτυπικής διακύμανσης εξηγείται από περιβαλλοντικούς, δηλαδή μη-γενετικούς,

παράγοντες και ονομάζεται περιβαλλοντική διακύμανση [σ_E^2] (198). Έτσι, τα ανωτέρω μπορούν να εκφραστούν από τη σχέση

$$\sigma_P^2 = \sigma_G^2 + \sigma_E^2$$

Συνοπτικά, το μέγεθος του αποτελέσματος (*effect size, ES*) ενός γενετικού πολυμορφισμού i εκφράζεται από τον τύπο

$$ES_i = 2 \times \beta_i^2 \times f_i \times (1 - f_i)$$

όπου β_i είναι το αποτέλεσμα της παλινδρόμησης (π.χ. ο λογάριθμος του λόγου αναλογιών σε μία λογαριθμική παλινδρόμηση) ανά αντίγραφο αλληλίου και f_i είναι η συχνότητα του ελάσσονος αλληλίου (197). Εκφραζόμενο από τον ανωτέρω τύπο το μέγεθος του αποτελέσματος αντιστοιχεί στη συνεισφορά του κάθε πολυμορφισμού i στη γενετική διακύμανση του φαινοτύπου. Για k πολυμορφισμούς, το άθροισμα

$$\sum_{i=1}^k ES_i$$

αντιστοιχεί στη συνολική γενετική διακύμανση που μπορεί να εξηγηθεί από το σύνολο των πολυμορφισμών αυτών. Έτσι υπολογίστηκε το σύνολο της γενετικής διακύμανσης, που μπορεί να εξηγηθεί από τους γενετικούς τόπους που έχουν ανιχνευθεί από μετα-αναλύσεις μελετών γονιδιακής σάρωσης με συνολικό μέγεθος δείγματος ≥ 10.000 .

2.1.3. ΑΠΟΤΕΛΕΣΜΑΤΑ

Έως τις 15/06/2012, ο ανωτέρω κατάλογος περιλάμβανε 1.271 δημοσιευμένα άρθρα. Από αυτά, 88 (**Παράρτημα 1**) περιλάμβαναν συνολικά 139 μετα-αναλύσεις που πληρούσαν τα κριτήρια εισαγωγής. Οι μετα-αναλύσεις αυτές είχαν δημοσιευθεί από το 2009 έως το 2012 σε 20 διαφορετικά περιοδικά (**Πίνακας 1**). Τα περισσότερα άρθρα είχαν δημοσιευθεί στο *Nature Genetics* (44%), ακολουθούμενο από τα περιοδικά *PLoS Genetics* (14%), *Circulation: Cardiovascular Genetics* (7%), *Human Molecular Genetics* (6%), *JAMA* (5%) και άλλα (24%). Από τις 139 μετα-αναλύσεις, 106 (76%) αφορούσαν σε νοσήματα και 33 (24%) σε ποσοτικούς φαινοτύπους. Το διάμεσο μέγεθος δείγματος για τις μετα-αναλύσεις αυτές ήταν 20.611 (ενδοτεταρτημοριακό εύρος, 14.392-38.238). Η πλειοψηφία (n=109) αφορούσε πληθυσμούς ευρωπαϊκής καταγωγής, με τις υπόλοιπες να αφορούν πληθυσμούς ασιατικής (n=20), αφρικανικής (n=2) ή μικτής καταγωγής (n=8). Επιπρόσθετες προσπάθειες επαλήθευσης είχαν πραγματοποιηθεί για τα ευρήματα 101 μετα-αναλύσεων. Το διάμεσο μέγεθος δείγματος για αυτές τις περιπτώσεις ήταν 16.378 (ενδοτεταρτημοριακό εύρος, 8.112-48.607). Από αυτές, 81 είχαν πραγματοποιηθεί σε πληθυσμούς ευρωπαϊκής καταγωγής και μόνο 20 είχαν πραγματοποιηθεί σε μη-ευρωπαϊκούς πληθυσμούς.

Πίνακας 1. Περιγραφή των 139 μετα-αναλύσεων μελετών γονιδιακής σάρωσης με μέγεθος δείγματος ≥10.000

Φαινότυπος *	Στάδιο ανακάλυψης		Στάδιο επαλήθευσης	
	Μέγεθος δείγματος **	Καταγωγή	Μέγεθος δείγματος **	Καταγωγή
Δοκιμασία ανοχής γλυκόζης (S63)	15.234	EYP	30.620	EYP
Επίπεδα αδιπονεκτίνης (S12)	35.355	EYP, ΑΣ, ΑΦΡ	M/E	M/E
Εναπόθεση λίπους: % λίπους σώματος (S38)	36.626	EYP, ΑΣ	39.576	EYP
Εναπόθεση λίπους: περιφέρεια μέσης (S43)	38.580	EYP	102.064	EYP
Γήρας: χρόνος θανάτου (S83)	25.007	EYP	10.411	EYP
Γήρας: χρόνος συμβάματος (S83)	16.995	EYP	M/E	M/E
Αμινοτρανσφεράση της αλανίνης (S10)	45.596	EYP, ΑΣ	8.112	EYP
Αμινοτρανσφεράση της αλανίνης (S33)	13.696	ΑΣ	M/E	M/E
Κατανάλωση αλκοόλ (S65)	26.316	EYP	21.185	EYP
Αλκαλική φωσφατάση (S10)	56.415	EYP, ΑΣ	8112	EYP
Αλλεργική ρινίτιδα (S60)	12.898 (3.933/8.965)	EYP	M/E	M/E
Νόσος Alzheimer (S28)	20.373 (6.688/13.685)	EYP	39.343 (13.182/26.161)	EYP
Σφυροβραχιόνιος δείκτης (S48)	41.692	EYP	16.717	EYP
Αγκυλοποιητική σπονδυλίτιδα (S21)	11.802 (3.023/8.779)	EYP	6.594 (2.111/4.483)	EYP
Αορτική σκληρία (S46)	20.634	EYP	5.306	EYP
Ασπαρτική αμινοτρανσφεράση (S33)	13.898	ΑΣ	M/E	M/E
Άσθμα (S47)	26.475 (10.365/16.110)	EYP	M/E	M/E
Ατοπική δερματίτιδα (S53)	26.171 (5.606/20.565)	EYP	25.252 (5.419/19.833)	EYP
Κολπική μαρμαρυγή (S4)	40.518 (3.413/37.105)	EYP	6.218 (2.145/4.073)	EYP
Διπολική διαταραχή (S64)	16.731 (7.481/9.250)	EYP	46.918 (4.496/42.422)	EYP
Βάρος γέννησης (S23)	10.623	EYP	27.591	EYP
Άζωτο ουρίας αίματος (S33)	14.347	ΑΣ	M/E	M/E

Δείκτης μάζας σώματος (S73)	123.865	EYP	125.931	EYP
Δείκτης μάζας σώματος (S85)	27.715	ΑΣ	55.333	ΑΣ
Οστική πυκνότητα κεφαλής μηριαίου (S19)	32.961	EYP, ΑΣ	50.933	EYP, ΑΣ
Οστική πυκνότητα οσφυικής μοίρας (S19)	31.800	EYP, ΑΣ	50.933	EYP, ΑΣ
Περιφέρεια βραχίονα(S7)	18.753	EYP	3.623	EYP
Κατανάλωση καφεΐνης (S11)	47.431	EYP	M/E	M/E
Επίπεδα ασβεστίου (S50)	20.611	EYP	M/E	M/E
Χρόνια νεφρική νόσος (S55)	68.678 (6.271/62.407)	EYP	64.356	EYP, ΑΦΡ
Παράγοντας πήξης VII (S67)	15422	EYP	3678	EYP
Παράγοντας πήξης VIII (S67)	15279	EYP	2381	EYP
Κατανάλωση καφέ (S1)	18.176	EYP	7.929	EYP
Πάχος έσω-μέσου χιτώνα κοινής καρωτίδας (S5)	31.211	EYP	10.403	EYP
Στεφανιαία νόσος (S66)	86.995 (22.233/64.762)	EYP	56.508 (28.834/27.674)	EYP
C-αντιδρώσα πρωτεΐνη (S14)	66.185	EYP	16.540	EYP
Ρυθμός σπειραματικής διήθησης βάσει κρεατινίνης (S55)	74.354	EYP	64.356	EYP, ΑΦΡ
Νόσος Crohn (S22)	21.389 (6.333/15.056)	EYP	29.720 (15.694/14.026)	EYP
Ρυθμός σπειραματικής διήθησης βάσει συστατίνης (S55)	22.937	EYP	3.709	EYP
D-διμερή (S69)	21.052	EYP	M/E	M/E
Δι-υδροεπιανδροστερόνη θειϊκή (S88)	14.846	EYP	M/E	M/E
Διαστολική αρτηριακή πίεση (S31)	69.395	EYP	133.361	EYP
Διαστολική αρτηριακή πίεση (S35)	19.608	ΑΣ	30.765	ΑΣ
Γλυκόζη νύστεως (S16)	46.186	EYP	76.558	EYP
Ινσουλίνη νύστεως (S16)	38.238	EYP	76.558	EYP
Ινωδογόνο (S15)	22.096	EYP	17.686	EYP
γ-GT (S10)	61.089	EYP, ΑΣ	8112	EYP
γ-GT (S33)	10.090	ΑΣ	M/E	M/E

Γλίωμα (S62)	11.582 (4.147/7.435)	EYP	M/E	M/E
Γλυκοζυλιωμένη αιμοσφαιρίνη (S71)	46.368	EYP	10448	EYP
Ουρική αρθρίτιδα (S86)	28.283	EYP	M/E	M/E
HDL χοληστερόλη (S78)	99.900	EYP	51.875	EYP, ΑΣ, ΑΦΡ
Περιφέρεια κεφαλής (νεογνά) (S77)	10.768	EYP	8.321	EYP
Καρδιακή ανεπάρκεια (S68)	20.926	EYP	M/E	M/E
Ύψος (S41)	133.653	EYP	50.074	EYP
Ύψος (S49)	20.427	ΑΦΡ	16.436	ΑΦΡ
Ύψος (S51)	19.633	ΑΣ	M/E	M/E
Αιματοκρίτης (S24)	24.731	EYP	9.456	EYP
Αιματοκρίτης (S33)	14.395	ΑΣ	M/E	M/E
Αιμοσφαιρίνη (S24)	24.731	EYP	9.456	EYP
Αιμοσφαιρίνη (S33)	14.402	ΑΣ	M/E	M/E
HOMA-B (S16)	36.466	EYP	76.558	EYP
HOMA-IR (S16)	37.037	EYP	76.558	EYP
IgE ευαισθητοποίηση σε χόρτο (S60)	12.347 (2.315/10.032)	EYP	M/E	M/E
IGF-1 (S34)	10.280	EYP	M/E	M/E
IGFBP-3 (S34)	10.018	EYP	M/E	M/E
Πάχος έσω-μέσου χιτώνα έσω καρωτίδας (S5)	10.962	EYP	M/E	M/E
Ενδοκράνιο ανεύρυσμα (S87)	15.295 (2.780/12.515)	EYP	4.777 (3.111/1.666)	ΑΣ
LDL χοληστερόλη (S78)	95.454	EYP	51.875	EYP, ΑΣ, ΑΦΡ
Εσωτερικές διαστάσεις αριστερής κοιλίας κατά τη διαστολή (S80)	12.612	EYP	4.094	EYP
Συστολική δυσλειτουργία αριστερής κοιλίας (S80)	12.612	EYP	4.094	EYP
Πάχος τοιχώματος αριστερής κοιλίας (S80)	12.612	EYP	4.094	EYP
Μάζα αριστερής κοιλίας (S80)	12.612	EYP	4.094	EYP
Λιποπρωτεϊνική φωσφολιπάση A2: δραστηριότητα (S26)	11.664	EYP	M/E	M/E

Λιποπρωτεϊνική φωσφολιπάση A2: μάζα (S26)	12.126	EYP	M/E	M/E
Επίπεδα μαγνησίου (S45)	15.366	EYP	8.463	EYP
Μείζονες διαταραχές συναισθήματος (S44)	15.754 (6.686/9.068)	EYP	4.677 (1.930/2.747)	EYP
Μέση αρτηριακή πίεση (S82)	74.064	EYP	48.607	EYP
Μέση περιεκτικότητα αιμοσφαιρίνης (S24)	13.329	EYP	9.456	EYP
Μέση περιεκτικότητα αιμοσφαιρίνης (S33)	14.362	ΑΣ	M/E	M/E
Μέση πυκνότητα αιμοσφαιρίνης (S24)	13.329	EYP	9.456	EYP
Μέση πυκνότητα αιμοσφαιρίνης (S33)	14.377	ΑΣ	M/E	M/E
Μέσος όγκος ερυθρών (S24)	21.456	EYP	9.456	EYP
Μέσος όγκος ερυθρών (S33)	14.364	ΑΣ	M/E	M/E
Μέσος όγκος αιμοπεταλίων (S25)	18.600	EYP	18.838	EYP
Εμμηναρχή (ηλικία έναρξης) (S18)	87.802	EYP	14.731	EYP
Εμμηνόπαυση (ηλικία έναρξης) (S75)	38.968	EYP	14.435	EYP
Μεταβολικό σύνδρομο (S40)	10.564 (2.637/7.927)	EYP	M/E	M/E
Ημικρανία (S42)	10.980 (2.446/8.534)	EYP	8.731 (2.957/5.774)	EYP
Πολλαπλή σκλήρυνση (S54)	17.698 (5.545/12.153)	EYP	M/E	M/E
Παχυσαρκία (παιδική) (S9)	13.848 (5.530/8.318)	EYP	6.901 (2.818/4.083)	EYP
Οστεοαρθρίτιδα (γόνατος) (S20)	38.280 (2.371/35.909)	EYP	15.296 (5.521/9.775)	EYP, ΑΣ
Νόσος Parkinson (S32)	17.352 (5.333/12.019)	EYP	16.060 (7.053/9.007)	EYP
Τύποι προσωπικότητας (S13)	17.375	EYP	3.294	EYP
Επίπεδα φωσφόρου (S36)	16.264	EYP	5.444	EYP
Πλάκα (αθηρωματική) καρωτίδας (S5)	25.179	EYP	6.013	EYP
Αριθμός αιμοπεταλίων (S25)	48.666	EYP	18.838	EYP
Αριθμός αιμοπεταλίων (S33)	14.806	ΑΣ	M/E	M/E
Αριθμός αιμοπεταλίων (S59)	16.388	ΑΦΡ	18.371	EYP, ΙΣΠ
Επίπεδα καλίου (S45)	13.683	EYP	8.463	EYP

Διάστημα PR (S57)	28.517	EYP	M/E	M/E
Επίπεδα προΐνσουλίνης (S76)	10.701	EYP	16.378	EYP
Πνευμονική λειτουργία (S70)	48.201	EYP	46.411	EYP
Πίεση σφυγμού (S82)	74.064	EYP	48.607	EYP
Διάρκεια QRS (S72)	40.407	EYP	7.170	EYP
Διάστημα QT (S56)	15.842	EYP	13.602	EYP
Αριθμός ερυθρών αιμοσφαιρίων (S24)	13.329	EYP	9.456	EYP
Αριθμός ερυθρών αιμοσφαιρίων (S33)	14.392	ΑΣ	M/E	M/E
Νεφροκυτταρικό καρκίνωμα (S58)	12.277 (3.772/8.505)	EYP	7.116 (2.198/4.918)	EYP
Διάστημα RR (S17)	38.991	EYP	M/E	M/E
Διάμετρος αρτηριολίων αμφιβληστροειδούς (S30)	15.358	EYP	6.673	EYP
Διάμετρος φλεβιδίων αμφιβληστροειδούς (S30)	15.358	EYP	6.673	EYP
Ρευματοειδής αρθρίτιδα (S52)	20.965 (4.074/16.891)	ΑΣ	52.669 (10.816/41.853)	EYP, ΑΣ
Ρευματοειδής αρθρίτιδα (S74)	25.708 (5.539/20.169)	EYP	15.574 (6.768/8.806)	EYP
Κρεατινίνη ορού (S33)	14345	ΑΣ	M/E	M/E
Κάπνισμα - Ηλικία έναρξης (S79)	22.438	EYP	M/E	M/E
Κάπνισμα - Αριθμός τσιγάρων ημερησίως (S79)	38.181	EYP	35.672	EYP
Κάπνισμα - καπνιστής έναντι μη-καπνιστής (S79)	69.409 (39.022/30.387)	EYP	68.988 (ΔΑ/ΔΑ)	EYP
Κάπνισμα - πρώην έναντι τωρινός καπνιστής (S79)	35.845 (20.619/15.226)	EYP	23.646	EYP
Επίπεδα νατρίου (S45)	11.552	EYP	8.463	EYP
Εγκεφαλικό (S29)	19.602 (3.810/15.792)	EYP	7.269 (935/6.334)	EYP, ΑΦΡ
Αιφνίδια καρδιακή ανακοπή (S3)	22.925 (1.283/21.642)	EYP	14.265 (3.119/11.146)	EYP
Συστολική αρτηριακή πίεση (S31)	69.395	EYP	133.361	EYP
Συστολική αρτηριακή πίεση (S35)	19.608	ΑΣ	30.765	ΑΣ
Ολική χοληστερόλη (S78)	100.184	EYP	51.875	EYP, ΑΣ, ΑΦΡ
Ολική πρωτεΐνη (S33)	10.047	ΑΣ	M/E	M/E

Τριγλυκερίδια (S78)	96.598	EYP	51.875	EYP, ΑΣ, ΑΦΡ
Σακχαρώδης διαβήτης τύπου 1 (S8)	26.890 (9.934/16.956)	EYP	3.360 (1.120/2.240)	EYP
Σακχαρώδης διαβήτης τύπου 2 (S39)	20.019 (5.561/14.458)	ΑΣ	85.685 (21.300/64.385)	EYP, ΑΣ
Σακχαρώδης διαβήτης τύπου 2 (S81)	47.117 (8.130/38.987)	EYP	94.337 (34.412/59.925)	EYP
Ελκώδης κολίτιδα (S2)	26.405 (6.687/19.718)	EYP	22.545 (9.628/12.917)	EYP
Επίπεδα ουρικού (S86)	28.283	EYP	M/E	M/E
Απέκκριση αλβουμίνης στα ούρα (S6)	31.580	EYP	38.258	EYP, ΑΦΡ
Καρκίνος ουροδόχου κύστης (S37)	41.199 (1.889/39.310)	EYP	8.650 (2.691/5.959)	EYP
Ανεπάρκεια βιταμίνης D (S84)	16.125	EYP	17.871	EYP
Παράγοντας von Willebrand (S67)	17596	EYP	6307	EYP
Λόγος περιφέρειας μέσης-ισχίου (S27)	77.167	EYP	113.636	EYP
Αριθμός λευκών αιμοσφαιρίων (S33)	14.677	ΑΣ	M/E	M/E
Αριθμός λευκών αιμοσφαιρίων (S61)	16.388	ΑΦΡ	37.827	EYP, ΑΣ, ΙΣΠ

Συντομογραφίες: ΑΦΡ: Αφρικανική, ΑΣ: Ασιατική, EYP: Ευρωπαϊκή, ΙΣΠ: Ισπανική (Λατινική Αμερική), HOMA-B: homeostatic model assessment beta cell function (ομοιοστατικό μοντέλο εκτίμησης λειτουργίας β-κυττάρων), HOMA-IR: homeostatic model assessment insulin resistance (ομοιοστατικό μοντέλο εκτίμησης ινσουλινο-αντίστασης), M/E: μη-εφαρμόσιμο γιατί δεν έγινε επαλήθευση, γ-GT: γ-γλουταμυλ-τρανσπεπτιδάση, ΔΑ: δεν αναφέρεται

* Για βιβλιογραφικές παραπομπές, βλ. Παράρτημα 1.

** Το μέγεθος δείγματος για τα στάδια ανακάλυψης και επαλήθευσης παρουσιάζεται ως συνολικός αριθμός συμμετεχόντων για συνεχείς φαινοτύπους και ως συνολικός αριθμός συμμετεχόντων (αριθμός δεικτών/αριθμός μαρτύρων) για διχότομες εκβάσεις.

Όπως φαίνεται στον **Πίνακα 2**, για όλες τις μετα-αναλύσεις εκτός από 2, οι επιμέρους μελέτες είχαν πραγματοποιήσει αναπλήρωση μη-γονοτυπημένων πολυμορφισμών. Η μέθοδος αναπλήρωσης χρησιμοποίησε διάφορους τύπους κρυπτών μοντέλων Markov (*hidden Markov models*), ανάλογα με το λογισμικό αναπλήρωσης. Επιπλέον, σε όλες σχεδόν τις μετα-αναλύσεις οι αντίστοιχες μελέτες χρησιμοποίησαν τα δεδομένα του προγράμματος HarMap για την αναπλήρωση των ελλειπόντων γονοτύπων, με εξαίρεση 2 μετα-αναλύσεις που χρησιμοποίησαν δεδομένα από το πρόγραμμα των 1000 Γονιδιωμάτων και μία επιπλέον μετα-ανάλυση που χρησιμοποίησε δεδομένα και από τα δύο προγράμματα. Οι εμπορικές πλατφόρμες της *Affymetrix* και της *Illumina* ήταν οι πιο συχνά χρησιμοποιούμενες πλατφόρμες γονοτύπησης, με τις πλατφόρμες της *Perlegen* να έχουν χρησιμοποιηθεί σε 23 μετα-αναλύσεις. Ο διάμεσος αριθμός πρωτογενών μελετών που συνδυάστηκαν μέσω μετα-ανάλυσης ήταν 9 (ενδοτεταρτημοριακό εύρος, 5-15), ενώ ο ελάχιστος αριθμός ήταν 2 και ο μέγιστος 46. Ένα μοντέλο σταθερών επιδράσεων (*fixed-effects model*) είχε χρησιμοποιηθεί ως η κύρια μέθοδος σύνθεσης των επιμέρους μελετών γονιδιακής σάρωσης και στις 139 μετα-αναλύσεις, ενώ επιπρόσθετα μοντέλα είχαν χρησιμοποιηθεί σε συνολικά 16 μετα-αναλύσεις και αφορούσαν μοντέλα τυχαίων επιδράσεων (*random-effects models*) [$n=13$], Μπεϋζιανές μεθόδους ($n=1$), μοντέλα τυχαίων επιδράσεων και Μπεϋζιανές μεθόδους ($n=1$), και μεθόδους που βασίζονται στη σύνθεση των τιμών P ($n=1$). Η μέθοδος στάθμισης αναφερόταν σε 137 μετα-αναλύσεις, από τις οποίες 127 είχαν χρησιμοποιήσει στάθμιση με βάση το αντίστροφο της διακύμανσης (*inverse-variance weighting*), 7 είχαν χρησιμοποιήσει στάθμιση με βάση το μέγεθος δείγματος, 1 είχε χρησιμοποιήσει και τις δύο μεθόδους, και 1 χρησιμοποίησε τη

μέθοδο Mantel-Haenzel. Ο διάμεσος αριθμός πολυμορφισμών που είχαν ανακαλυφθεί, δηλαδή που περνούσαν το αντίστοιχο όριο στατιστικής ή άλλης σημαντικότητας, ήταν 4 (ενδοτεταρτημοριακό εύρος, 2-9). Αξιοσημείωτο είναι ότι 18 μετα-αναλύσεις δεν είχαν κάνει καμία ανακάλυψη. Ο μέγιστος αριθμός ανακαλυφθέντων πολυμορφισμών ήταν 115 (S41).

Πίνακας 2. Πλατφόρμες γονοτύπησης, αριθμός μελετών, μέθοδοι μετα-ανάλυσης και μέθοδοι αναπλήρωσης ελλειπόντων γονοτύπων

Φαινότυπος *	Πλατφόρμες γονοτύπησης	SNPs που πληρούν έλεγχο ποιότητας	Μελέτες	Μέθοδοι μετα-ανάλυσης	Αριθμός SNPs **	Αναπληρούμενα SNPs **
Δοκιμασία ανοχής γλυκόζης (S63)	Affymetrix, Illumina	Δ/Α	9	F	3	Δ/Α
Επίπεδα αδιπονεκτίνης (S12)	Affymetrix, Illumina	Δ/Α	18	F, R, Μπεϋζιανές	10	Δ/Α
Εναπόθεση λίπους: % λίπους σώματος (S38)	Affymetrix, Illumina, Perlegen	~2.5 εκατομ.	15	F	2	Δ/Α
Εναπόθεση λίπους: περιφέρεια μέσης (S43)	Affymetrix, Illumina	2,573,738	16	F	0	~2.5 εκατομ.
Γήρας: χρόνος θανάτου (S83)	Affymetrix, Illumina	~2.5 εκατομ.	9	F	0	~2.5 εκατομ.
Γήρας: χρόνος συμβάματος (S83)	Affymetrix, Illumina	~2.5 εκατομ.	9	F	0	~2.5 εκατομ.
Αμινοτρανσφεράση της αλανίνης (S10)	Affymetrix, Illumina, Perlegen	~2.6 εκατομ.	14	F	4	Δ/Α
Αμινοτρανσφεράση της αλανίνης (S33)	Illumina	566,430	10	F	2	Δ/Α
Κατανάλωση αλκοόλ (S65)	Affymetrix, Illumina, Perlegen	~2.5 εκατομ.	12	F	0	Δ/Α
Αλκαλική φωσφατάση (S10)	Affymetrix, Illumina, Perlegen	~2.6 εκατομ.	19	F	14	Δ/Α
Αλλεργική ρινίτιδα (S60)	Affymetrix, Illumina	2,217,510	6	F	1	~2.2 εκατομ.
Νόσος Alzheimer (S28)	Affymetrix, Illumina	496,763	4	F, R	6	0.5 εκατομ.
Σφυροβραχιόνιος δείκτης (S48)	Affymetrix, Illumina, Perlegen	~2.5 εκατομ.	21	F	1	~2.5 εκατομ.
Αγκυλοποιητική σπονδυλίτιδα (S21)	Illumina	2,223,620	2	F	7	Δ/Α
Αορτική σκληρία (S46)	Affymetrix, Illumina	~2.5 εκατομ.	9	F	1	~2.5 εκατομ.
Ασπαρτική αμινοτρανσφεράση (S33)	Illumina	566,430	10	F	2	Δ/Α
Άσθμα (S47)	Illumina	582,892	23	F, R	6	Δ/Α
Ατοπική δερματίτιδα (S53)	Affymetrix, Illumina	~2.5 εκατομ.	16	F, R	2	Δ/Α
Κολπική μαρμαρυγή (S4)	Affymetrix, Illumina	~2.5 εκατομ.	5	F	2	~2.5 εκατομ.
Διπολική διαταραχή (S64)	Affymetrix, Illumina	2,415,422	11	F, R	4	~2.5 εκατομ.
Βάρος γέννησης (S23)	Affymetrix, Illumina	~2.4 εκατομ.	6	F	2	~2.4 εκατομ.

Άζωτο ουρίας αίματος (S33)	Illumina	566,430	10	F	3	Δ/Α
Δείκτης μάζας σώματος (S73)	Affymetrix, Illumina, Perlegen	~2.8 εκατομ.	46	F	19	Δ/Α
Δείκτης μάζας σώματος (S85)	Affymetrix, Illumina	2,474,474	8	F, βάεις τιμών <i>P</i>	2	Δ/Α
Οστική πυκνότητα κεφαλής μηριαίου (S19)	Affymetrix, Illumina	~2.5 εκατομ.	17	F	22	~2.5 εκατομ.
Οστική πυκνότητα οσφυικής μοίρας (S19)	Affymetrix, Illumina	~2.5 εκατομ.	16	F	27	~2.5 εκατομ.
Περιφέρεια βραχίονα(S7)	Affymetrix, Illumina	2,238,430	14	F	0	Δ/Α
Κατανάλωση καφεΐνης (S11)	Affymetrix, Illumina	~2.5 εκατομ.	10	F	2	~2.5 εκατομ.
Επίπεδα ασβεστίου (S50)	Affymetrix, Illumina	~2.5 εκατομ.	6	F	1	~2.5 εκατομ.
Χρόνια νεφρική νόσος (S55)	Affymetrix, Illumina	~2.5 εκατομ.	25	F	2	~2.5 εκατομ.
Παράγοντας πήξης VII (S67)	Affymetrix, Illumina	~2.6 εκατομ.	4	F	5	~2.6 εκατομ.
Παράγοντας πήξης VIII (S67)	Affymetrix, Illumina	~2.6 εκατομ.	3	F	5	~2.6 εκατομ.
Κατανάλωση καφέ (S1)	Affymetrix, Illumina, Perlegen	~2.6 εκατομ.	10	F	1	Δ/Α
Πάχος έσω-μέσου χιτώνα κοινής καρωτίδας (S5)	Affymetrix, Illumina	~2.5 εκατομ.	9	F	1	~2.5 εκατομ.
Στεφανιαία νόσος (S66)	Affymetrix, Illumina	~2.3 εκατομ.	14	F, R	13	~2.3 εκατομ.
C-αντιδρώσα πρωτεΐνη (S14)	Affymetrix, Illumina	~2.5 εκατομ.	15	F	17	~2.5 εκατομ.
Ρυθμός σπειραματικής διήθησης βάσει κρεατινίνης (S55)	Affymetrix, Illumina	~2.5 εκατομ.	26	F	27	~2.5 εκατομ.
Νόσος Crohn (S22)	Affymetrix, Illumina	953,241	6	F	52	~1 εκατομ.
Ρυθμός σπειραματικής διήθησης βάσει συστατίνης (S55)	Affymetrix, Illumina	~2.5 εκατομ.	26	F	3	~2.5 εκατομ.
D-διμερή (S69)	Affymetrix, Illumina	2,522,393	13	F	3	~2.6 εκατομ.
Διυδροεπιανδροστερόνη θειϊκή (S88)	Affymetrix, Illumina	~2.5 εκατομ.	7	F	5	~2.5 εκατομ.
Διαστολική αρτηριακή πίεση (S31)	Affymetrix, Illumina	~2.5 εκατομ.	29	F	9	Δ/Α
Διαστολική αρτηριακή πίεση (S35)	Affymetrix, Illumina	~1.7 εκατομ.	8	F	1	Δ/Α
Γλυκόζη νύστεως (S16)	Affymetrix, Illumina	~2.5 εκατομ.	21	F	12	Δ/Α
Ινσουλίνη νύστεως (S16)	Affymetrix, Illumina	~2.5 εκατομ.	21	F	0	Δ/Α
Ινωδογόνο (S15)	Affymetrix, Illumina	~2,661,766	6	F	4	~2.5 εκατομ.
γ-GT (S10)	Affymetrix, Illumina, Perlegen	~2.6 εκατομ.	21	F	26	Δ/Α
γ-GT (S33)	Illumina	566,430	10	F	2	Δ/Α
Γλίωμα (S62)	Illumina	424,460	4	F	6	Δ/Α

Γλυκοζυλιωμένη Αιμοσφαιρίνη (S71)	Affymetrix, Illumina	~2.5 εκατομ.	23	F	10	Δ/Α
Ουρική αρθρίτιδα (S86)	Affymetrix, Illumina	~2.5 εκατομ	5	F	2	~2.5 εκατομ.
HDL χοληστερόλη (S78)	Affymetrix, Illumina, Perlegen	~2.6 εκατομ.	46	F	47	Δ/Α
Περιφέρεια κεφαλής (νεογνά) (S77)	Affymetrix, Illumina	~2.4 εκατομ.	7	F	3	~2.4 εκατομ.
Καρδιακή ανεπάρκεια (S68)	Affymetrix, Illumina	2,478,304	4	F	1	~2.5 εκατομ.
Ύψος (S41)	Affymetrix, Illumina, Perlegen	2,834,208	46	F	115	~2.8 εκατομ.
Ύψος (S49)	Affymetrix, Illumina	3,310,998	9	F	2	Δ/Α
Ύψος (S51)	Illumina	420,885	23	F	8	Δ/Α
Αιματοκρίτης (S24)	Affymetrix, Illumina	~2.5 εκατομ.	6	F	8	~2.5 εκατομ.
Αιματοκρίτης (S33)	Illumina	566,430	10	F	2	Δ/Α
Αιμοσφαιρίνη (S24)	Affymetrix, Illumina	~2.5 εκατομ.	6	F	7	~2.5 εκατομ.
Αιμοσφαιρίνη (S33)	Illumina	566,430	10	F	2	Δ/Α
HOMA-B (S16)	Affymetrix, Illumina	~2.5 εκατομ.	21	F	3	Δ/Α
HOMA-IR (S16)	Affymetrix, Illumina	~2.5 εκατομ.	21	F	0	Δ/Α
IgE ευαισθητοποίηση σε χόρτο (S60)	Affymetrix, Illumina	2,217,510	6	F	3	~2.2 εκατομ.
IGF-1 (S34)	Affymetrix, Illumina	~2.5 εκατομ.	4	F	1	~2.5 εκατομ.
IGFBP-3 (S34)	Affymetrix, Illumina	~2.5 εκατομ.	4	F	4	~2.5 εκατομ.
Πάχος έσω-μέσου χιτώνα έσω καρωτίδας (S5)	Affymetrix, Illumina	~2.5 εκατομ	3	F	0	~2.5 εκατομ.
Ενδοκράνιο ανεύρυσμα (S87)	Illumina	831,534	4	F, Μπεϋζιανές	6	Δ/Α
LDL χοληστερόλη(S78)	Affymetrix, Illumina, Perlegen	~2.6 εκατομ.	46	F	37	Δ/Α
Εσωτερικές διαστάσεις αριστερής κοιλίας στη διαστολή (S80)	Affymetrix, Illumina	~2.5 εκατομ.	5	F	1	~2.5 εκατομ.
Συστολική δυσλειτουργία αριστερής κοιλίας (S80)	Affymetrix, Illumina	~2.5 εκατομ.	5	F	1	~2.5 εκατομ.
Πάχος τοιχώματος αριστερής κοιλίας (S80)	Affymetrix, Illumina	~2.5 εκατομ.	5	F	3	~2.5 εκατομ.
Μάζα αριστερής κοιλίας (S80)	Affymetrix, Illumina	~2.5 εκατομ.	5	F	3	~2.5 εκατομ.
Λιποπρωτεϊνική φωσφολιπάση A2: δραστηριότητα (S26)	Affymetrix, Illumina	2,661,766	3	F	6	~2.5 εκατομ.
Λιποπρωτεϊνική φωσφολιπάση A2: μάζα (S26)	Affymetrix, Illumina	2,661,766	4	F	2	~2.5 εκατομ.
Επίπεδα μαγνησίου (S45)	Affymetrix, Illumina	~2.5 εκατομ.	3	F	6	~2.5 εκατομ.
Μείζονες διαταραχές συναισθήματος (S44)	Affymetrix, Illumina, Perlegen	~2.1 εκατομ.	5	F, R	1	~2.1 εκατομ.

Μέση αρτηριακή πίεση (S82)	Affymetrix, Illumina, Perlegen	Δ/Α	35	F	9	~0.5 εκατομ.
Μέση περιεκτικότητα αιμοσφαιρίνης (S24)	Affymetrix, Illumina	~2.5 εκατομ.	4	F	9	~2.5 εκατομ.
Μέση περιεκτικότητα αιμοσφαιρίνης (S33)	Illumina	566,430	10	F	15	Δ/Α
Μέση πυκνότητα αιμοσφαιρίνης (S24)	Affymetrix, Illumina	~2.5 εκατομ.	4	F	2	~2.5 εκατομ.
Μέση πυκνότητα αιμοσφαιρίνης (S33)	Illumina	566,430	10	F	7	Δ/Α
Μέσος όγκος ερυθρών (S24)	Affymetrix, Illumina	~2.5 εκατομ.	5	F	16	~2.5 εκατομ.
Μέσος όγκος ερυθρών (S33)	Illumina	566,430	10	F	15	Δ/Α
Μέσος όγκος αιμοπεταλίων (S25)	Affymetrix, Illumina	~2.7 εκατομ.	13	F	26	Δ/Α
Εμμηναρχή (ηλικία έναρξης) (S18)	Affymetrix, Illumina	~2.5 εκατομ.	32	F	42	~2.5 εκατομ.
Εμμηνόπαυση (ηλικία έναρξης) (S75)	Affymetrix, Illumina	2,551,160	22	F	20	~2.5 εκατομ.
Μεταβολικό σύνδρομο (S40)	Illumina	1,257,079	4	F	1	~0.9 εκατομ.
Ημικρανία (S42)	Affymetrix, Illumina, Perlegen	~2.5 εκατομ.	6	F	0	~2.5 εκατομ.
Πολλαπλή σκλήρυνση (S54)	Affymetrix, Illumina	2,529,394	7	F, R	9	~2.5 εκατομ.
Παχυσαρκία (παιδική) (S9)	Affymetrix, Illumina	~2.7 εκατομ.	14	F	7	~2.54 εκατομ.
Οστεοαρθρίτιδα (knee) (S20)	Affymetrix, Illumina	2,335,627	4	F	0	Δ/Α
Νόσος Parkinson (S32)	Illumina	7,689,524	5	F, R	11	Δ/Α
Τύποι προσωπικότητας (S13)	Affymetrix, Illumina, Perlegen	~2.5 εκατομ.	10	F	2	~2.5 εκατομ.
Επίπεδα φωσφόρου (S36)	Affymetrix, Illumina	~2.5 εκατομ.	4	F	7	~2.5 εκατομ.
Πλάκα (αθηρωματική) καρωτίδας (S5)	Affymetrix, Illumina	~2.5 εκατομ.	7	F	3	~2.5 εκατομ.
Αριθμός αιμοπεταλίων (S25)	Affymetrix, Illumina	~2.7 εκατομ.	23	F	39	Δ/Α
Αριθμός αιμοπεταλίων (S33)	Illumina	566,430	10	F	6	Δ/Α
Αριθμός αιμοπεταλίων (S59)	Affymetrix, Illumina	~2.2 εκατομ.	7	F	15	~2.2 εκατομ.
Επίπεδα καλίου (S45)	Affymetrix, Illumina	~2.5 εκατομ.	4	F	0	~2.5 εκατομ.
Διάστημα PR (S57)	Affymetrix, Illumina	~2.5 εκατομ.	8	F	9	~2.5 εκατομ.
Επίπεδα προΐνσουλίνης (S76)	Affymetrix, Illumina	~2.5 εκατομ.	4	F	7	~2.5 εκατομ.
Πνευμονική λειτουργία (S70)	Affymetrix, Illumina	~2.5 εκατομ.	23	F	8	Δ/Α
Πίεση σφυγμού (S82)	Affymetrix, Illumina, Perlegen	Δ/Α	35	F	6	~0.5 εκατομ.
Διάρκεια QRS (S72)	Affymetrix, Illumina	~2.5 εκατομ.	14	F	25	~2.5 εκατομ.

Διάστημα QT (S56)	Affymetrix, Illumina	Δ/Α	5	F	12	~2.5 εκατομ.
Αριθμός ερυθρών αιμοσφαιρίων (S24)	Affymetrix, Illumina	~2.5 εκατομ.	4	F	2	~2.5 εκατομ.
Αριθμός ερυθρών αιμοσφαιρίων (S33)	Illumina	566,430	10	F	8	Δ/Α
Νεφροκυτταρικό καρκίνωμα (S58)	Illumina	586,069	2	F	6	Δ/Α
Διάστημα RR (S17)	Affymetrix, Illumina	~2.5 εκατομ.	15	F	9	~2.5 εκατομ.
Διάμετρος αρτηριολίων αμφιβληστροειδούς (S30)	Affymetrix, Illumina	~2.2 εκατομ.	4	F	0	~2.2 εκατομ.
Διάμετρος φλεβιδίων αμφιβληστροειδούς (S30)	Affymetrix, Illumina	~2.2 εκατομ.	4	F	5	~2.2 εκατομ.
Ρευματοειδής αρθρίτιδα (S52)	Affymetrix, Illumina	1,948,139	3	F	7	~2 εκατομ.
Ρευματοειδής αρθρίτιδα (S74)	Affymetrix, Illumina	~2,716,259	6	F	6	Δ/Α
Κρεατινίνη ορού (S33)	Illumina	566,430	10	F	1	Δ/Α
Κάπνισμα - ηλικία έναρξης (S79)	Affymetrix, Illumina, Perlegen	~2.5 εκατομ.	13	F, R	0	~2.5 εκατομ.
Κάπνισμα - αριθμός τσιγάρων ημερησίως (S79)	Affymetrix, Illumina, Perlegen	~2.5 εκατομ.	13	F, R	1	~2.5 εκατομ.
Κάπνισμα - καπνιστής έναντι μη-καπνιστής (S79)	Affymetrix, Illumina, Perlegen	~2.5 εκατομ.	16	F, R	0	~2.5 εκατομ.
Κάπνισμα - πρώην έναντι τωρινός καπνιστής (S79)	Affymetrix, Illumina, Perlegen	~2.5 εκατομ.	14	F, R	0	~2.5 εκατομ.
Επίπεδα νατρίου (S45)	Affymetrix, Illumina	~2.5 εκατομ.	3	F	0	~2.5 εκατομ.
Εγκεφαλικό (S29)	Affymetrix, Illumina	2,194,468	4	F	2	~2.5 εκατομ.
Αιφνίδια καρδιακή ανακοπή (S3)	Affymetrix, Illumina	~2.5 εκατομ.	5	F	2	~2.5 εκατομ.
Συστολική αρτηριακή πίεση (S31)	Affymetrix, Illumina	~2.5 εκατομ.	29	F	8	Δ/Α
Συστολική αρτηριακή πίεση (S35)	Affymetrix, Illumina	~1.7 εκατομ.	8	F	2	Δ/Α
Ολική χοληστερόλη (S78)	Affymetrix, Illumina, Perlegen	~2.6 εκατομ.	46	F	52	Δ/Α
Ολική πρωτεΐνη (S33)	Illumina	566,430	10	F	2	Δ/Α
Τριγλυκερίδια (S78)	Affymetrix, Illumina, Perlegen	~2.6 εκατομ.	46	F	32	Δ/Α
Σακχαρώδης διαβήτης τύπου 1 (S8)	Affymetrix, Illumina	~2.54 εκατομ.	5	F	19	~2.5 εκατομ.
Σακχαρώδης διαβήτης τύπου 2 (S39)	Illumina	568,976	3	F	1	M/E
Σακχαρώδης διαβήτης τύπου 2 (S81)	Affymetrix, Illumina	2,426,886	8	F	10	Δ/Α
Ελκώδης κολίτιδα (S2)	Affymetrix, Illumina	~1.1 εκατομ.	6	F	33	Δ/Α
Επίπεδα ουρικού (S86)	Affymetrix, Illumina	~2.5 εκατομ.	5	F	8	~2.5 εκατομ.
Απέκκριση αλβουμίνης στα ούρα (S6)	Affymetrix, Illumina	~2.5 εκατομ.	12	F	0	~2.5 εκατομ.

Καρκίνος ουροδόχου κύστης (S37)	Illumina	304,703	2	F	0	M/E
Ανεπάρκεια βιταμίνης D (S84)	Affymetrix, Illumina	2,548,976	5	F	3	Δ/A
Παράγοντας von Wilenbrand (S67)	Affymetrix, Illumina	~2.6 εκατομ.	5	F	8	~2.6 εκατομ.
Λόγος περιφέρειας μέσης-ισχίου (S27)	Affymetrix, Illumina, Perlegen	2,850,269	32	F, R	12	~2.8 εκατομ.
Αριθμός λευκών αιμοσφαιρίων (S33)	Illumina	566,430	10	F	5	Δ/A
Αριθμός λευκών αιμοσφαιρίων (S61)	Affymetrix, Illumina	2,489,215	7	F	3	~2.2 εκατομ.

Συντομογραφίες: F: fixed effect model (μοντέλο σταθερών επιδράσεων), HOMA-B: homeostatic model assessment beta cell function (ομοιοστατικό μοντέλο εκτίμησης λειτουργίας β-κυττάρων), HOMA-IR: homeostatic model assessment insulin resistance (ομοιοστατικό μοντέλο εκτίμησης ινσουλινο-αντίστασης), M/E: μη-εφαρμόσιμο γιατί δεν πραγματοποιήθηκε αναπλήρωση γονοτύπων (imputation), Δ/A: δεν αναφέρεται, R: random effects model (μοντέλο τυχαίων επιδράσεων), γ-GT: γ-γλουταμυλ-τρανσπεπτιδάση

* Για βιβλιογραφικές παραπομπές, βλ. Παράρτημα 1.

** Αριθμός πολυμορφισμών που περνούν το όριο στατιστικής σημαντικότητας που χρησιμοποιήθηκε σε κάθε μετα-ανάλυση

*** Όλες οι μελέτες πραγματοποίησαν αναπλήρωση των ελλειπουσών τιμών με βάση το HarMap εκτός από την παραπομπή S62 (που χρησιμοποίησε και το HarMap και το 1000 Genomes Project) και τις παραπομπές S32 και S58 (που χρησιμοποίησαν μόνο το 1000 Genomes Project). Όλες οι μελέτες χρησιμοποίησαν κάποιο είδος μοντέλου Markov ανάλογα με το είδος του λογισμικού (IMPUTE, MACH, BIMBAM, BEAGLE).

Ο Πίνακας 3 δείχνει τα κριτήρια ελέγχου ποιότητας για τους πολυμορφισμούς στα 88 άρθρα. Σε 69 από αυτά, οι επιμέρους μελέτες είχαν χρησιμοποιήσει μη-ταυτόσημα κριτήρια και μόνο σε 18 υπήρχε προφανής ομοιομορφία ανάμεσα στις επιμέρους μελέτες. Η συχνότητα ελάσσονος αλληλίου, η αναλογία ανάκλησης των πολυμορφισμών και η απόκλιση από την ισορροπία κατά Hardy-Weinberg ήταν οι πιο συχνά εφαρμοζόμενες παράμετροι. Το όριο του 1% για τη συχνότητα ελάσσονος αλληλίου χρησιμοποιήθηκε σε 41 άρθρα, ενώ 4 άρθρα συμπεριέλαβαν πολυμορφισμούς με συχνότητα $\geq 5\%$, 4 συμπεριέλαβαν πολυμορφισμούς με συχνότητα $\geq 2\%$, και ένα συμπεριέλαβε πολυμορφισμούς με συχνότητα $\geq 0.1\%$. Από τα 22 άρθρα που ανέφεραν το όριο για την αναλογία ανάκλησης των πολυμορφισμών, 11 χρησιμοποίησαν ως όριο το 95%, 4 το 90%, 1 το 97%, 4 το 98%, και 2 το 99%. Η τιμή P για τον έλεγχο απόκλισης από την ισορροπία κατά Hardy-Weinberg ποίκιλε από 10^{-4} έως 10^{-7} , με μία μελέτη να χρησιμοποιεί το 5×10^{-20} . Εικοσι ένα άρθρα ανέφεραν κάποιο επιπρόσθετο κριτήριο εκτός από τη συχνότητα ελάσσονος αλληλίου, την αναλογία ανάκλησης των πολυμορφισμών και την απόκλιση από την ισορροπία κατά Hardy-Weinberg. Τέλος, στα περισσότερα άρθρα οι επιμέρους μελέτες που συμπεριλήφθηκαν στη μετα-ανάλυση συγκράτησαν δεδομένα με υψηλή ποιότητα αναπλήρωσης των ελλειπόντων γονοτύπων, αποδεχόμενες διάφορα όρια και χρησιμοποιώντας διαφορετικό λογισμικό για την αναπλήρωση των ελλειπόντων γονοτύπων.

Πίνακας 3. Κριτήρια ελέγχου ποιότητας για τους πολυμορφισμούς στα 88 άρθρα

Παραπομπή*	Αριθμός μετα-αναλύσεων	Κριτήρια ελέγχου ποιότητας στις επιμέρους μελέτες	Κριτήρια ελέγχου ποιότητας για την εισαγωγή SNPs στη μετα-ανάλυση				Ποιότητα αναπλήρωσης γονοτύπων
			ΣΕΑ	% Ανάκλησης	HWE (<i>P</i>)	Άλλα	
S1	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.3
S2	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Π
S3	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Α
S4	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Α
S5	3	Διαφορετικά	$\geq 1\%$	Δ/Π	Δ/Π	SNPs διαθέσιμα σε ≥ 2 μελέτες	Μέτρα ποιότητας ≥ 0.3
S6	1	Διαφορετικά	$> 1\%$	Δ/Π	Δ/Π	Δ/Π	Δ/Π
S7	1	Διαφορετικά	$\geq 1\%$	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.3 ή 0.5 , ανάλογα με το λογισμικό
S8	1	Κοινά	$\geq 1\%$	$\geq 95\%$	10^{-5}	Κανένα	Μέτρα ποιότητας ≥ 0.3
S9	1	Κοινά	$\geq 1\%$	$\geq 95\%$	10^{-6}	Κανένα	Μέτρα ποιότητας ≥ 0.3 ή 0.4 , ανάλογα με το λογισμικό
S10	3	Διαφορετικά	$\geq 1\%$	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.3
S11	1	Διαφορετικά	$\geq 2\%$	Δ/Π	Δ/Π	SNPs διαθέσιμα σε ≥ 2 μελέτες	Μέτρα ποιότητας ≥ 0.7 ή 0.8 , ανάλογα με το λογισμικό
S12	1	Κοινά	$> 1\%$	$> 95\%$	10^{-6}	Κανένα	Μέτρα ποιότητας ≥ 0.3 ή 0.5 , ανάλογα με το λογισμικό
S13	1	Διαφορετικά	$\geq 1\%$	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.3
S14	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Π
S15	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Α
S16	4	Κοινά	$\geq 1\%$	$\geq 95\%$	10^{-4} -- 10^{-6}	Κανένα	Μέτρα ποιότητας ≥ 0.3 ή 0.4 , ανάλογα με το λογισμικό
S17	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Α
S18	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.3 ή 0.4 , ανάλογα με το λογισμικό
S19	2	Διαφορετικά	$\geq 1\%$	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.3 ή 0.4 , ανάλογα με το λογισμικό
S20	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Α

S21	1	Κοινά	≥0.1%	≥98%	5×10^{-20}	Πληροφορία Fisher ≥ 0.975	Μέτρα ποιότητας ≥ 0.3
S22	1	Διαφορετικά	Δ/Α	Δ/Α	Δ/Α	Δ/Α	Δ/Α
S23	1	Διαφορετικά	≥1%	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.3 ή 0.4, ανάλογα με το λογισμικό
S24	6	Διαφορετικά	≥1%	Δ/Π	Δ/Π	Δ/Π	Παρατηρούμενη προς αναμενόμενη διακύμανση > 1.1
S25	2	Διαφορετικά	≥1%	≥90%	10^{-6}	Κανένα	Μέτρα ποιότητας ≥ 0.4
S26	2	Διαφορετικά	≥1%	Δ/Π	Δ/Π	Δ/Π	Δ/Α
S27	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	AEA ≥ 3	Μέτρα ποιότητας ≥ 0.4
S28	1	Δ/Α	Δ/Α	Δ/Α	Δ/Α	Δ/Α	Δ/Α
S29	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	SNPs διαθέσιμα σε όλες τις μελέτες	Δ/Α
S30	2	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Α
S31	2	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	SNPs διαθέσιμα στο HarMap, SNPs με συχνότητες παρόμοιες με το HarMap	Μέτρα ποιότητας ≥ 0.3
S32	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.3
S33	14	Κοινά	≥1%	≥99%	10^{-7}	Σαφής ταξινόμηση των γραφημάτων στην οπτική ανασκόπηση	Μέτρα ποιότητας > 0.3
S34	2	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Α
S35	1	Διαφορετικά	≥5%	Δ/Π	Δ/Π	Δραστικό μέγεθος δείγματος $> 10,000$	Μέτρα ποιότητας ≥ 0.5
S36	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Α
S37	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	≥99.5% συνέπεια μεταξύ των μικροσυστοιχειών	M/E
S38	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.3 ή 0.4, ανάλογα με το λογισμικό
S39	1	Κοινά	$>1\%$	$>97\%$	10^{-6}	Κανένα	M/E
S40	1	Κοινά	≥2%	≥95%	10^{-4}	Κανένα	Δ/Α

S41	1	Διαφορετικά	Κανένα	Δ/Π	Δ/Π	AEA ≥ 3	Μέτρα ποιότητας ≥ 0.3 ή 0.4, ανάλογα με το λογισμικό
S42	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	SNPs διαθέσιμα σε ≥70% των ατόμων	Δ/Π
S43	1	Διαφορετικά	≥1%	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.3 ή 0.4, ανάλογα με το λογισμικό, αναπληρούμενοι γονότυποι μεταγενέστερη πιθανότητα > 0.9
S44	1	Διαφορετικά	≥5%	Δ/Π	Δ/Π	Δ/Π	Ποσοστό σφάλματος <0.01
S45	3	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Π
S46	1	Διαφορετικά	≥1%	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.3
S47	1	Κοινά	≥1%	Δ/Π	10 ⁻⁴	% έλλειψης γονοτύπων < 3%	Δ/Α
S48	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Α
S49	1	Διαφορετικά	≥1%	Δ/Π	Δ/Π	AEA ≥ 5	Μέτρα ποιότητας ≥ 0.3
S50	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Κανένα
S51	1	Κοινά	≥1%	≥99%	10 ⁻⁷	Σαφής ταξινόμηση των γραφημάτων στην οπτική ανασκόπηση	Μέτρα ποιότητας > 0.3
S52	2	Διαφορετικά	≥1%	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.5
S53	1	Διαφορετικά	≥1%	Δ/Π	Δ/Π	Κανένα	Μέτρα ποιότητας ≥ 0.4 ή 0.5, ανάλογα με το λογισμικό
S54	1	Κοινά	>1%	>95%	10 ⁻⁶	Κανένα	Σκορ ποιότητας ≥ 0.1
S55	3	Διαφορετικά	≥1%	Δ/Π	Δ/Π	Δ/Π	Δ/Π
S56	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.3
S57	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.3
S58	1	Κοινά	>5%	>90%	10 ⁻⁷	Κανένα	Μέτρα ποιότητας ≥ 0.3
S59	1	Διαφορετικά	>1%	>90%	Δ/Π	Πολυμορφικά SNPs, SNPs χαρτογραφημένα σε μία γονιδιακή περιοχή	Μέτρα ποιότητας ≥ 0.5
S60	2	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.4

S61	1	Κοινά	≥1%	≥95%	Δ/Α	SNPs χαρτογραφημένα σε μία γονιδιακή περιοχή	Μέτρα ποιότητας ≥ 0.5
S62	1	Κοινά	>5%	>98%	10 ⁻⁵	Κανένα	Δ/Α
S63	1	Διαφορετικά	>1%	>95%	10 ⁻⁶	Κανένα	Μέτρα ποιότητας ≥ 0.3 ή 0.4, ανάλογα με το λογισμικό
S64	1	Κοινά	≥1%	≥98%	10 ⁻⁶	Διαφορά συχνοτήτων με το HarMap <0.15, ποσοστό ελλειπόντων γονοτύπων μεταξύ ασθενών και μαρτύρων ανά SNP <0.02	Μέτρα ποιότητας ≥ 0.3
S65	1	Διαφορετικά	≥1%	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.5
S66	1	Διαφορετικά	≥1%	≥98%	10 ⁻⁴	Κανένα	Μέτρα ποιότητας ≥ 0.5
S67	3	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Α
S68	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Σταθμισμένη ΣΕΑ μετά τη μετα-ανάλυση > 1.5%	Δ/Α
S69	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Α
S70	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.3
S71	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Α
S72	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Α
S73	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	ΑΕΑ ≥ 6	Μέτρα ποιότητας ≥ 0.3 ή 0.4, ανάλογα με το λογισμικό
S74 **	1	Κοινά	≥1%	≥95%	10 ⁻⁶	Κανένα	Δ/Α
S75	1	Διαφορετικά	≥1%	Δ/Π	Δ/Π	Δ/Π	Σκορ ποιότητας ≥ 0.2
S76	1	Κοινά	>1%	>95%	10 ⁻⁶	Κανένα	Μέτρα ποιότητας ≥ 0.3 ή 0.4, ανάλογα με το λογισμικό
S77	1	Διαφορετικά	≥1%	Δ/Π	Δ/Π	SNPs διαθέσιμα σε ≥4 μελέτες	Μέτρα ποιότητας ≥ 0.3 ή 0.4, ανάλογα με το λογισμικό
S78	4	Διαφορετικά	≥1%	Δ/Π	Δ/Π	Δ/Π	Δ/Α
S79	4	Διαφορετικά	≥1%	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.7
S80	4	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Α
S81	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Α

S82	2	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.3
S83	2	Διαφορετικά	$\geq 1\%$	Δ/Π	Δ/Π	Δ/Π	Μέτρα ποιότητας ≥ 0.3
S84	1	Διαφορετικά	Δ/Π	Δ/Π	Δ/Π	Δ/Π	Δ/Α
S85	1	Διαφορετικά	$\geq 1\%$	$\geq 90\%$	10^{-6}	Κανένα	Μέτρα ποιότητας ≥ 0.3 ή 0.5 , ανάλογα με το λογισμικό
S86	2	Διαφορετικά	$\geq 2\%$	Δ/Π	Δ/Π	Δ/Π	Κανένα
S87	1	Κοινά	$\geq 1\%$	$\geq 95\%$	10^{-5}	Κανένα	Μεταγενέστερη πιθανότητα ≥ 0.9
S88	1	Διαφορετικά	$\geq 1\%$	Δ/Π	Δ/Π	Δ/Π	Σκορ ποιότητας ≥ 0.2

Συντομογραφίες: HWE (P): Τιμή P για την απόκλιση από την ισορροπία κατά Hardy-Weinberg , ΣΕΑ: συχνότητα ελάσσονος αλληλίου, M/E: μη-εφαρμόσιμο γιατί δεν πραγματοποιήθηκε αναπλήρωση γονοτύπων (imputation), Δ/Α: δεν αναφέρεται (δεν παρέχεται καμία πληροφορία σχετικά με το αν οι αντίστοιχες παράμετροι είχαν ληφθεί υπόψη), Δ/Π: δεν προσδιορίζεται (παρέχονται μόνο γενικές πληροφορίες για τις αντίστοιχες παραμέτρους, αλλά όχι συγκεκριμένα δεδομένα ή αριθμοί), ΑΕΑ: αριθμός ελάσσονων αλληλίων (υπολογίζεται ως το γινόμενο του μεγέθους δείγματος $\times$ ΣΕΑ).

* Για βιβλιογραφικές παραπομπές, βλ. Παράρτημα 1.

** Σε αυτή τη μελέτη, όλες οι επιμέρους μελέτες γονιδιακής σάρωσης εφάρμοσαν κοινά μέτρα ελέγχου ποιότητας εκτός από μία (τη μελέτη Wellcome Trust Case Control Consortium, WTCCC), η οποία χρησιμοποίησε μία παλαιότερη πλατφόρμα γονοτύπησης και εφάρμοσε τα εξής κριτήρια: ΣΕΑ $\geq 1\%$, ποσοστό ανάκλησης γονοτύπων $\geq 99\%$ και HWE $P > 10^{-5}$.

Ο Πίνακας 4 δείχνει τα όρια στατιστικής σημαντικότητας που επιλέχθηκαν στις μελέτες, τις μεθόδους διερεύνησης της ετερογένειας και τις διαδικασίες επαλήθευσης που εφαρμόστηκαν. Σε όλες τις μετα-αναλύσεις οι ερευνητές χρησιμοποίησαν κατά κύριο λόγο κάποια τιμή P προκειμένου να ορίσουν τη στατιστική σημαντικότητα ευρυγονιδιωμικά. Σε δύο μετα-αναλύσεις χρησιμοποίησαν επιπλέον Μπεϋζιανές μεθόδους: τον παράγοντα Bayes και την μετέπειτα πιθανότητα μιας σχέσης (131, 199). Τα όρια των τιμών P που χρησιμοποιήθηκαν ήταν 5×10^{-8} για 123 μετα-αναλύσεις, 5×10^{-7} για 8 μετα-αναλύσεις, 1×10^{-8} για 3 μετα-αναλύσεις, 7.2×10^{-8} για 2 μετα-αναλύσεις και 2.5×10^{-8} , 4×10^{-7} και 1.6×10^{-7} για καθεμία από τις υπόλοιπες 3 μετα-αναλύσεις. Μέθοδοι διερεύνησης της ετερογένειας αναφέρονταν σε 103 μετα-αναλύσεις. Αυτές αφορούσαν το στατιστικό *Cochran's Q* μόνο (37 μετα-αναλύσεις), το μέτρο I^2 μόνο (27 μετα-αναλύσεις), ή και τα δύο αυτά μέτρα (37 μετα-αναλύσεις), με 1 μετα-ανάλυση να χρησιμοποιεί τα μέτρα Q , I^2 και τ^2 και μία άλλη το τεστ των Breslow-Day. Το στατιστικό *Cochran's Q* ακολουθεί μία κατανομή κατά χ^2 και ελέγχει αν οι παρατηρούμενες διαφορές στα αποτελέσματα είναι συμβατές με την τύχη. Το I^2 ποσοτικοποιεί την ετερογένεια και μετρά το ποσοστό της διακύμανσης των μεγεθών του αποτελέσματος που μπορούν να αποδοθούν σε πραγματική ετερογένεια παρά σε τύχη (186).

Πίνακας 4. Όρια στατιστικής σημαντικότητας, μέθοδοι διερεύνησης της ετερογένειας και διαδικασίες επαλήθευσης

Φαινότυπος *	Όριο στατιστικής σημαντικότητας	Έλεγχος ετερογένειας	Διαδικασία επαλήθευσης
Δοκιμασία ανοχής γλυκόζης (S63)	5×10^{-8}	Q	Ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας και επιπλέον SNPs με βιολογική λειτουργικότητα
Επίπεδα αδιπονεκτίνης (S12)	5×10^{-7} , BF	Q, I^2	M/E
Εναπόθεση λίπους: % λίπους σώματος (S38)	5×10^{-8}	I^2	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Εναπόθεση λίπους: περιφέρεια μέσης (S43)	5×10^{-8}	Q, I^2	Ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας για τον ίδιο ή ένα σχετιζόμενο φαινότυπο και επιπλέον SNPs με βιολογική λειτουργικότητα
Γήρας: χρόνος θανάτου (S83)	5×10^{-8}	Δ/A	Ένα υποσύνολο των πιο ισχυρών SNPs από το στάδιο ανακάλυψης
Γήρας: χρόνος συμβάματος (S83)	5×10^{-8}	Δ/A	Κανένα SNP δεν πληρούσε τα κριτήρια για επαλήθευση
Αμινοτρανσφεράση της αλανίνης (S10)	1×10^{-8}	Q	Ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Αμινοτρανσφεράση της αλανίνης (S33)	5×10^{-8}	I^2	M/E
Κατανάλωση αλκοόλ (S65)	5×10^{-8}	Δ/A	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας και επιπλέον SNPs με βιολογική λειτουργικότητα
Αλκαλική φωσφατάση (S10)	1×10^{-8}	Q	Ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Αλλεργική ρινίτιδα (S60)	5×10^{-8}	Q	M/E
Νόσος Alzheimer (S28)	5×10^{-8}	Q, I^2	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Σφυροβραχιόνιος δείκτης (S48)	5×10^{-8}	Q	Ανεξάρτητα SNPs στατιστικά σημαντικά σε επίπεδο GWAS και επιπλέον ανεξάρτητα SNPs με χαμηλότερο επίπεδο στατιστικής σημαντικότητας και επιπλέον SNPs με υποστήριξη από αναλύσεις βιοπληροφορικής
Αγκυλοποιητική σπονδυλίτιδα (S21)	5×10^{-8}	Q	Δ/Π
Αορτική σκληρία (S46)	5×10^{-8}	Δ/A	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Ασπαρτική αμινοτρανσφεράση (S33)	5×10^{-8}	I^2	M/E
Άσθμα (S47)	7.2×10^{-8}	Q, I^2	M/E
Ατοπική δερματίτιδα (S53)	5×10^{-8}	Q	Ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Κολπική μαρμαρυγή (S4)	5×10^{-8}	Q	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας και ήταν διαθέσιμα σε έναν ελάχιστο αριθμό μελετών
Διπολική διαταραχή (S64)	5×10^{-8}	Q, I^2	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας

Βάρος γέννησης (S23)	5×10^{-8}	Q, I^2	Το ισχυρότερο SNP ανά γενετικό τόπο ή ένα ισχυρά συσχετιζόμενο SNP
Άζωτο ουρίας αίματος (S33)	5×10^{-8}	I^2	M/E
Δείκτης μάζας σώματος (S73)	5×10^{-8}	Q	Ανεξάρτητα SNPs που έφταναν σε στατιστική σημαντικότητα σε επίπεδο GWAS ή ένα ορισμένο επίπεδο στατιστικής σημαντικότητας
Δείκτης μάζας σώματος (S85)	5×10^{-8}	Q	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας για τον ίδιο φαινότυπο και προηγούμενως αναφερθέντα SNPs για ένα συσχετιζόμενο φαινότυπο ασχέτως στατιστικής σημαντικότητας
Οστική πυκνότητα κεφαλής μηριαίου (S19)	5×10^{-8}	Q, I^2	Ανεξάρτητα SNPs από το στάδιο ανακάλυψης
Οστική πυκνότητα οσφυικής μοίρας (S19)	5×10^{-8}	Q, I^2	Ανεξάρτητα SNPs από το στάδιο ανακάλυψης
Περιφέρεια βραχίονα(S7)	5×10^{-8}	I^2	Όλα τα νέα σήματα που έφταναν ένα ορισμένο επίπεδο στατιστικής σημαντικότητας και επιπλέον SNPs με βιολογική λειτουργικότητα
Κατανάλωση καφεΐνης (S11)	5×10^{-8}	Q, I^2	M/E
Επίπεδα ασβεστίου (S50)	5×10^{-8}	Δ/A	M/E
Χρόνια νεφρική νόσος (S55)	5×10^{-8}	I^2	SNPs που έφταναν σε στατιστική σημαντικότητα σε επίπεδο GWAS και επιπλέον νέα SNPs και επιπλέον SNPs βάσει ενός “τεστ κατεύθυνσης”
Παράγοντας πήξης VII (S67)	5×10^{-8}	Q	Νέα σήματα στατιστικά σημαντικά σε επίπεδο GWAS
Παράγοντας πήξης VIII (S67)	5×10^{-8}	Q	Κανένα SNP δεν πληρούσε τα κριτήρια για επαλήθευση
Κατανάλωση καφέ (S1)	5×10^{-8}	Q	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Πάχος έσω-μέσου χιτώνα κοινής καρωτίδας (S5)	5×10^{-8}	Q, I^2	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Στεφανιαία νόσος (S66)	5×10^{-8}	Q, I^2	Νέα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
C-αντιδρώσα πρωτεΐνη (S14)	5×10^{-8}	Q	SNPs που έφταναν σε στατιστική σημαντικότητα σε επίπεδο GWAS και επιπλέον SNPs που έφταναν σε χαμηλότερα επίπεδα στατιστικής σημαντικότητας
Ρυθμός σπειραματικής διήθησης κρεατινίνης (S55)	5×10^{-8}	I^2	SNPs που έφταναν σε στατιστική σημαντικότητα σε επίπεδο GWAS και επιπλέον νέα SNPs και επιπλέον SNPs βάσει ενός “τεστ κατεύθυνσης”
Νόσος Crohn (S22)	5×10^{-8}	Breslow-Day	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Ρυθμός σπειραματικής διήθησης συστατίνης (S55)	5×10^{-8}	I^2	SNPs που έφταναν σε στατιστική σημαντικότητα σε επίπεδο GWAS και επιπλέον νέα SNPs και επιπλέον SNPs βάσει ενός “τεστ κατεύθυνσης”
D-διμερή (S69)	5×10^{-8}	Δ/A	M/E
Διυδροεπιανδροστερόνη θεϊκή (S88)	5×10^{-8}	I^2	M/E
Διαστολική αρτηριακή πίεση (S31)	5×10^{-8}	Δ/A	Ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας

Διαστολική αρτηριακή πίεση (S35)	5×10^{-8}	Q	Ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας και επιπλέον ανεξάρτητα SNPs με χαμηλότερο όριο στατιστικής σημαντικότητας, αλλά με συνεπή κατεύθυνση του σήματος στις διάφορες μελέτες
Γλυκόζη νύστεως (S16)	5×10^{-8}	Q, I^2	Ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας και επιπλέον SNPs με βιολογική λειτουργικότητα
Ινσουλίνη νύστεως (S16)	5×10^{-8}	Q, I^2	Ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας και επιπλέον SNPs με βιολογική λειτουργικότητα
Ινωδογόνο (S15)	5×10^{-8}	Δ/A	Ανεξάρτητα SNPs στατιστικά σημαντικά σε επίπεδο GWAS
γ-GT (S10)	1×10^{-8}	Q	Ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
γ-GT (S33)	5×10^{-8}	I^2	M/E
Γλίωμα (S62)	5×10^{-7}	Q, I^2	M/E
Γλυκοζυλιωμένη Αιμοσφαιρίνη (S71)	5×10^{-8}	Q, I^2	SNPs επιλεγμένα προηγουμένως βάσει εσωτερικών αναλύσεων
Ουρική αρθρίτιδα (S86)	5×10^{-8}	Δ/A	M/E
HDL χοληστερόλη (S78)	5×10^{-8}	Q	Τα σημαντικότερα SNPs μελετήθηκαν σε Ευρωπαϊκούς και μη-Ευρωπαϊκούς πληθυσμούς
Περιφέρεια κεφαλής (νεογνά) (S77)	5×10^{-8}	Q, I^2	Ισχυρότερα SNPs ανά γενετικό τόπο ή ένα ισχυρά συσχετιζόμενο SNP
Καρδιακή ανεπάρκεια (S68)	5×10^{-7}	Δ/A	M/E
Ύψος (S41)	5×10^{-8}	Q, I^2	Νέα σήματα που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Ύψος (S49)	5×10^{-8}	Q, I^2	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Ύψος (S51)	5×10^{-8}	Q, I^2	M/E
Αιματοκρίτης (S24)	5×10^{-8}	Δ/A	SNPs που φτάνουν στατιστική σημαντικότητα σε επίπεδο GWAS στο στάδιο ανακάλυψης
Αιματοκρίτης (S33)	5×10^{-8}	I^2	M/E
Αιμοσφαιρίνη (S24)	5×10^{-8}	Δ/A	SNPs που φτάνουν στατιστική σημαντικότητα σε επίπεδο GWAS στο στάδιο ανακάλυψης
Αιμοσφαιρίνη (S33)	5×10^{-8}	I^2	M/E
HOMA-B (S16)	5×10^{-8}	Q, I^2	Ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας και επιπλέον SNPs με βιολογική λειτουργικότητα
HOMA-IR (S16)	5×10^{-8}	Q, I^2	Ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας με βιολογική λειτουργικότητα
IgE ευαισθητοποίηση σε χόρτο (S60)	5×10^{-8}	Q	M/E
IGF-1 (S34)	5×10^{-8}	Δ/A	M/E

IGFBP-3 (S34)	5×10^{-8}	Δ/A	M/E
Πάχος έσω-μέσου χιτώνα έσω καρωτίδας (S5)	5×10^{-8}	Q, I^2	Κανένα SNP δεν πληρούσε τα κριτήρια για επαλήθευση
Ενδοκράνιο ανεύρυσμα (S87)	5×10^{-8} , PPA, BF	Q, I^2, τ^2	Γονοτυπηθέντα SNPs με τη μεγαλύτερη απώτερη πιθανότητα σχέσης και επιπλέον δύο SNPs ανά γενετικό τόπο για κάθε ανεξάρτητο σήμα
LDL χοληστερόλη (S78)	5×10^{-8}	Q	Τα σημαντικότερα SNPs μελετήθηκαν σε Ευρωπαϊκούς και μη-Ευρωπαϊκούς πληθυσμούς
Εσωτερικές διαστάσεις αριστερής κοιλίας κατά τη διαστολή (S80)	5×10^{-7}	Δ/A	Ισχυρότερο SNP ανά γενετικό τόπο που έχει στατιστική σημαντικότητα σε επίπεδο GWAS
Συστολική δυσλειτουργία αριστερής κοιλίας (S80)	5×10^{-7}	Δ/A	Ισχυρότερο SNP ανά γενετικό τόπο που έχει στατιστική σημαντικότητα σε επίπεδο GWAS
Πάχος τοιχώματος αριστερής κοιλίας (S80)	5×10^{-7}	Δ/A	Ισχυρότερο SNP ανά γενετικό τόπο που έχει στατιστική σημαντικότητα σε επίπεδο GWAS
Μάζα αριστερής κοιλίας (S80)	5×10^{-7}	Δ/A	Ισχυρότερο SNP ανά γενετικό τόπο που έχει στατιστική σημαντικότητα σε επίπεδο GWAS
Λιποπρωτεϊνική φωσφολιπάση A2: δραστηριότητα (S26)	5×10^{-8}	I^2	M/E
Λιποπρωτεϊνική φωσφολιπάση A2: μάζα (S26)	5×10^{-8}	I^2	M/E
Επίπεδα μαγνησίου (S45)	5×10^{-8}	Q	SNPs με στατιστική σημαντικότητα σε επίπεδο GWAS και επιπλέον SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Μείζονες διαταραχές συναισθήματος (S44)	7.2×10^{-8}	Q, I^2	Καμία πληροφορία
Μέση αρτηριακή πίεση (S82)	5×10^{-8}	Δ/A	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας για ένα σχετιζόμενο φαινότυπο
Μέση περιεκτικότητα αιμοσφαιρίνης (S24)	5×10^{-8}	Δ/A	SNPs που φτάνουν στατιστική σημαντικότητα σε επίπεδο GWAS στο στάδιο ανακάλυψης
Μέση περιεκτικότητα αιμοσφαιρίνης (S33)	5×10^{-8}	I^2	M/E
Μέση πυκνότητα αιμοσφαιρίνης (S24)	5×10^{-8}	Δ/A	SNPs που φτάνουν στατιστική σημαντικότητα σε επίπεδο GWAS στο στάδιο ανακάλυψης
Μέση πυκνότητα αιμοσφαιρίνης (S33)	5×10^{-8}	I^2	M/E
Μέσος όγκος ερυθρών (S24)	5×10^{-8}	Δ/A	SNPs που φτάνουν στατιστική σημαντικότητα σε επίπεδο GWAS στο στάδιο ανακάλυψης
Μέσος όγκος ερυθρών (S33)	5×10^{-8}	I^2	M/E
Μέσος όγκος αιμοπεταλίων (S25)	5×10^{-8}	Q	Ισχυρότερα SNPs ανά γενετικό τόπο
Εμμηναρχή (ηλικία έναρξης) (S18)	5×10^{-8}	Q	Νέα ανακαλυφθέντα ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Εμμηνόπαυση (ηλικία έναρξης) (S75)	5×10^{-8}	Q	Ανεξάρτητα SNPs από το στάδιο ανακάλυψης
Μεταβολικό σύνδρομο (S40)	5×10^{-8}	I^2	M/E

Ημικρανία (S42)	5×10^{-8}	Δ/A	Υποσύνολο SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας και επιλεγμένα ώστε να είναι όσο το δυνατόν πιο ενημερωτικά δεδομένης της ανισορροπίας διασύνδεσης μεταξύ τους
Πολλαπλή σκλήρυνση (S54)	5×10^{-8}	Q, I^2	M/E
Παχυσαρκία (παιδική) (S9)	5×10^{-8}	Q	Όλα τα νέα σήματα που έφταναν ένα ορισμένο επίπεδο στατιστικής σημαντικότητας
Οστεοαρθρίτιδα (knee) (S20)	5×10^{-8}	Q, I^2	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Νόσος Parkinson (S32)	5×10^{-8}	Q, I^2	SNPs στατιστικά σημαντικά σε επίπεδο GWAS και που πληρούσαν τα κριτήρια ποιότητας
Τύποι προσωπικότητας (S13)	5×10^{-8}	Q	SNPs στατιστικά σημαντικά σε επίπεδο GWAS
Επίπεδα φωσφόρου (S36)	4×10^{-7}	Δ/A	Ισχυρότερα SNPs
Πλάκα (αθηρωματική) καρωτίδας (S5)	5×10^{-8}	Q, I^2	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Αριθμός αιμοπεταλίων (S25)	5×10^{-8}	Q	Ισχυρότερα SNPs ανά γενετικό τόπο
Αριθμός αιμοπεταλίων (S33)	5×10^{-8}	I^2	M/E
Αριθμός αιμοπεταλίων (S59)	5×10^{-8}	Q, I^2	Όλα τα νέα σήματα που έφταναν ένα ορισμένο επίπεδο στατιστικής σημαντικότητας
Επίπεδα καλίου (S45)	5×10^{-8}	Q	SNPs που έφταναν σε επίπεδο στατιστικής σημαντικότητας σε επίπεδο GWAS και επιπλέον SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Διάστημα PR (S57)	5×10^{-8}	Q, I^2	M/E
Επίπεδα προϊνσουλίνης (S76)	5×10^{-8}	Q, I^2	Ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας και επιπλέον SNPs με βιολογική λειτουργικότητα
Πνευμονική λειτουργία (S70)	5×10^{-8}	Q	Όλα τα νέα σήματα που έφταναν ένα ορισμένο επίπεδο στατιστικής σημαντικότητας
Πίεση σφυγμού (S82)	5×10^{-8}	Δ/A	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας για ένα σχετιζόμενο φαινότυπο
Διάρκεια QRS (S72)	5×10^{-8}	Q, I^2	SNPs που έφταναν σε επίπεδο στατιστικής σημαντικότητας σε επίπεδο GWAS και SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Διάστημα QT (S56)	5×10^{-8}	Δ/A	Το ισχυρότερο SNP ανά γενετικό τόπο που έφτανε σε ένα ορισμένο όριο στατιστικής σημαντικότητας
Αριθμός ερυθρών αιμοσφαιρίων (S24)	5×10^{-8}	Δ/A	SNPs που φτάνουν στατιστική σημαντικότητα σε επίπεδο GWAS στο στάδιο ανακάλυψης
Αριθμός ερυθρών αιμοσφαιρίων (S33)	5×10^{-8}	I^2	M/E
Νεφροκυτταρικό καρκίνωμα (S58)	5×10^{-7}	Q	Ανεξάρτητα SNPs στατιστικά σημαντικά σε επίπεδο GWAS
Διάστημα RR (S17)	5×10^{-8}	Δ/A	M/E

Διάμετρος αρτηριογλιών αμφιβληστροειδούς (S30)	5×10^{-8}	Δ/A	Ανεξάρτητα SNPs στατιστικά σημαντικά σε επίπεδο GWAS
Διάμετρος φλεβιδίων αμφιβληστροειδούς (S30)	5×10^{-8}	Δ/A	Ανεξάρτητα SNPs στατιστικά σημαντικά σε επίπεδο GWAS
Ρευματοειδής αρθρίτιδα (S52)	5×10^{-8}	Δ/A	Όλα τα νέα σήματα που έφταναν ένα ορισμένο επίπεδο στατιστικής σημαντικότητας
Ρευματοειδής αρθρίτιδα (S74)	5×10^{-8}	Q	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας και επιπλέον SNPs που είχαν προηγουμένως συσχετισθεί με σχετικούς φαινοτύπους καθώς και SNPs με βιολογική λειτουργικότητα
Κρεατινίνη ορού (S33)	5×10^{-8}	I^2	M/E
Κάπνισμα - Ηλικία έναρξης (S79)	5×10^{-8}	I^2	Κανένα SNP δεν πληρούσε τα κριτήρια για επαλήθευση
Κάπνισμα - Αριθμός τσιγάρων ημερισίως (S79)	5×10^{-8}	I^2	Υποσύνολο από τα πιο σημαντικά SNPs
Κάπνισμα - καπνιστής έναντι μη-καπνιστής (S79)	5×10^{-8}	I^2	Υποσύνολο από τα πιο σημαντικά SNPs
Κάπνισμα - πρώην καπνιστής έναντι επί του παρόντος καπνιστής (S79)	5×10^{-8}	I^2	Υποσύνολο από τα πιο σημαντικά SNPs
Επίπεδα νατρίου (S45)	5×10^{-8}	Q	SNPs που έφταναν το όριο στατιστικής σημαντικότητας σε επίπεδο GWAS και επιπλέον SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Εγκεφαλικό (S29)	5×10^{-8}	Δ/A	Ανεξάρτητα SNPs στατιστικά σημαντικά σε επίπεδο GWAS
Αιφνίδια καρδιακή ανακοπή (S3)	5×10^{-8}	Q, I^2	Ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Συστολική αρτηριακή πίεση (S31)	5×10^{-8}	Δ/A	Ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Συστολική αρτηριακή πίεση (S35)	5×10^{-8}	Q	Ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας και επιπλέον ανεξάρτητα SNPs που είχαν χαμηλότερο όριο στατιστικής σημαντικότητας αλλά είχαν συνεπή σχέση με τον φαινότυπο στις επιμέρους μελέτες
Ολική χοληστερόλη (S78)	5×10^{-8}	Q	Τα σημαντικότερα SNPs μελετήθηκαν σε Ευρωπαϊκούς και μη-Ευρωπαϊκούς πληθυσμούς
Ολική πρωτεΐνη (S33)	5×10^{-8}	I^2	M/E
Τριγλυκερίδια (S78)	5×10^{-8}	Q	Τα σημαντικότερα SNPs μελετήθηκαν σε Ευρωπαϊκούς και μη-Ευρωπαϊκούς πληθυσμούς
Σακχαρώδης διαβήτης τύπου 1 (S8)	5×10^{-8}	Q	Ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας και χωρίς ένδειξη ετερογένειας σε σχέση με προηγουμένως μη αναφερθέντες γενετικούς τόπους
Σακχαρώδης διαβήτης τύπου 2 (S39)	5×10^{-8}	Q	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Σακχαρώδης διαβήτης τύπου 2 (S81)	5×10^{-8}	Q, I^2	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Ελκώδης κολίτιδα (S2)	5×10^{-8}	Δ/A	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας και δεν είχαν επαληθευτεί προηγουμένως
Επίπεδα ουρικού (S86)	5×10^{-8}	Δ/A	M/E

Απέκκριση αλβουμίνης στα ούρα (S6)	5×10^{-8}	Q, I^2	Ισχυρότερα ανεξάρτητα SNPs που ξεπερνούσαν ένα ορισμένο όριο στατιστικής σημαντικότητας και συχνότητας ελάσσονος αλληλίου
Καρκίνος ουροδόχου κύστης (S37)	1.6×10^{-7}	Q, I^2	Μη-επαληθευμένα SNPs
Ανεπάρκεια βιταμίνης D (S84)	5×10^{-8}	Δ/A	SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Παράγοντας von Willebrand (S67)	5×10^{-8}	Q	Νέα σήματα στατιστικά σημαντικά σε επίπεδο GWAS
Λόγος περιφέρειας μέσης-ισχίου (S27)	5×10^{-8}	Q, I^2	Ανεξάρτητα SNPs που έφταναν ένα ορισμένο όριο στατιστικής σημαντικότητας
Αριθμός λευκών αιμοσφαιρίων (S33)	5×10^{-8}	I^2	M/E
Αριθμός λευκών αιμοσφαιρίων (S61)	2.5×10^{-8}	Q, I^2	SNPs ανιχνευόμενα για πρώτη φορά

Συντομογραφίες: BF, Bayes factor (παράγοντας Bayes); HOMA-B, homeostatic model assessment beta cell function (ομοιοστατικό μοντέλο εκτίμησης λειτουργίας β-κυττάρων); HOMA-IR, homeostatic model assessment insulin resistance (ομοιοστατικό μοντέλο εκτίμησης ινσουλινο-αντίστασης); ΣΕΑ: συχνότητα ελάσσονος αλληλίου, M/E: μη-εφαρμόσιμο γιατί δεν έγινε επαλήθευση σε επόμενο στάδιο, Δ/A: δεν αναφέρεται (δεν παρέχεται καμία πληροφορία σχετικά με τον αν οι αντίστοιχες παράμετροι είχαν ληφθεί υπόψη); Δ/Π: δεν προσδιορίζεται (παρέχονται μόνο γενικές πληροφορίες για τις αντίστοιχες παραμέτρους αλλά όχι συγκεκριμένα δεδομένα ή αριθμοί); Q , στατιστικό Q του Cochran; γ-GT, γ-γλουταμυλ-τρανσπεπτιδάση

* Για βιβλιογραφικές παραπομπές, βλ. Παράρτημα 1.

Προσπάθειες επαλήθευσης είχαν γίνει για τα αποτελέσματα 104 (75%) μετα-αναλύσεων, γεγονός που δείχνει την ευρεία αποδοχή της πρακτικής αυτής στο πεδίο των μελετών γονιδιακής σάρωσης (72, 200). Το σκεπτικό βάσει του οποίου επιλέχθηκαν οι πολυμορφισμοί προκειμένου να μελετηθούν στο στάδιο της επαλήθευσης, αναφέρεται στον **Πίνακα 4**. Στις περισσότερες περιπτώσεις αυτό αφορούσε κυρίως κριτήρια που σχετίζονται με τη στατιστική σημαντικότητα και σε μικρότερο βαθμό κριτήρια σχετικά με τη λειτουργία ή το βιολογικό ρόλο των γονιδίων.

Όπως φαίνεται στον **Πίνακα 5**, μόνο 27 μετα-αναλύσεις (19%) χρησιμοποίησαν πληθυσμούς μικτής καταγωγής στο στάδιο ανακάλυψης ή μελέτησαν τα ευρήματα της αντίστοιχης μετα-ανάλυσης σε πληθυσμούς διαφορετικής εθνικής καταγωγής από τον πληθυσμό ανακάλυψης. Για 79 μετα-αναλύσεις οι ερευνητές είχαν πραγματοποιήσει επιπλέον αναλύσεις συσχέτισης ανάμεσα σε πολυμορφισμούς που βρέθηκαν στη μετα-ανάλυση του σταδίου ανακάλυψης και φαινοτύπους ή νοσήματα διαφορετικά από αυτά που είχαν μελετηθεί στην αρχική μετα-ανάλυση. Επιπλέον, για 19 μετα-αναλύσεις (14%), οι ερευνητές είχαν επεκτείνει τον έλεγχο γενετικών συσχετίσεων σε ποικίλους φαινοτύπους. Αυτοί οι φαινότυποι προέρχονταν από το μετασχηματισμό ενός συνεχούς φαινοτύπου σε διχότομο νόσημα βάσει κάποιας τιμής (σφυροβραχιόνιος δείκτης και περιφερική αρτηριακή νόσος, συστολική πίεση και υπέρταση, διαστολική πίεση και υπέρταση, δείκτης μάζας σώματος και παχυσαρκία, απέκκριση αλβουμίνης στα ούρα και μικροαλβουμινουρία, πάχος μέσου χιτώνα της κοινής και έσω καρωτίδας και ύπαρξη πλάκας), από τη χρήση διαφορετικών ορισμών για τον μελετούμενο φαινότυπο (περιφέρεια μέσης και λόγος περιφέρειας μέσης προς περιφέρεια ισχύου για τη μέτρηση της

εναπόθεσης λίπους, FEV_1 και FEV_1/FVC για τη μέτρηση της πνευμονικής λειτουργίας, χρόνος έως το σύμβαμα και χρόνος έως το θάνατο για τη μέτρηση του γήρατος), από τη χρήση άλλων προϊόντων του ίδιου μεταβολικού μονοπατιού (προϊνσουλίνη και 32,33-split προϊνσουλίνη/C-δείκτης/ινσουλινογενής δείκτης, απέκκριση αλβουμίνης στα ούρα και λόγος αλβουμίνης προς κρεατινίνη), και τέλος, από τη χρήση ενδοφαινοτύπων που περιλαμβάνονταν στον αρχικό φαινότυπο (διαφορετικοί τύποι της λευκής σειράς αιμοσφαιρίων, διαφορετικοί τύποι εγκεφαλικού επεισοδίου, συνιστώσες του μεταβολικού συνδρόμου, τύποι προσωπικότητας, νόσος Alzheimer και ηλικία έναρξης της ίδιας νόσου).

Πίνακας 5. Φυλετικές καταγωγές συμμετεχόντων, ποικίλοι φαινότυποι και διαφαινοτυπικοί έλεγχοι

Φαινότυπος *	Διαφορετικές καταγωγές	Ποικίλοι φαινότυποι	Διαφαινοτυπικοί έλεγχοι
Δοκιμασία ανοχής γλυκόζης (S63)	Όχι	Όχι	Δείκτες έκκρισης ινσουλίνης, Σακχαρώδης διαβήτης τύπου 2
Επίπεδα αδιπονεκτίνης (S12)	Ναι	Όχι	Σακχαρώδης διαβήτης τύπου 2, HDL χοληστερόλη, LDL χοληστερόλη, Ολική χοληστερόλη, Τριγλυκερίδια, Λόγος περιφέρειας μέσης-ισχίου
Εναπόθεση λίπους: % λίπους σώματος (S38)	Ναι	Όχι	Λιπιδαιμικά προφίλ, Δείκτες ευαισθησίας ινσουλίνης, Κατανομή λίπους, Λεπτίνη, Αδιπονεκτίνη
Εναπόθεση λίπους: περιφέρεια μέσης (S43)	Όχι	Λόγος περιφέρειας μέσης-ισχίου	Δείκτης μάζας σώματος, Βάρος, Ύψος, bioimpedance data, DEXA, Υπερλιπιδαιμία, Σακχαρώδης διαβήτης τύπου 2
Γήρας: χρόνος θανάτου (S83)	Όχι	Γήρας: χρόνος συμβάματος	Όχι
Γήρας: χρόνος συμβάματος (S83)	Δ/Α	Γήρας: χρόνος θανάτου	Όχι
Αμινοτρανσφεράση της αλανίνης (S10)	Ναι	Όχι	Ηπατική στεάτωση
Αμινοτρανσφεράση της αλανίνης (S33)	Όχι	Όχι	Όχι
Κατανάλωση αλκοόλ (S65)	Όχι	Όχι	Όχι
Αλκαλική φωσφατάση (S10)	Ναι	Όχι	Ηπατική στεάτωση
Αλλεργική ρινίτιδα (S60)	Όχι	Όχι	Σειρά γέννησης
Νόσος Alzheimer (S28)	Όχι	Ηλικία έναρξης	Όχι
Σφυροβραχιόνιος δείκτης (S48)	Όχι	ΠΑΝ	Στεφανιαία νόσος, Έμφραγμα μυοκαρδίου
Αγκυλοποιητική σπονδυλίτιδα (S21)	Όχι	Όχι	Όχι
Αορτική σκληρία (S46)	Όχι	Όχι	Στεφανιαία νόσος, Καρδιακή ανεπάρκεια, Καρδιακός ρυθμός, Χρόνια νεφρική νόσος, Ρυθμός Σπειραματικής Διήθησης
Ασπαρτική αμινοτρανσφεράση (S33)	Όχι	Όχι	Όχι
Άσθμα (S47)	Όχι	Όχι	IgE ορού

Ατοπική δερματίτιδα (S53)	Όχι	Όχι	Άσθμα, Ολικά επίπεδα IgE ορού
Κολπική μαρμαρυγή (S4)	Όχι	Όχι	Όχι
Διπολική διαταραχή (S64)	Όχι	Όχι	Σχιζοφρένεια
Βάρος γέννησης (S23)	Ναι	Όχι	Μήκος γέννησης, Περίμετρος κεφαλής κατά τη γέννηση, ponderal index, Σακχαρώδης διαβήτης τύπου 2, Γλυκόζη νύστεως
Άζωτο ουρίας αίματος (S33)	Όχι	Όχι	Όχι
Δείκτης μάζας σώματος (S73)	Όχι	Υπέρβαρος/ παχύσαρκος	Βάρος, Ποσοστό λίπους σώματος, Σακχαρώδης διαβήτης τύπου 2, Γλυκόζη νύστεως, Δοκιμασία ανοχής γλυκόζης, Ινσουλίνη νύστεως, HOMA-B, HOMA-IR, HDL χοληστερόλη, LDL χοληστερόλη, Τριγλυκερίδια, Ύψος
Δείκτης μάζας σώματος (S85)	Όχι	Όχι	Όχι
Οστική πυκνότητα κεφαλής μηριαίου (S19)	Ναι	Όχι	Κατάγματα
Οστική πυκνότητα οσφυικής μοίρας (S19)	Ναι	Όχι	Κατάγματα
Περιφέρεια βραχίονα(S7)	Όχι	Όχι	Δείκτης μάζας σώματος, Λόγος περιφέρειας μέσης-ισχίου, Βάρος, Σακχαρώδης διαβήτης τύπου 2, Γλυκόζη νύστεως, Ινσουλίνη νύστεως, Γλυκοζυλιωμένη Αιμοσφαιρίνη, Δοκιμασία ανοχής γλυκόζης
Κατανάλωση καφεΐνης (S11)	Όχι	Όχι	Όχι
Επίπεδα ασβεστίου (S50)	Όχι	Όχι	Μαγνήσιο, φώσφορος, παραθορμόνη ορού, Οστική πυκνότητα οσφυικής μοίρας, Οστική πυκνότητα κεφαλής μηριαίου, Νεφρολιθίαση
Χρόνια νεφρική νόσος (S55)	Ναι	Όχι	Λόγος αλβουμίνης-κρεατινίνης ούρων, Μικροαλβουμινουρία, Αρτηριακή πίεση, Έμφραγμα μυοκαρδίου
Παράγοντας πήξης VII (S67)	Όχι	Όχι	Όχι
Παράγοντας πήξης VIII (S67)	Όχι	Όχι	Όχι
Κατανάλωση καφέ (S1)	Όχι	Όχι	Όχι
Πάχος έσω-μέσου χιτώνα κοινής καρωτίδας (S5)	Όχι	Πάχος έσω-μέσου χιτώνα έσω καρωτίδας, Πλάκα	Στεφανιαία νόσος
Στεφανιαία νόσος (S66)	Όχι	Όχι	HDL χοληστερόλη, LDL χοληστερόλη, Ολική χοληστερόλη, Υπέρταση, Δείκτης μάζας σώματος, Σακχαρώδης διαβήτης τύπου 2, Κάπνισμα

C-αντιδρώσα πρωτεΐνη (S14)	Όχι	Όχι	Έμφραγμα μυοκαρδίου, Στεφανιαία νόσος
Ρυθμός σπειραματικής διήθησης κρεατινίνης (S55)	Ναι	Όχι	Λόγος αλβουμίνης-κρεατινίνης ούρων, Μικροαλβουμινουρία, Αρτηριακή πίεση, Έμφραγμα μυοκαρδίου
Νόσος Crohn (S22)	Όχι	Όχι	Όχι
Ρυθμός σπειραματικής διήθησης βάσει συστατίνης (S55)	Όχι	Όχι	Λόγος αλβουμίνης-κρεατινίνης ούρων, Μικροαλβουμινουρία, Αρτηριακή πίεση, Έμφραγμα μυοκαρδίου
D-διμερή (S69)	Δ/Α	Όχι	Όχι
Διυδροεπιανδροστερόνη θειϊκή (S88)	Όχι	Όχι	Όχι
Διαστολική αρτηριακή πίεση (S31)	Ναι	Υπέρταση	Επίπεδα μεταβολιτών, Μάζα αριστερής κοιλίας, Πάχος τοιχώματος αριστερής κοιλίας, Καρδιακή ανεπάρκεια, Εγκεφαλικό, Στεφανιαία νόσος, Νεφρική νόσος και μέτρα νεφρικής λειτουργίας
Διαστολική αρτηριακή πίεση (S35)	Όχι	Υπέρταση	Δείκτης μάζας σώματος, Γλυκόζη και λιπίδια νύστεως, Στεφανιαία νόσος
Γλυκόζη νύστεως (S16)	Όχι	Όχι	HOMA-B, HOMA-IR, Γλυκοζυλιωμένη Αιμοσφαιρίνη, Δοκιμασία ανοχής γλυκόζης, Ινσουλίνη νύστεως, Δοκιμασία ανοχής γλυκόζης σε 2 ώρες, Σακχαρώδης διαβήτης τύπου 2, Διαστολική αρτηριακή πίεση, Συστολική αρτηριακή πίεση, Υπέρταση, HDL χοληστερόλη, LDL χοληστερόλη, Ολική χοληστερόλη, Τριγλυκερίδια
Ινσουλίνη νύστεως (S16)	Όχι	Όχι	HOMA-B, HOMA-IR, Γλυκοζυλιωμένη Αιμοσφαιρίνη, Δοκιμασία ανοχής γλυκόζης, Ινσουλίνη νύστεως, Δοκιμασία ανοχής γλυκόζης σε 2 ώρες, Σακχαρώδης διαβήτης τύπου 2, Διαστολική αρτηριακή πίεση, Συστολική αρτηριακή πίεση, Υπέρταση, HDL χοληστερόλη, LDL χοληστερόλη, Ολική χοληστερόλη, Τριγλυκερίδια
Ινωδογόνο (S15)	Όχι	Όχι	Όχι
γ-GT (S10)	Ναι	Όχι	Ηπατική στεάτωση
γ-GT (S33)	Όχι	Όχι	Όχι
Γλίωμα (S62)	Όχι	Όχι	Όχι
Γλυκοζυλιωμένη Αιμοσφαιρίνη (S71)	Όχι	Όχι	Σακχαρώδης διαβήτης τύπου 2, Στεφανιαία νόσος, Γλυκόζη νύστεως, Δοκιμασία ανοχής γλυκόζης, Ινσουλίνη νύστεως, Δοκιμασία ανοχής γλυκόζης σε 2 ώρες, HOMA-B, HOMA-IR, Δείκτης μάζας σώματος, Αιμοσφαιρίνη, Μέση περιεκτικότητα αιμοσφαιρίνης, Μέσος όγκος ερυθρών, τρανσφερρίνη, σίδηρος
Ουρική αρθρίτιδα (S86)	Όχι	Όχι	Διαστολική αρτηριακή πίεση, Συστολική αρτηριακή πίεση, Γλυκόζη, Ινσουλίνη, Ρυθμός Σπειραματικής Διήθησης, Χρόνια νεφρική νόσος, Στεφανιαία νόσος

HDL χοληστερόλη (S78)	Ναι	Όχι	Στεφανιαία νόσος, Υπερλιπιδαιμία, Ακραίες τιμές λιπιδίων
Περιφέρεια κεφαλής (νεογνά) (S77)	Όχι	Όχι	Ενδοκράνιος όγκος, Περιφέρεια κέφαλής τριμήνου, Περιφέρεια κεφαλής στη γέννηση, Ύψος
Καρδιακή ανεπάρκεια (S68)	Όχι	Όχι	Όχι
Ύψος (S41)	Όχι	Όχι	Όχι
Ύψος (S49)	Ναι	Όχι	Όχι
Ύψος (S51)	Όχι	Όχι	Όχι
Αιματοκρίτης (S24)	Όχι	Όχι	Διαστολική αρτηριακή πίεση, Συστολική αρτηριακή πίεση, Υπέρταση
Αιματοκρίτης (S33)	Όχι	Όχι	Όχι
Αιμοσφαιρίνη (S24)	Όχι	Όχι	Διαστολική αρτηριακή πίεση, Συστολική αρτηριακή πίεση, Υπέρταση
Αιμοσφαιρίνη (S33)	Όχι	Όχι	Όχι
HOMA-B (S16)	Όχι	Όχι	HOMA-B, HOMA-IR, Γλυκοζυλιωμένη Αιμοσφαιρίνη, Δοκιμασία ανοχής γλυκόζης, Ινσουλίνη νύστεως, Δοκιμασία ανοχής γλυκόζης σε 2 ώρες, Σακχαρώδης διαβήτης τύπου 2, Διαστολική αρτηριακή πίεση, Συστολική αρτηριακή πίεση, Υπέρταση, HDL χοληστερόλη, LDL χοληστερόλη, Ολική χοληστερόλη, Τριγλυκερίδια
HOMA-IR (S16)	Όχι	Όχι	HOMA-B, HOMA-IR, Γλυκοζυλιωμένη Αιμοσφαιρίνη, Δοκιμασία ανοχής γλυκόζης, Ινσουλίνη νύστεως, Δοκιμασία ανοχής γλυκόζης σε 2 ώρες, Σακχαρώδης διαβήτης τύπου 2, Διαστολική αρτηριακή πίεση, Συστολική αρτηριακή πίεση, Υπέρταση, HDL χοληστερόλη, LDL χοληστερόλη, Ολική χοληστερόλη, Τριγλυκερίδια
IgE ευαισθητοποίηση σε χόρτο (S60)	Όχι	Όχι	Σειρά γέννησης
IGF-1 (S34)	Δ/Α	Όχι	Όχι
IGFBP-3 (S34)	Δ/Α	Όχι	Όχι
Πάχος έσω-μέσου χιτώνα έσω καρωτίδας (S5)	Όχι	Όχι	Όχι
Ενδοκράνιο ανεύρυσμα (S87)	Ναι	Όχι	Όχι
LDL χοληστερόλη (S78)	Ναι	Όχι	Στεφανιαία νόσος, Υπερλιπιδαιμία, Ακραία επίπεδα λιπιδίων
Εσωτερικές διαστάσεις αριστερής κοιλίας κατά τη διαστολή (S80)	Όχι	Όχι	Όχι
Συστολική δυσλειτουργία αριστερής κοιλίας (S80)	Όχι	Όχι	Όχι
Πάχος τοιχώματος αριστερής κοιλίας (S80)	Όχι	Όχι	Όχι

Μάζα αριστερής κοιλίας (S80)	Όχι	Όχι	Όχι
Λιποπρωτεϊνική φωσφολιπάση A2: δραστηριότητα (S26)	Όχι	Όχι	Στεφανιαία νόσος
Λιποπρωτεϊνική φωσφολιπάση A2: μάζα (S26)	Όχι	Όχι	Στεφανιαία νόσος
Επίπεδα μαγνησίου (S45)	Όχι	Όχι	Ρυθμός Σπειραματικής Διήθησης, Διαστολική αρτηριακή πίεση, Συστολική αρτηριακή πίεση, Γλυκόζη νύστεως, Οστική πυκνότητα
Μείζονες διαταραχές συναισθήματος (S44)	Όχι	Όχι	Όχι
Μέση αρτηριακή πίεση (S82)	Όχι	Όχι	Όχι
Μέση περιεκτικότητα αιμοσφαιρίνης (S24)	Όχι	Όχι	Διαστολική αρτηριακή πίεση, Συστολική αρτηριακή πίεση, Υπέρταση
Μέση περιεκτικότητα αιμοσφαιρίνης (S33)	Όχι	Όχι	Όχι
Μέση πυκνότητα αιμοσφαιρίνης (S33)	Όχι	Όχι	Όχι
Μέση πυκνότητα αιμοσφαιρίνης (S24)	Όχι	Όχι	Διαστολική αρτηριακή πίεση, Συστολική αρτηριακή πίεση, Υπέρταση
Μέσος όγκος ερυθρών (S24)	Όχι	Όχι	Διαστολική αρτηριακή πίεση, Συστολική αρτηριακή πίεση, Υπέρταση
Μέσος όγκος ερυθρών (S33)	Όχι	Όχι	Όχι
Μέσος όγκος αιμοπεταλίων (S25)	Ναι	Όχι	Αιμοσφαιρίνη, Μέσος όγκος ερυθρών, Αριθμός ερυθρών αιμοσφαιρίων
Εμμηναρχή (ηλικία έναρξης) (S18)	Όχι	Όχι	Δείκτης μάζας σώματος, Λόγος περιφέρειας μέσης-ισχίου, Παχυσαρκία, Ύψος
Εμμηνόπαυση (ηλικία έναρξης) (S75)	Όχι	Όχι	Όχι
Μεταβολικό σύνδρομο (S40)	Όχι	Μεταβολικό σύνδρομο	Μεταβολίτες ορού
Ημικρανία (S42)	Όχι	Όχι	Όχι
Πολλαπλή σκλήρυνση (S54)	Δ/Α	Όχι	Νόσος Crohn, Ελκώδης κολίτιδα, Κοιλιοκάκη, Σακχαρώδης διαβήτης τύπου 1, Ρευματοειδής αρθρίτιδα, Συστηματικός ερυθηματώδης λύκος, Ψωρίαση
Παχυσαρκία (παιδική) (S9)	Όχι	Όχι	Δείκτης μάζας σώματος
Οστεοαρθρίτιδα (knee) (S20)	Ναι	Όχι	Όχι
Νόσος Parkinson (S32)	Όχι	Όχι	Όχι
Τύποι προσωπικότητας (S13)	Όχι	Τύποι προσωπικότητας	Όχι
Επίπεδα φωσφόρου (S36)	Όχι	Όχι	Όχι

		Πάχος μέσου χιτώνα κοινής καρωτίδας και έσω καρωτίδας	
Πλάκα (αθηρωματική) καρωτίδας (S5)	Όχι		Στεφανιαία νόσος
Αριθμός αιμοπεταλίων (S25)	Ναι	Όχι	Αιμοσφαιρίνη, Μέσος όγκος ερυθρών, Αριθμός ερυθρών αιμοσφαιρίων
Αριθμός αιμοπεταλίων (S33)	Όχι	Όχι	Όχι
Αριθμός αιμοπεταλίων (S59)	Ναι	Όχι	Συνάθροιση αιμοπεταλίων
Επίπεδα καλίου (S45)	Όχι	Όχι	Ρυθμός Σπειραματικής Διήθησης, Διαστολική αρτηριακή πίεση, Συστολική αρτηριακή πίεση, Γλυκόζη νύστεως, Οστική πυκνότητα
Διάστημα PR (S57)	Όχι	Όχι	Κολπική μαρμαρυγή
		32,33-split προ- ινσουλίνη, ινσουλινογενής δείκτης, C- πεπτίδιο	Γλυκόζη νύστεως, Δοκιμασία ανοχής γλυκόζης, Ινσουλίνη νύστεως, Δοκιμασία ανοχής γλυκόζης σε 2 ώρες, HOMA-B, HOMA-IR, Γλυκοζυλιωμένη Αιμοσφαιρίνη, Σακχαρώδης διαβήτης τύπου 2, Δείκτης μάζας σώματος, Στεφανιαία νόσος
Επίπεδα προΐνσουλίνης (S76)	Όχι		
Πνευμονική λειτουργία (S70)	Όχι	FEV ₁ , FEV ₁ /FVC	Ποσότητα καπνίσματος, Κάπνισμα ή όχι, Ύψος, Καρκίνος πνεύμονα
Πίεση σφυγμού (S82)	Όχι	Όχι	Διαστολική αρτηριακή πίεση, Συστολική αρτηριακή πίεση
Διάρκεια QRS (S72)	Όχι	Όχι	Έλλειμμα αγωγής καρδιακού ερεθίσματος, Διάστημα PR, Διάστημα QT
Διάστημα QT (S56)	Όχι	Όχι	Όχι
Αριθμός ερυθρών αιμοσφαιρίων (S24)	Όχι	Όχι	Διαστολική αρτηριακή πίεση, Συστολική αρτηριακή πίεση, Υπέρταση
Αριθμός ερυθρών αιμοσφαιρίων (S33)	Όχι	Όχι	Όχι
Νεφροκυτταρικό καρκίνωμα (S58)	Όχι	Όχι	Όχι
Διάστημα RR (S17)	Όχι	Όχι	Όχι
Διάμετρος αρτηριολίων αμφιβληστροειδούς (S30)	Όχι	Όχι	Central retinal venular equivalent, Στεφανιαία νόσος, Εγκεφαλικό, Έμφραγμα μυοκαρδίου, Υπέρταση, Σακχαρώδης διαβήτης τύπου 2
Διάμετρος φλεβιδίων αμφιβληστροειδούς (S30)	Όχι	Όχι	Central retinal venular equivalent, Στεφανιαία νόσος, Εγκεφαλικό, Έμφραγμα μυοκαρδίου, Υπέρταση, Σακχαρώδης διαβήτης τύπου 2
Ρευματοειδής αρθρίτιδα (S52)	Ναι	Όχι	Συστηματικός ερυθρεματώδης λύκος, Νόσος Graves'
Ρευματοειδής αρθρίτιδα (S74)	Όχι	Όχι	Όχι

Κρεατινίνη ορού (S33)	Όχι	Όχι	Όχι
Κάπνισμα - Ηλικία έναρξης (S79)	Όχι	Όχι	Όχι
Κάπνισμα - Αριθμός τσιγάρων ημερησίως (S79)	Όχι	Όχι	Όχι
Κάπνισμα - καπνιστής έναντι μη-καπνιστής (S79)	Όχι	Όχι	Όχι
Κάπνισμα - πρώην καπνιστής έναντι επί του παρόντος καπνιστής (S79)	Όχι	Όχι	Όχι
Επίπεδα νατρίου (S45)	Όχι	Όχι	Ρυθμός Σπειραματικής Διήθησης, Διαστολική αρτηριακή πίεση, Συστολική αρτηριακή πίεση, Γλυκόζη νύστεως, Οστική πυκνότητα
Εγκεφαλικό (S29)	Ναι	Τύποι εγκεφαλικού	Όχι
Αιφνίδια καρδιακή ανακοπή (S3)	Όχι	Όχι	Διάστημα QRS, Διάστημα QT, Διάστημα RR
Συστολική αρτηριακή πίεση (S31)	Ναι	Υπέρταση	Επίπεδα μεταβολιτών, Μάζα αριστερής κοιλίας, Πάχος τοιχώματος αριστερής κοιλίας, Καρδιακή ανεπάρκεια, Εγκεφαλικό, Στεφανιαία νόσος, Νεφρική νόσος και μέτρα νεφρικής λειτουργίας
Συστολική αρτηριακή πίεση (S35)	Όχι	Υπέρταση	Δείκτης μάζας σώματος, Γλυκόζη και λιπίδια νύστεως, Στεφανιαία νόσος
Ολική χοληστερόλη (S78)	Ναι	Όχι	Στεφανιαία νόσος, Υπερλιπιδαιμία, Ακραίες τιμές λιπιδίων
Ολική πρωτεΐνη (S33)	Όχι	Όχι	Όχι
Τριγλυκερίδια (S78)	Ναι	Όχι	Στεφανιαία νόσος, Υπερλιπιδαιμία, Ακραίες τιμές λιπιδίων
Σακχαρώδης διαβήτης τύπου 1 (S8)	Όχι	Όχι	Όχι
Σακχαρώδης διαβήτης τύπου 2 (S39)	Ναι	Όχι	Όχι
Σακχαρώδης διαβήτης τύπου 2 (S81)	Όχι	Όχι	Γλυκόζη νύστεως, Ινσουλίνη νύστεως, HOMA-B, HOMA-IR, Δείκτης μάζας σώματος
Ελκώδης κολίτιδα (S2)	Όχι	Όχι	Νόσος Crohn
Επίπεδα ουρικού (S86)	Όχι	Όχι	Διαστολική αρτηριακή πίεση, Συστολική αρτηριακή πίεση, Γλυκόζη, Ινσουλίνη, Ρυθμός σπειραματικής διήθησης, Χρόνια νεφρική νόσος, Στεφανιαία νόσος
Απέκκριση αλβουμίνης στα ούρα (S6)	Ναι	Μικροαλβουμινο υρία, Λόγος αλβουμίνης- κρεατινίνης ούρων	Χρόνος έως την επίμονη μικροαλβουμινουρία

Καρκίνος ουροδόχου κύστης (S37)	Όχι	Όχι	Επιβίωση ελεύθερης νόσου, Καπνιστική συμπεριφορά
Ανεπάρκεια βιταμίνης D (S84)	Όχι	Όχι	Κλινική ανεπάρκεια βιταμίνης D, Συγκεντρώσεις ορού της δεσμευτικής της βιταμίνης D πρωτεΐνης
Παράγοντας von Willebrand (S67)	Όχι	Όχι	Όχι
Λόγος περιφέρειας μέσης-ισχίου (S27)	Όχι	Όχι	Δείκτης μάζας σώματος, Περιφέρεια μέσης, Περιφέρεια ισχίων, Τριγλυκερίδια, LDL χοληστερόλη, HDL χοληστερόλη, Γλυκόζη νύστεως, Ινσουλίνη νύστεως, HOMA-IR, Δοκιμασία ανοχής γλυκόζης, Σακχαρώδης διαβήτης τύπου 2
Αριθμός λευκών αιμοσφαιρίων (S33)	Όχι	Όχι	Όχι
Αριθμός λευκών αιμοσφαιρίων (S61)	Ναι	Υπότυποι λευκής σειράς	Όχι

Συντομογραφίες: FEV₁: forced expiratory volume in 1 s, FVC: forced vital capacity, HOMA-B: homeostatic model assessment beta cell function (ομοιοστατικό μοντέλο εκτίμησης λειτουργίας β-κυττάρων), HOMA-IR: homeostatic model assessment insulin resistance, γ-GT: γ-γλουταμυλ-τρανσπεπτιδάση, Δ/Α: δεν αναφέρεται, ΠΑΝ: Περιφερική αρτηριακή νόσος

* Για βιβλιογραφικές παραπομπές, βλ. Παράρτημα 1.

Ο Πίνακας 6 περιγράφει 45 μελέτες γονιδιακής σάρωσης όπου είχε πραγματοποιηθεί τουλάχιστον ένα είδος λειτουργικής ανάλυσης που συνόδευε την αντίστοιχη μετα-ανάλυση. Αναλύσεις έκφρασης γενετικών τόπων ποσοτικών φαινοτύπων ήταν ο πιο συχνός τύπος λειτουργικής ανάλυσης (26 μελέτες). Άλλες αναλύσεις έκφρασης περιελάμβαναν γονιδιακή έκφραση (19 μελέτες) και έκφραση RNA (2 μελέτες). Μοντέλα ζώων (201) είχαν χρησιμοποιηθεί σε 9 μελέτες, και αναλύσεις μονοπατιών (*pathway analyses*) (202) είχαν πραγματοποιηθεί σε άλλες 13. Εξέταση των σχέσεων μεταξύ των ανακαλυφθέντων γονιδίων σε διαφορετικά νοσήματα με χρήση του προ-λογισμικού GRAIL (*Gene Relationships Across Implicated Loci*) (<http://www.broadinstitute.org/mpg/grail/>) (203) περιλαμβανόταν σε 8 μελέτες και ο ρόλος της κωδικοποιούσας γενετικής ποικιλομορφίας είχε εξεταστεί σε άλλες 8 μελέτες. Τέλος, άλλες λειτουργικές αναλύσεις περιλάμβαναν τη χρήση της ηλεκτρονικής βάσης OMIM (*Online Mendelian Inheritance in Man*) (<http://www.ncbi.nlm.nih.gov/omim>) (3 μελέτες), αλληλεπιδράσεις γονιδίων και γονιδιακή οντολογία (από 2 μελέτες), αλληλεπιδράσεις γονιδίων-περιβάλλοντος, μεθυλίωση DNA, ανάλυση μεταλλάξεων, και ανάλυση μεταγραφικού προφίλ (από 1 μελέτη). Περισσότερες από 1 λειτουργικές αναλύσεις είχαν πραγματοποιηθεί σε 28 μελέτες (62%), περισσότερες από 2 σε 15 μελέτες (33%), και περισσότερες από 3 σε 6 μελέτες (13%).

Πίνακας 6. Λειτουργικές αναλύσεις που συνόδευσαν τις μετα-αναλύσεις

Παραπομπή *	Γονιδιακή έκφραση	eQTLs	Έκφραση RNA	Μοντέλα ζώων	Αναλύσεις μονοπατιών	Αλληλεπιδράσεις πρωτεϊνών	GRAIL	Κωδικοποιούσα διακύμανση	Άλλες
S1	✓								
S2		✓					✓	✓	
S10		✓					✓	✓	
S12	✓								
S16	✓	✓							
S18		✓						✓	
S19		✓		✓	✓		✓		OMIM
S20	✓			✓					
S21									Αλληλεπιδράσεις γονιδίων
S22		✓					✓	✓	
S24	✓								
S25		✓		✓	✓	✓			Μεταγραφικό προφίλ
S27		✓	✓		✓		✓	✓	
S28		✓							
S31	✓				✓			✓	
S32	✓	✓							Μεθυλίωση DNA
S35		✓							
S37			✓						Μεταλλάξεις
S38	✓			✓					
S39		✓							
S40	✓								
S41		✓			✓		✓	✓	OMIM
S44	✓								

S46	✓									
S49		✓								
S50		✓								
S52		✓								
S53	✓									
S54	✓	✓			✓					
S55		✓		✓						
S64					✓					Οντολογία γονιδίων
S65	✓			✓						
S66	✓									
S70	✓				✓			✓		Αλληλεπιδράσεις γονιδίων-περιβάλλοντος
S72		✓		✓	✓					
S73		✓			✓					
S74								✓		
S75		✓			✓					
S76	✓	✓								
S77		✓							✓	
S78		✓		✓						
S81	✓	✓		✓			✓			
S83					✓		✓			Αλληλεπιδράσεις γονιδίων
S85		✓								
S88	✓				✓					Οντολογία γονιδίων

Συντομογραφίες: eQTLs, expression quantitative trait loci; GRAIL, Gene Relationships Across Implicated Loci (<http://www.broadinstitute.org/mpg/grail>); OMIM, Online Mendelian Inheritance in Man (<http://www.ncbi.nlm.nih.gov/omim>).

*Για βιβλιογραφικές παραπομπές, βλ. Παράρτημα 1.

Τέλος, ο **Πίνακας 7** δείχνει τη συνεισφορά των ανακαλυφθέντων γενετικών τόπων στην ολική γενετική διακύμανση των αντίστοιχων φαινοτύπων για τις 62 μετα-αναλύσεις που παρείχαν επαρκή δεδομένα για τους υπολογισμούς. Η αγκυλοποιητική σπονδυλίτιδα ήταν ο φαινότυπος με τη μεγαλύτερη εξηγούμενη γενετική διακύμανση, καθώς οι 7 γενετικοί τόποι που είχαν φτάσει σε όρια στατιστικής σημαντικότητας εξηγούν το 79% της νόσου. Αυτό οφείλεται κατά κύριο λόγο στην επίδραση του πολυμορφισμού rs4349859 στο *HLA-B* (λόγος αναλογιών=56, $P<10^{-200}$). Αυτός ο γενετικός τόπος είναι ένας καλά εδραιωμένος παράγοντας κινδύνου και ταυτόχρονα διαγνωστικό τεστ για τη συγκεκριμένη νόσο. Άλλοι 6 φαινότυποι είχαν εξηγούμενη γενετική διακύμανση >20%: η νόσος του Crohn (71%), η ρευματοειδής αρθρίτιδα (52% σε Ευρωπαίους και 33% σε Ασιάτες), η πολλαπλή σκλήρυνση (49%), η ελκώδης κολίτιδα (39%), ο σακχαρώδης διαβήτης τύπου I (36%), και η νόσος του Parkinson (20%). Υπήρχε ασθενής συσχέτιση ($r=0.28$, $P=0.03$) ανάμεσα στον αριθμό των ανακαλυφθέντων πολυμορφισμών και της εξηγούμενης γενετικής διακύμανσης, π.χ. 115 επικυρωμένοι πολυμορφισμοί εξηγούν λιγότερο από 7% της διακύμανσης του ύψους. Πολλοί φαινότυποι είχαν πολύ χαμηλά ποσοστά εξηγούμενης γενετικής διακύμανσης. Η μικρότερη εξηγούμενη γενετική διακύμανση αφορούσε τον σφυροβραχιόνιο δείκτη, όπου μόνο ένας πολυμορφισμός ήταν στατιστικά σημαντικός και εξηγούσε <0.1% της ολικής γενετικής διακύμανσης.

Πίνακας 7. Γενετική διακύμανση που εξηγείται από τον συνολικό αριθμό των ανακαλυφθέντων γενετικών τόπων στις μετα-αναλύσεις μελετών γονιδιακής σάρωσης.

Φαινότυπος *	Ανακαλυφθέντα SNPs	Εξηγούμενη γενετική διακύμανση (%)
Επίπεδα αδιπονεκτίνης (S12)	10	0.72
Εναπόθεση λίπους: % λίπους σώματος (S38)	2	0.27
Αμινοτρανσφεράση της αλανίνης (S33)	2	0.60
Αλλεργική ρινίτιδα (S60)	1	1.22
Νόσος Alzheimer (S28)	6	5.78
Σφυροβραχιόνιος δείκτης (S48)	1	0.002
Αγκυλοποιητική σπονδυλίτιδα (S21)	7	78.63
Αορτική σκληρία (S46)	1	0.26
Ασπαρτική αμινοτρανσφεράση (S33)	2	0.84
Άσθμα (S47)	6	5.71
Ατοπική δερματίτιδα (S53)	2	1.76
Κολπική μαρμαρυγή (S4)	2	4.23
Διπολική διαταραχή (S64)	4	3.70
Βάρος γέννησης (S23)	2	0.51
Άζωτο ουρίας αίματος (S33)	3	0.89
Οστική πυκνότητα κεφαλής μηριαίου (S19)	22	3.89
Οστική πυκνότητα οσφυικής μοίρας (S19)	26	5.62
Επίπεδα ασβεστίου (S50)	1	0.09
Χρόνια νεφρική νόσος (S55)	2	2.74
Κατανάλωση καφέ (S1)	1	0.005
Στεφανιαία νόσος (S66)	13	9.01
Νόσος Crohn (S22)	52	71.21
γ-GT (S33)	2	1.96
Γλίωμα (S62)	6	12.19
Ουρική αρθρίτιδα (S86)	2	14.34
Περιφέρεια κεφαλής (νεογνά) (S77)	3	0.77
Καρδιακή ανεπάρκεια (S68)	1	1.12
Ύψος (S41) **	115	6.56
Ύψος (S49) ***	2	0.30
Ύψος (S51) #	8	1.78
Αιματοκρίτης (S33)	2	0.56
Αιμοσφαιρίνη (S33)	2	0.61
IgE ευαισθητοποίηση σε χόρτο (S60)	3	5.54
Ενδοκράνιο ανεύρυσμα (S87)	6	12.20
Μείζονες διαταραχές συναισθήματος (S44)	1	0.94
Μέση περιεκτικότητα αιμοσφαιρίνης (S33)	15	6.98
Μέση πυκνότητα αιμοσφαιρίνης (S33)	7	2.07

Μέσος όγκος ερυθρών (S33)	15	6.56
Μεταβολικό σύνδρομο (S40)	1	2.07
Πολλαπλή σκλήρυνση (S54)	9	49.08
Παχυσαρκία (παιδική) (S9)	7	10.15
Νόσος Parkinson (S32)	11	20.17
Επίπεδα φωσφόρου (S36)	7	0.39
Πλάκα (αθηρωματική) καρωτίδας (S5)	3	2.48
Αριθμός αιμοπεταλίων (S33)	6	2.19
Επίπεδα προΐνσουλίνης (S76)	7	0.67
Πνευμονική λειτουργία (S70)	8	0.69
Αριθμός ερυθρών αιμοσφαιρίων (S33)	8	3.47
Νεφροκυτταρικό καρκίνωμα (S58)	6	5.38
Ρευματοειδής αρθρίτιδα (S52) #	7	32.90
Ρευματοειδής αρθρίτιδα (S74) **	6	51.65
Κρεατινίνη ορού (S33)	1	0.22
Εγκεφαλικό (S29)	2	4.97
Αιφνίδια καρδιακή ανακοπή (S3)	2	6.757
Ολική πρωτεΐνη (S33)	2	0.69
Σακχαρώδης διαβήτης τύπου 1 (S8)	19	36.42
Σακχαρώδης διαβήτης τύπου 2 (S39) #	1	2.09
Σακχαρώδης διαβήτης τύπου 2 (S81) **	10	13.09
Ελκώδης κολίτιδα (S2)	33	39.17
Ανεπάρκεια βιταμίνης D (S84)	3	12.86
Λόγος περιφέρειας μέσης-ισχίου (S27)	12	0.60
Αριθμός λευκών αιμοσφαιρίων (S33)	5	1.28

Συντομογραφίες: γ-GT, γ-γλουταμυλ-τρανσπεπτιδάση

* Για βιβλιογραφικές παραπομπές, βλ. Παράρτημα 1.

** Αναφέρεται σε πληθυσμούς Ευρωπαϊκής καταγωγής.

*** Αναφέρεται σε πληθυσμούς Αφρικανικής καταγωγής.

Αναφέρεται σε πληθυσμούς Ασιατικής καταγωγής.

2.1.4. ΣΥΖΗΤΗΣΗ

Μελετήθηκαν συνολικά 139 μετα-αναλύσεις μελετών γονιδιακής σάρωσης που είχαν πραγματοποιηθεί ως επί το πλείστον σε πληθυσμούς ευρωπαϊκής καταγωγής, με μεγέθη δείγματος πολύ μεγαλύτερα από τις κλασικές επιδημιολογικές μελέτες. Οι ανακαλυφθέντες πολυμορφισμοί για τις μετα-αναλύσεις αυτές κυμαίνονταν από 0 έως 115, ενώ κατά μέσο όρο κάθε μετα-ανάλυση ανακάλυψε 4 πολυμορφισμούς. Στις περισσότερες περιπτώσεις τα ευρήματα της αρχικής μετα-ανάλυσης διερευνήθηκαν σε στάδια επαλήθευσης, ενώ και οι λειτουργικές αναλύσεις ήταν ιδιαίτερα συχνές. Οι ανακαλυφθέντες πολυμορφισμοί εξηγούν ένα σημαντικό μέρος της γενετικής διακύμανσης του αντίστοιχου νοσήματος που σε ορισμένες περιπτώσεις ξεπερνά το 50%.

Στις περισσότερες μετα-αναλύσεις, οι επιμέρους μελέτες συντέθηκαν στατιστικά με τη μέθοδο των σταθερών επιδράσεων. Τα μοντέλα αυτά υποθέτουν ότι υπάρχει ένα κοινό γενετικό αποτέλεσμα σε όλες τις μελέτες και η όποια διακύμανση οφείλεται στο τυχαίο σφάλμα. Αντιθέτως, τα μοντέλα τυχαίων επιδράσεων υποθέτουν ότι οι επιμέρους μελέτες ανιχνεύουν διαφορετικά μεγέθη αποτελέσματος και η όποια διακύμανση, δηλαδή η ετερογένεια, οφείλεται σε αληθινές διαφορές μεταξύ των πληθυσμών (134, 187, 204-206). Κατά συνέπεια, στα πλαίσια αυτά η παρατηρούμενη επικράτηση των μοντέλων σταθερών επιδράσεων είναι αναμενόμενη, καθώς η μέθοδος αυτή έχει μεγαλύτερη ισχύ από τη μέθοδο των τυχαίων επιδράσεων για την ανακάλυψη νέων γενετικών σχέσεων, με την τελευταία να είναι προτιμητέα όταν στόχος είναι ο καθορισμός και η γενικευσιμότητα του μεγέθους του γενετικού αποτελέσματος (207). Η ετερογένεια στις μελέτες γονιδιακής σάρωσης μπορεί

να οφείλεται σε διαφορές των επιμέρους μελετών, όπως είναι οι διαφορετικοί πληθυσμοί, τα διαφορετικά πρότυπα ανισορροπίας διασύνδεσης, οι διαφορετικές περιβαλλοντικές εκθέσεις και οι διαφορετικές πλατφόρμες γονοτύπησης, ή μπορεί να αντιπροσωπεύει μη-εξηγούμενη στατιστική ετερογένεια (187, 207-209).

Οι περισσότερες μετα-αναλύσεις αφορούσαν μελέτες που είχαν πραγματοποιηθεί σε ευρωπαϊκούς πληθυσμούς, ενώ ορισμένες αφορούσαν μικτούς πληθυσμούς είτε στο στάδιο ανακάλυψης είτε στο στάδιο επαλήθευσης. Γενετικοί τόποι και πολύ περισσότερο μεγέθη γενετικών σχέσεων μπορεί να είναι ιδιαίτερα δύσκολο να γενικευθούν σε διαφορετικούς πληθυσμούς (210). Ωστόσο, τέτοιοι σχεδιασμοί μπορεί να είναι ιδιαίτερα χρήσιμοι για την ανακάλυψη πολυμορφισμών που έχουν χαμηλή συχνότητα σε έναν πληθυσμό, αλλά είναι κοινοί σε άλλους (56).

Οι μετα-αναλύσεις μελετών γονιδιακής σάρωσης έχουν συνεισφέρει σημαντικά στην ανακάλυψη της γενετικής αρχιτεκτονικής των πολυσύνθετων φαινοτύπων. Ένας μεγάλος αριθμός πολυμορφισμών έχει ανακαλυφθεί για τα περισσότερα νοσήματα, οι οποίοι όμως εξηγούν λιγότερο από το 20% της γενετικής διακύμανσης, με ορισμένες εξαιρέσεις. Δεδομένου ότι υπάρχει ασθενής συσχέτιση ανάμεσα στον αριθμό των πολυμορφισμών και το αντίστοιχο ποσοστό της εξηγούμενης γενετικής διακύμανσης, για τους περισσότερους φαινοτύπους υπάρχει προφανώς ένας σημαντικός αριθμός επιπλέον πολυμορφισμών που θα μπορούσε να αυξήσει αυτό το ποσοστό (197, 211), κάτι το οποίο θα μπορούσε να επιτευχθεί σε σημαντικό βαθμό με τις μελέτες αλληλούχισης (Nilanjan Chatterjee, προσωπική επικοινωνία).

Μετα-αναλύσεις με μεγαλύτερα μεγέθη δείγματος και λιγότερο αυστηρά κριτήρια στατιστικής σημαντικότητας (122) μπορούν αναμφίβολα να βοηθήσουν προς αυτή την κατεύθυνση, καθώς επίσης και στην ανακάλυψη πολυμορφισμών χαμηλής συχνότητας, ιδιαίτερα με τη χρήση αναπλήρωσης ελλειπόντων πολυμορφισμών, την αλληλούχιση χιλιάδων ατόμων, αλλά και τη χρήση νέων μετα-αναλυτικών τεχνικών (212-214). Κάποια πρώτα αποτελέσματα αναπλήρωσης ελλειπόντων πολυμορφισμών χαμηλής συχνότητας δείχνουν ότι αυτοί οι πολυμορφισμοί μπορούν να αναπληρωθούν και ότι έχουν σημαντικό ρόλο σε ορισμένα κοινά νοσήματα (215). Τα παραπάνω μπορούν να πραγματοποιηθούν πρωτίστως μέσω διεθνών συνεργατικών συνασπισμών (*consortia*) που επιτρέπουν τη συγκέντρωση μεγάλων δειγμάτων – συνεπώς και την αύξηση της ισχύος – αλλά και τον ελεύθερο διαμοιρασμό δεδομένων και την κεντρική και χωρίς σφάλματα ανάλυση και επαλήθευσή τους (141, 216).

2.2. ΑΝΑΣΚΟΠΗΣΗ ΤΩΝ ΠΟΛΥΜΟΡΦΙΣΜΩΝ ΧΑΜΗΛΗΣ ΣΥΧΝΟΤΗΤΑΣ ΣΕ ΜΕΛΕΤΕΣ ΓΟΝΙΔΙΑΚΗΣ ΣΑΡΩΣΗΣ

2.2.1. ΣΚΕΠΤΙΚΟ ΚΑΙ ΣΚΟΠΟΣ

Η πλειονότητα των ανακαλύψεων στη γενετική και γονιδιωματική επιδημιολογία τα τελευταία 5-10 χρόνια αφορά σε συσχετίσεις κοινών (*common*) γενετικών πολυμορφισμών με ποικίλους φαινοτύπους (217, 218). Πιο συγκεκριμένα, οι μελέτες σάρωσης του γονιδιώματος (*genome-wide association studies*) έχουν αυξήσει σημαντικά τον αριθμό των γενετικών σχέσεων με πολύ υψηλά επίπεδα στατιστικής σημαντικότητας (87, 90-92). Οι μελέτες σάρωσης του γονιδιώματος που έχουν πραγματοποιηθεί και δημοσιευθεί μέχρι σήμερα, έχουν χρησιμοποιήσει ως επί το πλείστον κοινούς γενετικούς δείκτες και έχουν ανακαλύψει χαμηλής διεισδυτικότητας (*low-penetrance*) πολυμορφισμούς με μικρά μεγέθη αποτελέσματος (50, 97). Οι πλατφόρμες γονοτύπησης σε αυτές τις μελέτες προσφέρουν πολύ καλή κάλυψη σε όλο το γονιδίωμα για πολυμορφισμούς με συχνότητα ελάσσονος αλληλίου μεγαλύτερη του 5% (97, 98). Ωστόσο, πολυμορφισμοί με χαμηλότερες συχνότητες ελάσσονος αλληλίου είτε αποκλείονται συστηματικά από τις εμπορικές πλατφόρμες είτε τυγχάνουν ανεπαρκούς κάλυψης (97, 99).

Για τις περισσότερες παθήσεις, οι γενετικές συσχετίσεις που έχουν ανακαλυφθεί από τις μελέτες σάρωσης του γονιδιώματος μπορούν να εξηγήσουν μόνο ένα μικρό ποσοστό της εκτιμώμενης ολικής κληρονομικότητας (*heritability*) (55, 110, 111). Έχουν διατυπωθεί πολλές υποθέσεις για τους λόγους που καθορίζουν αυτό το υπολειπόμενο άγνωστο συστατικό της γενετικής αρχιτεκτονικής – γνωστό επίσης και ως «γενετική σκοτεινή ύλη»

(*genetic dark matter*) (54, 55, 112). Μία εξήγηση είναι η ύπαρξη γενετικών συσχετίσεων που εμπεριέχουν μη-κοινούς (*uncommon*) (συχνότητα ελάσσονος αλληλίου $\leq 5\%$) και σπάνιους (*rare*) (συχνότητα ελάσσονος αλληλίου $<0.5\%$) πολυμορφισμούς (55, 97). Συσχετίσεις που αφορούν τέτοιους χαμηλής συχνότητας πολυμορφισμούς είναι πιθανό να έχουν ακόμα και σημαντικές γενετικές επιδράσεις, αλλά ήταν δύσκολο έως σήμερα να ανιχνευθούν κυρίως λόγω ανεπαρκούς κάλυψης στις μελέτες σάρωσης του γονιδιώματος, απαιτήσεων για μεγάλα μεγέθη δείγματος και ανεπαρκών αναλυτικών μεθοδολογιών (97, 160).

Νεότερες πλατφόρμες και τεχνικές γονοτύπησης, όπως η αλληλούχιση του συνόλου των εξονίων (*exome sequencing*) (161, 162) και η αλληλούχιση του πλήρους γονιδιώματος (*whole-genome sequencing*) (154, 163), καθώς και νέες μεθοδολογικές προσεγγίσεις (160, 164, 165) εφαρμόζονται ήδη για την αναζήτηση γενετικών συσχετίσεων που αφορούν μη-κοινούς και σπάνιους πολυμορφισμούς (166-169). Παρόλα αυτά, ακόμα και «παραδοσιακές» μελέτες σάρωσης του γονιδιώματος έχουν ανακαλύψει περιστασιακά γενετικές συσχετίσεις που αφορούν τέτοιους χαμηλής συχνότητας πολυμορφισμούς. Δεδομένου ότι περισσότερες από 1,500 μελέτες σάρωσης του γονιδιώματος έχουν δημοσιευθεί μέχρι σήμερα (<http://www.genome.gov/gwastudies/>) (89), μία ανασκόπηση αυτής της βιβλιογραφίας θα μπορούσε να αποφέρει ένα αξιοσημείωτο δείγμα συσχετίσεων για πολυμορφισμούς όπως οι παραπάνω. Μία τέτοια ανασκόπηση θα αναδείξει επίσης κάποιες πρώιμες πληροφορίες για αυτές τις γενετικές σχέσεις και τους αντίστοιχους γενετικούς τύπους.

Ερωτήματα όπως τα ακόλουθα θα μπορούσαν να τεθούν: Ποια είναι τα τυπικά μεγέθη αποτελέσματος για αυτές τις σχέσεις και πόσο ισχυρή είναι η

επαλήθευσή τους; Προέρχονται από μία ή περισσότερες μελέτες σάρωσης του γονιδιώματος για τον ίδιο φαινότυπο; Υπάρχουν κοινοί πολυμορφισμοί στους ίδιους γενετικούς τόπους; Αυτοί οι γενετικοί τόποι έχουν συσχετισθεί με κάποια βιολογική λειτουργία από την εποχή των υποψήφιων γονιδίων στόχων (*candidate genes*); Υπάρχουν γνωστές μεταλλάξεις σχετιζόμενες με παρόμοιους φαινοτύπους στους ίδιους γενετικούς τόπους; Ποιες είναι οι συχνότητες των πολυμορφισμών αυτών σε πληθυσμούς διαφορετικής καταγωγής;

Έν κατακλείδι, σκοπός ήταν η συστηματική αξιολόγηση αυτών των ερωτημάτων μέσω της ανάλυσης των γενετικών σχέσεων για μη-κοινούς πολυμορφισμούς με συχνότητα ελάσσονος αλληλίου 5% ή λιγότερο, που είχαν ανακαλυφθεί σε μελέτες GWAS με ισχυρή στατιστική υποστήριξη (56).

2.2.2. ΜΕΘΟΔΟΙ

Χρησιμοποιήσαμε τον Κατάλογο Δημοσιευμένων Μελετών Γονιδιακής Σάρωσης (*A Catalog of Published Genome-Wide Association Studies*) που διατηρείται από το Γραφείο Πληθυσμιακής Γενετικής (*Office of Population Genetics*) του National Human Genome Research Institute των Η.Π.Α (89) στον ιστότοπο <http://www.genome.gov/gwastudies/>. Ο Κατάλογος αυτός είναι μία διαδικτυακή βάση δεδομένων που ενημερώνεται σε τακτά χρονικά διαστήματα και καταλογογραφεί γενετικές σχέσεις μεταξύ διαφόρων φαινοτύπων και SNPs που εξάγονται από δημοσιευμένες μελέτες GWAS. Οι σχέσεις που περιέχει έχουν τιμές στατιστικής σημαντικότητας $P < 10^{-5}$. Για τη δική μας μελέτη επιλέξαμε γενετικές σχέσεις που ήταν στατιστικά σημαντικές σε επίπεδο GWAS με βάση μία τιμή $P \leq 10^{-7}$, ανεξαρτήτως της συχνότητας ελάσσονος αλληλίου του εμπλεκόμενου SNP. Επειδή ο Κατάλογος αναφέρει μόνο 1 SNP για κάθε γενετικό τύπο, ανατρέξαμε επίσης στα κύρια άρθρα και στα παραρτήματα όλων των μελετών GWAS που ήταν στατιστικά σημαντικές, προκειμένου να βρούμε επιπλέον σχέσεις που αφορούν πολυμορφισμούς χαμηλής συχνότητας, ανεξαρτήτως του αν περιλαμβάνονταν στον Κατάλογο ή όχι.

Κατάλληλες προς εισαγωγή στη μελέτη ήταν γενετικές σχέσεις στις οποίες το αλληλίο κινδύνου (*risk allele*) είχε συχνότητα $\leq 5\%$ ή $\geq 95\%$ (δηλαδή η συχνότητα του ελάσσονος αλληλίου ήταν $\leq 5\%$) και οι οποίες είχαν φτάσει σε σημείο στατιστικής σημαντικότητας σε επίπεδο GWAS με $P \leq 10^{-7}$ (122, 139) σε μία τουλάχιστον μελέτη, όταν τα δεδομένα από το στάδιο ανακάλυψης (*discovery stage*) και επαλήθευσης (*replication stage*) συντίθενταν με μετα-ανάλυση (142, 182, 187, 207). Οι σχέσεις αυτές με $P \leq 10^{-7}$ είναι πολύ πιθανό να αντιπροσωπεύουν αληθινές γενετικές συσχετίσεις (122). Το κριτήριο για τη

συχνότητα του αλληλίου κινδύνου αφορούσε την ομάδα ελέγχου (*control group*), όταν ο σχεδιασμός ήταν μία μελέτη δείκτου-ελέγχου ή μελέτη ασθενών-μαρτύρων (*case-control*), και όλο τον πληθυσμό όταν ο σχεδιασμός ήταν αυτός της κοόρτης (*cohort*). Συμπεριλάβαμε μόνο πολυμορφισμούς του ενός νουκλεοτιδίου και αποκλείσαμε απλοτύπους (*haplotypes*) και επαναλήψεις (*copy number variations*). Αν ο ίδιος πολυμορφισμός είχε βρεθεί σε περισσότερες από μία GWAS για τον ίδιο φαινότυπο, τότε αυτό το θεωρήσαμε ως μία γενετική σχέση, αλλά καταγράψαμε όλες τις άλλες μελέτες.

Αρχικά προβήκαμε σε μία περιγραφική ανάλυση των γενετικών σχέσεων που πληρούσαν τα κριτήρια εισαγωγής, την κατανομή των συχνοτήτων των αντίστοιχων πολυμορφισμών, τις τιμές στατιστικής σημαντικότητας P και τα μεγέθη του αποτελέσματος. Χρησιμοποιήσαμε κατάλληλους στατιστικούς μετασχηματισμούς (βλ. (219) για αναλυτική περιγραφή) προκειμένου να εκφράσουμε όλα τα μεγέθη στην ίδια κλίμακα, αυτή των λόγων αναλογιών, ΛΑ, (*odds ratios*).

Ακολούθως, υπολογίσαμε το τυπικό μέγεθος δείγματος των επιλεγμένων μελετών και με βάση αυτό το υπολογισθέν μέγεθος δείγματος εκτιμήσαμε την ισχύ τους, ώστε να ανακαλύψουν γενετικές σχέσεις για διάφορες συχνότητες ελάσσονος αλληλίου και διάφορους ΛΑ σε επίπεδο σφάλματος τύπου I $\alpha=10^{-7}$, υποθέτοντας ένα πολλαπλασιαστικό γενετικό μοντέλο (*multiplicative genetic model*) και ένα ιδανικό σενάριο χωρίς απώλεια ισχύος λόγω των πολλαπλών σταδίων στην επιλογή των SNPs. Για τον υπολογισμό της ισχύος χρησιμοποιήσαμε το πρόγραμμα QUANTO (220). Κατηγοριοποιήσαμε τις γενετικές σχέσεις με βάση τα τεταρτημόρια των ΛΑ και για συχνότητες ελάσσονος αλληλίου 1-2%, 3-4% και 5%. Για κάθε μία από τις προκύπτουσες 12

κατηγορίες υπολογίσαμε την ισχύ G μιας τυπικής GWAS να βρει μία στατιστικά σημαντική σχέση για το διάμεσο ΔA και τη μέση συχνότητα ελάσσονος αλληλίου. Πολλαπλασιάζοντας τον αριθμό των ανακαλυφθέντων πολυμορφισμών σε κάθε κατηγορία με το $1/G$ προκύπτει ο συνολικός αριθμός πολυμορφισμών (δηλαδή αυτών που έχουν και αυτών που δεν έχουν ανακαλυφθεί) για την αντίστοιχη κατηγορία.

Χρησιμοποιώντας το πρόγραμμα WGAViewer (Duke Institute for Genome Sciences & Policy; <http://compute1.lsrc.duke.edu/software/WGAViewer/>) (221) και τις βάσεις δεδομένων Human Genome Browser (University of California Santa Cruz; <http://genome.ucsc.edu/cgi-bin/hgGateway>) (222), dbSNP Build 130 (<http://www.ncbi.nlm.nih.gov/projects/SNP/>) και Ensembl (<http://www.ensembl.org/index.html>), βρήκαμε τη λειτουργική θέση των πολυμορφισμών πάνω στο γονιδίωμα, δηλαδή αν βρίσκονται σε ιντρόνια (*introns*), εξώνια (*exons*) ή επαγωγούς (*promoters*) των αντίστοιχων γονιδίων και αν προκαλούν συνώνυμες (*synonymous*) ή μη-συνώνυμες (*non-synonymous*) αλλαγές.

Επιπλέον, για τις γενετικές σχέσεις που συμπεριλάβαμε, εξετάσαμε αν οι αντίστοιχοι πολυμορφισμοί είχαν βρεθεί και σε άλλες GWAS πάνω στον ίδιο φαινότυπο, ανεξαρτήτως του επιπέδου στατιστικής σημαντικότητας που είχαν σε αυτές τις άλλες μελέτες. Επιπλέον, εξετάσαμε αν αυτές οι GWAS είχαν βρει άλλους πολυμορφισμούς στους ίδιους γενετικούς τόπους με τους πολυμορφισμούς χαμηλής συχνότητας. Όποτε αναφέρονταν τέτοιοι πολυμορφισμοί, καταγράψαμε τις συχνότητές τους, τα μεγέθη αποτελέσματος και τις τιμές στατιστικής σημαντικότητας P . Στη συνέχεια ελέγξαμε αν οι πολυμορφισμοί χαμηλής συχνότητας βρίσκονται σε υψηλή ανισορροπία

διασύνδεσης (*linkage disequilibrium*), δηλαδή $r^2 \geq 0.8$, με άλλους πολυμορφισμούς στους ίδιους γενετικούς τόπους χρησιμοποιώντας το πρόγραμμα SNAP (SNP Annotation and Proxy Server, Broad Institute, Cambridge, MA; <http://www.broadinstitute.org/mpg/snap/>) (223).

Επιπλέον, μέσω αναζητήσεων στη βιβλιογραφική βάση HuGE Pub Lit του συνασπισμού Human Genome Epidemiology, βρήκαμε αν οι γενετικοί τόποι που περιείχαν τους πολυμορφισμούς χαμηλής συχνότητας είχαν μελετηθεί και σε μελέτες υποψήφιων γονιδίων πριν από την ανακάλυψή τους σε μελέτες GWAS. Επίσης, για τους παραπάνω γενετικούς τόπους αναζητήσαμε (<http://www.ncbi.nlm.nih.gov/omim>) εάν έχουν αναφερθεί μεταλλάξεις μενδελιανού τύπου και ποιο φαινότυπο αυτές προκαλούν χρησιμοποιώντας τη βάση OMIM.

Τέλος, χρησιμοποιώντας τα δεδομένα του προγράμματος HapMap (*International HapMap Consortium*; <http://hapmap.ncbi.nlm.nih.gov/>) (76-78), καταγράψαμε τις συχνότητες των εξεταζόμενων πολυμορφισμών σε πληθυσμούς Ευρωπαϊκής, Αφρικανικής και Ασιατικής καταγωγής που έχουν αναλυθεί γενετικά στο πρόγραμμα HapMap.

2.2.3. ΑΠΟΤΕΛΕΣΜΑΤΑ

Εξετάστηκαν 806 μελέτες GWAS, οι οποίες ανασύρθηκαν μέσω αναζήτησης στον Κατάλογο του NHGRI. Συνολικά 43 γενετικές σχέσεις στατιστικά σημαντικές σε επίπεδο GWAS, στις οποίες εμπλέκονταν συνολικά 40 διαφορετικά SNPs χαμηλής συχνότητας από 28 μελέτες GWAS, πληρούσαν τα κριτήρια εισαγωγής (**Πίνακας 8** και **Παράρτημα 2**). Τα SNPs αυτά είχαν χαρτογραφηθεί από τις αντίστοιχες μελέτες σε ένα γονίδιο σε 31 περιπτώσεις, καθώς επίσης 1 SNP είχε χαρτογραφηθεί από μία μελέτη σε ένα γονίδιο και από μία άλλη σε περισσότερα από ένα γονίδια, 4 SNPs είχαν συσχετισθεί με περισσότερα από ένα γονίδια, ενώ 4 SNPs δεν είχαν συσχετισθεί με κάποιο συγκεκριμένο γονίδιο. Συνολικά είχαν αναφερθεί 30 διαφορετικοί γενετικοί τύποι για τις παραπάνω σχέσεις.

Πίνακας 8. Γενετικές σχέσεις με συχνότητα ελάσσονος αλληλίου έως 5% και $P \leq 10^{-7}$.

Νόσος/ Χαρακτηριστικό	Χρωμοσ. Περιοχή	Γενετικός τόπος	Πολυμορφισμός- Αλληλίο κινδύνου	Λειτουργία/ Εντόπιση	ΣΑΚ	Καταγωγή	Τιμή <i>P</i>	ΛΑ (95% ΔΕ)
Οξεία Λεμφοβλαστική Λευχαιμία	12q24.22	<i>KRTHB5</i>	rs2089222-A	Ιντρόνιο	0.03	Ευρωπαϊκή	8×10^{-08}	2.26 (1.60-3.00)
	6p21.33	<i>HCP5, MICB, MCCD1, BAT1, LTB, TNF</i>	rs2395029-G	Εξώνιο	0.03	Ευρωπαϊκή	3×10^{-19}	3.47 (2.39-5.04)
Εξέλιξη AIDS	6p21.3	<i>C6orf48</i>	rs9368699-C	5'-UTR	0.03	Ευρωπαϊκή	2×10^{-11}	Δεν αναφέρεται
Χρώμα ματιών (μπλέ έναντι πράσινο)	15q13.1	<i>OCA2</i>	rs1667394-A	Ιντρόνιο	0.97	Ευρωπαϊκή	2×10^{-53}	3.67 (2.67-5.05)
Φακίδες	16q24.3	<i>MC1R</i>	rs1805007-T	Εξώνιο	0.05	Ευρωπαϊκή	1×10^{-96}	3.33 (2.92-3.80)
	13q14	<i>AKAP11</i>	rs180851-C	Διαγονιδιακά	0.95	Ευρωπαϊκή	2×10^{-12}	1.46 (1.31-1.63)
	13q14	<i>AKAP11</i>	rs7326472-A	Διαγονιδιακά	0.95	Ευρωπαϊκή	1×10^{-10}	1.39 (1.24-1.54)
	13q14	<i>AKAP11</i>	rs12854504-T	Διαγονιδιακά	0.95	Ευρωπαϊκή	1×10^{-10}	1.39 (1.24-1.54)
	13q14	<i>AKAP11</i>	rs7998154-T	Διαγονιδιακά	0.02	Ευρωπαϊκή	2×10^{-08}	1.75 (1.46-2.10)
	13q14	<i>TNFSF11</i>	rs6561055-G	Διαγονιδιακά	0.95	Ευρωπαϊκή	3×10^{-10}	1.39 (1.24-1.54)
Μέση οστική πυνότητα στην ΟΜΣΣ	13q14	<i>TNFSF11</i>	rs17639156-T	Διαγονιδιακά	0.95	Ευρωπαϊκή	5×10^{-10}	1.39 (1.24-1.54)
Νοητικές λειτουργίες	Xp22.2	<i>HCCS</i>	rs5934953-C	Ιντρόνιο	0.02	Ευρωπαϊκή	1×10^{-07}	Δεν αναφέρεται
	16q12.1	<i>NOD2</i>	rs2066844-T	Εξώνιο	0.05	Ευρωπαϊκή	1×10^{-18}	2.48 (1.98-3.10)
	16q12.1	<i>NOD2</i>	rs2066845-C	Εξώνιο	0.01	Ευρωπαϊκή	8×10^{-10}	3.04 (2.09-4.42)
	16q12.1	<i>NOD2</i>	rs2066847-C	Frameshift coding	0.04	Ευρωπαϊκή	3×10^{-49}	4.30 (3.42-5.42)
Νόσος Crohn	12q12	<i>LRRK2, MUC19</i>	rs11175593-T	Ιντρόνιο	0.02	Ευρωπαϊκή	3×10^{-10}	1.54 (1.34-1.76)
	20q13.12	<i>HNF4A</i>	rs1800961-C	Εξώνιο	0.97	Ευρωπαϊκή	8×10^{-10}	1.41 (1.27-1.57)
HDL χοληστερόλη	9q31.1	<i>ABCA1</i>	rs9282541-T	Εξώνιο	0.03	Μικτή	5×10^{-08}	1.33 (1.21-1.45)
Αιματοκρίτης	6p22.1	<i>HFE</i>	rs1800562-A	Εξώνιο	0.04	Ευρωπαϊκή	2×10^{-09}	1.74 (1.45-2.09)
Αιμοσφαιρίνη	6p22.1	<i>HFE</i>	rs1800562-A	Εξώνιο	0.04	Ευρωπαϊκή	6×10^{-19}	1.33 (1.25-1.42)
LDL χοληστερόλη	1p32.3	<i>PCSK9</i>	rs11591147-G	Εξώνιο	0.99	Ευρωπαϊκή	2×10^{-44}	2.34 (2.07-2.64)
Μέση περιεκτικότητα ερυθρών	6p22.2	<i>SLC17A3</i>	rs1408272-G	Άγνωστη	0.03	Ευρωπαϊκή	4×10^{-39}	1.03 (1.02-1.04)
Μέσος όγκος ερυθρών	6p22.1	<i>HFE</i>	rs1800562-A	Εξώνιο	0.04	Ευρωπαϊκή	1×10^{-23}	12.83 (7.73-20.92)

Μη-συνδρομική λαγοχειλία	18q22.3	Διαγονιδιακά	rs17085106-T	Διαγονιδιακά	0.02	Ευρωπαϊκή	4×10 ⁻⁰⁸	4.07 (2.37-7.00)
	3p22.2	<i>ITGA9</i>	rs189897-A	Ιντρόνιο	0.03	Ασιατική	7×10 ⁻⁰⁸	3.18 (1.94-5.21)
Ρινοφαρυγγικό καρκίνωμα	3p22.2	<i>ITGA9</i>	rs197757-T	Ιντρόνιο	0.03	Ασιατική	1×10 ⁻⁰⁷	3.09 (1.89-5.05)
	12p13.31	<i>TMEM16B</i>	rs12579350-A	Ιντρόνιο	0.01	Ασιατική	4×10 ⁻⁰⁹	22.11 (5.30-92.14)
Διαταραχή πανικού	1q32.1	<i>PKP1</i>	rs860554-T	Ιντρόνιο	0.05	Ασιατική	5×10 ⁻⁰⁸	4.03 (2.40-6.76)
Καρκίνος προστάτη	8q24.21	Διαγονιδιακά	rs16901979-A	Διαγονιδιακά	0.03	Ευρωπαϊκή	1×10 ⁻¹²	1.79 (1.53-2.11)
Πρωτοπαθής χολική κίρρωση	6p21.3	<i>C6orf10</i>	rs2395148-A	Ιντρόνιο	0.02	Ευρωπαϊκή	4×10 ⁻¹⁴	2.87 (2.16-3.82)
Αλκαλική φωσφατάση	12q12	<i>PDZRN4, CNTN1</i>	rs1880887-C	Ιντρόνιο	0.03	Ευρωπαϊκή	1×10 ⁻¹⁰	Δεν αναφέρεται
Ελεύθερη Τ3	17p12	<i>HS3ST3B1</i>	rs3848445-C	Άγνωστη	0.05	Ευρωπαϊκή	8×10 ⁻⁰⁹	Δεν αναφέρεται
Ψωρίαση	6p21.33	<i>HLA-C</i>	rs2395029-C	Εξώνιο	0.03	Ευρωπαϊκή	2×10 ⁻²⁶	4.10 (3.10-5.30)
Ανταπόκριση στη θεραπεία για ΟΛΛ	10p12.33	<i>ST8SIA6</i>	rs359312-T	Ιντρόνιο	0.04	Μικτή	9×10 ⁻⁰⁸	3.91 (1.52-10.10)
	2p12	Διαγονιδιακά	rs17022444-G	Διαγονιδιακά	0.03	Μικτή	1×10 ⁻¹⁰	Δεν αναφέρεται
Ανταπόκριση σε αντιψυχωσικά	4q24	Διαγονιδιακά	rs7669317-C	Διαγονιδιακά	0.04	Μικτή	8×10 ⁻⁰⁸	Δεν αναφέρεται
	6q23.3	<i>TNFAIP3</i>	rs5029939-G	Ιντρόνιο	0.03	Ευρωπαϊκή	3×10 ⁻¹²	2.28 (1.80-2.88)
Συστηματικός ερυθματώδης λύκος	6q23.3	<i>TNFAIP3</i>	rs2230926-C	Εξώνιο	0.04	Ασιατική	1×10 ⁻¹⁷	1.72 (1.52-1.94)
Μελάχρωση δέρματος	5p13.3	<i>MATP</i>	rs35391-C	Ιντρόνιο	0.97	Ευρωπαϊκή	3×10 ⁻¹⁰	2.22 (1.72-2.86)
	11q23.3	<i>APOA1/C3/A4/A5</i>	rs662799-G	Upstream	0.05	Ευρωπαϊκή	2×10 ⁻¹⁵	1.31 (1.22-1.40)
Τριγλυκερίδια	11q23.3	<i>APOA1/C3/A4/A5, DSCAML1</i>	rs10892151-A	Ιντρόνιο	0.03	Ευρωπαϊκή	3×10 ⁻²⁹	Δεν αναφέρεται
Σακχαρώδης διαβήτης τύπου 1	7p12.1	<i>COBL</i>	rs4948088-C	Άγνωστη	0.95	Ευρωπαϊκή	4×10 ⁻⁰⁸	1.30 (1.11-1.49)
Σακχαρώδης διαβήτης τύπου 2	10q25.2	<i>TCF7L2</i>	rs7903146-T	Ιντρόνιο	0.04	Ασιατική	8×10 ⁻¹²	1.54 (1.36-1.74)

Συντομογραφίες: ΔΕ: διάστημα εμπιστοσύνης, ΟΛΛ: οξεία λεμφοβλαστική λευχαιμία, ΟΜΣΣ: οσφυική μοίρα σπονδυλικής στήλης, Τ3: τρι-ιωδο-θυρονίνη, ΣΑΚ: συχνότητα αλληλίου κινδύνου, ΛΑ: λόγος αναλογιών, ΔΕ: διάστημα εμπιστοσύνης

Οι φαινότυποι για αυτές τις 43 γενετικές σχέσεις ήταν οι ακόλουθοι: εξέλιξη AIDS (n=2 γενετικές σχέσεις), οστική πυκνότητα (n=6), νόσος Crohn (n=4), επίπεδα HDL χοληστερόλης (n=2), ρινοφαρυγγικό καρκίνωμα (n=2), διαταραχή πανικού (n=2), ανταπόκριση στην αντιψυχωσική θεραπεία (n=2), συστηματικός ερυθματώδης λύκος (n=2), επίπεδα τριγλυκεριδίων (n=2), οξεία λεμφοβλαστική λευχαιμία, χρώμα ματιών, νοητικές λειτουργίες, φακίδες, επίπεδα LDL χοληστερόλης, επίπεδα αιματοκρίτη, επίπεδα αιμοσφαιρίνης, μέση περιεκτικότητα ερυθρών, μέσος όγκος ερυθρών, μη-συνδρομική λαγοχειλία (non-syndromic cleft lip), καρκίνος προστάτη, επίπεδα αλκαλικής φωσφατάσης, επίπεδα ελεύθερης τριωδοθυρονίνης, πρωτοπαθής χολική κίρρωση, ψωρίαση, ανταπόκριση στη θεραπεία για οξεία λεμφοβλαστική λευχαιμία, μαύρισμα, σακχαρώδης διαβήτης τύπου 1 και σακχαρώδης διαβήτης τύπου 2.

Εντόπιση και Λειτουργία

Εννιά από τα 40 SNPs (22.5%) ήταν μη συνώνυμα, 15 (37.5%) βρίσκονταν σε ιντρόνια, 10 (25%) ήταν μεταξύ γονιδίων, 1 βρισκόταν στην 5'-αμετάφραστη περιοχή (5'-UTR), 1 βρισκόταν στο πλαίσιο ανάγνωσης του αντίστοιχου γονιδίου και για 3 δεν υπήρχε κάποια σχετική πληροφορία.

Συχνότητες και μεγέθη αποτελέσματος

Από τις 22 σχέσεις που αφορούσαν νοσήματα παρά ποσοτικά χαρακτηριστικά ή φυσιολογικούς φαινοτύπους, 21 είχαν αλληλία κινδύνου με συχνότητες των αλληλίων αυτών 5% ή λιγότερο και μόνο για μία σχέση το αλληλίο κινδύνου είχε συχνότητα 95%. Η τελευταία ήταν μάλιστα η σχέση με τον μικρότερο λόγο αναλογιών. Συνολικά 33 σχέσεις είχαν ανακαλυφθεί και επαληθευθεί αποκλειστικά σε πληθυσμούς ευρωπαϊκής καταγωγής, ενώ 10

είχαν ανακαλυφθεί ή/και επαληθευθεί σε μη-ευρωπαϊκούς ή μικτούς πληθυσμούς.

Έντεκα από τις 43 σχέσεις είχαν τιμές στατιστικής σημαντικότητας P μεταξύ 10^{-7} και 10^{-8} , ενώ 32 είχαν μεγαλύτερα επίπεδα στατιστικής σημαντικότητας. Οι λόγοι αναλογιών ήταν διαθέσιμοι ή υπολογίστηκαν για 36 γενετικές σχέσεις (για 7 σχέσεις δεν υπήρχαν διαθέσιμα δεδομένα). Οι λόγοι αναλογιών κυμαίνονταν από 1,03 (για το rs1408272 που καθορίζει τα επίπεδα της μέσης περιεκτικότητας αιμοσφαιρίνης) έως 22,11 (για το rs12579350 για τη διαταραχή πανικού). Ο διάμεσος λόγος αναλογιών ήταν 2,24 με ενδοτεταρτημοριακό εύρος 1,40-3,40).

Συχνότητες και μεγέθη αποτελέσματος

Το μέσο μέγεθος δείγματος που χρησιμοποιήθηκε στις 28 μελέτες GWAS ήταν 7.637 άτομα για τις μελέτες δείκτου-ελέγχου και 10.647 άτομα για όλα τα είδη μελετών (μελέτες κοόρτης και μελέτες δείκτου-ελέγχου). Στον **Πίνακα 9** φαίνονται οι ανακαλυφθείσες γενετικές σχέσεις για χαμηλής συχνότητας πολυμορφισμούς, χωρισμένες με βάση τα τεταρτημόρια του λόγου αναλογιών και τις ομάδες συχνότητας ελάσσονος αλληλίου 1-2%, 3-4% και 5%. Όπως φαίνεται, δεν περιλαμβάνεται κανένας πολυμορφισμός με λόγο αναλογιών $<1,40$ και συχνότητα ελάσσονος αλληλίου 1-2%, καθώς η ισχύς για να ανιχνευθούν τέτοιοι πολυμορφισμοί με το τυπικό μέγεθος δείγματος είναι ελάχιστη (0.37%). Οι υπολογισμοί ισχύος δείχνουν ότι μόνο 11% και 23% των πολυμορφισμών με παρόμοιους λόγους αναλογιών ($<1,40$) και συχνότητες ελάσσονος αλληλίου 3-4% και 5% αντίστοιχα, θα είχαν ανακαλυφθεί με το μέσο μέγεθος δείγματος των μελετών GWAS που εξετάσαμε. Πολυμορφισμοί με λόγους αναλογιών 1,40-2,24 και συχνότητες ελάσσονος αλληλίου 1-2% είχαν

56% πιθανότητα να ανακαλυφθούν. Σε όλες τις άλλες κατηγορίες λόγων αναλογιών και συχνοτήτων ελάσσονος αλληλίου η ισχύς ήταν >99%. Αυτό σημαίνει ότι με ένα μέγεθος δείγματος της τάξης του 10.647 θα ήταν δυνατόν να ανακαλυφθούν σχεδόν όλοι οι πολυμορφισμοί με λόγο αναλογιών >1,40 και συχνότητα ελάσσονος αλληλίου 3-5%, καθώς και αυτοί με λόγο αναλογιών >3 και συχνότητα ελάσσονος αλληλίου >1%. Τέλος, η ερμηνεία αυτών των υπολογισμών ισχύος υποδηλώνει ότι ο αριθμός των πολυμορφισμών με λόγο αναλογιών <1,40 και συχνότητα ελάσσονος αλληλίου 3-5% είναι πιθανόν τριπλάσιος αυτών των πολυμορφισμών με λόγο αναλογιών >1,40 και παρόμοια συχνότητα ελάσσονος αλληλίου, αλλά οι τελευταίοι είναι πολύ πιο εύκολο να ανακαλυφθούν με βάση το τυπικό μέγεθος δείγματος σε αυτές τις μελέτες GWAS.

Πίνακας 9. Αριθμός των ανακαλυφθεισών (Α) και προσδοκώμενων (Π) γενετικών σχέσεων* με πολυμορφισμούς χαμηλής συχνότητας με βάση τα τεταρτημόρια του λόγου αναλογιών και τις κατηγορίες της συχνότητας ελάσσονος αλληλίου.

Συχνότητα ελάσσονος αλληλίου	Εύρος λόγου αναλογιών σε τεταρτημόρια (διάμεσος)							
	<1,40 (1,33)		1,40-<2,24 (1,72)		2,24-3,40 (2,87)		>3,40 (4,07)	
	A	Π	A	Π	A	Π	A	Π
1-2%	0	– **	2	4	3	3	2	2
3-4%	3	28	6	6	4	4	6	6
5%	6	26	1	1	2	2	1	1

* Για 7 γενετικές σχέσεις δεν ήταν δυνατόν να αντληθούν δεδομένα για τα αντίστοιχα μεγέθη αποτελέσματος.

** Ο υπολογισμός δεν ήταν δυνατός

Πολυμορφισμοί στους ίδιους γενετικούς τόπους σε άλλες μελέτες GWAS

Για 37 από τις 43 σχέσεις βρέθηκε τουλάχιστον άλλη 1 μελέτη GWAS στον ίδιο φαινότυπο. Καμία επιπλέον μελέτη GWAS, πέραν αυτής που είχε ήδη περιληφθεί στην παρούσα μελέτη, δεν βρέθηκε για 6 γενετικές σχέσεις (φακίδες, διαταραχή πανικού [n=2], πρωτοπαθής χολική κίρρωση, ελεύθερη T3, ανταπόκριση στη θεραπεία για οξεία λεμφοβλαστική λευχαιμία).

Για 15 σχέσεις, επιπλέον μελέτες GWAS παρουσίαζαν δεδομένα για τον ίδιο χαμηλής συχνότητας πολυμορφισμό (n=16) ή/και για άλλο πολυμορφισμό στον ίδιο γενετικό τόπο (n=74), όπως φαίνεται στον **Πίνακα 10**. Για μία σχέση (καρκίνος προστάτη και rs16901979), δεν βρέθηκε άλλος επιπλέον πολυμορφισμός εκτός από εκείνον χαμηλής συχνότητας που συμπεριλήφθηκε στη μελέτη. Έτσι, για 14 γενετικές σχέσεις πολυμορφισμών χαμηλής συχνότητας (όπου εμπλέκονταν 10 γενετικοί τόποι) άλλες μελέτες GWAS είχαν βρει 1 ή περισσότερους κοινούς πολυμορφισμούς (συχνότητα ελάσσονος αλληλίου >5%) στον ίδιο γενετικό τόπο με τον πολυμορφισμό χαμηλής συχνότητας. Για 4 σχέσεις (χρώμα ματιών, LDL χοληστερόλη, τριγλυκερίδια, σακχαρώδης διαβήτης τύπου 2) οι ίδιες επιπλέον μελέτες GWAS παρουσίαζαν δεδομένα τόσο για τον πολυμορφισμό χαμηλής συχνότητας όσο και για έναν ή περισσότερους κοινούς πολυμορφισμούς.

Όποτε ο ίδιος πολυμορφισμός χαμηλής συχνότητας είχε βρεθεί από επιπρόσθετες μελέτες GWAS (8 τέτοιοι πολυμορφισμοί σε 16 επιπρόσθετες μελέτες GWAS), ο λόγος αναλογιών ήταν μεγαλύτερος από τον αντίστοιχο που είχε προταθεί στην πρώτη χρονολογικά μελέτη, φτάνοντας επίπεδα στατιστικής σημαντικότητας $P < 10^{-7}$ σε 6 περιπτώσεις, και μικρότερος σε 10 περιπτώσεις. Δώδεκα από τους 16 εκτιμητές του μεγέθους αποτελέσματος ήταν στατιστικά

σημαντικοί σε επίπεδο μελετών GWAS, ενώ όλοι οι 16 ήταν σημαντικοί σε επίπεδο $P=0.05$.

Όταν άλλες μελέτες GWAS παρουσίαζαν άλλους πολυμορφισμούς στον ίδιο γενετικό τόπο, σχεδόν όλοι (72/74) οι επιπλέον πολυμορφισμοί ήταν κοινοί (συχνότητα ελάσσονος αλληλίου $>5\%$). Οι λόγοι αναλογιών με βάση το αλληλίο κινδύνου ήταν μικρότεροι από αυτούς του πολυμορφισμού χαμηλής συχνότητας που συμπεριλήφθηκε στη μελέτη μας με 14 εξαιρέσεις. Πενήντα από τις 74 επιπρόσθετες σχέσεις ήταν στατιστικά σημαντικές σε επίπεδο GWAS και 65 ήταν στατιστικά σημαντικές σε επίπεδο $P=0.05$.

Ανάλυση αυτών των πολυμορφισμών στο εργαλείο SNAP έδειξε ότι οι 2 πολυμορφισμοί χαμηλής συχνότητας στο *TNFAIP3* που σχετίζονταν με συστηματικό ερυθματώδη λύκο σε 2 διαφορετικές μελέτες GWAS βρίσκονταν σε υψηλή ανισορροπία διασύνδεσης ($r^2=1$ και $D'=1$ τόσο σε Ευρωπαίους όσο και σε Ασιάτες). Επιπλέον, ο πολυμορφισμός rs180851 που σχετιζόταν με την οστική μάζα, βρισκόταν σε υψηλή ανισορροπία διασύνδεσης ($r^2=0.82$ και $D'=1$ για τις ανά ζεύγη συγκρίσεις) με τους πολυμορφισμούς χαμηλής συχνότητας rs7326472 και rs12854504. Επίσης, σε υψηλή ανισορροπία διασύνδεσης ($r^2=1$ και $D'=1$) μεταξύ τους βρίσκονταν οι πολυμορφισμοί χαμηλής συχνότητας rs7326472 & rs12854504 και rs6561055 & rs17639156. Ακόμη, ο σχετιζόμενος με σακχαρώδη διαβήτη τύπου 2 πολυμορφισμός rs7903146 στο γονίδιο *TCF7L2* βρισκόταν σε ανισορροπία διασύνδεσης με τον rs7901695 ($r^2=1$ και $D'=1$) και τον rs4506565 ($r^2=1$ και $D'=1$), που είχαν βρεθεί από 3 και 1 αντίστοιχα προηγούμενες μελέτες GWAS. Τόσο ο rs7901695 όσο και ο rs4506565 είναι κοινοί πολυμορφισμοί στους Ευρωπαίους που χρησιμοποιήθηκαν σε αυτές τις μελέτες, αλλά όχι στους Ιάπωνες όπου ο rs7903146 ήταν στατιστικά σημαντικός

σε επίπεδο GWAS. Κανένας από τους άλλους πολυμορφισμούς στους ίδιους γενετικούς τόπους με τους πολυμορφισμούς χαμηλής συχνότητας δεν ήταν σε ανισορροπία διασύνδεσης με αυτούς με βάση το r^2 . Εκτός από τις παραπάνω σχέσεις που είχαν $r^2=1$ και $D'=1$, άλλα 41 ζεύγη πολυμορφισμών είχαν $D'=1$ αλλά $r^2<1$.

Πίνακας 10. Πολυμορφισμοί στους ίδιους γενετικούς τόπους με τους πολυμορφισμούς χαμηλής συχνότητας, όπως περιγράφονται σε άλλες GWAS για τον ίδιο φαινότυπο

Φαινότυπος	Πολυμορφισμός Χαμηλής Συχνότητας	Γενετικός Τόπος	Άλλη Μελέτη GWAS με Πολυμορφισμούς στον Ίδιο Γενετικό Τόπο	Δημοσίευση της Άλλης Μελέτης GWAS	Άλλος Πολυμορφισμός	Συχνότητα Αλληλίου Κινδύνου	Τιμή <i>P</i>	Λόγος Αναλογιών
Μπλέ/ πράσινα μάτια	rs1667394-A	OCA2	(224)	Μεταγενέστερη	(ίδιος)	0.13 ^a	3×10 ⁻⁸⁷	1.82
					rs7495174	0.05 ^a	0.018	1.60
					(ίδιος) ^e	0.15	8.5×10 ⁻³¹	Δ/Α ^f
					rs11855019	0.19	8.6×10 ⁻²⁵	Δ/Α
					rs6497268	0.18	3.7×10 ⁻¹⁹	Δ/Α
Μέση οστική πυκνότητα (ΟΜΣΣ)	rs6561055-A rs17639156-G	TNFSF11	(226)	Πρότερη	rs7495174	0.10	2×10 ⁻²²	Δ/Α
					rs9533093	0.80	5.4×10 ⁻¹¹	1.22
					rs9594738	0.42 ^a	4×10 ⁻²³	1.34
					rs9594759	0.62	1.5×10 ⁻¹⁷	1.24
					rs9594759	0.49 ^a	Δ/Α	Δ/Α
			(227)	Πρότερη	rs9594738	0.42 ^a	Δ/Α	Δ/Α
					rs9594759	0.63	1.1×10 ⁻¹⁶	1.27
					rs10507507	0.82	1.6×10 ⁻⁰⁵	1.26
					rs7992970	0.78	8.5×10 ⁻⁰⁷	1.27
					rs9594738	0.56	2×10 ⁻²¹	1.36

Νόσος Crohn	rs2066844-T	<i>NOD2</i>	(229)	Πρότερη	rs2076756	0.35 ^a	5.1×10 ⁻¹⁰	Δ/Α
	rs2066845-C				rs2066843	0.36 ^a	2.9×10 ⁻⁰⁹	Δ/Α
	rs2066847-C		(230)	Πρότερη	rs2076756	0.24	7×10 ⁻¹⁴	Δ/Α
			(231)	Πρότερη	rs5743289	0.17	3.8×10 ⁻¹⁰	1.45
					rs17221417	0.29	9.4×10 ⁻¹²	1.29
			(232)	Μεταγενέστερη	(ίδιος: rs2066847)	0.02	3×10 ⁻²⁴	3.99
			(233)	Μεταγενέστερη	rs2076756	0.26	9.7×10 ⁻⁰⁸	1.33
HDL χοληστερόλη	rs9282541-T	<i>ABCA1</i>	(234)	Μεταγενέστερη	rs2076756	0.35 ^a	1×10 ⁻⁰⁹	Δ/Α
			(235)	Ταυτόχρονη	rs3890182	0.87	3×10 ⁻¹⁰	1.19 ^b
			(236)	Ταυτόχρονη	rs4149268	0.35	1.2×10 ⁻¹⁰	1.10 ^{b, c}
					rs4149274	0.69	7.4×10 ⁻⁰⁸	1.20 ^{b, c}
			(237)	Ταυτόχρονη	rs3890182	0.12	2×10 ⁻⁰⁶	5.58 ^{b, c}
			(238)	Μεταγενέστερη	rs3905000	0.86	8.6×10 ⁻¹³	1.24 ^b
					rs3847303	0.88	3.4×10 ⁻¹²	1.25 ^b
			(239)	Μεταγενέστερη	rs1883025	0.26	1×10 ⁻⁰⁹	1.16 ^b
			(240)	Μεταγενέστερη	rs4149268	0.27 ^a	0.69	1.20 ^b
			(241)	Μεταγενέστερη	rs2740491	0.36	3.1×10 ⁻⁰⁴	1.06 ^b
					rs3847303	0.13	3.2×10 ⁻⁰³	1.06 ^b
Αιμοσφαιρίνη	rs1800562-A	<i>HFE</i>	(242)	Ταυτόχρονη	(ίδιος)	0.04	1.6×10 ⁻⁰⁴	1.25 ^b
			(243)	Ταυτόχρονη	rs198833	0.08 ^a	1.4×10 ⁻⁰⁸	Δ/Α
					rs129128	0.08 ^a	3.3×10 ⁻⁰⁸	Δ/Α

					rs198851	0.08 ^a	3.4×10 ⁻⁰⁸	Δ/A
					rs1799945	0.14 ^a	4.3×10 ⁻⁰⁸	Δ/A
					rs198846	0.11 ^a	8.7×10 ⁻⁰⁶	1.23 ^b
LDL χοληστερόλη	rs11591147-G	<i>PCSK9</i>	(239)	Μεταγενέστερη	(ίδιος)	0.02	9×10 ⁻⁰⁶	2.65 ^b
					rs11206510	0.19	4×10 ⁻⁰⁸	1.17 ^b
			(240)	Μεταγενέστερη	(ίδιος)	0.02	1.6×10 ⁻⁰⁷	1.88 ^b
			(236)	Ταυτόχρονη	rs11206510	0.81	3.5×10 ⁻¹¹	1.15 ^{b, c}
			(241)	Μεταγενέστερη	rs11206510	0.01	2×10 ⁻¹²	1.33 ^b
			(244)	Ταυτόχρονη	(ίδιος)	0.04	1×10 ⁻⁴⁶	1.02 ^b
Μέσος όγκος ερυθρών	rs1800562-A	<i>HFE</i>	(243)	Ταυτόχρονη	rs198846	0.11	8.6×10 ⁻¹³	4.91 ^b
Καρκίνος προστάτη	rs16901979-A	<i>Διαγονιδιακά</i>	(245)	Μεταγενέστερη	(ίδιος)	0.04	2.5×10 ⁻¹⁴	1.80
Ψωρίαση	rs2395029-C	<i>HLA-C</i>	(246)	Πρότερη	rs3134792	0.15 ^a	1×10 ⁻⁰⁹	Δ/A
			(247)	Μεταγενέστερη	rs12191877	0.15	<1×10 ⁻¹⁰⁰	2.64
Τριγλυκερίδια	rs662799-G	<i>APOA1, APOC3, APOA4, APOA5</i>	(248)	Πρότερη	rs481843	0.11	3.3×10 ⁻⁰⁵	NR
			(235)	Ταυτόχρονη	rs28927680	0.07	2×10 ⁻¹⁷	1.60 ^b
			(249)	Ταυτόχρονη	rs2075292	0.16	5.3×10 ⁻⁰⁸	1.10 ^{b, c}
					rs7124741	0.17	8.6×10 ⁻⁰⁷	1.10 ^{b, c}
					rs17120139	0.17	2.3×10 ⁻⁰⁶	1.09 ^{b, c}
			(237)	Ταυτόχρονη	rs6589566	0.06	3×10 ⁻¹¹	10.90 ^{b, c}
			(240)	Μεταγενέστερη	(ίδιος)	0.06	2.9×10 ⁻¹⁵	1.60 ^b

Τριγλυκερίδια	rs10892151-A	APOA1, APOC3, APOA4, APOA5, DSCAML1	(238)	Μεταγενέστερη	rs3135506	0.06	5.5×10 ⁻¹²	1.57 ^b
					rs12272004	0.93	5.4×10 ⁻¹³	1.39 ^b
					rs480878	0.86	8×10 ⁻⁰⁹	1.19 ^b
					rs28927680	0.93	3.9×10 ⁻⁰⁹	1.64 ^b
					rs12292921	0.07	9.1×10 ⁻¹³	1.39 ^b
					rs35120633	0.93	2.3×10 ⁻¹⁰	1.75 ^b
					rs3135506	0.06	7.4×10 ⁻¹⁰	1.74 ^b
					rs2075292	0.13	5.7×10 ⁻¹²	1.23 ^b
					rs588918	0.87	4.9×10 ⁻⁰⁸	1.19 ^b
					rs1351452	0.86	7.4×10 ⁻¹⁰	1.23 ^b
			(239)	Μεταγενέστερη	rs964184	0.14	4×10 ⁻⁶²	1.72 ^b
			(241)	Μεταγενέστερη	rs12292921	0.06	1.4×10 ⁻⁰³	1.20 ^b
			(239)	Πρότερη	rs964184	0.14	4×10 ⁻⁶²	1.72 ^b
			(235)	Πρότερη	rs28927680	0.07	2×10 ⁻¹⁷	1.60 ^b
			(240)	Πρότερη	rs3135506	0.06	5.5×10 ⁻¹²	1.57 ^b
			(236)	Πρότερη	rs662799	0.06	3×10 ⁻¹⁵	1.60 ^b
					rs12286037	0.94	1×10 ⁻²⁶	1.51 ^{b, c}
					rs2075292	0.16	5.3×10 ⁻⁰⁸	1.10 ^{b, c}
			(249)	Πρότερη	rs7124741	0.17	8.6×10 ⁻⁰⁷	1.10 ^{b, c}
					rs17120139	0.17	2.3×10 ⁻⁰⁶	1.09 ^{b, c}
					rs481843	0.11	3.3×10 ⁻⁰⁵	NR
			(238)	Πρότερη	rs12272004	0.93	5.4×10 ⁻¹³	1.39 ^b

Σακχαρώδης Διαβήτης τύπου 2	rs7903146-T	TCF7L2		rs480878	0.86	8×10 ⁻⁰⁹	1.19 ^b	
				rs28927680	0.93	3.9×10 ⁻⁰⁹	1.64 ^b	
				rs12292921	0.07	9×10 ⁻¹³	1.39 ^b	
				rs35120633	0.93	2.3×10 ⁻¹⁰	1.75 ^b	
				rs3135506	0.06	7.4×10 ⁻¹⁰	1.74 ^b	
				rs2075292	0.13	5.7×10 ⁻¹²	1.23 ^b	
				rs588918	0.87	4.9×10 ⁻⁰⁸	1.19 ^b	
				rs1351452	0.86	7.4×10 ⁻¹⁰	1.23 ^b	
			(237)	Πρότερη	rs6589566	0.06	3×10 ⁻¹¹	10.90 ^{b, c}
			(241)	Ταυτόχρονη	rs12292921	0.06	1.4×10 ⁻⁰³	1.20 ^b
			(250)	Πρότερη	(ίδιος)	0.25 ^a	4.2×10 ⁻¹⁵	1.43
					rs7901695	0.28 ^a	8.3×10 ⁻¹³	1.37
			(251)	Πρότερη	(ίδιος)	0.25 ^a	3×10 ⁻²³	1.37
			(252)	Πρότερη	(ίδιος)	0.25 ^a	5.5×10 ⁻⁰⁸	1.71
					rs7901695	0.28 ^a	3.4×10 ⁻⁰⁷	1.66
					rs12255372	0.22 ^a	5.3×10 ⁻⁰⁷	1.64
			(253)	Πρότερη	(ίδιος)	0.18	1×10 ⁻⁴⁸	1.37
			(254)	Πρότερη	(ίδιος)	0.29	1.5×10 ⁻³⁴	1.65
			(255)	Πρότερη	(ίδιος)	0.49	0.005 ^d	1.28 ^d
					rs7100927	0.49	0.007 ^d	1.56 ^d
			(256)	Μεταγενέστερη	(ίδιος)	0.27	1.2×10 ⁻³⁰	1.48
			(231)	Πρότερη	rs4506565	0.32	5.7×10 ⁻¹³	NR

(257)	Πρότερη	rs7901695	0.28 ^a	1×10 ⁻⁴⁸	1.37
(258)	Πρότερη	rs7100927	0.40 ^a	0.007 ^d	1.56 ^d
		rs10509966	0.25 ^a	0.64 ^d	1.09 ^d
		rs10509969	0.20 ^a	0.60 ^d	1.12 ^d
		rs290483	0.42 ^a	0.93 ^d	1.01 ^d
		rs7917983	0.39 ^a	0.17 ^d	1.22 ^d
		rs10509970	0.23 ^a	0.51 ^d	1.14 ^d
		rs10509967	0.26 ^a	0.82 ^d	1.04 ^d

Συντομογραφίες: GWAS: genome-wide association study, Δ/A: δεν αναφέρεται

^a Η συχνότητα αλληλίου κινδύνου προέρχεται από το HarMap γιατί δεν αναφερόταν στη μελέτη

^b Το ισοδύναμο του λόγου αναλογιών υπολογίστηκε από τη σταθμισμένη μέση διαφορά

^c Το ισοδύναμο του λόγου αναλογιών υπολογίστηκε από τη μέση διαφορά χρησιμοποιώντας την τυπική απόκλιση από τη μελέτη NHANES για τα επίπεδα της LDL χοληστερόλης, της HDL χοληστερόλης και των τριγλυκεριδίων γιατί οι τυπικές αποκλίσεις δεν παρέχονταν στις δημοσιευμένες μελέτες.

^d Οι τιμές P και οι αναλογίες κινδύνου προέρχονται από μοντέλα αναλογικών κινδύνων Cox για αναλύσεις επιβίωσης.

^e Το rs1667394 αναφερόταν ότι εντοπίζεται στο γονίδιο *HERC2*.

^f Ο λόγος αναλογιών δεν ήταν διαθέσιμος από τα αναφερόμενα δεδομένα, αλλά με βάση την τιμή P και το μέγεθος δείγματος συγκρίναμε τα δύο μεγέθη αποτελέσματος και συμπεράναμε ότι ο μη-αναφερόμενος όγκος αναλογιών είναι μάλλον μικρότερος από αυτόν του μη-συχνού πολυμορφισμού.

Προηγούμενη βιβλιογραφία

Στη βάση του HuGE Navigator βρέθηκαν 2 προηγούμενες μελέτες για τη σχέση ανάμεσα στην εξέλιξη του AIDS και το γενετικό τόπο *HCP-TNF*, 3 μελέτες για τη σχέση ανάμεσα στην οστική πυκνότητα και τον γενετικό τόπο *TNFSF11*, 176 μελέτες για τη σχέση ανάμεσα στη νόσο του Crohn και το *NOD2*, 1 μελέτη για τη σχέση της νόσου του Crohn με το *MUC19*, 3 μελέτες για την HDL χοληστερόλη και το *HNF4A*, 34 μελέτες για την HDL χοληστερόλη και το *ABCA1*, 3 μελέτες για τη σχέση ανάμεσα στο *HFE* και τον αιματοκρίτη, 17 μελέτες για το *HFE* και την αιμοσφαιρίνη, 10 μελέτες για την LDL χοληστερόλη και το *PCSK9*, 3 μελέτες για τη σχέση του *HFE* και του μέσου όγκου ερυθρών αιμοσφαιρίων, 57 μελέτες για τη σχέση ψωρίασης και *HLA-C*, 118 μελέτες για τα διάφορα γονίδια του συμπλέγματος *APOA1-APOC3-APOA4-APOA5* και τα επίπεδα τριγλυκεριδίων, και τέλος, 14 μελέτες για τη σχέση του σακχαρώδους διαβήτη τύπου 2 με το *TCF7L2*. Τα ανωτέρω ευρήματα συνοψίζονται στον **Πίνακα 11**.

Με βάση τα σχόλια των συγγραφέων των μελετών GWAS, πολλοί από τους γενετικούς τόπους με πολυμορφισμούς χαμηλής συχνότητας είχαν κάποια τεκμηρίωση από προηγούμενες μελέτες, αν και όχι απαραίτητα από μελέτες υποψηφίων γονιδίων σε ανθρώπινους πληθυσμούς – αλλά π.χ. από κυτταροκαλλιέργειες ή μελέτες σε πειραματόζωα.

Πίνακας 11. Αριθμός μελετών γενετικής συσχέτισης υποψήφιων γονιδίων για τους γενετικούς τόπους που περιείχαν πολυμορφισμούς χαμηλής συχνότητας, καθώς και τα σχόλια των συγγραφέων των μελετών GWAS για την πρότερη τεκμηρίωσή τους.

Φαινότυπος	Γενετικός Τόπος	SNP	Αριθμός Μελετών Μέχρι το Έτος Πριν τη Μελέτη GWAS	Αριθμός Μελετών	Σχόλια στις Μελέτες GWAS για το Γενετικό Τόπο*
ΟΛΛ	<i>KRTHB5</i>	rs2089222-A	0	0	<i>Κανένα</i>
Πρόοδος AIDS	<i>HCP5, MICB, MCCD1, BAT1, LTB, TNF</i>	rs2395029-G	1 για <i>HCP</i> , 1 για <i>TNF</i> , 0 για τα υπόλοιπα	1	HCP5 was previously identified by a GWAS based the Euro-CHAVI cohort, and also proposed by candidate-gene association studies
Μπλέ έναντι πράσινα μάτια	<i>C6orf48</i>	rs9368699-C	0	0	<i>Κανένα</i>
	<i>OCA2</i>	rs1667394-A	0	1	Previously reported to be associated with albinism, eye color, hair color and skin pigmentation/ <i>OCA2</i> mutations are known to be a major cause for albinism/ <i>OCA2</i> has been discovered in linkage studies
Φακίδες	<i>MC1R</i>	rs1805007-T	0	3	It was known by previous reports/ previously documented mutations in <i>MC1R</i>
Μέση οστική πυκνότητα (ΟΜΣΣ)	<i>AKAP11</i>	rs180851-G rs7326472-G rs12854504-G rs7998154-T	0	0	<i>Κανένα</i>
	<i>TNFSF11</i>	rs6561055-A rs17639156-G	3	11	<i>Κανένα</i>
Νοητική λειτουργία	<i>HCCS</i>	rs5934953-C	0	0	<i>Κανένα</i>
Νόσος Crohn	<i>NOD2</i>	rs2066844-T rs2066845-C rs2066847-C	176	301	(<i>NOD2</i> is a) Previously known Νόσος Crohn locus
HDL χοληστερόλη	<i>LRRK2, MUC19</i>	rs11175593-T	1 για <i>MUC19</i> , 0 για <i>LRRK2</i>	1 για <i>MUC19</i> , 0 για <i>LRRK2</i>	LRRK2: evidence from a previous cell study/ <i>MUC19</i> : evidence from a previous animal study
	<i>HNF4A</i>	rs1800961-C	3	5	Function in humans has previously been studied/ Although mice lacking either <i>Hnf4a</i> or <i>Hnf1a</i> have altered plasma cholesterol levels, there has been only modest evidence to date connecting these genes to either HDL or LDL cholesterol concentrations in humans.

Αιματοκρίτης	<i>ABCA1</i>	rs9282541-T	34	53	It is a well-recognized association
	<i>HFE</i>	rs1800562-A	3	4	
Αιμοσφαιρίνη	<i>HFE</i>	rs1800562-A	17	21	Mutations in the HFE gene are already known to underlie hereditary hemochromatosis. The HFE gene induces expression of the iron regulatory hormone hepcidin.
LDL χοληστερόλη	<i>PCSK9</i>	rs11591147-G	10	26	Mutations in the HFE gene are already known to underlie hereditary hemochromatosis. The HFE gene induces expression of the iron regulatory hormone hepcidin.
Μέση περιεκτικότητα ερυθρών	<i>SLC17A3</i>	rs1408272-G	0	0	Prior evidence for association with LDL cholesterol concentrations/ has also been shown to cause mendelian syndromes or harbor multiple rare alleles that contribute to trait variation
Μέσος όγκος ερυθρών	<i>HFE</i>	rs1800562-A	3	5	HFE is known to be associated with iron homeostasis.
Μη-μικροκυτταρικός καρκίνος πνεύμονα	Διαγονιδιακά	rs17085106-T	M/E	M/E	
Πινοφαρυγγικός καρκίνος	<i>ITGA9</i>	rs189897-A	0	1	The gene is located at the chromosomal 3p22–21.3 segment, which is known to be commonly deleted in various types of carcinoma including NPC. A linkage study also mapped an NPC susceptibility locus to chromosome 3p21.31–21.2, indicating that the genes in this region are crucial for the formation of NPC.
		rs197757-T	0	1	
Διαταραχή πανικού	<i>TMEM16B^a</i>	rs12579350-A	0	1	Kavένα
	<i>PKP1</i>	rs860554-T	0	1	
Καρκίνος προστάτη	Διαγονιδιακά	rs16901979-A	M/E	M/E	The gene has an important role in the cytoskeleton/cell membrane interaction. The protein of PKP1, plackoglobin, acts as linker molecules at adherence junctions and desmosome at the plasma membrane.
Πρωτοπαθής χολική κίρρωση	<i>C6orf10</i>	rs2395148-A	0	0	
ALP	<i>PDZRN4, CNTN1</i>	rs1880887-C	0	1	Kavένα
fT3	<i>HS3ST3B1</i>	rs3848445-C	0	1	Kavένα
Ψωρίαση	<i>HLA-C</i>	rs2395029-C	57	65	Strongest association with this region is consistent with previous results from our group and others
Ανταπόκριση στη θεραπεία για ΟΑΛ	<i>ST8SIA6</i>	rs359312-T	0	0	Kavένα

Ανταπόκριση στην αντιψυχωσική θερπαπεία ΣΕΛ	Διαγονιδιακά	rs17022444-G	M/E	M/E	
	Διαγονιδιακά	rs7669317-C	M/E	M/E	
	<i>TNFAIP3</i>	rs5029939-G	0	10	Previously unreported for SLE susceptibility/ recent reports for influencing RA risk/ This GWAS identifies TNFAIP3 as a new susceptibility locus in SLE.
ΣΕΛ	<i>TNFAIP3</i>	rs2230926-C	0	10	Reported by previous GWAS
Tanning	<i>MATP</i>	rs35391-T	0	0	SNPs in MATP were previously evaluated in the GWAS of natural hair color by our group. Three SNPs in the MATP gene have been associated with human pigmentation
Τριγλυκερίδια	<i>APOA1, APOC3, APOA4, APOA5,</i>	rs662799-G	118 για <i>APOA1, APOC3, APOA4, APOA5</i>	165 συνολικά για <i>APOA1, APOC3, APOA4, APOA5</i>	These loci have been previously implicated in lipid metabolism.
Τριγλυκερίδια	<i>APOA1, APOC3, APOA4, APOA5, DSCAML1</i>	rs10892151-A	118 για <i>APOA1, APOC3, APOA4, APOA5</i>	165 συνολικά για <i>APOA1, APOC3, APOA4, APOA5</i>	APOA1, APOC3, APOA4, APOA5 is a cluster of more likely candidate genes, given the established key roles of their products in lipid metabolism
Σακχαρώδης διαβήτης τύπου 1	<i>COBL</i>	rs4948088-C	0	0	Κανένα
Σακχαρώδης διαβήτης τύπου 2	<i>TCF7L2</i>	rs7903146-T	14	140	It was reported by previous studies.

Συντομογραφίες: M/E: Μη εφαρμόσιμο γιατί οι πολυμορφισμοί βρίσκονται σε διαγονιδιακές περιοχές

^a Το TMEM16B αναγνωρίστηκε ως *ANO2* στη βάση του HuGE Navigator

* Τα σχόλια παρατίθενται στα Αγγλικά αυτούσια όπως αναφέρονταν στις μελέτες GWAS.

Γνωστές μεταλλάξεις στους ίδιους γενετικούς τόπους

Σύμφωνα με τη βάση OMIM (Online Mendelian Inheritance in Man), για 11 γενετικούς τόπους (που αντιστοιχούσαν σε 13 γενετικές σχέσεις), όπου μη-κοινοί πολυμορφισμοί είχαν βρεθεί από μελέτες GWAS (*OCA2* και χρώμα ματιών, *TNFSF11* μέση οστική πυκνότητα, *NOD2* και νόσος του Crohn's, *MC1R* και φακίδες, *HNF4A* και HDL χοληστερόλη, *ABCA1* και HDL χοληστερόλη, *HFE* και αιματοκρίτης, αιμοσφαιρίνη και μέσος όγκος ερυθρών, *PCSK9* και LDL χοληστερόλη, *MATP* και μαύρισμα, *HLA-C* και ψωρίαση, και *APOA1/C3/A4/A5* και τριγλυκερίδια), υπήρχαν γνωστές μεταλλάξεις που συνεισέφεραν τα ίδια ή παρόμοια φαινοτυπικά αποτελέσματα, όπως φαίνεται στον **Πίνακα 12**.

Για 5 γενετικούς τόπους (*HCCS*, *LRRK2*, *PKP1*, *CNTN1*, *HLA-C*), είχαν περιγραφεί μεταλλάξεις με φαινοτυπικά αποτελέσματα (συνδρομική μικροφθαλμία, νόσος Parkinson, εξωδερμική δυσπλασία/σύνδρομο ευθραύστου δέρματος, μυοπάθεια Compton-North, και ιαμμία HIV-1, αντίστοιχα), τα οποία δεν ήταν παρόμοια με αυτά που προκαλούσαν οι πολυμορφισμοί χαμηλής συχνότητας που είχαν βρεθεί στις μελέτες GWAS.

Πίνακας 12. Μεταλλάξεις στους ίδιους γενετικούς τόπους με τους πολυμορφισμούς χαμηλής συχνότητας που προκαλούν τα ίδια ή παρόμοια φαινοτυπικά αποτελέσματα σύμφωνα με τη βάση OMIM

Γονίδια	Περιοχή	Πολυμορφισμός	Φαινότυπος	Μεταλλάξεις με το ίδιο/ παρόμοιο φαινοτυπικό αποτέλεσμα	Φαινοτυπικά αποτελέσματα των μεταλλάξεων
<i>OCA2</i>	15q13.1	rs1667394-A	Μπλέ/ πράσινα μάτια	2.7-kb del, ex7del IVS17DS, G-T, +1 Pro743Leu 1-bp del Ala334Val 122.5-kb del Trp679Cys Asn489Asp Met394Ile	Δερματοβλενογόνιος αλφισμός τύπου 2
<i>TNFSF11</i>	13q14	rs6561055-A rs17639156-G	Μέση οστική πυκνότητα	5-bp del, IVS7+4 Met199Lys 2-bp del, 828CG	Οστεοπέτρωση, αυτοσωμική υπολειπόμενη τύπου 2
<i>NOD2</i>	16q12.1	rs2066844-T rs2066845-C rs2066847-C	Νόσος Crohn	3020insC Gly881Arg Arg675Trp IVS8+158	Νόσος Crohn
<i>MC1R</i>	16q24.3	rs1805007-T	Φακίδες	3-bp del 439TTC Thr157Ile Pro159Thr	Βλάβη δέρματος επαγόμενη από υπεριώδη ακτινοβολία
<i>HNF4A</i>	20q13.12	rs1800961-C	HDL χοληστερόλη	Gln268Ter Arg154Ter Arg127Trp 1-bp del Phe75T IVS5, Del A, -2 Met364Arg Val393Ile	Νεανικός διαβήτης όψιμης έναρξης, τύπου 1 Σακχαρώδης διαβήτης τύπου 2

<i>ABCA1</i>	9q31.1	rs9282541-T	HDL χοληστερόλη	Cys1417Arg IVS24DS, G-C Gln537Arg 110-bp in/14-bp del 2-bp del, 3283TC 1-bp del, 2665C Ser1446Leu int12-14 del, int16-31 del Arg1680Trp Asp1229Asn Arg2021Trp 1-bp del, 1764G Asn875Ser Ala877Val Trp530Ser 1-bp del, 1764G Tyr573Ter 3-bp del	Νόσος Tangier (ανεπάρκεια HDL τύπου 1)
<i>HFE</i>	6p22	rs1800562-A	Αιματοκρίτης Αιμοσφαιρίνη Μέσος όγκος ερυθρών	Cys282Tyr His63Asp Arg330Met Gln283Pro	Ανεπάρκεια HDL τύπου 2 Αιμοχρωμάτωση
<i>PCSK9</i>	1p32.3	rs11591147-G	LDL χοληστερόλη	Asp374Tyr Tyr142Ter Cys679Ter 3-bp del 290_292delGCC	Οικογενής υπερχοληστερολαιμία, τύπου 3 Επίπεδα LDL χοληστερόλης
<i>HLA-C</i> <i>MATP</i>	6p21.33 5p13.3	rs2395029-C rs35391-T	Ψωρίαση Μελάνωση	HLA-C, HLA-Cw6 allele IVS2, G-A, -1 1-bp del, 986C 3-bp del	Ψωρίαση Δερματοβλενογόνιος αλφισμός τύπου 4

<i>APOA1, APOC3, APOA4, APOA5</i>	11q23.3	rs662799-G rs10892151-A	Τριγλυκερίδια	Ala486Val	Ανεπάρκεια απολιποπρωτεΐνης A-I
				Asp157Asn	
				1-bp del, 1121T	
				Gln84Ter (<i>APOA1/APOC3</i>)	
				Val156Glu (<i>APOA1/APOC3</i>)	
				Gln-2Ter (<i>APOA1/APOC3</i>)	
				1-bp ins (<i>APOA1/APOC3</i>)	
				Gln32Ter (<i>APOA1/APOC3</i>)	Αν-αλφα-λιποπρωτεΐναιμία
				Gln139Ter (<i>APOA5</i>)	Πρωτοπαθής υπο-αλφα-λιποπρωτεΐναιμία
					Περικογχικό ξανθέλασμα
					Υπερλιποπρωτεΐναιμία τύπου 4

Συρροή κοινών πολυμορφισμών, προηγούμενων υποψήφιων πολυμορφισμών και μεταλλάξεων στους ίδιους γενετικούς τόπους με πολυμορφισμούς χαμηλής συχνότητας

Συνολικά, οι μελέτες GWAS που συμπεριλήφθηκαν στην παρούσα διατριβή είχαν ανακαλύψει 30 διαφορετικές σχέσεις ανάμεσα σε φαινοτύπους και γενετικούς τόπους που περιλάμβαναν μη-κοινούς πολυμορφισμούς. Από αυτές, για 16 σχέσεις άλλοι κοινοί πολυμορφισμοί είχαν ανακαλυφθεί από μελέτες GWAS (n=10), πολυμορφισμοί είχαν προταθεί από μελέτες υποψήφιων γονιδίων πριν από την πρώτη μελέτη GWAS που είχε προτείνει τον αντίστοιχο γενετικό τόπο (n=13), ή είχαν περιγραφεί μεταλλάξεις που συνεισφέρουν τα ίδια ή παρόμοια φαινοτυπικά αποτελέσματα (n=13). Για 4 από τις 16 σχέσεις ίσχυαν δύο από τα τρία παραπάνω χαρακτηριστικά, ενώ για άλλες 8 ίσχυαν και τα τρία.

Για τις υπόλοιπες 14 σχέσεις ανάμεσα στους γενετικούς τόπους και τους αντίστοιχους φαινοτύπους (*KRTHB5* και οξεία λεμφοβλαστική λευχαιμία, *C6orf48* και πρόοδος AIDS, *AKAP11* και μέση οστική πυκνότητα, *HCCS* και γνωσιακή επίδοση, *SLC17A3* και μέση πυκνότητα αιμοσφαιρίνης, *ITGA9* και ρινοφαρυγγικός καρκίνος, *TMEM16B* και διαταραχή πανικού, *PKP1* διαταραχή πανικού, *C6orf10* και πρωτοπαθής χολική κίρρωση, *PDZRN4/CNTN1* και αλκαλική φωσφατάση, *HS3ST3B1* και ελεύθερη T3, *ST8SIA6* και ανταπόκριση σε θεραπεία για οξεία λεμφοβλαστική λευχαιμία, *TNFAIP3* και λύκος, *COBL* και σακχαρώδης διαβήτης τύπου 1), δεν βρέθηκαν κοινοί πολυμορφισμοί, μεταλλάξεις ή δεδομένα από μελέτες υποψήφιων γονιδίων.

Συχνότητες αλληλίων σε πληθυσμούς διαφορετικής εθνικής καταγωγής

Τρεις πολυμορφισμοί (rs11591147-T, rs9282541-T και rs2066847-C) δεν βρέθηκαν σε κανένα από τα 4 δείγματα πληθυσμών (Ευρωπαίοι, Κινέζοι, Ιάπωνες, Αφρικανοί). Άλλοι 12 πολυμορφισμοί ήταν μη-κοινοί και στα 4 δείγματα. Κατά συνέπεια, 25 από τους συνολικά 40 μη-κοινούς πολυμορφισμούς ήταν κοινοί (συχνότητα ελάσσονος αλληλίου >5%) σε τουλάχιστον 1 πληθυσμιακό δείγμα του HarMap. Τα ευρήματα αυτά παρουσιάζονται στον **Πίνακα 13**.

Πίνακας 13. Συχνότητες ελασσόνων αλληλίων των μη-κοινών πολυμορφισμών που αναλύθηκαν στη διατριβή στα 4 δείγματα πληθυσμών του HarMap.

SNP	Γονίδιο	Περιοχή	Πληθυσμός GWAS	Συχνότητα Ελάσσονος Αλληλίου				
				Μελέτη GWAS	CEU HarMap	CHB HarMap	JPT HarMap	YRI HarMap
rs2089222-A	<i>KRTHB5</i>	12q24.22	Ευρωπαϊκός	0.03	0.04	0.26	0.33	0.19
rs2395029-G	<i>HCP5, MICB, MCCD1, BAT1, LTB, TNF</i>	6p21.33	Ευρωπαϊκός	0.03	0.05	0.01	0	0
	<i>HLA-C</i>		Ευρωπαϊκός	0.03				
rs9368699-C	<i>C6orf48</i>	6p21.3	Ευρωπαϊκός	0.03	0.06	0.18	0.10	0
rs1667394-A	<i>OCA2</i>	15q13.1	Ευρωπαϊκός	0.02	0.13	0.20	0.14	0.05
rs1805007-T	<i>MC1R</i>	16q24.3	Ευρωπαϊκός	0.05	0.15	0	0	0
rs5934953-C	<i>HCCS</i>	Xp22.2	Ευρωπαϊκός	0.02	0.04	0	0	0
rs180851-G	<i>AKAP11</i>	13q14	Ευρωπαϊκός	0.05	0.05	0.08	0.03	0.14
rs7326472-G	<i>AKAP11</i>	13q14	Ευρωπαϊκός	0.05	0.04	0.11	0.09	0.16
rs12854504-G	<i>AKAP11</i>	13q14	Ευρωπαϊκός	0.05	0.04	0.10	0.07	0.03
rs7998154-T	<i>AKAP11</i>	13q14	Ευρωπαϊκός	0.02	0.02	0	0	0
rs6561055-A	<i>TNFSF11</i>	13q14	Ευρωπαϊκός	0.05	0.04	0	0	0
rs17639156-G	<i>TNFSF11</i>	13q14	Ευρωπαϊκός	0.05	0.04	0.07	0.09	0
rs2066844-T	<i>NOD2</i>	16q12.1	Ευρωπαϊκός	0.05	0.11	0	0	0
rs2066845-C	<i>NOD2</i>	16q12.1	Ευρωπαϊκός	0.01	0.02	0	0	0
rs2066847-C	<i>NOD2</i>	16q12.1	Ευρωπαϊκός	0.04	0	0	0	0
rs11175593-T	<i>LRRK2, MUC19</i>	12q12	Ευρωπαϊκός	0.02	0.02	0.03	0.01	0
rs1800961-C	<i>HNF4A</i>	20q13.12	Ευρωπαϊκός	0.03	0.05	0.01	0	0
rs9282541-T	<i>ABCA1</i>	9q31.1	Μικτός	0.03	0	0	0	0
			Ευρωπαϊκός	0.04				
rs1800562-A	<i>HFE</i>	6p22.1	Ευρωπαϊκός	0.04	0.04	0	0	0
			Ευρωπαϊκός	0.04				
rs11591147-G	<i>PCSK9</i>	1p32.3	Ευρωπαϊκός	0.01	0	0	0	0
rs1408272-G	<i>SLC17A3</i>	6p22.1	Ευρωπαϊκός	0.03	0.03	0	0	0
rs17085106-T	<i>Διαγονιδιακά</i>	18q22.3	Ευρωπαϊκός	0.02	0	0	0	0.16
rs189897-A	<i>ITGA9</i>	3p22.2	Ασιατικός	0.03	0.27	0.08	0.12	0.01
rs197757-T	<i>ITGA9</i>	3p22.2	Ασιατικός	0.03	0	0.07	0.17	0.02

rs12579350-A	<i>TMEM16B</i>	12p13.31	Ασιατικός	0.01	0	0.01	0	0
rs860554-T	<i>PKP1</i>	1q32.1	Ασιατικός	0.05	0.20	0.12	0.06	0.01
rs16901979-A	<i>Διαγονιδιακά</i>	8q24.21	Ευρωπαϊκός	0.03	0.02	0.29	0.16	0.46
rs2395148-A	<i>C6orf10</i>	6p21.3	Ευρωπαϊκός	0.02	0.05	0.21	0.07	0.09
rs1880887-C	<i>PDZRN4, CNTN1</i>	12q12	Ευρωπαϊκός	0.03	0.03	0.14	0.08	0.38
rs3848445-C	<i>HS3ST3B1</i>	17p12	Ευρωπαϊκός	0.05	0.05	0.32	0.26	0.17
rs359312-T	<i>ST8SIA6</i>	10p12.33	Ευρωπαϊκός, Αφρικανικός, Άλλος	0.04	0	0.47	0.42	0
rs17022444-G	<i>Διαγονιδιακά</i>	2p12	Ευρωπαϊκός, Αφρικανικός, Άλλος	0.03	0	0	0	0.07
rs7669317-C	<i>Διαγονιδιακά</i>	4q24	Ευρωπαϊκός, Αφρικανικός, Άλλος	0.04	0.05	0	0	0
rs5029939-G	<i>TNFAIP3</i>	6q23.3	Ευρωπαϊκός	0.03	0.04	0.09	0.20	0.50
rs2230926-C	<i>TNFAIP3</i>	6q23.3	Ασιατικός	0.04	0.01	0.09	0.18	0.48
rs35391-T	<i>MATP</i>	5p13.3	Ευρωπαϊκός	0.03	0	0.38	0.35	0.47
rs662799-G	<i>APOA1, APOC3, APOA4, APOA5</i>	11q23.3	Ευρωπαϊκός	0.05	0.02	0.27	0.29	0.13
rs10892151-A	<i>APOA1, APOC3, APOA4, APOA5, DSCAML1</i>	11q23.3	Ευρωπαϊκός	0.03	0.02	0.06	0	0.41
rs4948088-C	<i>COBL</i>	7p12.1	Ευρωπαϊκός	0.05	0.02	0	0	0.04
rs7903146-T	<i>TCF7L2</i>	10q25.2	Ασιατικός	0.04	0.25	0.02	0.02	0.29

Συντομογραφίες: GWAS: genome-wide association study – μελέτη σάρωσης του γονιδιώματος, YRI: Yoruba στο Ibada της Νιγηρίας, JPT: Ιάπωνες στο Tokyo της Ιαπωνίας, CHB: Han Chinese στο Πεκίνο της Κίνας, CEU: κάτοικοι της Utah με βόρεια και δυτική ευρωπαϊκή καταγωγή από τη συλλογή CEPH (Centre d'Étude du Polymorphisme Humain)

2.2.4. ΣΥΖΗΤΗΣΗ

Στην παρούσα διατριβή ανασκοπήθηκαν τα χαρακτηριστικά πολυμορφισμών με συχνότητα ελάσσονος αλληλίου 5% ή λιγότερο, οι οποίοι ήταν επιπλέον στατιστικά σημαντικοί σε επίπεδο GWAS ($P \leq 10^{-7}$). Βρέθηκαν 43 γενετικές σχέσεις, σε 12 από τις οποίες εμπλέκονταν πολυμορφισμοί που βρίσκονταν σε εξόνια. Οι περισσότερες σχέσεις είχαν ανακαλυφθεί σε πληθυσμούς ευρωπαϊκής καταγωγής. Τα μεγέθη αποτελέσματος ήταν κατά βάση μεγάλα. Μερικοί πολυμορφισμοί είχαν προταθεί σε περισσότερες από μία μελέτες GWAS και για 14 σχέσεις οι μελέτες αυτές είχαν ανακαλύψει και κοινούς πολυμορφισμούς για τον ίδιο φαινότυπο στους ίδιους γενετικούς τόπους. Οι γενετικοί τόποι που εμπλέκονταν σε 13 γενετικές σχέσεις είχαν κάποια τεκμηρίωση από προηγούμενες μελέτες υποψήφιων γονιδίων. Επιπλέον, για 13 γενετικές σχέσεις οι αντίστοιχοι γενετικοί τόποι περιλάμβαναν μεταλλάξεις με το ίδιο ή παρόμοιο φαινοτυπικό αποτέλεσμα. Τέλος, οι περισσότεροι από τους μη-κοινούς πολυμορφισμούς ήταν κοινοί σε τουλάχιστον ένα πληθυσμιακό δείγμα του HarMap.

Εξακολουθεί να υπάρχει διαμάχη και κάποιες πρώιμες ενδείξεις αναφορικά με το μοντέλο «σπάνιοι πολυμορφισμοί – κοινά νοσήματα» (*rare variants – common disease model*) για τη γενετική ευαλωτότητα σε διάφορα νοσήματα, όπως ο καρκίνος, ο διαβήτης, η σχιζοφρένεια κ.ά. (42, 43, 48, 50-52, 259, 260). Σύμφωνα με αυτή την υπόθεση, η πολλαπλασιαστική δράση των μη-κοινών (48, 55) και σπάνιων πολυμορφισμών (55) με μέσους και μεγάλους λόγους αναλογιών, μπορεί να εξηγήσει ένα σημαντικό μέρος της γενετικής διακύμανσης σε πολλούς κοινούς φαινοτύπους και νόσους (42, 43, 48). Σχεδόν σε όλες τις γενετικές σχέσεις που ανασκοπήθηκαν στην παρούσα μελέτη και που

αφορούσαν νοσολογικές οντότητες, το αλλήλιο κινδύνου είχε συχνότητα $\leq 5\%$ ή $\geq 95\%$. Η μόνη εξαίρεση ήταν ένας πολυμορφισμός στο *COBL* που προδιέθετε για σακχαρώδη διαβήτη τύπου 1, όπου το μέγεθος του αποτελέσματος ήταν εξαιρετικά μικρό και η στατιστική υποστήριξη από τις πιο αδύναμες. Πιθανόν οι πολυμορφισμοί χαμηλής συχνότητας να έχουν εξελικτικό μειονέκτημα και αυτό να μην τους επιτρέπει να γίνουν πιο επικρατείς σε πληθυσμιακό επίπεδο, ή επίσης πιθανόν να είναι πιο πρόσφατοι, ακόμα και αν η εξελικτική τους επίδραση είναι ουδέτερη (113, 157). Επιπλέον, οι περισσότερες γενετικές σχέσεις που μελετήθηκαν είχαν λόγο αναλογιών >2 που είναι το τυπικό μέγεθος αποτελέσματος που αναμένεται για τους μη-κοινούς πολυμορφισμούς (42, 43, 48, 52, 260). Ειδικά όσον αφορά τους φαινοτύπους που σχετίζονται με φαρμακογενωμική (π.χ. ανταπόκριση σε μια θεραπεία) οι μεγάλοι λόγοι αναλογιών μπορεί να οφείλονται στην έλλειψη εξελικτικής ανταπόκρισης σε πρόσφατες εξωγενείς εκθέσεις, όπως οι φαρμακολογικοί παράγοντες, καθώς το γονιδίωμα δεν έχει τον απαιτούμενο χρόνο να προσαρμοστεί σε αυτού του είδους τις εκθέσεις (261). Παρόλα αυτά, λόγοι αναλογιών που ξεπερνούν κατά πολύ τα μικρά μεγέθη αποτελέσματος που είναι τυπικά για τις μελέτες GWAS (50) δεν σημαίνουν απαραίτητα ότι οι μη κοινοί πολυμορφισμοί θα έχουν και τέτοια μεγάλα μεγέθη αποτελέσματος. Λόγω θεμάτων ισχύος, οι τυπικές μελέτες GWAS αναμένεται να ανιχνεύσουν πρωτίστως εκείνους τους πολυμορφισμούς χαμηλής συχνότητας που έχουν μεγάλα μεγέθη αποτελέσματος (48, 55, 91). Αυτό υποστηρίζεται από την παρούσα μελέτη η οποία έδειξε ότι τα τυπικά μεγέθη δείγματος στις περισσότερες μελέτες GWAS μέχρι πρόσφατα ήταν ανεπαρκή για να ανιχνεύσουν την πλειοψηφία των μη-κοινών πολυμορφισμών με λόγους αναλογιών <1.40 . Πιθανώς υπάρχουν πολύ περισσότερες σχέσεις

πολυμορφισμών χαμηλής συχνότητας με μεσαία παρά με μεγάλα μεγέθη αποτελέσματος στη γενετική αρχιτεκτονική των πολυσύνθετων νοσημάτων. Η πλειοψηφία των σχέσεων σε αυτή την τελευταία κατηγορία έχει μάλλον ήδη ανακαλυφθεί, ειδικά όταν έχουν αναλυθεί μεγάλα μεγέθη δείγματος σε μελέτες GWAS.

Παρόλο που οι μη-κοινοί και σπάνιοι πολυμορφισμοί μπορεί να αποτελούν έως και το 60% της ποικιλομορφίας του ανθρώπινου γονιδιώματος (48, 262), η κάλυψή τους στις μελέτες GWAS είναι ως επί το πλείστον φτωχή (43, 97, 263) και συχνά αποκλείονται εξ ορισμού από τις αντίστοιχες αναλύσεις, καθώς πολλές φορές υιοθετούνται όρια συχνότητας ελάσσονος αλληλίου της τάξης του 1% ή του 5% στα πλαίσια των κριτηρίων ποιότητας. Αυτό ίσως να εξηγεί και το γεγονός ότι όλοι οι πολυμορφισμοί που συμπεριλήφθηκαν στη μελέτη ήταν μη-κοινοί παρά σπάνιοι, δηλαδή είχαν συχνότητα ελάσσονος αλληλίου μεταξύ 0.5-5%. Πράγματι, στην παρούσα μελέτη μόνο ένα μικρό μέρος των καταλογογραφημένων πολυμορφισμών στον *Κατάλογο Δημοσιευμένων Μελετών GWAS* είχε συχνότητες ελάσσονος αλληλίου 5% ή λιγότερο, και ακόμα ένα μικρότερο ποσοστό ήταν στατιστικά σημαντικό σε επίπεδο GWAS. Η ανίχνευση των μη κοινών πολυμορφισμών απαιτεί πολύ μεγαλύτερα μεγέθη δείγματος (91, 264) από αυτά των περισσότερων μελετών GWAS μέχρι πρόσφατα (55). Η κατάσταση αυτή φαίνεται να βελτιώνεται τα τελευταία χρόνια με τη χρησιμοποίηση μεγάλων δειγμάτων (265), τη δημιουργία διεθνών συνασπισμών (140, 141), καθώς και τη μετα-ανάλυση πολλαπλών μελετών GWAS (137, 142, 181, 182, 192).

Το γεγονός ότι οι περισσότεροι πολυμορφισμοί χαμηλής συχνότητας που μελετήθηκαν εδώ είχαν ανακαλυφθεί σε πληθυσμούς ευρωπαϊκής καταγωγής,

αντανακλά απλώς το γεγονός ότι η μεγάλη πλειοψηφία των μελετών GWAS έχει πραγματοποιηθεί σε αυτήν την εθνική ομάδα, ενώ πρόσφατα αυξήθηκε αισθητά ο αριθμός τέτοιων μελετών σε άλλους πληθυσμούς (266). Όπως φαίνεται και από τα αποτελέσματα, σχετικά λίγοι από τους μη-κοινούς πολυμορφισμούς είναι μη-κοινοί στις διάφορες εθνικές ομάδες. Αντιστρόφως, πολλοί από τους κοινούς πολυμορφισμούς που βρέθηκαν σε μελέτες GWAS είναι μη-κοινοί σε άλλες διαφορετικές εθνικές ομάδες. Συνεπώς, η μελέτη σε πληθυσμούς άλλης εθνικής καταγωγής γενετικών τόπων που δίνουν στατιστικά σημαντικά σήματα σε έναν πληθυσμό, θα μπορούσε να είναι ένας δυνητικός μηχανισμός για την ανακάλυψη περαιτέρω πολυμορφισμών χαμηλής συχνότητας.

Τέλος, στην παρούσα μελέτη βρέθηκαν αρκετά γονίδια που περιλαμβάνουν τόσο κοινούς όσο και μη-κοινούς πολυμορφισμούς με στατιστική σημαντικότητα σε επίπεδο GWAS. Οι εκτιμητές του αποτελέσματος για τους πολυμορφισμούς χαμηλής συχνότητας ήταν κατά κανόνα μεγαλύτεροι από αυτούς για τους κοινούς πολυμορφισμούς, με ελάχιστες εξαιρέσεις. Αυτό υποστηρίζει την υπόθεση ότι γονίδια που περιέχουν κοινούς πολυμορφισμούς με μεσαία μεγέθη αποτελέσματος για κοινά νοσήματα, είναι πιθανόν να περιλαμβάνουν και μη κοινούς πολυμορφισμούς με πολύ μεγαλύτερα μεγέθη αποτελέσματος (55). Εναλλακτικά, οι μη-κοινοί και σπάνιοι πολυμορφισμοί μπορεί να δημιουργούν «συνθετικές συσχετίσεις» με το να συνυπάρχουν στοχαστικά πιο συχνά σε συσχέτιση με ένα ή περισσότερα αλληλία σε μία θέση κοινού πολυμορφισμού (267). Ωστόσο, στην παρούσα μελέτη παρατηρήθηκαν λίγα παραδείγματα όπου κοινοί και μη-κοινοί πολυμορφισμοί βρίσκονταν σε υψηλή ανισορροπία διασύνδεσης. Επιπλέον, μερικοί από τους ίδιους γενετικούς τόπους φέρουν γνωστές μεταλλάξεις που οδηγούν σε σχετικούς φαινοτύπους.

Συνολικά, η ανωτέρω εικόνα είναι περισσότερο συμβατή με μία «συσσώρευση» σπάνιων, μη-κοινών και κοινών πολυμορφισμών, καθώς και μεταλλάξεων στον ίδιο γενετικό τόπο, πιθανόν συνεισφέροντας ανεξάρτητες επιδράσεις στη διαμόρφωση των πολυσύνθετων νοσημάτων (54).

Περιορισμοί της παρούσας μελέτης

Πρώτον, η συχνότητα ελάσσονος αλληλίου για έναν συγκεκριμένο πολυμορφισμό μπορεί να διαφέρει σημαντικά μεταξύ διαφορετικών μελετών αναλόγως των πληθυσμών που μελετούνται. Έτσι, το ίδιο αλληλίο μπορεί να χαρακτηριστεί ως μη-κοινό σε έναν πληθυσμό αλλά ως κοινό σε έναν άλλον (268). Η ολοκλήρωση του διεθνούς προγράμματος 1000 Genomes Project (<http://www.1000genomes.org/>) και η δημοσίευση των αποτελεσμάτων του προσέφερε πολύ καλύτερη και μεγαλύτερη ακρίβεια όσο αφορά τις συχνότητες αλληλίων, καθώς και καλύτερο χαρακτηρισμό των μη-κοινών και σπάνιων πολυμορφισμών σε σχέση με την πληροφορία του HapMap Project όπου βασίστηκαν οι μελέτες GWAS της παρούσας διατριβής (173, 174, 269-271).

Δεύτερον, δεν βρέθηκαν όλα τα απαιτούμενα δεδομένα για την ανάλυση των πολυμορφισμών που συμπεριλήφθηκαν στη μελέτη για ορισμένους φαινοτύπους, καθώς για μερικούς από αυτούς έλειπαν, και έτσι δεν κατέστη δυνατόν να βρεθούν οι εκτιμητές των αποτελεσμάτων, οι τιμές στατιστικής σημαντικότητας ή οι συχνότητες ελάσσονος αλληλίου. Τα μεγέθη αποτελέσματος μπορεί να είναι μικρότερα από αυτά που παρατηρήθηκαν στη μελέτη, με βάση το γεγονός ότι τα δημοσιευμένα δεδομένα ίσως υπόκεινται σε κάποιο βαθμό στο σφάλμα της «κατάρας του νικητή» (*winner's curse*) (72, 178, 185, 272, 273).

Συμπεράσματα

Ο αριθμός των γενετικών σχέσεων με πολυμορφισμούς χαμηλής συχνότητας που ανακαλύπτονται με αγνωστικιστικές μεθόδους γονοτύπησης αυξάνεται διαρκώς, καθώς διευρύνεται η χρήση νέων τεχνολογιών αλληλούχισης του συνόλου των εξονίων ή/και ολόκληρου του γονιδιώματος (146, 167-169). Μια διαμάχη που ήταν έντονη κατά τα προηγούμενα χρόνια – και που σε κάποιο μικρότερο βαθμό συνεχίζεται ακόμα – αφορούσε το αν οι προσπάθειες αλληλούχισης θα έπρεπε να εστιάζουν στα εξόνια ή σε ολόκληρο το γονιδίωμα, προκειμένου να ανακαλυφθεί ένα μεγάλο μέρος της υπολειπόμενης γενετικής σκοτεινής ύλης. Με βάση τα αποτελέσματα της παρούσας διατριβής, τα εξόνια πιθανώς περιλαμβάνουν μόνο ένα μικρό μέρος των πολυμορφισμών χαμηλής συχνότητας και συνεπώς η αλληλούχιση του πλήρους γονιδιώματος καθίσταται μάλλον αναπόφευκτη για την επιτυχή ανίχνευση πολυμορφισμών με ιδιαίτερο ενδιαφέρον για τη γενετική των πολυσύνθετων νοσημάτων (153, 163). Επιπλέον, δεδομένων των τεχνικών δυσκολιών και των θεμάτων ισχύος, οι μελέτες GWAS μέχρι σήμερα κατάφεραν να μας αποκαλύψουν λίγα πράγματα για τους πολυμορφισμούς με συχνότητες ελάσσονος αλληλίου $<0.5\%$. Ακόμα και με τις νέες τεχνολογίες, αυτοί οι πολυμορφισμοί θα εντοπιστούν μόνο εφόσον έχουν ιδιαίτερα μεγάλα μεγέθη αποτελέσματος ή αναλυθούν μεγάλα μεγέθη δείγματος (146). Μάλιστα, σε ορισμένες περιπτώσεις η μετα-ανάλυση πολλαπλών μελετών GWAS (137, 181) με 10.000 ασθενείς και μάρτυρες έχει μεγαλύτερη ισχύ σε σχέση με μία μελέτη αλληλούχισης με 1.000 ασθενείς και μάρτυρες (146).

2.3. ΤΟ ΠΑΡΑΔΕΙΓΜΑ ΤΟΥ ΜΕΛΑΝΩΜΑΤΟΣ

2.3.1. ΣΚΕΠΤΙΚΟ ΚΑΙ ΣΚΟΠΟΣ

Πληθώρα μελετών έχει δείξει ότι η υπεριώδης ακτινοβολία, καθώς και συγκεκριμένα φαινοτυπικά χαρακτηριστικά, όπως κόκκινα ή ξανθά μαλλιά, ανοιχτόχρωμα μάτια, ανοιχτή χροιά δέρματος κ.ά., σχετίζονται με αυξημένο κίνδυνο εμφάνισης δερματικού καρκίνου του τύπου του μελανώματος (274-277). Επιπλέον, ένα ισχυρό γενετικό υπόβαθρο έχει υποστηριχτεί από μελέτες διδύμων που δείχνουν μία συνεισφορά της τάξης του 55%.

Ο πιο συχνός γενετικός τύπος είναι το γονίδιο του υποδοχέα 1 της μελανοκορτίνης (melanocortin receptor gene, *MC1R*), πολυμορφισμοί στον οποίο έχουν συσχετισθεί τόσο με μελάνωμα (278-283) όσο και με συναφείς φαινοτύπους (278, 284, 285). Επιπλέον, οι μελέτες γονιδιωματικής σάρωσης έχουν ανακαλύψει έναν σημαντικό αριθμό κοινών και σπάνιων πολυμορφισμών σε διάφορους γενετικούς τόπους που σχετίζονται με μελάνωμα (286). Μία λεπτομερής καταλογογράφηση που προέκυψε από μία σύνοψη πεδίου (*field synopsis*) όλων των δημοσιευμένων γενετικών σχέσεων που αφορούν σε μελάνωμα, είναι διαθέσιμη στην ηλεκτρονική βάση MelGene (<http://www.melgene.org>).

Η πλειοψηφία των μελετών γενετικής συσχέτισης για μελάνωμα έχει πραγματοποιηθεί σε πληθυσμούς με ανοιχτό χρώμα δέρματος. Κατά συνέπεια η επίδραση πολυμορφισμών που σχετίζονται με μελάνωμα σε πληθυσμούς με σκούρο χρώμα δέρματος και μεγαλύτερη έκθεση σε υπεριώδη ακτινοβολία, είναι λιγότερο μελετημένη. Λόγω της γεωγραφικής της θέσης η Ελλάδα χαρακτηρίζεται από αυξημένη έκθεση στον ήλιο για μεγάλα χρονικά

διαστήματα, ενώ επιδημιολογικές μελέτες έχουν δείξει ότι ο ελληνικός πληθυσμός έχει κατά μέσο όρο πιο σκούρα χροιά δέρματος σε σχέση με τους πληθυσμούς της Βόρειας Ευρώπης και χαμηλότερη συχνότητα μελανώματος (287-289). Επιπλέον, αναλύσεις μοριακής γενετικής σε επίπεδο μεταλλάξεων έχουν βρει μεγαλύτερη συχνότητα μεταλλαγμένων μορφών των γονιδίων *CDKN2A/CDK4*, γεγονός που υποδηλώνει έναν ενισχυμένο ρόλο της γενετικής ποικιλομορφίας, όσον αφορά την αιτιολογία του μελανώματος σε περιοχές με χαμηλή συχνότητα της νόσου (288, 290).

Τα τελευταία χρόνια οι μελέτες σάρωσης του γονιδιώματος έχουν ανακαλύψει ένα σημαντικό αριθμό γενετικών πολυμορφισμών που έχουν ισχυρές συσχετίσεις με το μελάνωμα, ορισμένοι από τους οποίους έχουν χαμηλή συχνότητα, που τους κατατάσσει στους μη-κοινούς πολυμορφισμούς. Δεδομένου ότι οι περισσότερες τέτοιες σχέσεις αφορούν βορειοευρωπαϊκούς πληθυσμούς, στην παρούσα μελέτη επιδιώξαμε να εξετάσουμε τις γενετικές αυτές σχέσεις σε έναν ελληνικό πληθυσμό και να διερευνήσουμε πόσοι και ποιοι από τους αντίστοιχους κοινούς αλλά και μη κοινούς πολυμορφισμούς μπορούν να επαληθευθούν. Επιπλέον, μελετήσαμε την προγνωστική ικανότητα της γενετικής πληροφορίας όσον αφορά την πρόβλεψη της νόσου πριν ακόμα αυτή εκδηλωθεί κλινικά.

2.3.2. ΜΕΘΟΔΟΙ

Πληθυσμός

Η μελέτη αυτή είναι μία μελέτη δείκτου-ελέγχου (*case-control study*) (291). Ο πληθυσμός της μελέτης αποτελούνταν από ασθενείς με μελάνωμα και φυσιολογικούς μάρτυρες χωρίς μελάνωμα ηλικίας 18 ετών και άνω. Οι ασθενείς είχαν διαγνωσθεί ιστολογικά με μελάνωμα μη-οικογενούς μορφής στο νοσοκομείο «Ανδρέας Συγγρός», στην Αθήνα, και σε συμμετέχοντα κέντρα μεταξύ των ετών 2003-2009. Οι μάρτυρες ήταν υγιείς αιμοδότες από ένα κέντρο αιμοδοσίας, καθώς και άτομα που παρακολουθούνταν στο «Ανδρέας Συγγρός» για δερματικές μη-νεοπλασματικές παθήσεις. Οι ασθενείς και οι μάρτυρες είχαν ταυτιστεί (*matched*) μεταξύ τους ως προς το φύλο και την ηλικία (± 2 έτη). Άτομα με ιστορικό μελανώματος, άλλους δερματικούς καρκίνους ή άλλη μη-δερματική νεοπλασία αποκλείστηκαν από την ομάδα ελέγχου.

Επιλογή πολυμορφισμών προς γονοτύπηση

Όλοι οι πολυμορφισμοί που μελετήθηκαν επιλέχθηκαν από την ηλεκτρονική βάση MelGene (www.melgene.org) (286) τον Οκτώβριο του 2011. Η βάση αυτή καταλογογραφεί όλες τις γενετικές σχέσεις για μελάνωμα που δημοσιεύονται στη διεθνή βιβλιογραφία.

Επιλέχθηκαν 34 πολυμορφισμοί που διακρίνονται σε δύο ομάδες. Η πρώτη ομάδα περιλάμβανε 28 πολυμορφισμούς που είχαν συσχετισθεί με μελάνωμα σε επίπεδο στατιστικής σημαντικότητας $P < 0.05$, έπειτα από μετα-ανάλυση 3 ή περισσότερων μελετών ασθενών-μαρτύρων. Η δεύτερη ομάδα περιλάμβανε πολυμορφισμούς που είχαν κάποια βιολογική συσχέτιση με αιτιολογικά μονοπάτια και είχαν βρεθεί είτε σε μελέτες γονιδιακής σάρωσης (3 πολυμορφισμοί) είτε σε μελέτες υποψήφιων γονιδίων (3 πολυμορφισμοί), βάσει

αντίστοιχων ορίων στατιστικής σημαντικότητας ($P < 10^{-7}$ σε μελέτες γονιδιακής σάρωσης ή $P < 0.05$ σε μελέτες υποψήφιων γονιδίων). Συνολικά 19 πολυμορφισμοί ήταν στατιστικά σημαντικοί σε επίπεδο γονιδιώματος (*genome-wide significant*) σε προηγούμενες μελέτες γονιδιακής σάρωσης ή στη βάση MelGene.

Απομόνωση DNA, γονοτύπηση και έλεγχος ποιότητας

Το γονιδιωματικό DNA απομονώθηκε από το περιφερικό αίμα με το QIAamp DNA blood mini kit της εταιρίας Qiagen. Η συγκέντρωση του DNA ποσοτικοποιήθηκε πριν τη γονοτύπηση με το Quant-iT dsDNA HS Assay kit της Invitrogen. 50 ng από κάθε δείγμα DNA χρησιμοποιήθηκαν για τη γονοτύπηση των πολυμορφισμών που επιλέχθηκαν με τη χρήση του Sequenom iPLEX assay (Sequenom, Hamburg, Germany). Ο προσδιορισμός των αλληλίων έγινε με τη χρήση time-of-flight φασματοσκοπείας μάζας, όπως έχει περιγραφεί προηγουμένως (292). Τα κριτήρια του ελέγχου ποιότητας αφορούσαν πολυμορφισμούς με ρυθμό ανάκλησης $\geq 95\%$ καθώς και πολυμορφισμούς που δεν είχαν απόκλιση από την ισορροπία κατά Hardy-Weinberg (*Hardy-Weinberg equilibrium [HWE]*) με βάση ένα τεστ χ^2 ($P < 0.05$ αντιπροσωπεύει απόκλιση από την ισορροπία).

Στατιστική Ανάλυση

Η σχέση κάθε πολυμορφισμού με την έκβαση (μελάνωμα) ελέγχθηκε με τη χρήση μονοπαραγοντικής λογαριθμικής παλινδρόμησης για ταυτισμένα δείγματα (*univariate conditional logistic regression*). Υπολογίσαμε το λόγο αναλογιών (ΛΑ) για κάθε έλασσον αλληλίο (*per-allele odds ratio*). Προκειμένου να ελέγξουμε για την επίδραση άλλων παραγόντων κινδύνου, κάθε πολυμορφισμός μελετήθηκε στη συνέχεια με μοντέλα πολυπαραγοντικής

λογαριθμικής παλινδρόμησης. Οι παράγοντες κινδύνου που συνυπολογίστηκαν ήταν το χρώμα ματιών (ανοιχτό: μπλε/πράσινο/γκρι/ανοιχτό καφέ, ή σκούρο: σκούρο καφέ/μαύρο), το χρώμα μαλλιών (ανοιχτό: ξανθά/κόκκινα/ανοιχτό καφέ, ή σκούρο: σκούρο καφέ/μαύρο), το χρώμα δέρματος (ανοιχτό: άσπρο/ανοιχτό καφέ, ή σκούρο: σκούρο καφέ), ο φωτότυπος (τύπος I, II, III ή IV σύμφωνα με την κλίμακα Fitzpatrick), η ικανότητα μελάγχρωσης (ηλιακό έγκαυμα, ελάχιστη μελάγχρωση, ηλιακό έγκαυμα και έπειτα μελάγχρωση, έντονη μελάγχρωση) και το ιστορικό ηλιακού εγκαύματος (ναι ή όχι). Για την αναπλήρωση ελλειπουσών τιμών σε μία ή περισσότερες από τις συµμεταβλητές αυτές έγινε χρήση μεθόδων πολλαπλής αναπλήρωσης (*multiple imputation*).

Επιπλέον, υπολογίστηκε η συσχέτιση ανάμεσα στα μεγέθη αποτελέσματος στον ελληνικό πληθυσμό και στις αρχικές μελέτες που είχαν ανακαλύψει τους αντίστοιχους πολυμορφισμούς ή στα μεγέθη αποτελέσματος στη βάση MelGene, ανάλογα με την πηγή από την οποία είχαν επιλεγεί οι πολυμορφισμοί, ενώ εξετάσαμε και για πόσες γενετικές σχέσεις τα μεγέθη αποτελέσματος ήταν στην ίδια ή σε διαφορετική κατεύθυνση. Επίσης, υπολογίσαμε τη συσχέτιση ανάμεσα στις συχνότητες αλληλίων κινδύνου στον ελληνικό πληθυσμό και στις αρχικές μελέτες ή τη βάση MelGene από όπου προέρχονταν οι πολυμορφισμοί.

Επιπρόσθετα, πραγματοποιήσαμε αναλύσεις ισχύος στις οποίες υπολογίσαμε την ισχύ της συγκεκριμένης μελέτης, δοθέντων του μεγέθους δείγματος και της συχνότητας ελάσσονος αλληλίου κάθε πολυμορφισμού στον ελληνικό πληθυσμό, προκειμένου να ανιχνεύσει προηγούμενους πολυμορφισμούς σε επίπεδο σφάλματος τύπου I (*type I error*) $\alpha=0.05$. Το άθροισμα των εκτιμητών ισχύος για κάθε πολυμορφισμό αντιστοιχεί στον

αριθμό των πολυμορφισμών που αναμένονται να επαληθευθούν στη συγκεκριμένη μελέτη. Ακολουθώντας, χρησιμοποιώντας ένα διωνυμικό τεστ ελέγξαμε αν ο αριθμός των επαληθευμένων πολυμορφισμών ήταν στατιστικά σημαντικά διαφορετικός από αυτά που αναμέναμε με βάση την ισχύ.

Τέλος, υπολογίσαμε την προγνωστική ικανότητα των πολυμορφισμών αυτών (293-295) όσον αφορά την πρόβλεψη του μελανώματος, και κατασκευάσαμε τις καμπύλες ROC (*receiver operating characteristic curves*). Στις καμπύλες αυτές παρατίθενται η ευαισθησία ενός μοντέλου (δηλαδή η ικανότητά του να ανιχνεύει τα αληθώς θετικά) και το συμπληρωματικό της ειδικότητας, 1-ειδικότητα (δηλαδή η ικανότητα του μοντέλου να ανιχνεύει τα αληθώς αρνητικά) (296, 297), ενώ η επιφάνεια κάτω από την αντίστοιχη καμπύλη (*area under the curve – AUC*) δείχνει την ικανότητα του μοντέλου να διακρίνει αυτούς που θα αναπτύξουν τη νόσο από αυτούς που δεν θα την αναπτύξουν. Η AUC παίρνει τιμές από 0.5 έως 1, με το 1 να δηλώνει τέλεια διακριτική ικανότητα και το 0.5 να δηλώνει πλήρη έλλειψη διακριτικής ικανότητας (το μοντέλο δεν διαφέρει από την τύχη) (297, 298). Κατασκευάσαμε τις καμπύλες ROC και υπολογίσαμε την AUC για 3 μοντέλα. Το πρώτο μοντέλο περιλάμβανε μόνο τα φαινοτυπικά χαρακτηριστικά που είναι διαθέσιμα στον κλινικό μέσω απλών ερωτήσεων στα πλαίσια λήψης του ιστορικού του ασθενούς ή μέσω της φυσικής εξέτασης. Το δεύτερο μοντέλο περιλάμβανε αυτούς τους παράγοντες κινδύνου συν τους πολυμορφισμούς που ήταν στατιστικά σημαντικοί μετά τη διόρθωση για πολλαπλές συγκρίσεις. Τέλος, το τρίτο μοντέλο περιλάμβανε τους μη-γενετικούς παράγοντες κινδύνου συν όλους τους στατιστικά σημαντικούς πολυμορφισμούς σε επίπεδο $P=0.05$.

2.3.3. ΑΠΟΤΕΛΕΣΜΑΤΑ

Συνολικά μελετήθηκαν 284 ασθενείς (*cases*) με μελάνωμα και 284 μάρτυρες (*controls*, δηλαδή υγιή άτομα), ταυτισμένοι ως προς το φύλο και την ηλικία. Από αυτούς 270 (48%) ήταν άνδρες. Η διάμεση ηλικία ήταν τα 48 έτη (εύρος 18-85 έτη) για τους ασθενείς και τα 42 έτη (εύρος 18-81 έτη) για τους μάρτυρες. Ο **Πίνακας 14** δείχνει την κατανομή των δημογραφικών και φαινοτυπικών χαρακτηριστικών στις δύο ομάδες. Συνολικά επιλέχθηκαν 34 πολυμορφισμοί για την ανάλυση, όπως φαίνεται στον **Πίνακα 15**, και όλοι γονοτυπύθηκαν επιτυχώς με ποσοστό ανάκλησης τουλάχιστον 95%. Μόνο ένας πολυμορφισμός (rs4636294) βρισκόταν σε απόκλιση από την ισοροπία κατά Hardy-Weinberg στην ομάδα των μαρτύρων και κατά συνέπεια αποκλείστηκε από την παρούσα μελέτη.

Από τους πολυμορφισμούς που συμπεριλήφθηκαν, 4 βρίσκονταν στις 3'-αμετάφραστες περιοχές (*3'-untranslated regions; 3'-UTRs*) των αντίστοιχων γονιδίων και 1 στην 5'-αμετάφραστη περιοχή (5'-UTR), 13 βρίσκονταν σε ιντρόνια και 10 σε εξώνια, ενώ οι άλλοι 6 πολυμορφισμοί βρίσκονταν σε διαγονιδιακές περιοχές. Επιπλέον, ορισμένοι πολυμορφισμοί βρίσκονταν σε ανισορροπία διασύνδεσης: το rs2218220 με το rs4636294 ($r^2=0.95$), το rs1393350 με το rs1126809 ($r^2=0.94$), και τέλος τα rs1885120, rs910873 και rs17305573 ($r^2=0.90$) αναμεταξύ τους.

Πίνακας 14. Δημογραφικά χαρακτηριστικά και κατανομή των φαινοτύπων χρώματος στους ασθενείς με μελάνωμα και στους υγιείς μάρτυρες

Χαρακτηριστικά	Ασθενείς (n=284)*	Μάρτυρες (n=284)*	Τιμή P
Διάμεση ηλικία (έτη)	44	42	
Άντρες, N (%)	135 (48%)	135 (48%)	
Χρώμα μαλλιών			0.104
Ανοιχτό	31	27	
Σκούρο	186	255	
Χρώμα ματιών			0.073
Ανοιχτό	40	71	
Σκούρο	177	211	
Χρώμα δέρματος			<0.0001
Ανοιχτό	122	265	
Σκούρο	95	17	
Φωτότυπος			<0.001
Φωτότυπος I	9	12	
Φωτότυπος II	109	69	
Φωτότυπος III	75	141	
Φωτότυπος IV	23	55	
Ικανότητα μελάγχρωσης			<0.0001
Έγκαυμα	35	34	
Ελάχιστη μελάγχρωση	98	66	
Έγκαυμα έπειτα μελάγχρωση	70	104	
Βαθιά μελάγχρωση	12	73	
Ιστορικό ηλιακού εγκαύματος			0.301
Ναι	91	136	
Όχι	115	142	

* Το άθροισμα των επιμέρους αριθμητικών δεδομένων σε ορισμένα στρώματα είναι μικρότερο από τον συνολικό αριθμό ατόμων λόγω ελλειπουσών τιμών. Η αναπλήρωση των τιμών αυτών έγινε με τη χρήση μεθόδων πολλαπλής αναπλήρωσης (multiple imputation).

Συσχετίσεις πολυμορφισμών με μελάνωμα

Στον **Πίνακα 15** φαίνονται οι 33 πολυμορφισμοί που τελικά εξετάστηκαν στη μελέτη καθώς και τα μεγέθη αποτελέσματος, τα ελάσσονα και τα μείζονα αλληλία και οι συχνότητές τους στον ελληνικό πληθυσμό. Όλα τα αλληλία που ανιχνεύθηκαν ως ελάσσονα στον ελληνικό πληθυσμό ήταν ελάσσονα και στον ευρωπαϊκό πληθυσμό του HarMap, εκτός από το rs6001027 του οποίου το έλασσον αλληλίο στον ελληνικό πληθυσμό ήταν το T ενώ στο HarMap ήταν το C.

Πίνακας 15. Πολυμορφισμοί που επιλέχθηκαν για επικύρωση στον Ελληνικό πληθυσμό

SNP	Αλληλία Ελληνικού Πληθυσμού	Γενετικός τύπος	Εντόπιση/ Λειτουργία	ΣΕΑ (95% ΔΕ)		Λόγος Αναλογιών (95% ΔΕ)		Στατιστική σημαντικότητα σε επίπεδο GWAS	Πληθυσμός αναφοράς*
				Ελληνικός πληθυσμός	Πληθυσμός αναφοράς*	Ελληνικός πληθυσμός	Πληθυσμός αναφοράς*		
rs16891982	C/G	<i>SLC45A2</i>	Εξώνιο	0.14 (0.11-0.17)	0.068	0.51 (0.34-0.76)	0.40 (0.33-0.47)	Ναι	Melgene
rs401681	T/C	<i>CLPTM1L</i>	Ιντρόνιο	0.40 (0.36-0.44)	0.450	1.60 (1.22-2.10)	1.15 (1.08-1.22)	Ναι	Melgene
rs12203592	T/C	<i>IRF4</i>	Ιντρόνιο	0.04 (0.03-0.07)	0.2259	1.08 (0.62-1.89)	1.04 (0.99-1.08)	Όχι	MYT
rs7023329	G/A	<i>MTAP</i>	Ιντρόνιο	0.40 (0.36-0.44)	0.497	0.95 (0.73-1.25)	0.84 (0.80-0.89)	Ναι	Melgene
rs11515	G/C	<i>CDKN2A</i>	3' UTR	0.18 (0.15-0.21)	0.127	0.97 (0.70- 1.35)	1.15 (1.01-1.32)	Όχι	Melgene
rs3088440	A/G	<i>CDKN2A</i>	3' UTR	0.07 (0.05-0.09)	0.077	1.41 (0.89-2.21)	1.25 (1.03-1.51)	Όχι	Melgene
rs4636294**	G/A	<i>MTAP region</i>	Διαγονιδιακά	0.41 (0.36-0.45)	0.485	M/E	0.82 (0.75-0.90)	Ναι	Melgene
rs1335510	G/T	<i>MTAP region</i>	Διαγονιδιακά	0.31 (0.27-0.35)	0.417	0.86 (0.64-1.15)	0.83 (0.78-0.89)	Ναι	Melgene
rs2218220	T/C	<i>MTAP region</i>	Διαγονιδιακά	0.41 (0.37-0.46)	0.487	0.74 (0.56-0.97)	0.84 (0.80-0.89)	Ναι	Melgene
rs10757257	A/G	<i>MTAP</i>	Ιντρόνιο	0.31 (0.27-0.35)	0.415	0.87 (0.65-1.16)	0.81 (0.76-0.86)	Ναι	Melgene
rs1011970	T/G	<i>CDKN2B</i>	5' UTR	0.16 (0.13-0.20)	M/E	1.27 (0.91-1.78)	M/E	Όχι	GWAS
rs1408799	T/C	<i>TYRP-1region</i>	Διαγονιδιακά	0.32 (0.28-0.36)	0.258	1.16 (0.87-1.54)	0.86 (0.80-0.93)	Όχι	Melgene
rs1042602	A/C	<i>TYR</i>	Εξώνιο	0.46 (0.42-0.51)	0.316	1.11 (0.85-1.46)	0.95 (0.90-0.99)	Όχι	Melgene
rs1126809	A/G	<i>TYR</i>	Εξώνιο	0.20 (0.16-0.23)	0.295	1.19 (0.88-1.62)	1.22 (1.14-1.31)	Ναι	Melgene
rs1393350	A/G	<i>TYR</i>	Ιντρόνιο	0.19 (0.16-0.23)	0.269	1.16 (0.85-1.58)	1.27 (1.16-1.39)	Ναι	Melgene
rs1544410	A /G	<i>VDR</i>	Ιντρόνιο	0.42 (0.38-0.46)	0.400	0.95 (0.72-1.24)	0.89 (0.82-0.97)	Όχι	Melgene
rs1800407	A/G	<i>OCA2</i>	Εξώνιο	0.06 (0.04-0.08)	0.070	1.32 (0.81-2.16)	1.4 (1.07-1.82)	Όχι	Melgene
rs258322	T/C	<i>CDK10</i>	Ιντρόνιο	0.05 (0.03-0.07)	0.095	2.26 (1.32-3.88)	1.66 (1.48-1.86)	Ναι	Melgene
rs1805005	T/G	<i>MC1R</i>	Εξώνιο	0.13 (0.11-0.16)	0.114	1.59 (1.09-2.32)	1.13 (1.02-1.26)	Όχι	Melgene
rs1805007	T/C	<i>MC1R</i>	Εξώνιο	0.02 (0.01-0.03)	0.078	4.38 (2.03-9.43)	1.83 (1.56-2.15)	Ναι	Melgene
rs1805008	T/C	<i>MC1R</i>	Εξώνιο	0.02 (0.01-0.04)	0.098	1.64 (0.85-3.19)	1.54 (1.33-1.79)	Ναι	Melgene
rs1805009#	C/G	<i>MC1R</i>	Εξώνιο	0	0.015	M/Y	1.89 (1.51-2.38)	Ναι	Melgene
rs1805006	A/C	<i>MC1R</i>	Εξώνιο	0.005 (0.001-0.02)	0.009	0.33 (0.03-3.20)	1.47 (1.18-1.83)	Όχι	Melgene
rs11547464#	T/G	<i>MC1R</i>	Εξώνιο	0	0.010	M/Y	1.67 (1.26-2.21)	Όχι	Melgene
rs4785763	A/C	<i>AFG3L1</i>	3'UTR	0.28 (0.24-0.32)	0.328	1.29 (0.97-1.72)	1.36 (1.27-1.45)	Ναι	Melgene
rs6058017	G/A	<i>ASIP</i>	3' UTR	0.14 (0.12-0.18)	0.089	1.10 (0.78-1.54)	0.89 (0.81-0.99)	Όχι	Melgene

rs4911414	T/G	<i>ASIP region</i>	Διαγονιδιακά	0.24 (0.21-0.28)	0.31	0.86 (0.64-1.16)	1.21 (0.96-1.51)	Όχι	MYG
rs1015362	A/G	<i>ASIP region</i>	Διαγονιδιακά	0.27 (0.24-0.31)	0.27	0.91 (0.68-1.21)	0.89 (0.69-1.13)	Όχι	MYG
rs910873	A/G	<i>PIGU</i>	Ιντρόνιο	0.02 (0.01-0.03)	0.076	2.11 (0.96-4.67)	1.52 (1.36-1.70)	Ναι	Melgene
rs17305573	C/T	<i>PIGU</i>	Ιντρόνιο	0.02 (0.01-0.03)	0.09	2.11 (0.96-4.67)	1.53 (1.25-1.87)	Όχι	GWAS
rs1885120	C/G	<i>MYH7B</i>	Ιντρόνιο	0.02 (0.01-0.03)	0.073	2.22 (1.01-4.88)	1.59 (1.41-1.79)	Ναι	Melgene
rs4911442	G/A	<i>NCOA6</i>	Ιντρόνιο	0.04 (0.03-0.07)	0.15	1.79 (1.02-3.14)	1.2 (0.99-1.46)	Ναι	GWAS
rs2284063	G/A	<i>PLA2G6</i>	Ιντρόνιο	0.36 (0.32-0.40)	0.366	1.12 (0.87-1.43)	0.85 (0.75-0.96)	Ναι	Melgene
rs6001027	T/C	<i>PLA2G6</i>	Ιντρόνιο	0.30 (0.27-0.34)	0.363	0.88 (0.67-1.17)	0.86 (0.77-0.96)	Ναι	Melgene

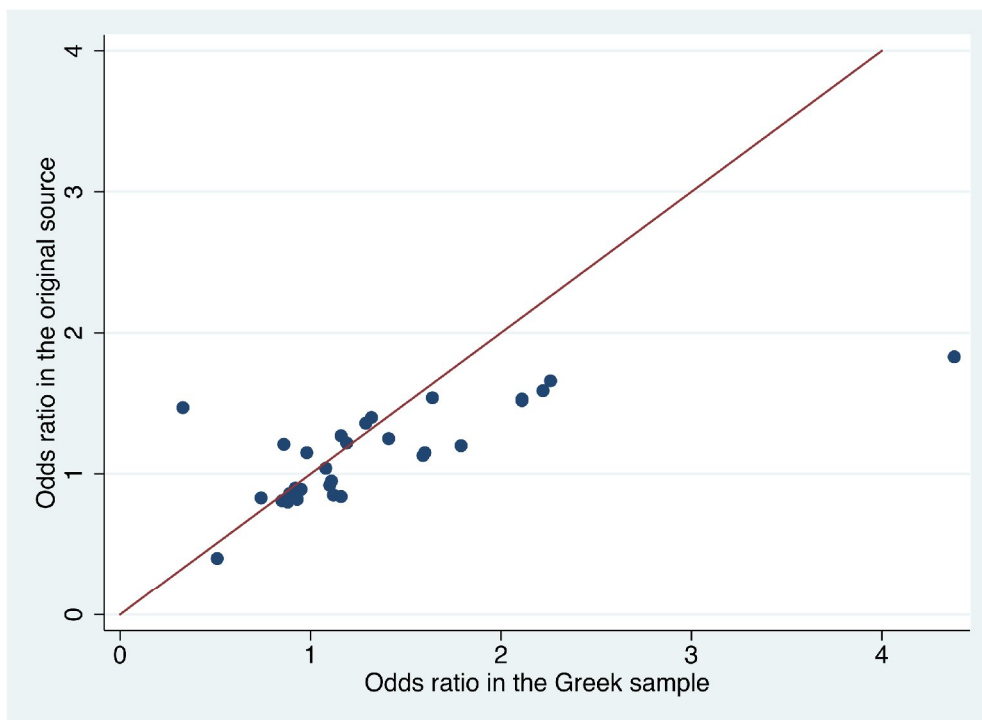
Συντομογραφίες: ΔΕ: διάστημα εμπιστοσύνης, MYG: μελέτη υποψήφιων γονιδίων, ΣΕΑ: συχνότητα ελάσσονος αλληλίου

* Πληθυσμός αναφοράς = Melgene: στατιστικά σημαντικές συσχετίσεις με μελάνωμα μετά από μετα-ανάλυση των δεδομένων για αυτόν τον πολυμορφισμό από τουλάχιστον 3 μελέτες ή δεδομένα που προέρχονται από την αρχική μελέτη στο Melgene χωρίς να έχει γίνει μετα-ανάλυση

** Απόκλιση από την ισορροπία κατά Hardy-Weinberg. Ως εκ τούτου αποκλείστηκαν από τη μελέτη, και η συχνότητα και ο λόγος αναλογιών στον ελληνικό πληθυσμό ήταν μη-εφαρμόσιμα (M/E).

Όλα τα άτομα ήταν ομόζυγα για τα μείζονα αλληλία. Ο λόγος αναλογιών στον ελληνικό πληθυσμό ήταν μη-υπολογίσιμος (M/Y).

Στην **Εικόνα 2** απεικονίζεται η συσχέτιση ανάμεσα στους λόγους αναλογιών για τους 33 πολυμορφισμούς στον ελληνικό πληθυσμό και στην αρχική πηγή από την οποία προήλθε αυτός ο πολυμορφισμός. Υπήρξε μία μέτρια συσχέτιση ($r=0.72$, $P<0.0001$) χωρίς να υπάρχουν διαφορές στους λόγους αναλογιών πέρα από την τύχη (τα 95% διαστήματα εμπιστοσύνης για τους δύο πληθυσμούς αλληλοεπικαλύπτονταν για κάθε έναν πολυμορφισμό), ενώ δεν υπήρξε και στατιστικά σημαντική διαφορά στους λόγους αναλογιών στο σύνολό τους ($P = 0.411$ for Mann-Whitney U).



Εικόνα 2. Συσχέτιση των μεγεθών αποτελέσματος στον Ελληνικό πληθυσμό και στις αρχικές μελέτες.

Δεν περιλαμβάνονται οι πολυμορφισμοί rs4636294 (αποκλείστηκε από τις αναλύσεις λόγω απόκλισης από τη ισορροπία Hardy-Weinberg), rs1805009 και rs11547464 (γιατί όλοι οι συμμετέχοντες στη μελέτη ήταν ομόζυγοι για τα αλληλία αυτά και κατά συνέπεια δεν μπορούσε να προσδιοριστεί το αλληλίο κινδύνου) και rs1011970 (γιατί το μέγεθος του αποτελέσματος δεν αναφερόταν στην αρχική δημοσίευση).

Όταν η ανάλυση περιορίστηκε σε πολυμορφισμούς που είχαν προηγουμένως φτάσει σε επίπεδα στατιστικής σημαντικότητας σε επίπεδο γονιδιώματος είτε στο Melgene είτε σε μελέτη γονιδιακής σάρωσης, η συσχέτιση των λόγων αναλογιών ήταν $r=0.83$ ($P<0.0001$) και η συσχέτιση των συχνοτήτων των αλληλίων κινδύνου ήταν $r=0.98$ ($P<0.0001$). Αντιθέτως, για τους 14 πολυμορφισμούς που δεν ήταν στατιστικά σημαντικοί σε επίπεδο γονιδιώματος, οι αντίστοιχοι συντελεστές συσχέτισης ήταν $r=0.24$ ($P=0.43$) και $r=0.72$ ($P=0.003$).

Στην μονοπαραγοντική ανάλυση και με βάση ένα πολλαπλασιαστικό γενετικό μοντέλο, 8 πολυμορφισμοί είχαν στατιστικά σημαντικές σχέσεις με το μελάνωμα σε επίπεδο $P=0.05$ (**Πίνακας 16**). Όλοι αυτοί οι πολυμορφισμοί, εκτός από το rs1805005, είχαν φτάσει και σε επίπεδα στατιστικής σημαντικότητας σε επίπεδο γονιδιώματος σε προηγούμενες μελέτες. Οι ισχυρότερες γενετικές σχέσεις αφορούσαν τους πολυμορφισμούς rs401681-T στο γονίδιο *CLPTM1L* ($\Delta A=1.60$, 95% ΔΕ 1.22-2.10; $P=0.001$), rs16891982-C στο γονίδιο *SCL45A2* ($\Delta A=0.51$, 95% ΔΕ 0.34-0.76; $P=0.001$), και rs1805007-T στο γονίδιο *MC1R* ($\Delta A=4.38$, 95% ΔΕ 2.03-9.43; $P=2\times 10^{-5}$). Οι 3 αυτές σχέσεις παρέμειναν σημαντικές και μετά τη διόρθωση για πολλαπλές συγκρίσεις κατά Bonferroni για 33 συγκρίσεις. Οι υπόλοιποι στατιστικά σημαντικοί πολυμορφισμοί ήταν rs258322-T στο *CDK10* ($\Delta A=2.26$, 95% ΔΕ 1.32-3.88; $P=0.003$), rs1805005-T στο *MC1R* ($\Delta A=1.59$, 95% ΔΕ 1.09-2.32; $P=0.016$), rs1885120-C στο *MYH7B* ($\Delta A=2.22$, 95% ΔΕ 1.01-4.88; $P=0.047$), rs2218220-T στο *MTAP* ($\Delta A=0.74$, 95% ΔΕ 0.56-0.97; $P=0.032$) και rs4911442-G στην περιοχή *ASIP* ($\Delta A=1.79$, 95% ΔΕ 1.02-3.14; $P=0.042$).

Στις πολυπαραγοντικές αναλύσεις, 5 συνολικά πολυμορφισμοί είχαν στατιστικά σημαντικές σχέσεις με μελάνωμα μετά από έλεγχο για τις επιδράσεις του χρώματος μαλλιών, του χρώματος δέρματος, του χρώματος ματιών, του φωτοτύπου, του ιστορικού εγκαύματος και της ικανότητας μελάγχρωσης του δέρματος (**Πίνακας 16**).

Πίνακας 16. Αποτελέσματα μονοπαραγοντικής και πολυπαραγοντικής ανάλυσης για τη σχέση μεταξύ πολυμορφισμών και μελανώματος

Πολυμορφισμός- Έλασσον αλληλίο	ΣΕΑ	Γενετικός τόπος	Μονοπαραγοντική ανάλυση		Πολυπαραγοντική ανάλυση		MelGene **		Γνωστές γενετικές σχέσεις	
			ΛΑ (95% ΔΕ)	P	ΛΑ (95% ΔΕ)	P	ΛΑ (95% ΔΕ)	P	Σπίλοι	Μελάχρωση
rs258322-T *	0.05	<i>CDK10</i>	2.26 (1.32-3.88)	0.003	1.77 (0.68-4.62)	0.241	1.66 (1.48-1.86)	4×10 ⁻¹⁸	Όχι	Ναι
rs401681-T	0.4	<i>CLPTM1L</i>	1.60 (1.22-2.10)	0.001	1.99 (1.21-3.26)	0.006	1.15 (1.08-1.22)	9.6×10 ⁻⁶	Ασθενής	Όχι
rs1805005-T	0.13	<i>MC1R</i>	1.59 (1.09-2.32)	0.016	1.61 (0.81-3.20)	0.179	1.13 (1.02-1.26)	0.024	Όχι	Ναι
rs1805007-T	0.02	<i>MC1R</i>	4.38 (2.03-9.43)	2×10 ⁻⁵	5.50 (1.37-22.15)	0.016	1.83 (1.56-2.15)	2.7×10 ⁻¹³	Όχι	Ναι
rs1885120-C	0.02	<i>MYH7B</i>	2.22 (1.01-4.88)	0.047	3.10 (0.89-10.82)	0.0176	1.59 (1.41-1.79)	7.4×10 ⁻¹⁵	Όχι	Ναι
rs2218220-T	0.41	<i>MTAP</i>	0.74 (0.56-0.97)	0.032	0.54 (0.33-0.90)	0.05	0.84 (0.80-0.89)	5.5×10 ⁻¹¹	Ναι	Όχι
rs4911442-G #	0.05	<i>(NCOA6) ASIP region</i>	1.79 (1.02-3.14)	0.042	3.29 (1.21-8.93)	0.02	1.2 (0.99-1.46)	1.03×10 ⁻⁸	Όχι	Ναι
rs16891982-C	0.14	<i>SCL45A2</i>	0.51 (0.34-0.76)	0.001	0.39 (0.17-0.89)	0.042	0.40 (0.33-0.47)	4×10 ⁻²⁷	Όχι	Ναι

* Για την ανάλυση χρησιμοποιήθηκε η αρνητική αλυσίδα DNA.

** MelGene: Δεδομένα από τη βάση MelGene (<http://www.melgene.org/>). Ο ΛΑ (95% ΔΕ) και η τιμή P αντιστοιχούν σε στατιστικά σημαντικές σχέσεις για μελάνωμα μετά από μετα-ανάλυση για κάθε πολυμορφισμό

Για αυτόν τον πολυμορφισμό δεν είχε πραγματοποιηθεί μετα-ανάλυση στο MelGene λόγω ανεπαρκούς αριθμού μελετών. Τα δεδομένα εδώ προέρχονται από την αρχική μελέτη γονιδιακής σάρωσης που ανέφερε τη γενετική σχέση για μελάνωμα.

Αναλύσεις Ισχύος

Η παρούσα μελέτη είχε ισχύ να ανακαλύψει πολυμορφισμούς με μεγέθη αποτελέσματος παρόμοια με αυτά που είχαν βρεθεί σε προηγούμενες μελέτες, δεδομένων και των συχνοτήτων των αλληλίων στον ελληνικό πληθυσμό, που κυμαίνονταν από 5.2% για το rs12203592 έως 100% για το rs16891982 σε επίπεδο $\alpha=0.05$. Αθροίζοντας του εκτιμητές της ισχύος για όλους τους πολυμορφισμούς, υπολογίσαμε ότι αν οι λόγοι αναλογιών ήταν πανομοιότυποι στον ελληνικό πληθυσμό, τότε η παρούσα μελέτη θα έβρισκε 8 στατιστικά σημαντικές σχέσεις από τις 33 που εξετάστηκαν. Από τους 18 πολυμορφισμούς που είχαν στατιστική σημαντικότητα σε επίπεδο γονιδιώματος σε προηγούμενες μελέτες, η παρούσα μελέτη αναμενόταν να βρει 6 στατιστικά σημαντικές σχέσεις – εν προκειμένω βρήκε 7 τέτοιες σχέσεις.

Συχνότητες αλληλίων κινδύνου στον ελληνικό πληθυσμό και στο HarMap

Για 20 πολυμορφισμούς, τα ελάσσονα αλλήλια ήταν και τα αλλήλια κινδύνου για μελάνωμα ($LA>1$). Στον **Πίνακα 17** φαίνονται τα αλλήλια κινδύνου στον ελληνικό πληθυσμό και στον ευρωπαϊκής καταγωγής πληθυσμό του HarMap (HarMap CEU).

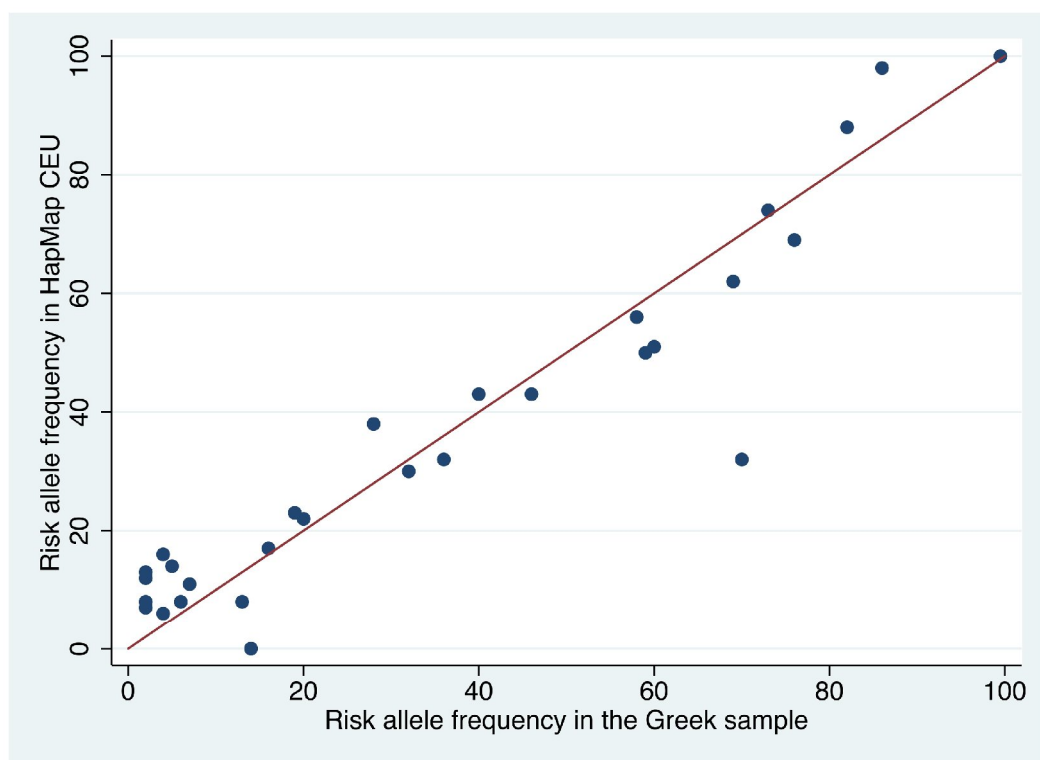
Πίνακας 17. Γονοτυπηθέντα SNPs, αλληλία κινδύνου και συχνότητές τους στον Ελληνικό πληθυσμό και στην Ευρωπαϊκή ομάδα του HarMap (HarMap CEU)

SNP	Αλληλία κινδύνου στον Ελληνικό πληθυσμό	Συχνότητα αλληλίου κινδύνου (95% ΔΕ)	
		Ελληνικός πληθυσμός	HarMap CEU
rs16891982	G	0.86 (0.83-0.89)	0.98 (0.94-0.99)
rs401681	T	0.40 (0.36-0.44)	0.43 (0.36-0.50)
rs12203592	T	0.04 (0.03-0.07)	0.16 (0.11-0.21)
rs7023329	A	0.60 (0.56-0.64)	0.51 (0.44-0.57)
rs11515	C	0.82 (0.79-0.85)	0.88 (0.80-0.92)
rs3088440	A	0.07 (0.05-0.09)	0.11 (0.07-0.15)
rs4636294 *	M/E	M/E	0.50 (0.43-0.57)
rs1335510	T	0.69 (0.65-0.73)	0.62 (0.54-0.67)
rs2218220	C	0.59 (0.54-0.63)	0.50 (0.43-0.57)
rs10757257	G	0.69 (0.65-0.73)	0.62 (0.55-0.68)
rs1011970	T	0.16 (0.13-0.2)	0.17 (0.12-0.23)
rs1408799	T	0.32 (0.28-0.36)	0.30 (0.24-0.36)
rs1042602	A	0.46 (0.42-0.51)	0.43 (0.36-0.49)
rs1126809	A	0.20 (0.16-0.23)	0.22 (0.14-0.29)
rs1393350	A	0.19 (0.16-0.23)	0.23 (0.17-0.28)
rs1544410	G	0.58 (0.54-0.62)	0.56 (0.49-0.62)
rs1800407	A	0.06 (0.04-0.08)	0.08 (0.04-0.11)
rs258322	T	0.05 (0.03-0.07)	0.14 (0.09-0.18)
rs1805005	T	0.13 (0.11-0.16)	0.08 (0.04-0.12)
rs1805007	T	0.02 (0.01-0.03)	0.12 (0.08-0.17)
rs1805008	T	0.02 (0.01-0.04)	0.13 (0.08-0.17)
rs1805009	G	M/E	1 (M/E)
rs1805006	C	0.995 (0.98-0.99)	1 (M/E)
rs11547464	G	M/E	1 (M/E)
rs4785763	A	0.28 (0.24-0.32)	0.38 (0.31-0.44)
rs6058017	G	0.14 (0.12-0.18)	0 (M/E)
rs4911414	G	0.76 (0.72-0.79)	0.69 (0.62-0.74)
rs1015362	G	0.73 (0.69-0.76)	0.74 (0.67-0.79)
rs910873	A	0.02 (0.01-0.03)	0.08 (0.05-0.12)
rs17305573	C	0.02 (0.01-0.03)	0.08 (0.05-0.12)
rs1885120	C	0.02 (0.01-0.03)	0.07 (0.04-0.11)
rs4911442	G	0.04 (0.03-0.07)	0.06 (0.02-0.12)
rs2284063	G	0.36 (0.32-0.40)	0.32 (0.25-0.38)
rs6001027	C	0.70 (0.66-0.73)	0.32 (0.25-0.38)

Συντομογραφίες: M/E: μη εφαρμόσιμο γιατί όλα τα άτομα ήταν ομόζυγα για τα μείζονα αλληλία.

* Απόκλιση από την ισορροπία κατά Hardy-Weinberg. Δεν συμπεριλήφθηκαν στην ανάλυση.

Η διάμεση συχνότητα των αλληλίων κινδύνου στον ελληνικό πληθυσμό ήταν 20% (ενδοτεταρτημοριακό εύρος, 4-60%) και στον πληθυσμό HarMap CEU ήταν 32% (ενδοτεταρτημοριακό εύρος, 12-62%) ($P=0.243$ για τη σύγκρισή τους με τεστ Mann-Whitney U), ενώ η συσχέτιση των συχνοτήτων των 2 πληθυσμών ήταν ιδιαίτερα υψηλή ($r=0.95$, $P<0.0001$) (Εικόνα 3).



Εικόνα 3. Συσχέτιση ανάμεσα στις συχνότητες των αλληλίων κινδύνου στον Ελληνικό πληθυσμό και τις συχνότητες των αντίστοιχων αλληλίων στην Ευρωπαϊκή ομάδα του HarMap.

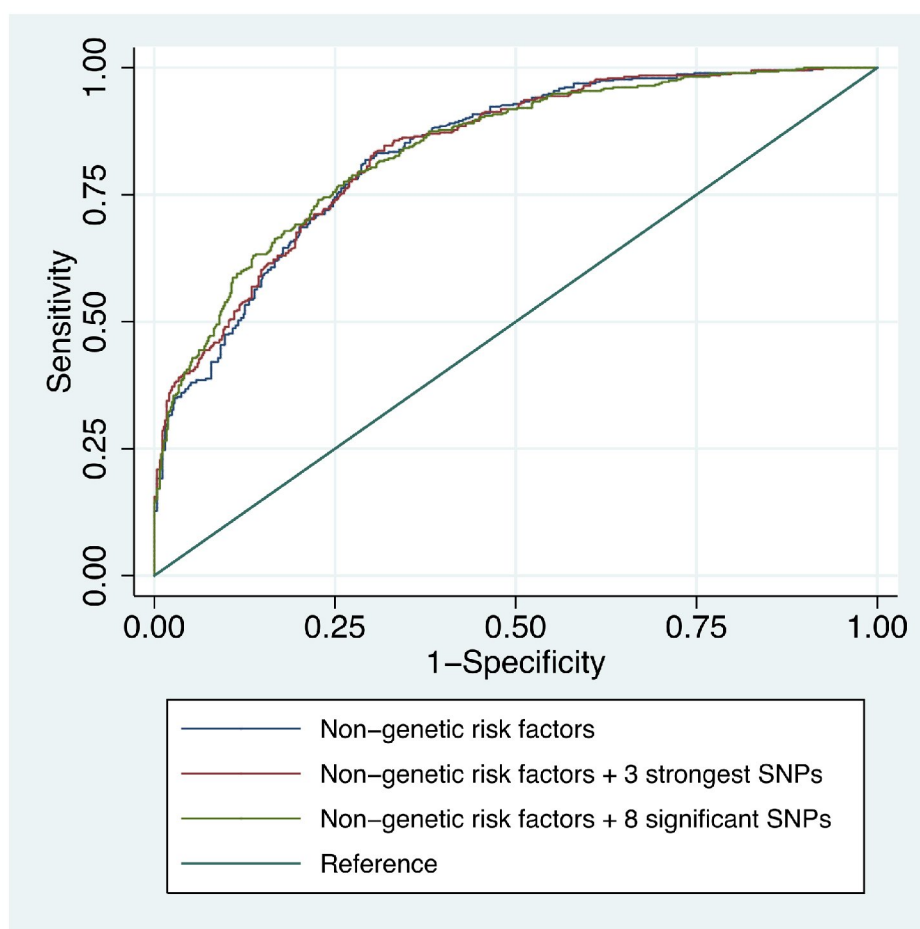
Δεν περιλαμβάνονται τα SNPs rs4636294 (αποκλείστηκε από τις αναλύσεις λόγω απόκλισης από την ισορροπία Hardy-Weinberg), rs1805009 και rs11547464, γιατί όλοι οι συμμετέχοντες στη μελέτη ήταν ομόζυγοι για τα αλλήλια αυτά και κατά συνέπεια δεν μπορούσε να προσδιοριστεί το αλλήλιο κινδύνου.

Οι συχνότητες των αλληλίων κινδύνου για 9 πολυμορφισμούς (rs6001027-C, rs16891982-G, rs12203592-T, rs258322-T, rs1805007-T, rs1805008-T, rs910873-A, rs17305573-C, and rs1885120-C) διέφεραν πέρα από την τύχη ανάμεσα στον ελληνικό πληθυσμό και τον πληθυσμό HarMap CEU (δηλαδή τα 95% διαστήματα εμπιστοσύνης των συχνοτήτων στους δύο πληθυσμούς δεν αλληλοκαλύπτονταν). Για όλους αυτούς τους πολυμορφισμούς (με εξαίρεση τον rs6001027 στο *PLA2G6* που σχετίζεται με τους σπίλους) τα αλληλία κινδύνου στον ελληνικό πληθυσμό είχαν σημαντικά χαμηλότερες συχνότητες από ό,τι στον πληθυσμό HarMap CEU, ενώ 6 από αυτούς βρίσκονται σε γονίδια που έχουν εδραιωμένες σχέσεις με φαινοτύπους που σχετίζονται με το χρώμα του δέρματος, των μαλλιών ή των ματιών (rs16891982 στο *SCL45A2*, rs12203592 στο *IRF4*, rs258322 στο *CDK10*, rs1885120 στο *MYH7B*, rs1805007 and rs1805008 στο *MC1R*).

Προγνωστική αξία των πολυμορφισμών για μελάνωμα

Στην **Εικόνα 4** φαίνονται οι καμπύλες ROC και οι αντίστοιχες περιοχές AUC για 3 προγνωστικά μοντέλα με διαφορετικό βαθμό γενετικής πληροφορίας. Συγκρινόμενα με τα φαινοτυπικά χαρακτηριστικά μόνα τους, τα μοντέλα που περιλάμβαναν πολυμορφισμούς που σχετίζονταν με μελάνωμα, βελτίωναν την περιοχή AUC κατά ένα μικρό ποσοστό. Το προγνωστικό μοντέλο M1 που περιλάμβανε μόνο τα φαινοτυπικά χαρακτηριστικά (χρώμα ματιών, χρώμα δέρματος, ηλιακό έγκαυμα και ικανότητα μελάγχρωσης) είχε AUC=83.3%, ενώ το μοντέλο M2 που περιλάμβανε επιπλέον των φαινοτυπικών χαρακτηριστικών και τους 3 πολυμορφισμούς που ήταν σημαντικοί μετά τη διόρθωση κατά Bonferroni είχε AUC=83.7%, και το μοντέλο M3 που περιλάμβανε τα

φαινοτυπικά χαρακτηριστικά και όλους τους 8 σημαντικούς πολυμορφισμούς είχε AUC=83.9%. Σε σχέση με το μη-γενετικό μοντέλο, τα 2 γενετικά μοντέλα δεν συνεισέφεραν στατιστικά σημαντικές βελτιώσεις όσον αφορά την πρόγνωση του μελανώματος ($P=0.42$ για τη σύγκριση M1-M2, και $P=0.66$ για τη σύγκριση M1-M3).



Εικόνα 4. Καμπύλες ROC όπου φαίνεται η διακριτική ικανότητα των 3 μοντέλων που μελετήθηκαν.

2.3.4. ΣΥΖΗΤΗΣΗ

Η παρούσα μελέτη επαλήθευσε στον ελληνικό πληθυσμό γενετικές σχέσεις για μελάνωμα, οι οποίες αφορούσαν τόσο κοινούς όσο και χαμηλής συχνότητας πολυμορφισμούς με συχνότητες των ελασσόνων αλληλίων που κυμαίνονταν από 2% έως 41%. Οχτώ γενετικές σχέσεις ήταν στατιστικά σημαντικές στον ελληνικό πληθυσμό, οι περισσότερες εκ των οποίων ήταν στατιστικά σημαντικές σε επίπεδο γονιδιώματος σε προηγούμενες μελέτες. Η ισχυρή επαλήθευση γενετικών περιοχών, όπως 20q11.2 (περιοχή *ASIP*), 9p21 (περιοχή *MTAP*), 16q24 (περιοχή *MC1R*) και 5p13 (περιοχή *CLPTM1L*), που είχαν αρχικά ανακαλυφθεί σε μελέτες γονιδιακής σάρωσης, καταδεικνύει τη μεγάλη συνεισφορά αυτών των μελετών με την αγνωστικιστική προσέγγιση για την ανακάλυψη νέων γενετικών σχέσεων για τα πολυσύνθετα νοσήματα. Για 8 πολυμορφισμούς τα αλληλία κινδύνου είχαν σημαντικά χαμηλότερες συχνότητες στον ελληνικό πληθυσμό σε σχέση με τον πληθυσμό αναφοράς του HarMap, ενώ για μόνο 1 πολυμορφισμό ίσχυε το αντίθετο. Τα γενετικά μοντέλα που περιλάμβαναν πολυμορφισμούς που προδιαθέτουν για μελάνωμα, προσέφεραν μικρή μόνο βελτίωση στην προγνωστική ικανότητα των φαινοτυπικών χαρακτηριστικών του δέρματος.

Η μελέτη αυτή στόχευε να επικυρώσει έναν αριθμό πολυμορφισμών στον ελληνικό πληθυσμό δεδομένων πρόσφατων αναφορών σχετικά με την αυξημένη γενετική συνεισφορά μεταλλάξεων στο *CDKN2A/CDK4* (288). Πρόσφατες μελέτες γονιδιακής σάρωσης για μελάνωμα ανακάλυψαν έναν αριθμό γονιδίων που σχετίζονται με τον έλεγχο του κυτταρικού κύκλου, τη διατήρηση των τελομερών και την αντίδραση σε βλάβη του DNA, όπως τα *MITF*, *ATM*, *PARP-1*, *TERT*, *CASP8*, *CCND1*, καθώς και τα *MX2*, *SETDB1* και *ARNT/LASS2/ANXA9* (299-

302). Τα ευρήματα της παρούσας μελέτης υπογραμμίζουν τη σημασία γονιδίων που σχετίζονται με φαινοτύπους χρώσης του δέρματος και την αντίδραση στη βλάβη του DNA όσον αφορά τον κίνδυνο εμφάνισης μελανώματος στον ελληνικό πληθυσμό. Κατ' επέκταση, αυτό πιθανόν να αντανακλά τη σημασία αυτών των βιολογικών λειτουργιών για την ανάπτυξη του μελανώματος σε πληθυσμούς με πιο σκούρα χροιά δέρματος, που κατοικούν σε περιοχές με μεγάλης διάρκειας έκθεση σε ηλιακή ακτινοβολία. Οι περισσότεροι πολυμορφισμοί με σημαντικά χαμηλότερες συχνότητες στον ελληνικό πληθυσμό βρίσκονται σε γονιδιακούς τόπους που σχετίζονται με τα χαρακτηριστικά του χρώματος (*SCL45A2*, *IRF4*, *CDK10*, *MYH7B*, *MC1R*), και όλοι, εκτός από τους rs16891982, rs258322, rs1805007 και rs1885120, επαληθεύθηκαν στον ελληνικό πληθυσμό. Μάλιστα στο γονίδιο *MC1R* έχουν βρεθεί τόσο κοινοί όσο και μη κοινοί πολυμορφισμοί που σχετίζονται με τους ανωτέρω φαινότυπους, καταδεικνύοντας το ρόλο των πολυμορφισμών αυτών στην αιτιολογία του μελανώματος (56). Ειδικότερα, όσον αφορά τους πολυμορφισμούς χαμηλής συχνότητας, πρόσφατα ανακαλύφθηκαν με τη χρήση αλληλούχισης σπάνιοι πολυμορφισμοί στο γονίδιο *POT1* που προδιαθέτουν σε οικογενές μελάνωμα (303).

Αυτά τα ευρήματα υποδηλώνουν ότι πιθανόν να υπάρχουν διαφορές στο γενετικό υπόβαθρο που καθορίζει τις φαινοτυπικές διαφορές μεταξύ του ελληνικού πληθυσμού και άλλων ευρωπαϊκών πληθυσμών, και θα μπορούσαν να εξηγήσουν σε κάποιο βαθμό τη χαμηλότερη συχνότητα του μελανώματος σε πληθυσμούς με πιο σκούρα χροιά δέρματος, που κατοικούν σε περιοχές με μεγάλης διάρκειας έκθεση σε ηλιακή ακτινοβολία. Επιπλέον, τα αποτελέσματα της παρούσας μελέτης υποδεικνύουν το ρόλο της φυσικής επιλογής, που

περιορίζει τη συχνότητα των αλληλίων κινδύνου σε έναν πληθυσμό με μεγάλη έκθεση σε ηλιακή ακτινοβολία και αυξάνει τη συχνότητα προστατευτικών αλληλίων που δρουν επίσης μέσω των προστατευτικών μονοπατιών της χρώσης του δέρματος. Ωστόσο, άτομα από τον ελληνικό πληθυσμό που έχουν συγκεκριμένα αλλήλια κινδύνου σε γονίδια που καθορίζουν χαρακτηριστικά του χρώματος, μπορεί να βρίσκονται σε αυξημένο κίνδυνο για την εμφάνιση μελανώματος.

ΠΑΡΑΡΤΗΜΑ 1

**Άρθρα που πραγματοποίησαν μετα-ανάλυση μελετών γονιδιακής
σάρωσης στο στάδιο της ανακάλυψης, τα οποία συμπεριλήφθηκαν στη
διατριβή**

- S1. Amin N, Byrne E, Johnson J, Chenevix-Trench G, Walter S, et al. 2012. Genome-wide association analysis of coffee drinking suggests association with CYP1A1/CYP1A2 and NRCAM. *Mol Psychiatry* 17: 1116-29
- S2. Anderson CA, Boucher G, Lees CW, Franke A, D'Amato M, et al. 2011. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet* 43: 246-52
- S3. Arking DE, Juntila MJ, Goyette P, Huertas-Vazquez A, Eijgelsheim M, et al. 2011. Identification of a sudden cardiac death susceptibility locus at 2q24.2 through genome-wide association in European ancestry individuals. *PLoS Genet* 7: e1002158
- S4. Benjamin EJ, Rice KM, Arking DE, Pfeufer A, van Noord C, et al. 2009. Variants in ZFHX3 are associated with atrial fibrillation in individuals of European ancestry. *Nat Genet* 41: 879-81
- S5. Bis JC, Kavousi M, Franceschini N, Isaacs A, Abecasis GR, et al. 2011. Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and Πλάκα (αθηρωματική) καρωτίδας. *Nat Genet* 43: 940-7
- S6. Boger CA, Chen MH, Tin A, Olden M, Kottgen A, et al. 2011. CUBN is a gene locus for albuminuria. *J Am Soc Nephrol* 22: 555-70
- S7. Boraska V, Day-Williams A, Franklin CS, Elliott KS, Panoutsopoulou K, et al. 2012. Genome-wide association study to identify common variants associated with brachial circumference: a meta-analysis of 14 cohorts. *PLoS One* 7: e31369
- S8. Bradfield JP, Qu HQ, Wang K, Zhang H, Sleiman PM, et al. 2011. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLoS Genet* 7: e1002293
- S9. Bradfield JP, Taal HR, Timpson NJ, Scherag A, Lecoecur C, et al. 2012. A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat Genet* 44: 526-31

- S10. Chambers JC, Zhang W, Sehmi J, Li X, Wass MN, et al. 2011. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet* 43: 1131-8
- S11. Cornelis MC, Monda KL, Yu K, Paynter N, Azzato EM, et al. 2011. Genome-wide meta-analysis identifies regions on 7p21 (AHR) and 15q24 (CYP1A2) as determinants of habitual caffeine consumption. *PLoS Genet* 7: e1002033
- S12. Dastani Z, Hivert MF, Timpson N, Perry JR, Yuan X, et al. 2012. Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet* 8: e1002607
- S13. de Moor MH, Costa PT, Terracciano A, Krueger RF, de Geus EJ, et al. 2012. Meta-analysis of genome-wide association studies for personality. *Mol Psychiatry* 17: 337-49
- S14. Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, et al. 2011. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation* 123: 731-8
- S15. Dehghan A, Yang Q, Peters A, Basu S, Bis JC, et al. 2009. Association of novel genetic Loci with circulating fibrinogen levels: a genome-wide association study in 6 population-based cohorts. *Circ Cardiovasc Genet* 2: 125-33
- S16. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, et al. 2010. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 42: 105-16
- S17. Eijgelsheim M, Newton-Cheh C, Sotoodehnia N, de Bakker PI, Muller M, et al. 2010. Genome-wide association analysis identifies multiple loci related to resting heart rate. *Hum Mol Genet* 19: 3885-94
- S18. Elks CE, Perry JR, Sulem P, Chasman DI, Franceschini N, et al. 2010. Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat Genet* 42: 1077-85
- S19. Estrada K, Styrkarsdottir U, Evangelou E, Hsu YH, Duncan EL, et al. 2012. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat Genet* 44: 491-501
- S20. Evangelou E, Valdes AM, Kerkhof HJ, Styrkarsdottir U, Zhu Y, et al. 2011. Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. *Ann Rheum Dis* 70: 349-55
- S21. Evans DM, Spencer CC, Pointon JJ, Su Z, Harvey D, et al. 2011. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat Genet* 43: 761-7

- S22. Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, et al. 2010. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 42: 1118-25
- S23. Freathy RM, Mook-Kanamori DO, Sovio U, Prokopenko I, Timpson NJ, et al. 2010. Variants in ADCY5 and near CCNL1 are associated with fetal growth and birth weight. *Nat Genet* 42: 430-5
- S24. Ganesh SK, Zakai NA, van Rooij FJ, Soranzo N, Smith AV, et al. 2009. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat Genet* 41: 1191-8
- S25. Gieger C, Radhakrishnan A, Cvejic A, Tang W, Porcu E, et al. 2011. New gene functions in megakaryopoiesis and platelet formation. *Nature* 480: 201-8
- S26. Grallert H, Dupuis J, Bis JC, Dehghan A, Barbalic M, et al. 2012. Eight genetic loci associated with variation in lipoprotein-associated phospholipase A2 mass and activity and coronary heart disease: meta-analysis of genome-wide association studies from five community-based studies. *Eur Heart J* 33: 238-51
- S27. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, et al. 2010. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet* 42: 949-60
- S28. Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, et al. 2011. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 43: 429-35
- S29. Ikram MA, Seshadri S, Bis JC, Fornage M, DeStefano AL, et al. 2009. Genomewide association studies of stroke. *N Engl J Med* 360: 1718-28
- S30. Ikram MK, Sim X, Jensen RA, Cotch MF, Hewitt AW, et al. 2010. Four novel Loci (19q13, 6q24, 12q24, and 5q14) influence the microcirculation in vivo. *PLoS Genet* 6: e1001184
- S31. International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret GB, Munroe PB, Rice KM, Bochud M, et al. 2011. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 478: 103-9
- S32. International Parkinson Disease Genomics Consortium, Nalls MA, Plagnol V, Hernandez DG, Sharma M, et al. 2011. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* 377: 641-9
- S33. Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, et al. 2010. Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet* 42: 210-5
- S34. Kaplan RC, Petersen AK, Chen MH, Teumer A, Glazer NL, et al. 2011. A genome-wide association study identifies novel loci associated with circulating IGF-I and IGFBP-3. *Hum Mol Genet* 20: 1241-51

- S35. Kato N, Takeuchi F, Tabara Y, Kelly TN, Go MJ, et al. 2011. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat Genet* 43: 531-8
- S36. Kestenbaum B, Glazer NL, Kottgen A, Felix JF, Hwang SJ, et al. 2010. Common genetic variants associate with serum phosphorus concentration. *J Am Soc Nephrol* 21: 1223-32
- S37. Kiemeny LA, Sulem P, Besenbacher S, Vermeulen SH, Sigurdsson A, et al. 2010. A sequence variant at 4p16.3 confers susceptibility to urinary bladder cancer. *Nat Genet* 42: 415-9
- S38. Kilpelainen TO, Zillikens MC, Stancakova A, Finucane FM, Ried JS, et al. 2011. Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. *Nat Genet* 43: 753-60
- S39. Kooner JS, Saleheen D, Sim X, Sehmi J, Zhang W, et al. 2011. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat Genet* 43: 984-9
- S40. Kristiansson K, Perola M, Tikkanen E, Kettunen J, Surakka I, et al. 2012. Genome-wide screen for metabolic syndrome susceptibility Loci reveals strong lipid gene contribution but no evidence for common genetic basis for clustering of metabolic syndrome traits. *Circ Cardiovasc Genet* 5: 242-9
- S41. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, et al. 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467: 832-8
- S42. Ligthart L, de Vries B, Smith AV, Ikram MA, Amin N, et al. 2011. Meta-analysis of genome-wide association for migraine in six population-based European cohorts. *Eur J Hum Genet* 19: 901-7
- S43. Lindgren CM, Heid IM, Randall JC, Lamina C, Steinthorsdottir V, et al. 2009. Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. *PLoS Genet* 5: e1000508
- S44. McMahon FJ, Akula N, Schulze TG, Muglia P, Tozzi F, et al. 2010. Meta-analysis of genome-wide association data identifies a risk locus for major mood disorders on 3p21.1. *Nat Genet* 42: 128-31
- S45. Meyer TE, Verwoert GC, Hwang SJ, Glazer NL, Smith AV, et al. 2010. Genome-wide association studies of serum magnesium, potassium, and sodium concentrations identify six Loci influencing serum magnesium levels. *PLoS Genet* 6
- S46. Mitchell GF, Verwoert GC, Tarasov KV, Isaacs A, Smith AV, et al. 2012. Common genetic variation in the 3'-BCL11B gene desert is associated with carotid-femoral pulse wave velocity and excess cardiovascular disease risk: the AortaGen Consortium. *Circ Cardiovasc Genet* 5: 81-90
- S47. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, et al. 2010. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 363: 1211-21

- S48. Murabito JM, White CC, Kavousi M, Sun YV, Feitosa MF, et al. 2012. Association between chromosome 9p21 variants and the ankle-brachial index identified by a meta-analysis of 21 genome-wide association studies. *Circ Cardiovasc Genet* 5: 100-12
- S49. N'Diaye A, Chen GK, Palmer CD, Ge B, Tayo B, et al. 2011. Identification, replication, and fine-mapping of Loci associated with adult height in individuals of african ancestry. *PLoS Genet* 7: e1002298
- S50. O'Seaghdha CM, Yang Q, Glazer NL, Leak TS, Dehghan A, et al. 2010. Common variants in the calcium-sensing receptor gene are associated with total serum calcium levels. *Hum Mol Genet* 19: 4296-303
- S51. Okada Y, Kamatani Y, Takahashi A, Matsuda K, Hosono N, et al. 2010. A genome-wide association study in 19 633 Japanese subjects identified LHX3-QSOX2 and IGF1 as adult height loci. *Hum Mol Genet* 19: 2303-12
- S52. Okada Y, Terao C, Ikari K, Kochi Y, Ohmura K, et al. 2012. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nat Genet* 44: 511-6
- S53. Paternoster L, Standl M, Chen CM, Ramasamy A, Bonnelykke K, et al. 2012. Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. *Nat Genet* 44: 187-92
- S54. Patsopoulos NA, Bayer Pharma MSGWG, Steering Committees of Studies Evaluating I-b, a CCRA, Consortium AN, et al. 2011. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann Neurol* 70: 897-912
- S55. Pattaro C, Kottgen A, Teumer A, Garnaas M, Boger CA, et al. 2012. Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS Genet* 8: e1002584
- S56. Pfeufer A, Sanna S, Arking DE, Muller M, Gateva V, et al. 2009. Common variants at ten loci modulate the QT interval duration in the QTSCD Study. *Nat Genet* 41: 407-14
- S57. Pfeufer A, van Noord C, Marciante KD, Arking DE, Larson MG, et al. 2010. Genome-wide association study of PR interval. *Nat Genet* 42: 153-9
- S58. Purdue MP, Johansson M, Zelenika D, Toro JR, Scelo G, et al. 2011. Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. *Nat Genet* 43: 60-5
- S59. Qayyum R, Snively BM, Ziv E, Nalls MA, Liu Y, et al. 2012. A meta-analysis and genome-wide association study of platelet count and mean platelet volume in african americans. *PLoS Genet* 8: e1002491
- S60. Ramasamy A, Curjuric I, Coin LJ, Kumar A, McArdle WL, et al. 2011. A genome-wide meta-analysis of genetic variants associated with allergic rhinitis and grass sensitization and their interaction with birth order. *J Allergy Clin Immunol* 128: 996-1005
- S61. Reiner AP, Lettre G, Nalls MA, Ganesh SK, Mathias R, et al. 2011. Genome-wide association study of white blood cell count in 16,388 African

- Americans: the continental origins and genetic epidemiology network (COGENT). *PLoS Genet* 7: e1002108
- S62. Sanson M, Hosking FJ, Shete S, Zelenika D, Dobbins SE, et al. 2011. Chromosome 7p11.2 (EGFR) variation influences glioma risk. *Hum Mol Genet* 20: 2897-904
- S63. Saxena R, Hivert MF, Langenberg C, Tanaka T, Pankow JS, et al. 2010. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* 42: 142-8
- S64. Schizophrenia Psychiatric Genome-Wide Association Study Consortium. 2011. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 43: 969-76
- S65. Schumann G, Coin LJ, Lourdasamy A, Charoen P, Berger KH, et al. 2011. Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. *Proc Natl Acad Sci USA* 108: 7119-24
- S66. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, et al. 2011. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 43: 333-8
- S67. Smith NL, Chen MH, Dehghan A, Strachan DP, Basu S, et al. 2010. Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium. *Circulation* 121: 1382-92
- S68. Smith NL, Felix JF, Morrison AC, Demissie S, Glazer NL, et al. 2010. Association of genome-wide variation with the risk of incident heart failure in adults of European and African ancestry: a prospective meta-analysis from the cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium. *Circ Cardiovasc Genet* 3: 256-66
- S69. Smith NL, Huffman JE, Strachan DP, Huang J, Dehghan A, et al. 2011. Genetic predictors of fibrin D-dimer levels in healthy adults. *Circulation* 123: 1864-72
- S70. Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, et al. 2011. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 43: 1082-90
- S71. Soranzo N, Sanna S, Wheeler E, Gieger C, Radke D, et al. 2010. Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemic and nonglycemic pathways. *Diabetes* 59: 3229-39
- S72. Sotoodehnia N, Isaacs A, de Bakker PI, Dorr M, Newton-Cheh C, et al. 2010. Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. *Nat Genet* 42: 1068-76
- S73. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, et al. 2010. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 42: 937-48

- S74. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, et al. 2010. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet* 42: 508-14
- S75. Stolk L, Perry JR, Chasman DI, He C, Mangino M, et al. 2012. Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nat Genet* 44: 260-8
- S76. Strawbridge RJ, Dupuis J, Prokopenko I, Barker A, Ahlqvist E, et al. 2011. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* 60: 2624-34
- S77. Taal HR, St Pourcain B, Thiering E, Das S, Mook-Kanamori DO, et al. 2012. Common variants at 12q15 and 12q24 are associated with infant head circumference. *Nat Genet* 44: 532-8
- S78. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, et al. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466: 707-13
- S79. Tobacco Genetics Consortium. 2010. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* 42: 441-7
- S80. Vasan RS, Glazer NL, Felix JF, Lieb W, Wild PS, et al. 2009. Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. *JAMA* 302: 168-78
- S81. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, et al. 2010. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 42: 579-89
- S82. Wain LV, Verwoert GC, O'Reilly PF, Shi G, Johnson T, et al. 2011. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet* 43: 1005-11
- S83. Walter S, Atzmon G, Demerath EW, Garcia ME, Kaplan RC, et al. 2011. A genome-wide association study of aging. *Neurobiol Aging* 32: 2109 e15-28
- S84. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, et al. 2010. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. *Lancet* 376: 180-8
- S85. Wen W, Cho YS, Zheng W, Dorajoo R, Kato N, et al. 2012. Meta-analysis identifies common variants associated with body mass index in east Asians. *Nat Genet* 44: 307-11
- S86. Yang Q, Kottgen A, Dehghan A, Smith AV, Glazer NL, et al. 2010. Multiple genetic loci influence serum urate levels and their relationship with gout and cardiovascular disease risk factors. *Circ Cardiovasc Genet* 3: 523-30
- S87. Yasuno K, Bilguvar K, Bijlenga P, Low SK, Krischek B, et al. 2010. Genome-wide association study of intracranial aneurysm identifies three new risk loci. *Nat Genet* 42: 420-5

- S88. Zhai G, Teumer A, Stolk L, Perry JR, Vandenput L, et al. 2011. Eight common genetic variants associated with serum DHEAS levels suggest a key role in ageing mechanisms. *PLoS Genet* 7: e1002025

ΠΑΡΑΡΤΗΜΑ 2

Μελέτες σάρωσης του γονιδιώματος που είχαν ανακαλύψει πολυμορφισμούς χαμηλής συχνότητας

1. Trevino LR, Yang W, French D, et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat Genet* 2009;41:1001-5.
2. Limou S, Le Clerc S, Coulonges C, et al. Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). *J Infect Dis* 2009;199:419-26.
3. Sulem P, Gudbjartsson DF, Stacey SN, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet* 2007;39:1443-52.
4. Need AC, Attix DK, McEvoy JM, et al. A genome-wide study of common SNPs and CNVs in cognitive performance in the CANTAB. *Hum Mol Genet* 2009;18:4650-61.
5. Rivadeneira F, Styrkarsdottir U, Estrada K, et al. Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. *Nat Genet* 2009;41:1199-206.
6. Franke A, Hampe J, Rosenstiel P, et al. Systematic association mapping identifies NELL1 as a novel IBD disease gene. *PLoS One* 2007;2:e691.
7. Barrett JC, Hansoul S, Nicolae DL, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 2008;40:955-62.
8. Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* 2009;41:56-65.
9. Kooner JS, Chambers JC, Aguilar-Salinas CA, et al. Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nat Genet* 2008;40:149-51.
10. Ganesh SK, Zakai NA, van Rooij FJ, et al. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat Genet* 2009;41:1191-8.
11. Kathiresan S, Melander O, Guiducci C, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* 2008;40:189-97.
12. Soranzo N, Spector TD, Mangino M, et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat Genet* 2009;41:1182-90.
13. Ng CC, Yew PY, Puah SM, et al. A genome-wide association study identifies ITGA9 conferring risk of nasopharyngeal carcinoma. *J Hum Genet* 2009;54:392-7.

14. Grant SF, Wang K, Zhang H, et al. A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. *J Pediatr* 2009;155:909-13.
15. Otowa T, Yoshida E, Sugaya N, et al. Genome-wide association study of panic disorder in the Japanese population. *J Hum Genet* 2009;54:122-6.
16. Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 2007;39:631-7.
17. Hirschfield GM, Liu X, Xu C, et al. Primary biliary cirrhosis associated with HLA, IL12A, and IL12RB2 variants. *N Engl J Med* 2009;360:2544-55.
18. Melzer D, Perry JR, Hernandez D, et al. A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet* 2008;4:e1000072.
19. Liu Y, Helms C, Liao W, et al. A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. *PLoS Genet* 2008;4:e1000041.
20. Aberg K, Adkins DE, Bukszar J, et al. Genomewide association study of movement-related adverse antipsychotic effects. *Biol Psychiatry*;67:279-82.
21. Yang JJ, Cheng C, Yang W, et al. Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia. *JAMA* 2009;301:393-403.
22. Han JW, Zheng HF, Cui Y, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat Genet* 2009;41:1234-7.
23. Graham RR, Cotsapas C, Davies L, et al. Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat Genet* 2008;40:1059-61.
24. Nan H, Kraft P, Qureshi AA, et al. Genome-wide association study of tanning phenotype in a population of European ancestry. *J Invest Dermatol* 2009;129:2250-7.
25. Willer CJ, Sanna S, Jackson AU, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 2008;40:161-9.
26. Pollin TI, Damcott CM, Shen H, et al. A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. *Science* 2008;322:1702-5.
27. Barrett JC, Clayton DG, Concannon P, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* 2009.
28. Takeuchi F, Serizawa M, Yamamoto K, et al. Confirmation of multiple risk Loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population. *Diabetes* 2009;58:1690-9.

ΣΥΜΠΕΡΑΣΜΑΤΑ ΤΗΣ ΔΙΑΤΡΙΒΗΣ

Από την παρούσα διατριβή προκύπτουν ορισμένα συμπεράσματα. Πρώτον, οι μετα-αναλύσεις μελετών γονιδιακής σάρωσης έχουν ανακαλύψει έναν ικανοποιητικό αριθμό κοινών αλλά και μη-κοινών πολυμορφισμών με ισχυρή στατιστική σημαντικότητα, που εξηγούν ένα σημαντικό μέρος της γενετικής διακύμανσης των πολυσύνθετων νοσημάτων. Η μέθοδος μετα-ανάλυσης με τη μεγαλύτερη στατιστική ισχύ είναι αυτή των σταθερών επιδράσεων, που είναι και η πιο συχνά εφαρμοζόμενη μέθοδος. Δεύτερον, για πολλούς σύνθετους φαινοτύπους υπάρχουν γενετικοί τόποι που περιέχουν πολυμορφισμούς από όλο το αλληλικό φάσμα (κοινούς, μη-κοινούς και μεταλλάξεις). Οι λόγοι αναλογιών για πολυμορφισμούς χαμηλής συχνότητας είναι κατά βάση μεγάλοι (>2) και με βάση την ισχύ των μελετών γονιδιακής σάρωσης αναμένουμε ότι οι περισσότεροι πολυμορφισμοί με μεγάλα μεγέθη έχουν ήδη ανακαλυφθεί. Τέλος, παρά την ισχυρή στατιστική σημαντικότητα, η κλινική χρησιμότητα των πολυμορφισμών που έχουν ανακαλυφθεί και επικυρωθεί για σύνθετους φαινοτύπους, είναι μάλλον περιορισμένη όσον αφορά την προγνωστική ικανότητά τους.

ΠΕΡΙΛΗΨΗ

Οι μελέτες σάρωσης του γονιδιώματος έχουν οδηγήσει στην ανακάλυψη εκατοντάδων κοινών πολυμορφισμών, αλλά και πολυμορφισμών χαμηλής συχνότητας, που σχετίζονται με σύνθετους φαινοτύπους και νοσήματα, ειδικά μέσω μετα-ανάλυσης των επιμέρους δεδομένων. Στην παρούσα διατριβή έγινε αρχικά μία συστηματική ανασκόπηση της ισχύος τέτοιων μετα-αναλύσεων. Ο διάμεσος αριθμός πολυμορφισμών που είχαν ανακαλυφθεί στις 139 μετα-αναλύσεις που εξετάστηκαν ήταν 4, με το μέγιστο αριθμό να είναι 115, ενώ κάποιες μετα-αναλύσεις δεν είχαν ανακαλύψει κάποιον πολυμορφισμό. Οι ανακαλυφθέντες πολυμορφισμοί εξηγούν ένα σημαντικό μέρος της γενετικής διακύμανσης του αντίστοιχου νοσήματος, που σε ορισμένες περιπτώσεις ξεπερνά το 50%, ενώ για τα περισσότερα νοσήματα κυμαίνεται <50%. Ακολούθως, πραγματοποιήθηκε μία συστηματική ανασκόπηση των πολυμορφισμών χαμηλής συχνότητας που έχουν βρεθεί σε μελέτες γονιδιακής σάρωσης. Εντοπίστηκαν 43 γενετικές σχέσεις για πολυμορφισμούς χαμηλής συχνότητας με κατά βάση μεγάλα μεγέθη αποτελέσματος. Πολλοί από τους πολυμορφισμούς αυτούς ήταν κοινοί σε πληθυσμούς άλλης εθνικής καταγωγής στο HarMap, ενώ για κάποιους φαινοτύπους ο ίδιος γενετικός τύπος έφερε τόσο κοινούς όσο και μη-κοινούς πολυμορφισμούς, αλλά και μεταλλάξεις που σχετίζονταν με τον ίδιο φαινότυπο. Τέλος, εξετάστηκε ο ρόλος κοινών και μη-κοινών πολυμορφισμών όσον αφορά τη γενετική βάση του μελανώματος μέσω μιας μελέτης δείκτου-ελέγχου, η οποία επικύρωσε 8 τέτοιους πολυμορφισμούς σε έναν ελληνικό πληθυσμό. Ωστόσο, η προγνωστική ικανότητα αυτών των πολυμορφισμών ήταν πολύ μικρή και κλινικά, αλλά και στατιστικά μη-

σημαντική συγκρινόμενη με γνωστούς φαινοτυπικούς παράγοντες κινδύνου, όπως το χρώμα του δέρματος.

ABSTRACT

Author: Orestis Panagiotou, MD

Title: «Genetic associations for low-frequency genetic polymorphisms in genome-wide association studies»

Genome-wide association studies (GWAS) have discovered hundreds of common and uncommon single nucleotide polymorphisms (SNPs) associated with common traits and diseases, especially through meta-analysis of individual studies within international consortia. In the current thesis, we initially conducted a systematic appraisal of the power of published genome-wide meta-analyses. The 139 eligible meta-analyses had discovered on average 4 SNPs each, while the maximum number of discoveries was 115. These SNPs explain a substantial portion of the genetic variance of the respective traits which exceeds 50% in some cases. Next, we systematically reviewed the low-frequency/uncommon SNPs discovered in GWAS. A total of 43 associations were identified and most of them had rather large effect sizes. Many of these uncommon SNPs were common in different HapMap populations than those discovered, while for some phenotypes the same genetic locus harbors both common and uncommon SNPs as well mutations associated with the same phenotype. Finally, we examined the role of common and uncommon SNPs in the genetic susceptibility of melanoma conducting a case-control study. This study replicated 8 such SNPs in the Greek population. However, the predictive ability of the replicated SNPs was small and clinically non-significant when compared to established phenotypic risk factors such as skin color.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Watson JD, Crick FH. Genetical implications of the structure of deoxyribonucleic acid. *Nature* 1953;171(4361):964-7.
2. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 1953;171(4356):737-8.
3. Avery OT, Macleod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types : Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* type III. *J Exp Med* 1944;79(2):137-58.
4. Clamp M, Fry B, Kamal M, et al. Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci U S A* 2007;104(49):19428-33.
5. Feero WG, Guttmacher AE, Collins FS. Genomic medicine--an updated primer. *N Engl J Med* 2010;362(21):2001-11.
6. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409(6822):860-921.
7. Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. *Science* 2003;300(5617):286-90.
8. Guttmacher AE, Collins FS. Welcome to the genomic era. *N Engl J Med* 2003;349(10):996-8.
9. Manolio TA, Collins FS. The HapMap and genome-wide association studies in diagnosis and therapy. *Annu Rev Med* 2009;60:443-56.
10. Hamburg MA, Collins FS. The path to personalized medicine. *N Engl J Med* 2010;363(4):301-4.
11. Burke W, Psaty BM. Personalized medicine in the era of genomics. *JAMA* 2007;298(14):1682-4.
12. Green ED, Guyer MS, National Human Genome Research I. Charting a course for genomic medicine from base pairs to bedside. *Nature* 2011;470(7333):204-13.
13. Kruglyak L, Nickerson DA. Variation is the spice of life. *Nat Genet* 2001;27(3):234-6.
14. Rosenberg NA, Pritchard JK, Weber JL, et al. Genetic structure of human populations. *Science* 2002;298(5602):2381-5.

15. Pearson TA, Manolio TA. How to interpret a genome-wide association study. *JAMA* 2008;299(11):1335-44.
16. Kidd JM, Cooper GM, Donahue WF, et al. Mapping and sequencing of structural variation from eight human genomes. *Nature* 2008;453(7191):56-64.
17. Attia J, Ioannidis JP, Thakkeinstian A, et al. How to use an article about genetic association: A: Background concepts. *JAMA* 2009;301(1):74-81.
18. Scherer SW, Lee C, Birney E, et al. Challenges and standards in integrating surveys of structural variation. *Nat Genet* 2007;39(7 Suppl):S7-15.
19. Wang DG, Fan JB, Siao CJ, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998;280(5366):1077-82.
20. Chakravarti A. To a future of genetic medicine. *Nature* 2001;409(6822):822-3.
21. Sachidanandam R, Weissman D, Schmidt SC, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001;409(6822):928-33.
22. Reich DE, Gabriel SB, Altshuler D. Quality and completeness of SNP databases. *Nat Genet* 2003;33(4):457-8.
23. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Annu Rev Med* 2010;61:437-55.
24. Lee C, Morton CC. Structural genomic variation and personalized medicine. *N Engl J Med* 2008;358(7):740-1.
25. MacMahon B, Pugh TF. *Epidemiology: principles and methods*. Boston: Little & Brown; 1970.
26. Ellsworth DL, Manolio TA. The emerging importance of genetics in epidemiologic research. I. Basic concepts in human genetics and laboratory technology. *Ann Epidemiol* 1999;9(1):1-16.
27. Burton PR, Tobin MD, Hopper JL. Key concepts in genetic epidemiology. *Lancet* 2005;366(9489):941-51.
28. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet* 2003;33 Suppl:228-37.

29. Badano JL, Katsanis N. Beyond Mendel: an evolving view of human genetic disease transmission. *Nat Rev Genet* 2002;3(10):779-89.
30. Dawn Teare M, Barrett JH. Genetic linkage studies. *Lancet* 2005;366(9490):1036-44.
31. Jimenez-Sanchez G, Childs B, Valle D. Human disease genes. *Nature* 2001;409(6822):853-5.
32. Kerem B, Rommens JM, Buchanan JA, et al. Identification of the cystic fibrosis gene: genetic analysis. *Science* 1989;245(4922):1073-80.
33. Riordan JR, Rommens JM, Kerem B, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 1989;245(4922):1066-73.
34. Altmuller J, Palmer LJ, Fischer G, et al. Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet* 2001;69(5):936-50.
35. Risch NJ. Searching for genetic determinants in the new millennium. *Nature* 2000;405(6788):847-56.
36. Murray CJL, Lopez AD. *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. Cambridge, MA: Harvard School of Public Health, Harvard University Press on behalf of the World Health Organization and The World Bank; 1996.
37. Schork NJ. Genetics of complex disease: approaches, problems, and solutions. *Am J Respir Crit Care Med* 1997;156(4 Pt 2):S103-9.
38. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science* 1994;265(5181):2037-48.
39. Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet* 2001;17(9):502-10.
40. Lander ES. The new genomics: global views of biology. *Science* 1996;274(5287):536-9.
41. Chakravarti A. Population genetics--making sense out of sequence. *Nat Genet* 1999;21(1 Suppl):56-60.
42. Schork NJ, Murray SS, Frazer KA, et al. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev* 2009;19(3):212-9.

43. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 2008;40(6):695-701.
44. Peng B, Kimmel M. Simulations provide support for the common disease-common variant hypothesis. *Genetics* 2007;175(2):763-76.
45. Altshuler D, Hirschhorn JN, Klannemark M, et al. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 2000;26(1):76-80.
46. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 1993;261(5123):921-3.
47. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 2001;69(1):124-37.
48. Gorlov IP, Gorlova OY, Sunyaev SR, et al. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am J Hum Genet* 2008;82(1):100-12.
49. Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant...or not? *Hum Mol Genet* 2002;11(20):2417-23.
50. Campbell H, Manolio T. Commentary: rare alleles, modest genetic effects and the need for collaboration. *Int J Epidemiol* 2007;36(2):445-8.
51. McClellan JM, Susser E, King MC. Schizophrenia: a common disease caused by multiple rare alleles. *Br J Psychiatry* 2007;190:194-9.
52. Cohen JC, Pertsemlidis A, Fahmi S, et al. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc Natl Acad Sci U S A* 2006;103(6):1810-5.
53. Nejentsev S, Walker N, Riches D, et al. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 2009;324(5925):387-9.
54. Galvan A, Ioannidis JP, Dragani TA. Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer. *Trends Genet* 2010;26(3):132-41.

55. Mechanic LE, Chen HS, Amos CI, et al. Next generation analytic tools for large scale genetic epidemiology studies of complex diseases. *Genet Epidemiol* 2012;36(1):22-35.
56. Panagiotou OA, Evangelou E, Ioannidis JP. Genome-wide significant associations for variants with minor allele frequency of 5% or less--an overview: A HuGE review. *Am J Epidemiol* 2010;172(8):869-89.
57. Cordell HJ, Clayton DG. Genetic association studies. *Lancet* 2005;366(9491):1121-31.
58. Lewis CM. Genetic association studies: design, analysis and interpretation. *Brief Bioinform* 2002;3(2):146-53.
59. Newton-Cheh C, Hirschhorn JN. Genetic association studies of complex traits: design and analysis issues. *Mutat Res* 2005;573(1-2):54-69.
60. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996;273(5281):1516-7.
61. Lewis CM, Knight J. Introduction to genetic association studies. *Cold Spring Harb Protoc* 2012;2012(3):297-306.
62. Teare MD. Candidate gene association studies. *Methods Mol Biol* 2011;713:105-17.
63. Rebbeck TR, Spitz M, Wu X. Assessing the function of genetic variants in candidate gene association studies. *Nat Rev Genet* 2004;5(8):589-97.
64. Bao L, Cui Y. Prediction of the phenotypic effects of non-synonymous single nucleotide polymorphisms using structural and evolutionary information. *Bioinformatics* 2005;21(10):2185-90.
65. Tian J, Wu N, Guo X, et al. Predicting the phenotypic effects of non-synonymous single nucleotide polymorphisms based on support vector machines. *BMC Bioinformatics* 2007;8:450.
66. Wacholder S, Chanock S, Garcia-Closas M, et al. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst* 2004;96(6):434-42.
67. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003;361(9357):598-604.

68. Hutchison KE, Stallings M, McGeary J, et al. Population stratification in the candidate gene study: fatal threat or red herring? *Psychol Bull* 2004;130(1):66-79.
69. Pfeiffer T, Bertram L, Ioannidis JP. Quantifying selective reporting and the Proteus phenomenon for multiple datasets with similar bias. *PLoS One* 2011;6(3):e18362.
70. Ioannidis JP, Tarone R, McLaughlin JK. The false-positive to false-negative ratio in epidemiologic studies. *Epidemiology* 2011;22(4):450-6.
71. Ioannidis JP, Trikalinos TA. Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. *J Clin Epidemiol* 2005;58(6):543-9.
72. Ioannidis JP, Ntzani EE, Trikalinos TA, et al. Replication validity of genetic association studies. *Nat Genet* 2001;29(3):306-9.
73. Khoury MJ, Wacholder S. Invited commentary: from genome-wide association studies to gene-environment-wide interaction studies--challenges and opportunities. *Am J Epidemiol* 2009;169(2):227-30; discussion 34-5.
74. Ioannidis JP, Panagiotou OA. Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. *JAMA* 2011;305(21):2200-10.
75. Ragoussis J. Genotyping technologies for genetic research. *Annu Rev Genomics Hum Genet* 2009;10:117-33.
76. International HapMap Consortium. The International HapMap Project. *Nature* 2003;426(6968):789-96.
77. International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;437(7063):1299-320.
78. International HapMap Consortium, Frazer KA, Ballinger DG, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;449(7164):851-61.
79. Slatkin M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* 2008;9(6):477-85.

80. Wall JD, Pritchard JK. Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet* 2003;4(8):587-97.
81. Reich DE, Cargill M, Bolk S, et al. Linkage disequilibrium in the human genome. *Nature* 2001;411(6834):199-204.
82. Sabeti PC, Reich DE, Higgins JM, et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002;419(6909):832-7.
83. Cardon LR, Abecasis GR. Using haplotype blocks to map human complex trait loci. *Trends Genet* 2003;19(3):135-40.
84. Johnson GC, Esposito L, Barratt BJ, et al. Haplotype tagging for the identification of common disease genes. *Nat Genet* 2001;29(2):233-7.
85. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005;6(2):95-108.
86. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296(5576):2225-9.
87. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science* 2008;322(5903):881-8.
88. Hindorff LA, Junkins HA, Mehta JP, et al. A Catalog of Published Genome-Wide Association Studies. National Human Genome Research Institute. (<http://www.genome.gov/26525384>). (Accessed).
89. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America* 2009;106(23):9362-7.
90. Vineis P, Brennan P, Canzian F, et al. Expectations and challenges stemming from genome-wide association studies. *Mutagenesis* 2008;23(6):439-44.
91. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008;9(5):356-69.
92. Ioannidis JP, Thomas G, Daly MJ. Validating, augmenting and refining genome-wide association signals. *Nat Rev Genet* 2009;10(5):318-29.

93. Christensen K, Murray JC. What genome-wide association studies can do for medicine. *N Engl J Med* 2007;356(11):1094-7.
94. Manolio TA. Genomewide association studies and assessment of the risk of disease. *The New England journal of medicine* 2010;363(2):166-76.
95. Ioannidis JP. Calibration of credibility of agnostic genome-wide associations. *Am J Med Genet B Neuropsychiatr Genet* 2008;147B(6):964-72.
96. Witte JS. Genome-wide association studies and beyond. *Annu Rev Public Health* 2010;31:9-20 4 p following
97. Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. *Nat Genet* 2006;38(6):659-62.
98. Zondervan KT, Cardon LR. The complex interplay among factors that influence allelic association. *Nat Rev Genet* 2004;5(2):89-100.
99. Anderson CA, Pettersson FH, Barrett JC, et al. Evaluating the effects of imputation on the power, coverage, and cost efficiency of genome-wide SNP platforms. *Am J Hum Genet* 2008;83(1):112-9.
100. Anderson CA, Pettersson FH, Clarke GM, et al. Data quality control in genetic case-control association studies. *Nature protocols* 2010;5(9):1564-73.
101. Garcia-Closas M, Jacobs K, Kraft P, et al. Analysis of epidemiologic studies of genetic effects and gene-environment interactions. *IARC scientific publications* 2011(163):281-301.
102. Gondro C, Lee SH, Lee HK, et al. Quality control for genome-wide association studies. *Methods Mol Biol* 2013;1019:129-47.
103. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. *The Journal of clinical investigation* 2008;118(5):1590-605.
104. Wacholder S, Yeager M, Liao LM. Invited commentary: more surprises from a gene desert. *American journal of epidemiology* 2012;175(6):488-91.
105. Chung CC, Chanock SJ. Current status of genome-wide association studies in cancer. *Hum Genet* 2011;130(1):59-78.

106. Sur I, Tuupanen S, Whittington T, et al. Lessons from functional analysis of genome-wide association studies. *Cancer Res* 2013;73(14):4180-4.
107. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. *Science* 2005;308(5720):385-9.
108. Klein R, Knudtson MD, Klein BE, et al. Inflammation, complement factor h, and age-related macular degeneration: the Multi-ethnic Study of Atherosclerosis. *Ophthalmology* 2008;115(10):1742-9.
109. Amos CI, Wu X, Broderick P, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet* 2008;40(5):616-22.
110. Goldstein DB. Common genetic variation and human traits. *N Engl J Med* 2009;360(17):1696-8.
111. Ioannidis JP. Prediction of cardiovascular disease outcomes and established cardiovascular risk factors by genome-wide association markers. *Circ Cardiovasc Genet* 2009;2(1):7-15.
112. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009;461(7265):747-53.
113. Boyko AR, Williamson SH, Indap AR, et al. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet* 2008;4(5):e1000083.
114. Lander ES. Initial impact of the sequencing of the human genome. *Nature* 2011;470(7333):187-97.
115. Eichler EE, Flint J, Gibson G, et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 2010;11(6):446-50.
116. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet* 2009;10(6):392-404.
117. Thomas D. Gene--environment-wide association studies: emerging approaches. *Nat Rev Genet* 2010;11(4):259-72.
118. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81(3):559-75.

119. Efron B. Simultaneous inference: When should hypothesis testing problems be combined? *The Annals of Applied Statistics* 2008;2(1):197-223.
120. Ionita-Laza I, McQueen MB, Laird NM, et al. Genomewide weighted hypothesis testing in family-based association studies, with an application to a 100K scan. *Am J Hum Genet* 2007;81(3):607-14.
121. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003;100(16):9440-5.
122. Panagiotou OA, Ioannidis JP, for the Genome-Wide Significance Project. What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *International journal of epidemiology* 2012;41(1):273-86.
123. Lazzeroni LC, Lu Y, Belitskaya-Levy I. P-values in genomics: Apparent precision masks high uncertainty. *Mol Psychiatry* 2014.
124. Martin LJ, Woo JG, Avery CL, et al. Multiple testing in the genomics era: findings from Genetic Analysis Workshop 15, Group 15. *Genet Epidemiol* 2007;31 Suppl 1:S124-31.
125. Song K, Lu Q, Lin X, et al. Genome-wide association studies using an adaptive two-stage analysis for a case-control design. *BMC Proc* 2007;1 Suppl 1:S147.
126. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ* 1995;310(6973):170.
127. Ludbrook J. Multiple comparison procedures updated. *Clin Exp Pharmacol Physiol* 1998;25(12):1032-7.
128. Aschard H, Guedj M, Demenais F. A two-step multiple-marker strategy for genome-wide association studies. *BMC Proc* 2007;1 Suppl 1:S134.
129. Wakefield J. Commentary: Genome-wide significance thresholds via Bayes factors. *International journal of epidemiology* 2012;41(1):286-91.
130. Li SM, Wakefield J, Self S. A transdimensional Bayesian model for pattern recognition in DNA sequences. *Biostatistics* 2008;9(4):668-85.
131. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann Intern Med* 1999;130(12):1005-13.

132. Wakefield J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet Epidemiol* 2009;33(1):79-86.
133. Panagiotou OA, Ioannidis JP. Primary study authors of significant studies are more likely to believe that a strong association exists in a heterogeneous meta-analysis compared with methodologists. *J Clin Epidemiol* 2012;65(7):740-7.
134. Trikalinos TA, Salanti G, Zintzaras E, et al. Meta-analysis methods. *Adv Genet* 2008;60:311-34.
135. Kraft P, Cox DG. Study designs for genome-wide association studies. *Adv Genet* 2008;60:465-504.
136. Skol AD, Scott LJ, Abecasis GR, et al. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006;38(2):209-13.
137. Panagiotou OA, Willer CJ, Hirschhorn JN, et al. The power of meta-analysis in genome-wide association studies. *Annual review of genomics and human genetics* 2013;14:441-65.
138. Pe'er I, Yelensky R, Altshuler D, et al. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* 2008;32(4):381-5.
139. Hoggart CJ, Clark TG, De Iorio M, et al. Genome-wide significance for dense SNP and resequencing data. *Genet Epidemiol* 2008;32(2):179-85.
140. Ingelsson E. Large-scale genome-wide association studies consortia: blessing, burden, or necessity? *Circ Cardiovasc Genet* 2010;3(5):396-8.
141. Austin MA, Hair MS, Fullerton SM. Research guidelines in the era of large-scale collaborations: an analysis of Genome-wide Association Study Consortia. *Am J Epidemiol* 2012;175(9):962-9.
142. Zeggini E, Ioannidis JP. Meta-analysis in genome-wide association studies. *Pharmacogenomics* 2009;10(2):191-201.
143. Estrada K, Styrkarsdottir U, Evangelou E, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat Genet* 2012;44(5):491-501.

144. Evangelou E, Valdes AM, Kerkhof HJ, et al. Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. *Ann Rheum Dis* 2011;70(2):349-55.
145. Kraft P, Zeggini E, Ioannidis JP. Replication in genome-wide association studies. *Stat Sci* 2009;24(4):561-73.
146. Kilpinen H, Barrett JC. How next-generation sequencing is transforming complex disease genetics. *Trends Genet* 2012.
147. Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genet* 2008;24(3):133-41.
148. Shendure J, Ji H. Next-generation DNA sequencing. *Nature biotechnology* 2008;26(10):1135-45.
149. The IcgC Mmml-Seq Project, Richter J, Schlesner M, et al. Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. *Nat Genet* 2012;44(12):1316-20.
150. Gudmundsson J, Sulem P, Gudbjartsson DF, et al. A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer. *Nat Genet* 2012;44(12):1326-9.
151. Holm H, Gudbjartsson DF, Sulem P, et al. A rare variant in MYH6 is associated with high risk of sick sinus syndrome. *Nat Genet* 2011;43(4):316-20.
152. Romeo S, Pennacchio LA, Fu Y, et al. Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nat Genet* 2007;39(4):513-6.
153. Day-Williams AG, Zeggini E. The effect of next-generation sequencing technology on complex trait research. *Eur J Clin Invest* 2011;41(5):561-7.
154. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* 2010;11(6):415-25.
155. Tennessen JA, Bigham AW, O'Connor TD, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 2012;337(6090):64-9.

156. Nelson MR, Wegmann D, Ehm MG, et al. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* 2012;337(6090):100-4.
157. Marth GT, Yu F, Indap AR, et al. The functional spectrum of low-frequency coding variation. *Genome Biol* 2011;12(9):R84.
158. Ji W, Foo JN, O'Roak BJ, et al. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet* 2008;40(5):592-9.
159. Cohen JC, Kiss RS, Pertsemlidis A, et al. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 2004;305(5685):869-72.
160. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* 2008;83(3):311-21.
161. Singleton AB. Exome sequencing: a transformative technology. *Lancet Neurol* 2011;10(10):942-6.
162. Teer JK, Mullikin JC. Exome sequencing: the sweet spot before whole genomes. *Hum Mol Genet* 2010;19(R2):R145-51.
163. Zeggini E. Next-generation association studies for complex traits. *Nat Genet* 2011;43(4):287-8.
164. Zhu X, Feng T, Li Y, et al. Detecting rare variants for complex traits using family and unrelated data. *Genet Epidemiol* 2010;34(2):171-87.
165. Asimit J, Zeggini E. Rare variant association analysis methods for complex traits. *Annu Rev Genet* 2010;44:293-308.
166. Morris AP, Zeggini E, Lindgren CM. Identification of novel putative rheumatoid arthritis susceptibility genes via analysis of rare variants. *BMC Proc* 2009;3 Suppl 7:S131.
167. Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* 2009;106(45):19096-101.
168. Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;461(7261):272-6.
169. Hodges E, Xuan Z, Baliya V, et al. Genome-wide in situ exon capture for selective resequencing. *Nat Genet* 2007;39(12):1522-7.

170. Sanna S, Jackson AU, Nagaraja R, et al. Common variants in the GDF5-UQCC region are associated with variation in human height. *Nat Genet* 2008;40(2):198-203.
171. Stefansson H, Rujescu D, Cichon S, et al. Large recurrent microdeletions associated with schizophrenia. *Nature* 2008;455(7210):232-6.
172. Stefanaki I, Panagiotou OA, Kodela E, et al. Replication and predictive value of SNPs associated with melanoma and pigmentation traits in a Southern European case-control study. *PloS one* 2013;8(2):e55712.
173. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010;467(7319):1061-73.
174. 1000 Genomes Project Consortium, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491(7422):56-65.
175. Elbaz A, Nelson LM, Payami H, et al. Lack of replication of thirteen single-nucleotide polymorphisms implicated in Parkinson's disease: a large-scale international study. *Lancet Neurol* 2006;5(11):917-23.
176. Ioannidis JP. Non-replication and inconsistency in the genome-wide association setting. *Hum Hered* 2007;64(4):203-13.
177. Maraganore DM, de Andrade M, Lesnick TG, et al. High-resolution whole-genome association study of Parkinson disease. *Am J Hum Genet* 2005;77(5):685-93.
178. Lohmueller KE, Pearce CL, Pike M, et al. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003;33(2):177-82.
179. Willer CJ, Bonnycastle LL, Conneely KN, et al. Screening of 134 single nucleotide polymorphisms (SNPs) previously associated with type 2 diabetes replicates association with 12 SNPs in nine genes. *Diabetes* 2007;56(1):256-64.
180. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003;361(9360):865-72.
181. Evangelou E, Ioannidis JP. Meta-analysis methods for genome-wide association studies and beyond. *Nat Rev Genet* 2013;14(6):379-89.

182. Evangelou E, Maraganore DM, Ioannidis JP. Meta-analysis in genome-wide association datasets: strategies and application in Parkinson disease. *PLoS One* 2007;2(2):e196.
183. Skol AD, Scott LJ, Abecasis GR, et al. Optimal designs for two-stage genome-wide association studies. *Genet Epidemiol* 2007;31(7):776-88.
184. Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. *Ann Intern Med* 1997;127(9):820-6.
185. Young NS, Ioannidis JP, Al-Ubaydli O. Why current publication practices may distort science. *PLoS Med* 2008;5(10):e201.
186. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *Bmj* 2003;327(7414):557-60.
187. Ioannidis JP, Patsopoulos NA, Evangelou E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS One* 2007;2(9):e841.
188. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in medicine* 2002;21(11):1539-58.
189. Cantor RM, Lange K, Sinsheimer JS. Prioritizing GWAS results: A review of statistical methods and recommendations for their application. *Am J Hum Genet* 2010;86(1):6-22.
190. Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2004;33(1):30-42.
191. Ebrahim S, Davey Smith G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum Genet* 2008;123(1):15-33.
192. Gogele M, Minelli C, Thakkinstian A, et al. Methods for meta-analyses of genome-wide association studies: critical assessment of empirical evidence. *Am J Epidemiol* 2012;175(8):739-49.
193. Li Y, Willer C, Sanna S, et al. Genotype imputation. *Annu Rev Genomics Hum Genet* 2009;10:387-406.
194. de Bakker PI, Ferreira MA, Jia X, et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet* 2008;17(R2):R122-8.

195. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* 2010;11(7):499-511.
196. Marchini J, Donnelly P, Cardon LR. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat Genet* 2005;37(4):413-7.
197. Park JH, Wacholder S, Gail MH, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet* 2010;42(7):570-5.
198. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet* 2008;9(4):255-66.
199. Glickman ME, van Dyk DA. Basic Bayesian methods. *Methods Mol Biol* 2007;404:319-38.
200. NCI-NHGRI Working Group on Replication in Association Studies, Chanock SJ, Manolio T, et al. Replicating genotype-phenotype associations. *Nature* 2007;447(7145):655-60.
201. Kitsios GD, Tangri N, Castaldi PJ, et al. Laboratory mouse models for the human genome-wide associations. *PLoS One* 2010;5(11):e13782.
202. Ramanan VK, Shen L, Moore JH, et al. Pathway analysis of genomic data: concepts, methods, and prospects for future development. *Trends Genet* 2012;28(7):323-32.
203. Raychaudhuri S, Plenge RM, Rossin EJ, et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet* 2009;5(6):e1000534.
204. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7(3):177-88.
205. Kavvoura FK, Ioannidis JP. Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls. *Hum Genet* 2008;123(1):1-14.
206. Ntzani EE, Khoury MJ, Ioannidis JP. Combining molecular and genetic data from different sources. *IARC scientific publications* 2011(163):323-36.
207. Pereira TV, Patsopoulos NA, Salanti G, et al. Discovery properties of genome-wide association signals from cumulatively combined data sets. *Am J Epidemiol* 2009;170(10):1197-206.

208. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet* 2011;88(5):586-98.
209. Zaitlen N, Eskin E. Imputation aware meta-analysis of genome-wide association studies. *Genet Epidemiol* 2010;34(6):537-42.
210. Ntzani EE, Liberopoulos G, Manolio TA, et al. Consistency of genome-wide associations across major ancestral groups. *Hum Genet* 2012;131(7):1057-71.
211. Park JH, Gail MH, Weinberg CR, et al. Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. *Proc Natl Acad Sci U S A* 2011;108(44):18026-31.
212. Panoutsopoulou K, Tachmazidou I, Zeggini E. In search of low-frequency and rare variants affecting complex traits. *Hum Mol Genet* 2013;22(R1):R16-21.
213. Lee S, Teslovich TM, Boehnke M, et al. General framework for meta-analysis of rare variants in sequencing association studies. *Am J Hum Genet* 2013;93(1):42-53.
214. Hu YJ, Berndt SI, Gustafsson S, et al. Meta-analysis of gene-level associations for rare variants based on single-variant statistics. *Am J Hum Genet* 2013;93(2):236-48.
215. Magi R, Asimit JL, Day-Williams AG, et al. Genome-Wide Association Analysis of Imputed Rare Variants: Application to Seven Common Complex Diseases. *Genet Epidemiol* 2012.
216. Burgio MR, Jr., Ioannidis JP, Kaminski BM, et al. Collaborative Cancer Epidemiology in the 21st Century: the Model of Cancer Consortia. *Cancer Epidemiol Biomarkers Prev* 2013.
217. Lin BK, Clyne M, Walsh M, et al. Tracking the epidemiology of human genes in the literature: the HuGE Published Literature database. *Am J Epidemiol* 2006;164(1):1-4.
218. Yu W, Gwinn M, Clyne M, et al. A navigator for human genome epidemiology. *Nat Genet* 2008;40(2):124-5.
219. Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med* 2000;19(22):3127-31.

220. Gauderman W, Morrison J. QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies, <http://hydra.usc.edu/gxe>. 2006.
221. Ge D, Zhang K, Need AC, et al. WGAViewer: software for genomic annotation of whole genome association studies. *Genome Res* 2008;18(4):640-3.
222. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res* 2002;12(6):996-1006.
223. Johnson AD, Handsaker RE, Pulit SL, et al. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008;24(24):2938-9.
224. Sulem P, Gudbjartsson DF, Stacey SN, et al. Two newly identified genetic determinants of pigmentation in Europeans. *Nat Genet* 2008;40(7):835-7.
225. Kayser M, Liu F, Janssens AC, et al. Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *Am J Hum Genet* 2008;82(2):411-23.
226. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, et al. New sequence variants associated with bone mineral density. *Nat Genet* 2009;41(1):15-7.
227. Timpson NJ, Tobias JH, Richards JB, et al. Common variants in the region around Osterix are associated with bone mineral density and growth in childhood. *Hum Mol Genet* 2009;18(8):1510-7.
228. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, et al. Multiple genetic loci for bone mineral density and fractures. *N Engl J Med* 2008;358(22):2355-65.
229. Duerr RH, Taylor KD, Brant SR, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 2006;314(5804):1461-3.
230. Rioux JD, Xavier RJ, Taylor KD, et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet* 2007;39(5):596-604.

231. The Wellcome Trust Case-Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447(7145):661-78.
232. Barrett JC, Hansoul S, Nicolae DL, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 2008;40(8):955-62.
233. Kugathasan S, Baldassano RN, Bradfield JP, et al. Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease. *Nature genetics* 2008;40(10):1211-5.
234. Raelson JV, Little RD, Ruether A, et al. Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci. *Proc Natl Acad Sci U S A* 2007;104(37):14747-52.
235. Kathiresan S, Melander O, Guiducci C, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* 2008;40(2):189-97.
236. Willer CJ, Sanna S, Jackson AU, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 2008;40(2):161-9.
237. Wallace C, Newhouse SJ, Braund P, et al. Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. *Am J Hum Genet* 2008;82(1):139-49.
238. Aulchenko YS, Ripatti S, Lindqvist I, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nature genetics* 2009;41(1):47-55.
239. Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* 2009;41(1):56-65.
240. Chasman DI, Pare G, Zee RY, et al. Genetic loci associated with plasma concentration of low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglycerides, apolipoprotein A1, and Apolipoprotein B among 6382 white women in genome-wide analysis with replication. *Circ Cardiovasc Genet* 2008;1(1):21-30.

241. Sabatti C, Service SK, Hartikainen AL, et al. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet* 2009;41(1):35-46.
242. Soranzo N, Spector TD, Mangino M, et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat Genet* 2009;41(11):1182-90.
243. Chambers JC, Zhang W, Li Y, et al. Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels. *Nat Genet* 2009;41(11):1170-2.
244. Ganesh SK, Zakai NA, van Rooij FJ, et al. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat Genet* 2009;41(11):1191-8.
245. Gudmundsson J, Sulem P, Gudbjartsson DF, et al. Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat Genet* 2009;41(10):1122-6.
246. Capon F, Bijlmakers MJ, Wolf N, et al. Identification of ZNF313/RNF114 as a novel psoriasis susceptibility gene. *Hum Mol Genet* 2008;17(13):1938-45.
247. Nair RP, Duffin KC, Helms C, et al. Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. *Nat Genet* 2009;41(2):199-204.
248. Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007;316(5829):1331-6.
249. Kooner JS, Chambers JC, Aguilar-Salinas CA, et al. Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nat Genet* 2008;40(2):149-51.
250. Timpson NJ, Lindgren CM, Weedon MN, et al. Adiposity-related heterogeneity in patterns of type 2 diabetes susceptibility observed in genome-wide association data. *Diabetes* 2009;58(2):505-10.
251. Zeggini E, Scott LJ, Saxena R, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 2008;40(5):638-45.

252. Salonen JT, Uimari P, Aalto JM, et al. Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium. *Am J Hum Genet* 2007;81(2):338-45.
253. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007;316(5829):1341-5.
254. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007;445(7130):881-5.
255. Florez JC, Manning AK, Dupuis J, et al. A 100K genome-wide association scan for diabetes and related traits in the Framingham Heart Study: replication and integration with other genome-wide datasets. *Diabetes* 2007;56(12):3063-74.
256. Rung J, Cauchi S, Albrechtsen A, et al. Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat Genet* 2009;41(10):1110-5.
257. Zeggini E, Weedon MN, Lindgren CM, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007;316(5829):1336-41.
258. Meigs JB, Manning AK, Fox CS, et al. Genome-wide association with diabetes-related traits in the Framingham Heart Study. *BMC Med Genet* 2007;8 Suppl 1:S16.
259. Khoury MJ, Little J, Gwinn M, et al. On the synthesis and interpretation of consistent but weak gene-disease associations in the era of genome-wide association studies. *Int J Epidemiol* 2007;36(2):439-45.
260. Benn M, Stene MC, Nordestgaard BG, et al. Common and rare alleles in apolipoprotein B contribute to plasma levels of low-density lipoprotein cholesterol in the general population. *J Clin Endocrinol Metab* 2008;93(3):1038-45.
261. Harper AR, Topol EJ. Pharmacogenomics in clinical practice and drug development. *Nature biotechnology* 2012;30(11):1117-24.
262. Wong GK, Yang Z, Passey DA, et al. A population threshold for functional polymorphisms. *Genome Res* 2003;13(8):1873-9.

263. Ku CS, Loy EY, Pawitan Y, et al. The pursuit of genome-wide association studies: where are we now? *J Hum Genet* 2010;55(4):195-206.
264. Zeggini E, Rayner W, Morris AP, et al. An evaluation of HapMap sample size and tagging SNP performance in large-scale empirical and simulated data sets. *Nat Genet* 2005;37(12):1320-2.
265. Collins FS, Manolio TA. Merging and emerging cohorts: necessary but not sufficient. *Nature* 2007;445(7125):259.
266. Need AC, Goldstein DB. Next generation disparities in human genomics: concerns and remedies. *Trends Genet* 2009;25(11):489-94.
267. Dickson SP, Wang K, Krantz I, et al. Rare variants create synthetic genome-wide associations. *PLoS Biol* 2010;8(1):e1000294.
268. Ioannidis JP. Population-wide generalizability of genome-wide discovered associations. *J Natl Cancer Inst* 2009;101(19):1297-9.
269. Via M, Gignoux C, Burchard EG. The 1000 Genomes Project: new opportunities for research and social challenges. *Genome Med* 2010;2(1):3.
270. Nielsen R. Genomics: In search of rare human variants. *Nature* 2010;467(7319):1050-1.
271. Gunter C. Genomics: A picture worth 1000 Genomes. *Nat Rev Genet* 2010;11(12):814.
272. Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology* 2008;19(5):640-8.
273. Kraft P. Curses--winner's and otherwise--in genetic epidemiology. *Epidemiology* 2008;19(5):649-51; discussion 57-8.
274. Gandini S, Sera F, Cattaruzza MS, et al. Meta-analysis of risk factors for cutaneous melanoma: III. Family history, actinic damage and phenotypic factors. *Eur J Cancer* 2005;41(14):2040-59.
275. Tucker MA, Goldstein AM. Melanoma etiology: where are we? *Oncogene* 2003;22(20):3042-52.
276. Han J, Colditz GA, Hunter DJ. Risk factors for skin cancers: a nested case-control study within the Nurses' Health Study. *Int J Epidemiol* 2006;35(6):1514-21.

277. Bataille V. Genetic epidemiology of melanoma. *Eur J Cancer* 2003;39(10):1341-7.
278. Bastiaens M, ter Huurne J, Gruis N, et al. The melanocortin-1-receptor gene is the major freckle gene. *Hum Mol Genet* 2001;10(16):1701-8.
279. Bastiaens MT, ter Huurne JA, Kielich C, et al. Melanocortin-1 receptor gene variants determine the risk of nonmelanoma skin cancer independently of fair skin and red hair. *Am J Hum Genet* 2001;68(4):884-94.
280. Raimondi S, Sera F, Gandini S, et al. MC1R variants, melanoma and red hair color phenotype: a meta-analysis. *Int J Cancer* 2008;122(12):2753-60.
281. Landi MT, Kanetsky PA, Tsang S, et al. MC1R, ASIP, and DNA repair in sporadic and familial melanoma in a Mediterranean population. *J Natl Cancer Inst* 2005;97(13):998-1007.
282. Kennedy C, ter Huurne J, Berkhout M, et al. Melanocortin 1 receptor (MC1R) gene variants are associated with an increased risk for cutaneous melanoma which is largely independent of skin type and hair color. *J Invest Dermatol* 2001;117(2):294-300.
283. Stratigos AJ, Dimisianos G, Nikolaou V, et al. Melanocortin receptor-1 gene polymorphisms and the risk of cutaneous melanoma in a low-risk southern European population. *J Invest Dermatol* 2006;126(8):1842-9.
284. Naysmith L, Waterston K, Ha T, et al. Quantitative measures of the effect of the melanocortin 1 receptor on human pigmentary status. *J Invest Dermatol* 2004;122(2):423-8.
285. Duffy DL, Box NF, Chen W, et al. Interactive effects of MC1R and OCA2 on melanoma risk phenotypes. *Hum Mol Genet* 2004;13(4):447-61.
286. Chatzinasiou F, Lill CM, Kypreou K, et al. Comprehensive field synopsis and systematic meta-analyses of genetic association studies in cutaneous melanoma. *J Natl Cancer Inst* 2011;103(16):1227-35.
287. Lasithiotakis K, Kruger-Krasagakis S, Manousaki A, et al. The incidence of cutaneous melanoma on Crete, Greece. *International journal of dermatology* 2006;45(4):397-401.

288. Nikolaou V, Kang X, Stratigos A, et al. Comprehensive mutational analysis of CDKN2A and CDK4 in Greek patients with cutaneous melanoma. *The British journal of dermatology* 2011;165(6):1219-22.
289. Lasithiotakis K, Kruger-Krasagakis S, Ioannidou D, et al. Epidemiological differences for cutaneous melanoma in a relatively dark-skinned Caucasian population with chronic sun exposure. *Eur J Cancer* 2004;40(16):2502-7.
290. Stratigos AJ, Yang G, Dimisianos R, et al. Germline CDKN2A mutations among Greek patients with early-onset and multiple primary cutaneous melanoma. *J Invest Dermatol* 2006;126(2):399-401.
291. Wacholder S, McLaughlin JK, Silverman DT, et al. Selection of controls in case-control studies. I. Principles. *Am J Epidemiol* 1992;135(9):1019-28.
292. Gabriel S, Ziaugra L, Tabbaa D. SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Current protocols in human genetics / editorial board, Jonathan L Haines [et al]* 2009;Chapter 2:Unit 2 12.
293. Chatterjee N, Wheeler B, Sampson J, et al. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nature genetics* 2013;45(4):400-5, 5e1-3.
294. Wu J, Pfeiffer RM, Gail MH. Strategies for developing prediction models from genome-wide association studies. *Genet Epidemiol* 2013;37(8):768-77.
295. Wray NR, Yang J, Goddard ME, et al. The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet* 2010;6(2):e1000864.
296. Wentzensen N, Wacholder S. From differences in means between cases and controls to risk stratification: a business plan for biomarker development. *Cancer discovery* 2013;3(2):148-57.
297. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clinical chemistry* 2008;54(1):17-23.
298. Alonzo TA, Pepe MS. Development and evaluation of classifiers. *Methods Mol Biol* 2007;404:89-116.

299. Barrett JH, Iles MM, Harland M, et al. Genome-wide association study identifies three new melanoma susceptibility loci. *Nat Genet* 2011;43(11):1108-13.
300. Macgregor S, Montgomery GW, Liu JZ, et al. Genome-wide association study identifies a new melanoma susceptibility locus at 1q21.3. *Nat Genet* 2011;43(11):1114-8.
301. Amos CI, Wang LE, Lee JE, et al. Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. *Hum Mol Genet* 2011;20(24):5012-23.
302. Law MH, Macgregor S, Hayward NK. Melanoma genetics: recent findings take us beyond well-traveled pathways. *J Invest Dermatol* 2012;132(7):1763-74.
303. Shi J, Yang XR, Ballew B, et al. Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma. *Nat Genet* 2014.