

**Γεφυρώνοντας μεγάλα ομικά, γενετικά και ιατρικά
δεδομένα για την ευρεία εφαρμογή της Ιατρικής Ακριβείας
στην Ελλάδα**

ΠΑΡΑΔΟΤΕΟ Π7.1

**«Τεχνική αναφορά με τις στατιστικές τεχνικές ανάλυσης
δεδομένων από GWAS»**

Φορέας	Πανεπιστήμιο Θεσσαλίας
Τύπος Παραδοτέου	Έκθεση
Ημερομηνία Υποβολής Παραδοτέου	31 Δεκεμβρίου 2025
Ενότητα Εργασίας	Ενότητα Εργασίας 7: Μεθοδολογίες Ανάλυσης Δεδομένων από GWAS

1. Εισαγωγή	4
2. METACP	6
3. PRED-LD: Ένα αποδοτικό εργαλείο αναπλήρωσης μη-γονοτυπημένων πολυμορφισμών για συνοπτικά δεδομένα μελετών γενετικής συσχέτισης	9
4. fastMETA: Ένα γρήγορο και αποδοτικό εργαλείο για πολυμεταβλητή μετα-ανάλυση μελετών γενετικής συσχέτισης	12
5. PYRAMA: Ένα εργαλείο ανοικτού κώδικα για μετα-αναλύσεις μελετών γενετικής συσχέτισης	15
6. Ανακατασκευή συνοπτικών γονοτύπων	18
7. Βιβλιογραφία	24



Πίνακας Εικόνων

Εικόνα 1: Σύγκριση των στατιστικών μεθόδων που υλοποιούνται στο πακέτο metacp για τον συνδυασμό εξαρτημένων τιμών p (CCT, MinP, MCM, CMC) έναντι των μεθόδων που υλοποιούνται στο GVAR και χρησιμοποιούν αριθμητική ολοκλήρωση. Χρησιμοποιήθηκαν οι μέθοδοι MAX και MIN2 του GVAR, οι οποίες εφαρμόστηκαν σε 67 SNPs από μια μελέτη GWAS που διενεργήθηκε για τον εντοπισμό γενετικών θέσεων ευπάθειας για τον διαβήτη τύπου 2. Χρησιμοποιείται ο $-\log(10)$ των τιμών p που προέκυψαν από (Α) τη μέθοδο MIN2 του πακέτου GVAR και (Β) τη μέθοδο MAX του πακέτου GVAR. Για κάθε σύγκριση μεταξύ των μεθόδων, παρατίθενται οι τιμές R^2 οι οποίες κυμαίνονται από 0,9818 έως 1,000, αποδεικνύοντας τη συνεπή συμφωνία μεταξύ των διαφορετικών προσεγγίσεων.

Εικόνα 2: Radar Plot που συγκρίνει τη συνολική απόδοση του imputation του PRED-LD με τα αντίστοιχα εργαλεία.

Εικόνα 3: Στιγμιότυπο από την διαδικτυακή υποδομή του PRED-LD.

Εικόνα 4: Σύγκριση των συνολικών εκτιμήσεων επίδρασης (A,B,D,E) και των $-\log_{10}(p\text{-values})$ από τα global test (C,F) που προέκυψαν από το fastMETA, έναντι εκείνων που αναφέρονται στα αποτελέσματα της διμεταβλητής μετα-ανάλυσης από τους Medina-Gomez et al.

Εικόνα 5: Στιγμιότυπο από την διαδικτυακή υποδομή του fastMETA.

Εικόνα 6: Στιγμιότυπο από την διαδικτυακή υποδομή του PYRAMA.

Εικόνα 7: Σύγκριση των 3 μεθόδων ανακατασκευής με το δείκτη $\Delta_{\log p}$ αναφορικά με τις τιμές $R_{\log p \text{value}}^2$ και το ποσοστό SNPs που κρατήθηκαν για κάθε κατώφλι.

Εικόνα 8: Σύγκριση των 3 μεθόδων ανακατασκευής με το δείκτη $\Delta_{\log p}$ αναφορικά με τις τιμές R_z^2 και το ποσοστό SNPs που κρατήθηκαν για κάθε κατώφλι.

Εικόνα 9: Συσχέτιση των παρατηρούμενων $p\text{-value}$ και των ανακατασκευασμένων με τη μέθοδο Direct και το δείκτη $\Delta_{\log p}$ παρουσιάζοντας τα SNPs που κρατήθηκαν και τα SNPs που διαγράφηκαν για το κατώφλι 20%.

Εικόνα 10: Συσχέτιση των παρατηρούμενων $p\text{-value}$ και των ανακατασκευασμένων με τη μέθοδο Di Pietrantonij και το δείκτη $\Delta_{\log p}$ παρουσιάζοντας τα SNPs που κρατήθηκαν και τα SNPs που διαγράφηκαν για το κατώφλι 20%.

Εικόνα 11: Συσχέτιση των παρατηρούμενων $p\text{-value}$ και των ανακατασκευασμένων με τη μέθοδο Yang και το δείκτη $\Delta_{\log p}$ παρουσιάζοντας τα SNPs που κρατήθηκαν και τα SNPs που διαγράφηκαν για το κατώφλι 20%.

1. Εισαγωγή

Στην ΕΕ αυτή (ΕΕ7), σύμφωνα και με το τεχνικό παράρτημα επικεντρωθήκαμε σε μεθοδολογίες ανάλυσης δεδομένων από GWAS. Για τον σκοπό αυτό, σε αυτό το παραδοτέο (Π7.1) κινηθήκαμε προς την κατεύθυνση της ανάπτυξης νέας μεθοδολογίας με την οποία θα ενοποιούνται παρόμοιας φύσης δεδομένα από διαφορετικές πηγές, όπως για παράδειγμα, μετα-ανάλυση, πολυδιάστατη μετα-ανάλυση με πολλαπλές εκβάσεις, robust methods κ.ο.κ., αλλά και επιχειρήθηκε η ενοποίηση δεδομένων διαφορετικής προέλευσης, όπως για παράδειγμα, δεδομένα διαφορετικών ασθενειών με σκοπό την εύρεση κοινών παραγόντων που τις επηρεάζουν. Τα συνοπτικά στατιστικά δεδομένα (summary statistics) είναι πολύ σημαντικά γιατί επιτρέπουν αμεσο διαμοιρασμό χωρίς τους περιορισμούς της χρήσης ατομικών δεδομένων και επιτρέπουν, υπο προϋποθέσεις, την ανάπτυξη μεθόδων αντίστοιχων με αυτές που χρησιμοποιούν ατομικά δεδομένα. Σε πολλές περιπτώσεις, οι μέθοδοι αυτές και τα πορίσματα που δίνουν, μπορούν να χρησιμοποιηθούν και στην προσωποποιημένη ιατρική, όπως στην περίπτωση των PRS την οποία θα αναπτύξουμε για τον ιατρικό φάκελο. Στο Π7.1 αναπτύξαμε και περιγράφονται παρακάτω μια σειρά νέων μεθόδων για ανάλυση και μετα-ανάλυση δεδομένων από GWAS. Στην αρχή αναπτύχθηκε το **METACP** (Ενότητα 2) το οποίο είναι ένα ευέλικτο πρόγραμμα με το οποίο μπορεί να γίνει υπολογισμός συνολικής στατιστικής σημαντικότητας (p-value) συνδυάζοντας διαφορετικά p-value, με χρήση μιας πλειάδας διαφορετικών μεθόδων (κατάλληλες τόσο για συσχετισμένες παρατηρήσεις, όσο και για ανεξάρτητες). Έπειτα, αναπτύχθηκε το **PRED-LD** (Ενότητα 3), το οποίο αποτελεί μια απλή, αποδοτική και πολύ γρήγορη μέθοδο αναπλήρωσης μη-γονοτυπημένων πολυμορφισμών σε συνοπτικά στατιστικά (summary statistics imputation) με προυπολογισμένα reference panels για την ανισορροπία σύνδεσης, διαθέσιμη ως open-source εργαλείο PRED-LD και ως διαδικτυακό εργαλείο. Έπειτα, στην Ενότητα 4 αναπτύχθηκε το **PYRAMA** (Python tool for Robust Analysis and Meta-Analysis), ένα πακέτο ανοικτού κώδικα για την ανάλυση και τη μετα-ανάλυση GWAS, το οποίο προσφέρει δυνατότητες που δεν υπάρχουν σε αντίστοιχα εργαλεία. Το PYRAMA είναι δημόσια διαθέσιμο ως εργαλείο γραμμής εντολών αλλά και ως διαδικτυακό εργαλείο, υποστηρίζοντας μοντέλα σταθερών και τυχαίων επιδράσεων, robust tests, Μπεϋζιανή μετα-ανάλυση, imputation συνοπτικών στατιστικών με χρήση προ-υπολογισμένων panels ανισορροπίας σύνδεσης (LD reference panels), καθώς και ανάλυση εμπλουτισμού. Το **fastMETA** (Ενότητα 5), αποτελεί με τη σειρά του, ένα νέο και υπολογιστικά αποδοτικό λογισμικό για πολυμεταβλητή μετα-ανάλυση συνοπτικών στατιστικών GWAS. Το fastMETA υλοποιεί μια παραλλαγή της marginal method of moments (MMoM), αποφεύγοντας το υπολογιστικό κόστος των ιεραρχικών πολυμεταβλητών μοντέλων, διατηρώντας παράλληλα τη στατιστική αυστηρότητα. Τέλος, υλοποιήσαμε μια **μέθοδο ανακατασκευής γονοτύπων** (Ενότητα 6) από τα συνήθη συνοπτικά στατιστικά μιας μελέτης, κάτι που είναι ιδιαίτερα χρήσιμο σε περαιτέρω αναλύσεις δεδομένων όπως ο υπολογισμός PRS αλλά και άλλες εφαρμογές ιατρικής ακριβείας.

2. METACP

Η τεχνική του συνδυασμού των τιμών p (p -values) αποτελεί μια θεμελιώδη προσέγγιση στη σύγχρονη επιστημονική έρευνα και στις μεθόδους μετα-ανάλυσης, επιτρέποντας τη σύνθεση αποτελεσμάτων από πολλαπλές στατιστικές δοκιμές σε ένα ενιαίο και ισχυρότερο συμπέρασμα. Η ανάγκη για τέτοιες μεθοδολογίες είναι ιδιαίτερα έντονη σε πεδία που διαχειρίζονται μεγάλο όγκο δεδομένων, όπως η ψυχολογία, οι κοινωνικές επιστήμες και, κατεξοχήν, η βιοϊατρική έρευνα και η γενετική. Παρά την κρισιμότητα αυτών των τεχνικών, τα υπάρχοντα στατιστικά πακέτα προσφέρουν συχνά μόνο ένα περιορισμένο υποσύνολο των διαθέσιμων μεθόδων, περιορίζοντας την ευελιξία των ερευνητών στην ανάλυση σύνθετων δεδομένων.

Για την κάλυψη αυτού του κενού, η ερευνητική ομάδα προχώρησε στην ανάπτυξη του metacp, ενός ευέλικτου λογισμικού ανοιχτού κώδικα που υλοποιεί μια εξαιρετικά ευρεία γκάμα στατιστικών μεθόδων για τον συνδυασμό τόσο ανεξάρτητων όσο και εξαρτημένων τιμών p . Το λογισμικό σχεδιάστηκε για να ανταποκρίνεται σε προηγμένες αναλυτικές ανάγκες, όπως οι μελέτες συσχέτισης σε ολόκληρο το γονιδίωμα (GWAS), όπου απαιτείται ο συνδυασμός διαφορετικών SNPs εντός ενός γονιδίου (gene-based testing), καθώς και σε πολυ-ωμικές αναλύσεις (multi-omics). Στις τελευταίες, το metacp επιτρέπει την αποτελεσματική ενσωμάτωση ετερογενών δεδομένων από μελέτες TWAS, RNA-seq, colocalization ή Μεντελιανής Τυχαιοποίησης, ενισχύοντας τη δυνατότητα εξαγωγής έγκυρων βιολογικών συμπερασμάτων.

Η καινοτομία του metacp έγκειται στο γεγονός ότι ενσωματώνει τη μεγαλύτερη συλλογή αναλυτικών προσεγγίσεων παγκοσμίως, συμπεριλαμβανομένων κλασικών μεθόδων όπως των Fisher και Stouffer, αλλά και σύγχρονων τεχνικών για εξαρτημένες δοκιμές, όπως το Cauchy Combination Test (CCT) και η Empirical Brown's Method (EBM). Επιπλέον, το εργαλείο προσφέρει τη δυνατότητα προσαρμογής για τον αποτελεσματικό αριθμό ελέγχων, εξασφαλίζοντας υψηλή ακρίβεια στις αναλύσεις πολλαπλών χαρακτηριστικών.

Το λογισμικό απαιτεί ελάχιστα δεδομένα εισόδου (ένα αρχείο κειμένου με n γραμμές και k στήλες τιμών p ή z -scores) και διακρίνεται για την ταχύτητα εκτέλεσής του. Είναι διαθέσιμο τόσο στη γλώσσα προγραμματισμού **Python** όσο και ως εντολή στο περιβάλλον **STATA**, γεγονός που το καθιστά προσβάσιμο σε ένα ευρύ φάσμα επιστημόνων. Η πρόσβαση στο εργαλείο και την απαραίτητη τεκμηρίωση παρέχεται μέσω της διεύθυνσης: www.github.com/pbagos/metacp.

Η αξιοπιστία του metacp αποδείχθηκε μέσω της επανάλυσης δεδομένων από 67 SNPs μιας μελέτης GWAS για τον διαβήτη τύπου 2. Τα αποτελέσματα συγκρίθηκαν με το λογισμικό GWAR (στατιστικές MAX και MIN2), δείχνοντας εντυπωσιακή συμφωνία με συντελεστές προσδιορισμού R^2 που κυμαίνονται από 0,98 έως 1,00. Το κύριο πλεονέκτημα του metacp έναντι άλλων εργαλείων (όπως το GWAR) είναι ότι υπολογίζει τους συνδυασμένους δείκτες

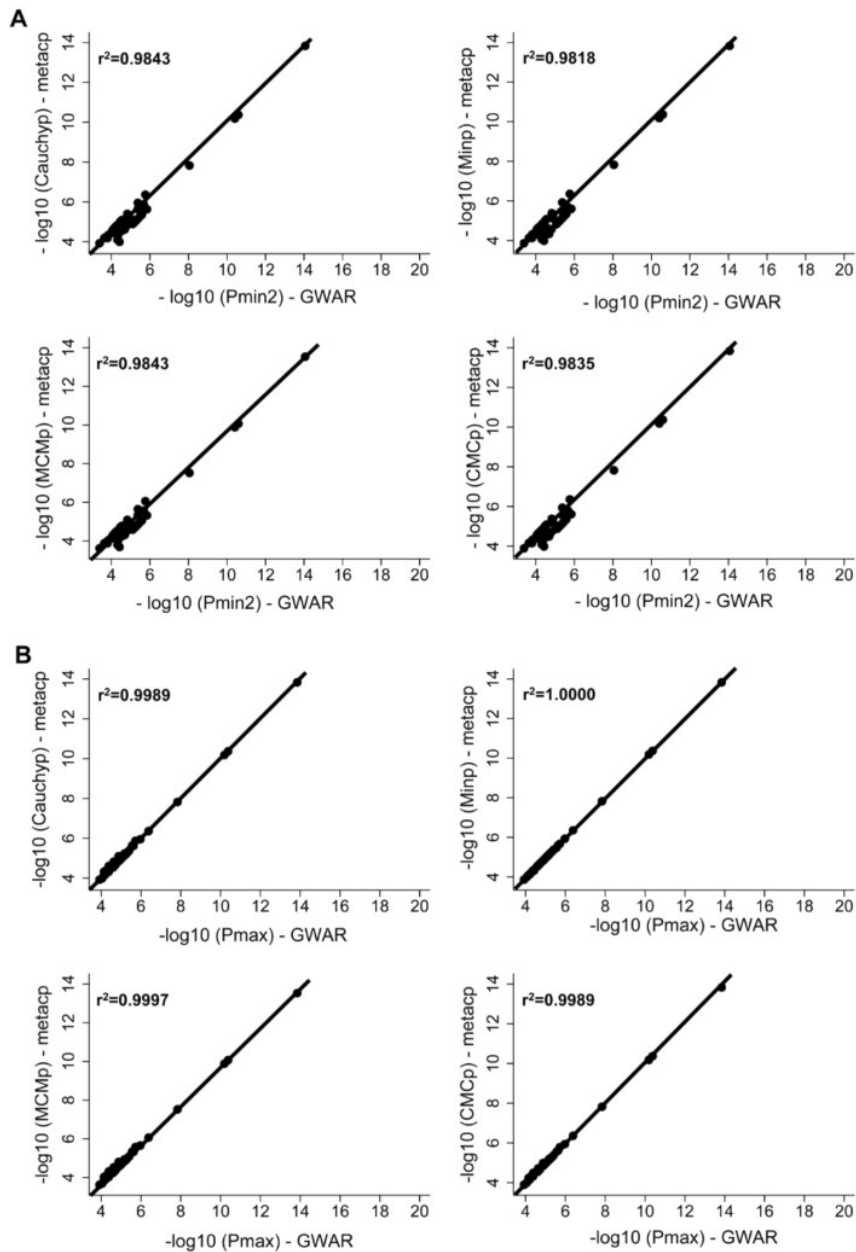
χωρίς να βασίζεται σε χρονοβόρες αριθμητικές ολοκληρώσεις, καθιστώντας το πολύ ταχύτερο (Εικόνα 1). Επιπλέον, σε σύγκριση με άλλα πακέτα όπως το poolr ή το metar, το metacr προσφέρει τη μεγαλύτερη συλλογή αναλυτικών μεθόδων.

Τα ευρήματα και το λογισμικό της παρούσας έρευνας έχουν παρουσιαστεί σε διεθνή και εθνικά επιστημονικά συνέδρια, όπως:

- Στο **45ο Ετήσιο Συνέδριο της Διεθνούς Εταιρείας Κλινικής Βιοστατιστικής (ISCB)**, που διεξήχθη στη Θεσσαλονίκη τον Ιούλιο του 2024-Nikolitsa EK, Kontou PI, Bagos PG. A versatile software package for combining dependent or independent p-values. 45th Annual Conference of the International Society for Clinical Biostatistics; 2024/07/21; Thessaloniki Concert Hall, Greece2024. p. P-A09-16.
- Στο **37ο Πανελλήνιο και 3ο Διεθνές Συνέδριο Στατιστικής** στη Λάρισα (Απρίλιος 2025) - Nikolitsa, E., Kontou, P., & Bagos, P. (2025, April 24–27). *Metacr: A versatile software package for combining dependent or independent p-values*. 37th Panhellenic & 3rd International Statistics Conference, Larissa, Greece. p. 53.

Επιπλέον, η εργασία έχει δημοσιευθεί στο έγκριτο διεθνές επιστημονικό περιοδικό **BMC Bioinformatics** (2025), με τίτλο «*metacr: a versatile software package for combining dependent or independent p-values*».

Συνολικά, το metacr αποτελεί έναν ολοκληρωμένο πόρο που ανταποκρίνεται στις σύγχρονες προκλήσεις της βιοστατιστικής, προσφέροντας ταχύτητα, ευκολία στη χρήση και απaráμιλλη μεθοδολογική πληρότητα.



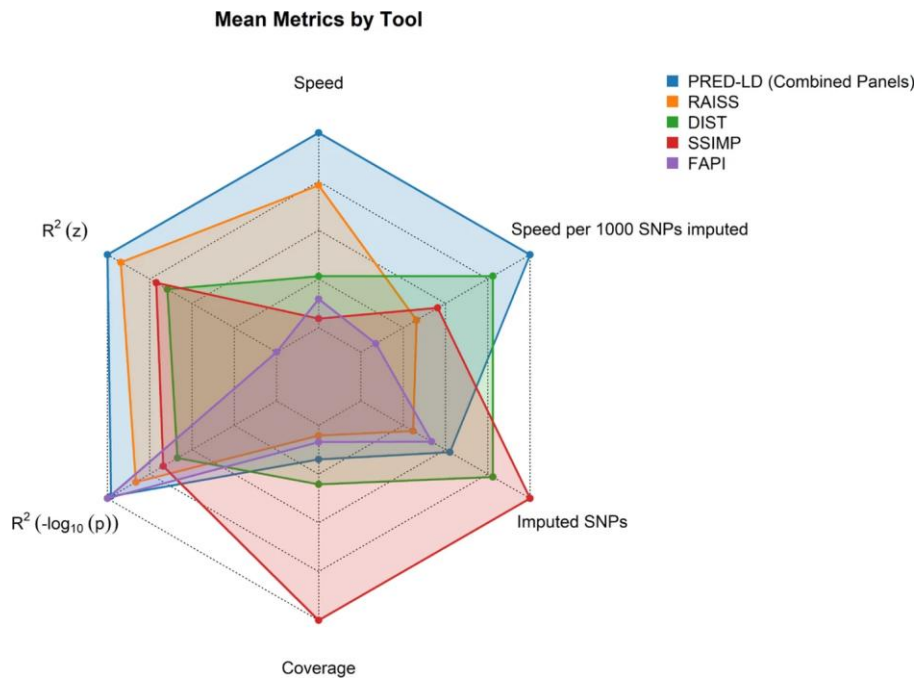
Εικόνα 1. Σύγκριση των στατιστικών μεθόδων που υλοποιούνται στο πακέτο metacp για τον συνδυασμό εξαρτημένων τιμών p (CCT, MinP, MCM, CMC) έναντι των μεθόδων που υλοποιούνται στο GWAR και χρησιμοποιούν αριθμητική ολοκλήρωση. Χρησιμοποιήθηκαν οι μέθοδοι MAX και MIN2 του GWAR, οι οποίες εφαρμόστηκαν σε 67 SNPs από μια μελέτη GWAS που διενεργήθηκε για τον εντοπισμό γενετικών θέσεων ευπάθειας για τον διαβήτη τύπου 2. Χρησιμοποιείται ο $-\log(10)$ των τιμών p που προέκυψαν από (A) τη μέθοδο MIN2 του πακέτου GWAR και (B) τη μέθοδο MAX του πακέτου GWAR. Για κάθε σύγκριση μεταξύ των μεθόδων, παρατίθενται οι τιμές R^2 οι οποίες κυμαίνονται από 0,9818 έως 1,000, αποδεικνύοντας τη συνεπή συμφωνία μεταξύ των διαφορετικών προσεγγίσεων.

3. PRED-LD: Ένα αποδοτικό εργαλείο αναπλήρωσης μη-γονοτυπημένων πολυμορφισμών για συνοπτικά δεδομένα μελετών γενετικής συσχέτισης

Οι μελέτες γενετικής συσχέτισης (Genome-Wide Association Studies - GWAS) έχουν βοηθήσει πολύ στον εντοπισμό γενετικών παραλλαγών που σχετίζονται με ασθένειες, αλλά συνήθως καλύπτουν μόνο μέρος των SNPs και συχνά διαφέρουν ως προς τους γονοτυπημένους πολυμορφισμούς (SNPs) ανάλογα με την πλατφόρμα. Για να αυξηθεί η κάλυψη και η ισχύς, προτείνεται η αναπλήρωση μη-γονοτυπημένων πολυμορφισμών (imputation) σε ατομικά δεδομένα. Όταν δεν είναι εφικτή η αναπλήρωση γονοτύπων λόγω των περιορισμών που συνοδεύουν τον διαμοιρασμό των ατομικών δεδομένων, υπάρχει η εναλλακτική του imputation σε συνοπτικά στατιστικά (summary statistics imputation), με εργαλεία όπως DIST (Lee et al., 2013), ImpG (Pasaniuc et al., 2014), FAPI (Kwan et al., 2016), SSIMP (Rüeger et al., 2018), RAISS (Julienne et al., 2019), GAUSS (Lee & Bacanu, 2024), αξιοποιώντας reference panels όπως είναι το 1000 Genomes Project (Siva, 2008). Για μεικτούς (admixed) πληθυσμούς υπάρχουν μέθοδοι όπως το DISTMIX (Lee et al., 2015).

Σε αυτή την ενότητα, παρουσιάζουμε μια πιο απλή, αποδοτική και πολύ γρήγορη μέθοδο αναπλήρωσης μη-γονοτυπημένων πολυμορφισμών σε συνοπτικά στατιστικά (summary statistics imputation) με προυπολογισμένα reference panels για την ανισορροπία σύνδεσης, διαθέσιμη ως open-source εργαλείο PRED-LD και ως διαδικτυακό εργαλείο.

Η μέθοδος του PRED-LD κάνει imputation συνοπτικών στατιστικών αξιοποιώντας την προ-υπολογισμένη πληροφορία ανισορροπίας σύνδεσης από τρία LD reference panels, το HarMap (Gibbs et al., 2003), PhenoScanner (Staley et al., 2016) και το TOP-LD (Huang et al., 2022). Λαμβάνοντας αυτή την πληροφορία και τις συχνότητες των αλληλομόρφων από αυτά τα reference panels, είτε μεμονωμένα, είτε από τον συνδυασμό τους, είναι εύκολο να υπολογιστεί το Odds Ratio και το τυπικό σφάλμα (Standard Error) για έναν μη-γονοτυπημένο πολυμορφισμό, από έναν γνωστό τύπο των Zondervan και Cardon (Zondervan & Cardon, 2004) και μέσω της Delta Method (Oehlert, 1992), αντίστοιχα. Στην συνέχεια, η προτεινόμενη μέθοδος του PRED-LD συγκρίθηκε σε 8 ελεύθερες GWAS με συνοπτικά στατιστικά από διαφορετικούς πληθυσμούς από την GWAS Atlas (Watanabe et al., 2019) με τα DIST, SSIMP, FAPI και RAISS. Το PRED-LD είχε την πιο γρήγορη εκτέλεση και ταυτόχρονα την πιο αποδοτική σε επίπεδο imputation από τα υπόλοιπα εργαλεία (Εικόνα 2).



Εικόνα 2: Radar Plot που συγκρίνει τη συνολική απόδοση του imputation συνοπτικών στατιστικών σε όλα τα σύνολα δεδομένων GWAS. Το PRED-LD παρουσιάζει τη μεγαλύτερη ακρίβεια και τον υψηλότερο λόγο κάλυψης και είναι ταχύτερο σε σύγκριση με τα άλλα εργαλεία. Μόνο τα DIST και SSIMP το ξεπερνούν ως προς την κάλυψη του imputation και τον αριθμό των SNPs που αναπληρώνονται, λόγω ότι διαθέτουν αρκετά μεγαλύτερα reference panels.

Συνοψίζοντας, το PRED-LD αποτελεί μια βέλτιστη επιλογή για εργασίες imputation μεγάλης κλίμακας σε μελέτες γενετικής συσχέτισης, όπου τόσο η υπολογιστική αποδοτικότητα όσο και η ακρίβεια του imputation έχουν σημαντικό ρόλο. Η διαδικτυακή έκδοση του PRED-LD μπορεί να βοηθήσει τους χρήστες να αντλούν πληροφορία ανισορροπίας σύνδεσης από διάφορες πηγές και να εκτελούν εύκολα imputation σε συνοπτικά στατιστικά, χωρίς να χρειάζεται να κατεβάζουν reference panels για πολλούς πληθυσμούς και χρωμοσώματα. Το PRED-LD θα ενημερώνεται συνεχώς, για παράδειγμα με την προσθήκη νέων reference panels ή με βελτιστοποιήσεις στην ταχύτητα (παραλληλοποίηση κ.λπ.). Ειδικότερα, σχεδιάζουμε να ενσωματώσουμε το PRED-LD σε διάφορα εργαλεία που θα διευκολύνουν, για παράδειγμα, τη μετα-ανάλυση με μη επικαλυπτόμενα σύνολα παραλλαγών, την ανάλυση πολλαπλών χαρακτηριστικών (multiple traits), ή το στατιστικό fine-mapping αιτιωδών παραλλαγών σε GWAS. Η έκδοση γραμμής εντολών διατίθεται ελεύθερα στο GitHub (<https://github.com/rbagos/PRED-LD>) και η διαδικτυακή υποδομή είναι διαθέσιμη στην ακόλουθη διεύθυνση: compgen.dib.uth.gr/PRED-LD/ (Εικόνα 3). Η εργασία αυτή δημοσιεύτηκε στο περιοδικό **BMC Bioinformatics** τον Μάρτιο του 2025 (Manios, Michailidi, et al., 2025).

Εικόνα 3: Στιγμιότυπο από την διαδικτυακή υποδομή του PRED-LD.

4. fastMETA: Ένα γρήγορο και αποδοτικό εργαλείο για πολυμεταβλητή μετα-ανάλυση μελετών γενετικής συσχέτισης

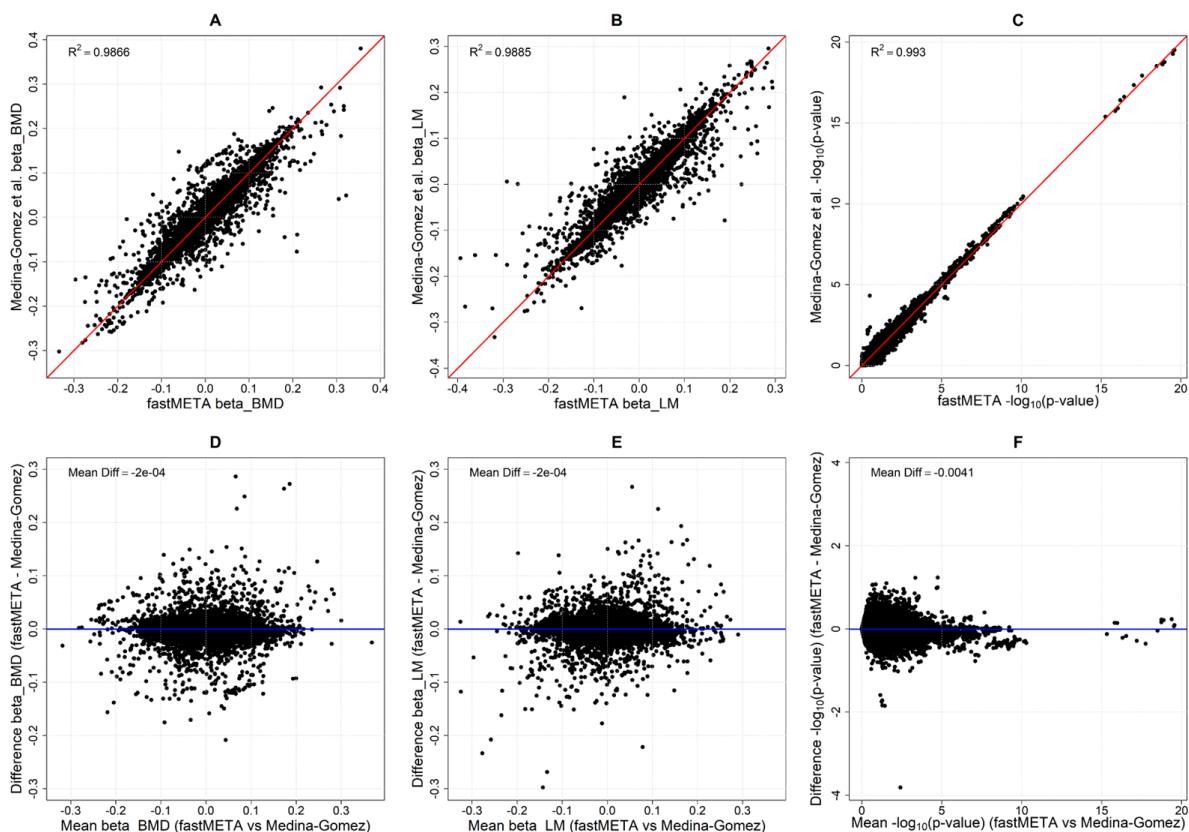
Οι μελέτες γενετικής συσχέτισης έχουν μετασηματίσει τη γενετική του ανθρώπου, εντοπίζοντας χιλιάδες γονιδιακούς τόπους που σχετίζονται με σύνθετα χαρακτηριστικά και ασθένειες (Ziegler et al., 2008, Uffelmann et al., 2021). Ωστόσο, οι μεμονωμένες GWAS συχνά έχουν περιορισμένη στατιστική ισχύ, ενώ οι παραδοσιακές μέθοδοι μετα-ανάλυσης, αν και χρησιμοποιούνται ευρέως σε εργαλεία όπως τα PLINK (Purcell et al., 2007), METAL (Willer et al., 2010) και GWAMA (Mägi and Morris, 2010), συνήθως αναλύουν ένα χαρακτηριστικό κάθε φορά. Αυτή η μονομεταβλητή προσέγγιση ενέχει τον κίνδυνο να παραληφθούν φαινόμενα πλειοτροπίας και οι συσχετίσεις μεταξύ χαρακτηριστικών που υποκρύπτουν πολύπλοκες γενετικές αρχιτεκτονικές.

Σε αυτήν την ενότητα, παρουσιάζουμε το fastMETA, ένα νέο και υπολογιστικά αποδοτικό λογισμικό για πολυμεταβλητή μετα-ανάλυση συνοπτικών στατιστικών GWAS. Το fastMETA υλοποιεί μια παραλλαγή της marginal method of moments (MMoM) (Chen et al., 2016), αποφεύγοντας το υπολογιστικό κόστος των ιεραρχικών πολυμεταβλητών μοντέλων, διατηρώντας παράλληλα τη στατιστική αυστηρότητα.

Το fastMETA υλοποιεί τρεις διακριτές στρατηγικές εκτίμησης: (i) άμεση αναπαραγωγή της κλασικής προσέγγισης MMoM, (ii) προσέγγιση βασισμένη στον συντελεστή συσχέτισης

Pearson και (iii) μια νέα μέθοδο που αξιοποιεί τις συσχετίσεις μεταξύ των υπό μελέτη χαρακτηριστικών, όπως αυτές εκτιμώνται από τα σύνολα των πολυμορφισμών, με σκοπό την εκτίμηση ενός σταθερού πίνακα συσχέτισης μεταξύ χαρακτηριστικών. Η τελευταία στρατηγική είναι ιδιαίτερα κατάλληλη για GWAS μεγάλης κλίμακας, όπου οι συσχετίσεις μεταξύ των μελετών σπάνια είναι διαθέσιμες.

Στην συνέχεια έγινε σύγκριση της προτεινόμενης μεθόδου με υπάρχοντα πακέτα πολυμεταβλητής μετα-ανάλυσης, συγκεκριμένα με τα `mvmeta` σε R και `Stata` και `xmeta` σε R, χρησιμοποιώντας τόσο πραγματικά όσο και συνθετικά σύνολα δεδομένων. Σε όλα τα σενάρια, το `fastMETA` πέτυχε σταθερά χρόνους εκτέλεσης 15-20 φορές ταχύτερους, διατηρώντας υψηλό βαθμό συμφωνίας με τις καθιερωμένες μεθόδους. Συγκεκριμένα, το `fastMETA` εφαρμόστηκε σε (α) μια διμεταβλητή μετα-ανάλυση παιδιατρικών μυοσκελετικών χαρακτηριστικών, που αναπαρήγαγε πλειοτροπικά αποτελέσματα στον τόπο `TOM1L2/SREBF1` (Medina-Gomez, 2017), (β) και σε μια πολυμεταβλητή μετα-ανάλυση δεδομένων γονιδιακής έκφρασης για τη φλεγμονώδη νόσο του εντέρου (Venou et al., 2020), με σχεδόν ταυτόσημα αποτελέσματα σε σχέση με δημοσιευμένα ευρήματα και (γ) σε ένα μεγάλο σύνολο μετα-αναλύσεων γενετικών συσχετίσεων, που και εκεί υπήρξε μεγάλη συμφωνία των αποτελεσμάτων, ακόμη και όταν οι συσχετίσεις των μελετών αγνοήθηκαν.



Εικόνα 4: Σύγκριση των συνολικών εκτιμήσεων επίδρασης (A, B, D, E) και των $-\log_{10}(p\text{-values})$ από τα global test (C, F) που προέκυψαν από το `fastMETA`, έναντι εκείνων που αναφέρονται στα αποτελέσματα της διμεταβλητής μετα-ανάλυσης από τους Medina-Gomez et al.

Συνοψίζοντας, μέσω του fastMETA καθίσταται δυνατή η αποτελεσματική διερεύνηση της πλειοτροπίας και των συσχετίσεων μεταξύ σύνθετων χαρακτηριστικών στις σύγχρονες μελέτες γενετικής συσχέτισης. Το fastMETA έχει υλοποιηθεί στην γλώσσα προγραμματισμού Python και διατίθεται τόσο ως αυτόνομο εργαλείο (<https://github.com/pbagos/fastMETA>) όσο και ως διαδικτυακή υποδομή (<https://compngen.dib.uth.gr/fastMETA/>) (Εικόνα 5). Η εργασία αυτή δημοσιεύτηκε στο περιοδικό **Frontiers in Genetics** τον Δεκέμβριο του 2025 (Manios, Kandyias, et al., 2025).

The screenshot displays the fastMETA web application interface, which is divided into three main sections: Settings, Uploaded data preview, and Results.

Settings: This section allows users to configure the analysis. It includes an 'Upload input file (tsv / .txt, tab-separated)' field with a 'Browse...' button and a file named 'medina_gomez_data_input_catalog.txt'. Below this, there are options for 'Required columns: variable, BETA1, SE1, BETA2, SE2, ...', a 'Method' dropdown set to 'method 3 — shared correlation coefficient', and a 'Heterogeneity estimator (only for method 2/method 3)' dropdown set to 'DerSimonian-Laird'. A 'Run fastMETA' button is located at the bottom.

Uploaded data preview: This section shows a table of input data. The table has columns for 'variable', 'BETA1', 'BETA2', 'SE1', and 'SE2'. The data is displayed in a grid format with search and pagination options.

Results: This section displays the output of the analysis. It includes a 'Download results' button and a table with columns for 'variable', 'meta_BETA1', 'meta_SE1', 'meta_BETA2', 'meta_SE2', and 'Wald'. The results are presented in a grid format with search and pagination options.

At the bottom of the interface, the text 'Computational Genetics Group | Dpt. of Computer Science and Biomedical Informatics, University of Thessaly' is visible.

Εικόνα 5: Στιγμιότυπο από την διαδικτυακή υποδομή του fastMETA.

5. PYRAMA: Ένα εργαλείο ανοικτού κώδικα για μετα-αναλύσεις μελετών γενετικής συσχέτισης

Στις μελέτες γενετικής συσχέτισης χρησιμοποιούνται διάφορες στατιστικές μέθοδοι για τον εντοπισμό γενετικών παραλλαγών που σχετίζονται με πολύπλοκα νοσήματα (Ziegler et al., 2008, Uffelmann et al., 2021). Πολλές σημαντικές ανακαλύψεις έχουν προκύψει από την ανάλυση των συνοπτικών στατιστικών (summary statistics) των μελετών γενετικής συσχέτισης (Kontou and Bagos, 2024), ενώ η διαθεσιμότητά τους έχει βελτιώσει την προσβασιμότητα στα γενετικά δεδομένα και έχει προωθήσει τη γονιδιωματική έρευνα. Η μετα-ανάλυση, μέσω του συνδυασμού πολλών μελετών, αυξάνει τη στατιστική ισχύ για την ανίχνευση γενετικών παραλλαγών (Evangelou and Ioannidis, 2013). Έχουν αναπτυχθεί αρκετά λογισμικά για την διεξαγωγή μετα-αναλύσεων χρησιμοποιώντας συνοπτικά στατιστικά, με χαρακτηριστικά παραδείγματα το PLINK (Purcell et al., 2007), το METAL (Willer et al., 2010) και το GWAMA (Mägi and Morris, 2010). Τα εργαλεία αυτά έχουν χρησιμοποιηθεί εκτενώς στην επιστημονική βιβλιογραφία και έχουν συμβάλει σημαντικά σε γενετικές ανακαλύψεις. Ωστόσο, περιλαμβάνουν κυρίως τις τυπικές μεθόδους μετα-ανάλυσης, υποθέτοντας εκ των προτέρων το γενετικό μοντέλο (συνήθως το συνεπικρατές μοντέλο κληρονομικότητας) και δέχονται μόνο συνοπτικά στατιστικά μελετών γενετικής συσχέτισης ως είσοδο ή στερούνται άλλων σημαντικών λειτουργιών. Επομένως, υπάρχει ανάγκη για ένα εργαλείο με το οποίο οι χρήστες θα μπορούν να πραγματοποιούν GWAS μετα-αναλύσεις και να έχουν την δυνατότητα να επιλέγουν ανάμεσα σε πολλές μεθόδους. Επιπλέον, η εφαρμογή αναπλήρωσης μη-γονοτυπημένων πολυμορφισμών (imputation) στα συνοπτικά στατιστικά είναι απαραίτητη σε μια μελέτη, ενισχύοντας περαιτέρω τη στατιστική ισχύ στη μετα-ανάλυση.

Στο πλαίσιο αυτό, αναπτύχθηκε το PYRAMA (Python tool for Robust Analysis and Meta-Analysis), ένα πακέτο ανοικτού κώδικα για την ανάλυση και τη μετα-ανάλυση GWAS. Το PYRAMA είναι δημόσια διαθέσιμο ως εργαλείο γραμμής εντολών αλλά και ως διαδικτυακό εργαλείο, υποστηρίζοντας μοντέλα σταθερών και τυχαίων επιδράσεων, robust tests, Μπεϋζιανή μετα-ανάλυση, imputation συνοπτικών στατιστικών με χρήση προ-υπολογισμένων panels ανισορροπίας σύνδεσης (LD reference panels), καθώς και ανάλυση εμπλουτισμού.

Ως σενάριο χρήσης, αξιοποιήθηκαν τρεις δημόσιες διαθέσιμες GWAS για τη νόσο Parkinson (Mayo Tier 1, Mayo Tier 2 και NINDS), όπως περιγράφονται στη μετα-ανάλυση των Evangelou και συνεργατών (Evangelou et al., 2007). Αρχικά, χρησιμοποιήθηκε το σύνολο δεδομένων NINDS, το οποίο περιλαμβάνει διακριτές κατανομές φαινοτύπου, για να εφαρμόσουμε τα robust tests MIN2 και MAX με το Cochran Armitage Trend Test ως μέγεθος επίδρασης. Παρότι με το κλασικό συνεπικρατές μοντέλο κληρονομικότητας καμία παραλλαγή δεν έφτανε το κατώφλι της στατιστικής σημαντικότητας, οι robust μέθοδοι ανέδειξαν μία (rs10963676, με το MAX) και τρεις παραλλαγές (rs9952724, rs850084, rs1504212, με το MIN2) στατιστικά σημαντικές υπό το υπολειπόμενο μοντέλο κληρονομικότητας, οι οποίες είχαν ήδη χαμηλά p-

values και έχουν εντοπιστεί και σε προηγμένες μεθοδολογικές προσεγγίσεις, αν και ο βιολογικός τους ρόλος παραμένει ασαφής. Στη συνέχεια, αναπαράξαμε τα αποτελέσματα της μετα-ανάλυσης των Evangelou και συνεργατών, εφαρμόζοντας μοντέλα σταθερών και τυχαίων επιδράσεων σε όλες τις κοινές παραλλαγές των τριών μελετών, όπως και στην αρχική δημοσίευση. Τέλος, διερευνήσαμε αν η μέθοδος για imputation συνοπτικών στατιστικών που έχει υλοποιηθεί στο PYRAMA αυξάνει την ισχύ της μετα-ανάλυσης, πραγματοποιώντας imputation σε παραλλαγές που έλειπαν από μία ή δύο μελέτες, αυξάνοντας έτσι την κάλυψη των παραλλαγών. Πραγματοποιώντας μετα-ανάλυση με το μοντέλο των τυχαίων επιδράσεων στις μεγαλύτερες και imputed πλέον μελέτες, εντοπίστηκαν επτά στατιστικά σημαντικές παραλλαγές, που δεν θα μπορούσαν να ανιχνευθούν χωρίς imputation. Η ανάλυση εμπλουτισμού αυτών των rsIDs με το g:Profiler (Raudvere et al., 2019) ανέδειξε όρους σχετιζόμενους με νευροανάπτυξη, ρύθμιση της μεταγραφής και κυτταρικές αποκρίσεις σε περιβαλλοντικά ερεθίσματα, αναδεικνύοντας, μεταξύ άλλων, παραλλαγές στα γονίδια CRB1, ELP4 και RBFOX1, τα οποία έχουν συνδεθεί με νευρολογικές και νευροαναπτυξιακές διαταραχές.

Συνοψίζοντας, το PYRAMA είναι ένα γρήγορο και ευέλικτο εργαλείο που αντιμετωπίζει βασικούς περιορισμούς των κλασικών λογισμικών GWAS μετα-ανάλυσης. Από όσο γνωρίζουμε, είναι το μόνο ενεργά συντηρούμενο πρόγραμμα που συνδυάζει την αναπλήρωση μη-γονοτυπημένων πολυμορφισμών με μετα-ανάλυση, όπως φάνηκε σε εφαρμογή στη νόσο Parkinson όπου αρκετά SNPs έφτασαν το επίπεδο genome-wide significance. Το PYRAMA υποστηρίζει την δυνατότητα πολυνηματικής (multi-threading) εκτέλεσης και υπερτερεί σε ταχύτητα έναντι των αντίστοιχων εργαλείων, καθιστώντας το κατάλληλο για μελέτες μεγάλης κλίμακας. Η έκδοση γραμμής εντολών διατίθεται ελεύθερα στο GitHub (<https://github.com/rbagos/PYRAMA>) και η διαδικτυακή υποδομή είναι διαθέσιμη στην ακόλουθη διεύθυνση: <https://compgen.dib.uth.gr/PYRAMA> (Εικόνα 6). Το λογισμικό θα ενημερώνεται τακτικά με νέες μεθόδους μετα-ανάλυσης, εκτιμητές τυχαίων επιδράσεων, επιπλέον LD reference panels και διαδραστικές απεικονίσεις. Η εργασία αυτή έχει γίνει δεκτή προς δημοσίευση στο περιοδικό **Bioinformatics** (Manios et al., 2026)

PYRAMA: A tool for Meta-analysis of Genome Wide Association Studies



©PYRAMA - Computational Genetics Group, Department of Computer Science and Biomedical Informatics, University of Thessaly

Εικόνα 6: Στιγμιότυπο από την διαδικτυακή υποδομή του PYRAMA.

6. Ανακατασκευή συνοπτικών γονοτύπων

Συνήθως, τα συνοπτικά στατιστικά από μια μελέτη GWAS περιέχουν, εκτός από το αναγνωριστικό ID και την γονιδιωματική θέση, το μέγεθος επίδρασης (log Odds Ratio), το p-value, το 95% διάστημα εμπιστοσύνης (CI) ή το τυπικό σφάλμα. Η ανακατασκευή των συνοπτικών αριθμών γονοτύπων από τέτοια συνοπτικά στατιστικά μπορεί να είναι χρήσιμη σε πολλές περιπτώσεις, για παράδειγμα, όταν θέλουμε να αποκρυπτογραφήσουμε τον γενετικό τρόπο κληρονομικότητας ή να εκτελέσουμε άλλες στατιστικές δοκιμές που απαιτούν ομαδοποιημένους γονοτύπους.

Δύο μέθοδοι είναι διαθέσιμες αυτή τη στιγμή για την ανακατασκευή του αριθμού των γονοτύπων, η μέθοδος του (Di Pietrantonj, 2006), η οποία αρχικά διατυπώθηκε για γενικούς Πίνακες 2x2, και η πρόσφατα προταθείσα μέθοδος των (Yang et al., 2022). Οι δύο μέθοδοι είναι παρόμοιες. Και οι δύο κάνουν την υπόθεση της ισορροπίας HWE και απαιτούν ως δεδομένα εισόδου τα μεγέθη του δείγματος, το Odds Ratio και το τυπικό σφάλμα (se) ή το CI του, μαζί με μια εκτίμηση των συχνοτήτων των αλληλόμορφων. Μια άλλη πιο εύκολα εφαρμόσιμη μέθοδος (την ονομάζουμε Direct μέθοδο) μπορεί να διατυπωθεί λαμβάνοντας ως δεδομένα εισόδου το μέγεθος του δείγματος, το odds ratio και την ακριβή τιμή της συχνότητας των αλληλόμορφων στην ομάδα των control. Στην εργασία μας, διαπιστώσαμε ότι είναι κρίσιμο να υπολογίσουμε επίσης τα Διαστήματα Εμπιστοσύνης που προκύπτουν από τις τιμές p-value, δεδομένου ότι οι αναφερόμενες τιμές των CI στο σύνολο δεδομένων της GWAS Catalog αναφέρονται μόνο σε λίγες περιπτώσεις με τα επιθυμητά δεκαδικά ψηφία (δύο ή περισσότερα) εάν δεν αναφέρονται καθόλου. Ο συγγραφέας της εργασίας (Di Pietrantonj, 2006) καθιστά σαφές ότι η μέθοδός του λειτουργεί καλύτερα όταν το CI αναφέρεται με δύο ή περισσότερα δεκαδικά ψηφία. Αρχικά, χρησιμοποιήσαμε 5 σύνολα δεδομένων GWAS για τα οποία ήταν διαθέσιμοι οι αριθμοί γονοτύπων και δοκιμάσαμε τη διαδικασία ανακατασκευής. Επίσης, δημιουργήσαμε μια διαδικασία ελέγχου ποιότητας για δεδομένα που ελήφθησαν σε πραγματικές καταστάσεις. Αποφασίσαμε να συγκρίνουμε τα παρατηρούμενα συνοπτικά στατιστικά στοιχεία (p-value, z, se) με αυτά που παράγονται από την ανακατασκευή και να διατηρήσουμε τα SNP με μικρές αποκλίσεις. Χρησιμοποιήσαμε το ποσοστό σχετικής αλλαγής σε κάθε στατιστικό στοιχείο ($\Delta_{\log p}$, Δ_z και Δ_{se}) και εφαρμόσαμε διαφορετικά κατώφλια αποκοπής (20%, 10%, 5%, 1%) προκειμένου να αξιολογήσουμε την καλύτερη στρατηγική με σκοπό τη χρήση της σε προβλήματα του πραγματικού κόσμου.

Συγκρίναμε τις τρεις μεθόδους στις δημοσιευμένες GWAS για τις οποίες είναι διαθέσιμοι οι αριθμοί των γονοτύπων (Πίνακας 1) και στα συνοπτικά στατιστικά στοιχεία από τις συσχετίσεις που είναι διαθέσιμες στη βάση δεδομένων της GWAS Catalog (Πίνακας 2, Εικόνες 7-11). Ο Πίνακας 1 δείχνει ότι σχεδόν σε όλες τις περιπτώσεις περισσότερο από το 99% των SNPs παρουσιάζουν καλή προσαρμογή των παρατηρούμενων αριθμών γονοτύπων με τους ανακατασκευασμένους σε όλες τις μεθόδους. Αυτό υπογραμμίζει το γεγονός ότι η

μέθοδος Direct, η οποία είναι η απλούστερη, λειτουργεί καλά, εάν τα CI και οι συχνότητες των αλληλομόρφων αναφέρονται με ακρίβεια. Από τον Πίνακα 2 και τις Εικόνες 7-11, παρατηρούμε ότι οι καλύτερες μέθοδοι ανακατασκευής με οποιοδήποτε κατώφλι αποκοπής και οποιονδήποτε δείκτη είναι η μέθοδος Di Pietrantonj με τις υπολογισμένες από εμάς τιμές CI και η μέθοδος Yang. Η μέθοδος Yang μπορεί να υπερβαίνει αυτήν του Di Pietrantonj σε ορισμένες μεταβλητές, αλλά συνολικά, η μέθοδος Di Pietrantonj φαίνεται πιο αποτελεσματική. Οι δύο μέθοδοι όχι μόνο δίνουν πολύ κοντινά αποτελέσματα - κοντά στο 1 όσον αφορά το R^2 κάθε μεταβλητής, αλλά περισσότερο από το 95% των SNPs διατηρούνται σε όλες τις περιπτώσεις. Στις Εικόνα 7 και Εικόνα 8 παρουσιάζονται οι δείκτες Δ_{logr} και Δ_z , ενώ τα αποτελέσματα είναι αρκετά παρόμοια και για τον δείκτη Δ_{se} . Επιπλέον, παρατηρώντας τις τιμές του $R_{logrvalue}^2$ και τα SNPs που διατηρήθηκαν, πιστεύουμε ότι το καλύτερο κατώφλι αποκοπής για τέτοιες αναλύσεις είναι το 20%, διατηρώντας τόσο υψηλή ακρίβεια όσο και υψηλό ποσοστό διατηρούμενων SNPs.

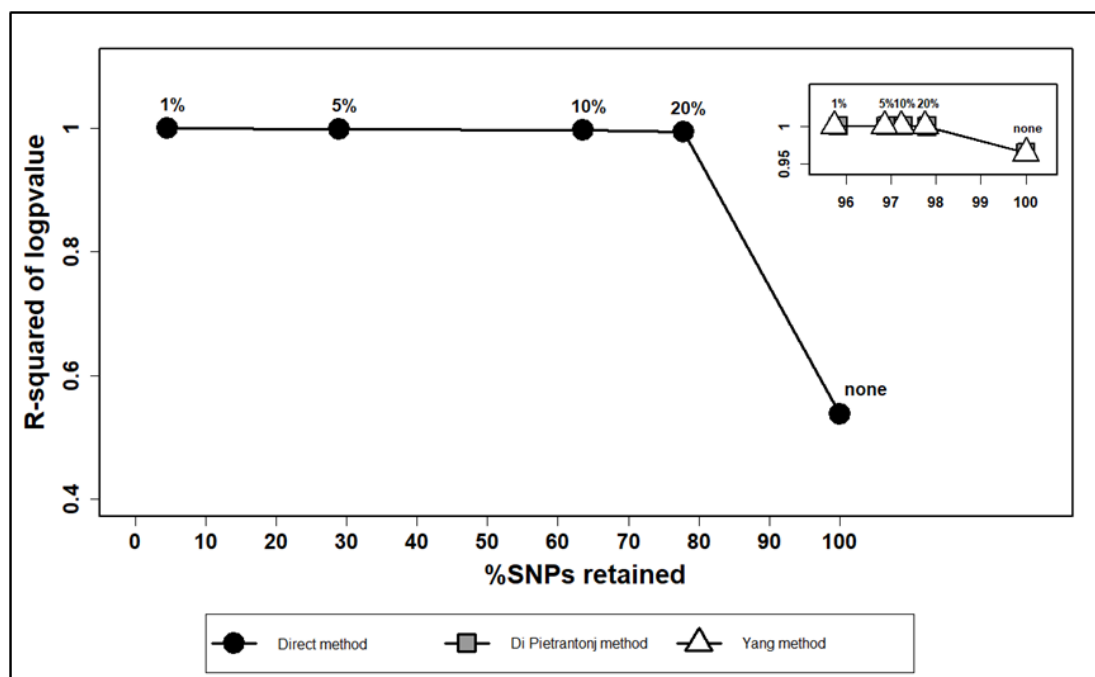
Πίνακας 1: Αποτελέσματα του χ^2 τεστ καλής προσαρμογής των 3 μεθόδων ανακατασκευής για τα 5 σύνολα δεδομένων.

Σύνολο δεδομένων	Μέθοδος	SNPs χωρίς καλή προσαρμογή	SNPs με καλή προσαρμογή/ SNPs συνολικά (%)
ADHD	Direct	1689 (0.55%)	306621/308677 (99.33%)
	Di Pietrantonj	1673 (0.54%)	306636/308677 (99.34%)
	Yang	1671 (0.54%)	306639/308677 (99.34%)
Alzheimer	Direct	3696 (1.06%)	341445/349525 (97.69%)
	Di Pietrantonj	3715 (1.06%)	341425/349525 (97.68%)
	Yang	3705 (1.06%)	341436/349525 (97.69%)
Parkinson	Direct	3898 (0.99%)	391746/395644 (99.02%)
	Di Pietrantonj	3913 (0.99%)	391726/395644 (99.01%)
	Yang	3891 (0.98%)	391753/395644 (99.02%)
Sigma_Exome	Direct	932 (0.58%)	158502/159443 (99.41%)
	Di Pietrantonj	948 (0.59%)	158486/159443 (99.40%)
	Yang	925 (0.58%)	158509/159443 (99.41%)

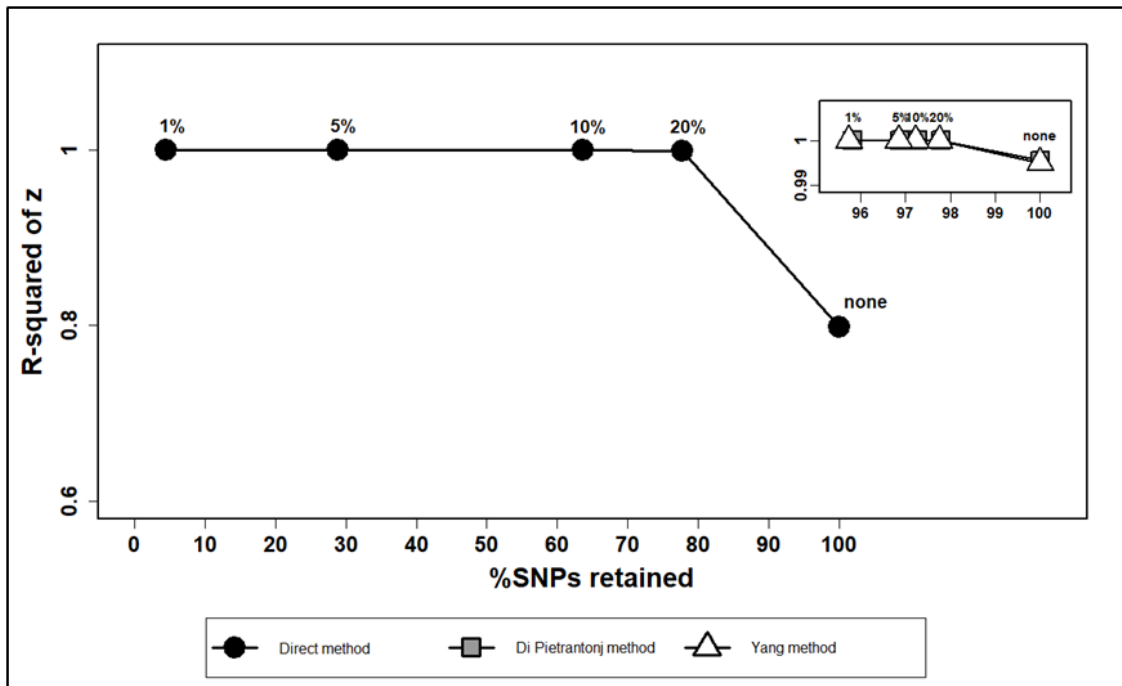
Sigma_Omni	Direct	2393 (0.68%)	347131/349525 (99.32%)
	Di Pietrantonj	2356 (0.67%)	347167/349525 (99.33%)
	Yang	2384 (0.68%)	347140/349525 (99.32%)

Πίνακας 2: Τιμές R^2 των 3 μεθόδων ανακατασκευής με κατώφλι 20% για το δείκτη Δ_{logp} .

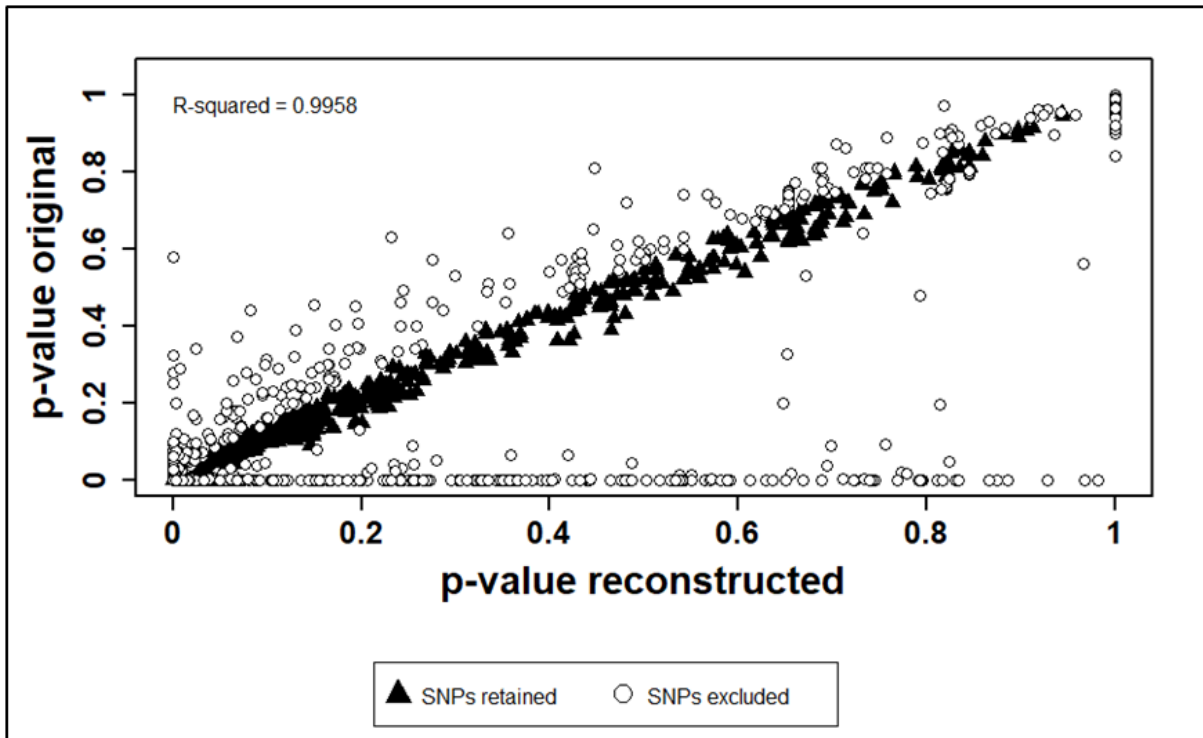
Μέθοδος	SNPs που κρατήθηκαν (%)	R^2 (z)	R^2 (p)	R^2 (logp)
Direct	77.80%	.9990	.9958	.9928
Di Pietrantonj	97.80%	1	.9999	.9996
Di Pietrantonj – reported CI	53.11%	.9975	.9974	.9890
Yang	97.76%	1	.9998	.9998



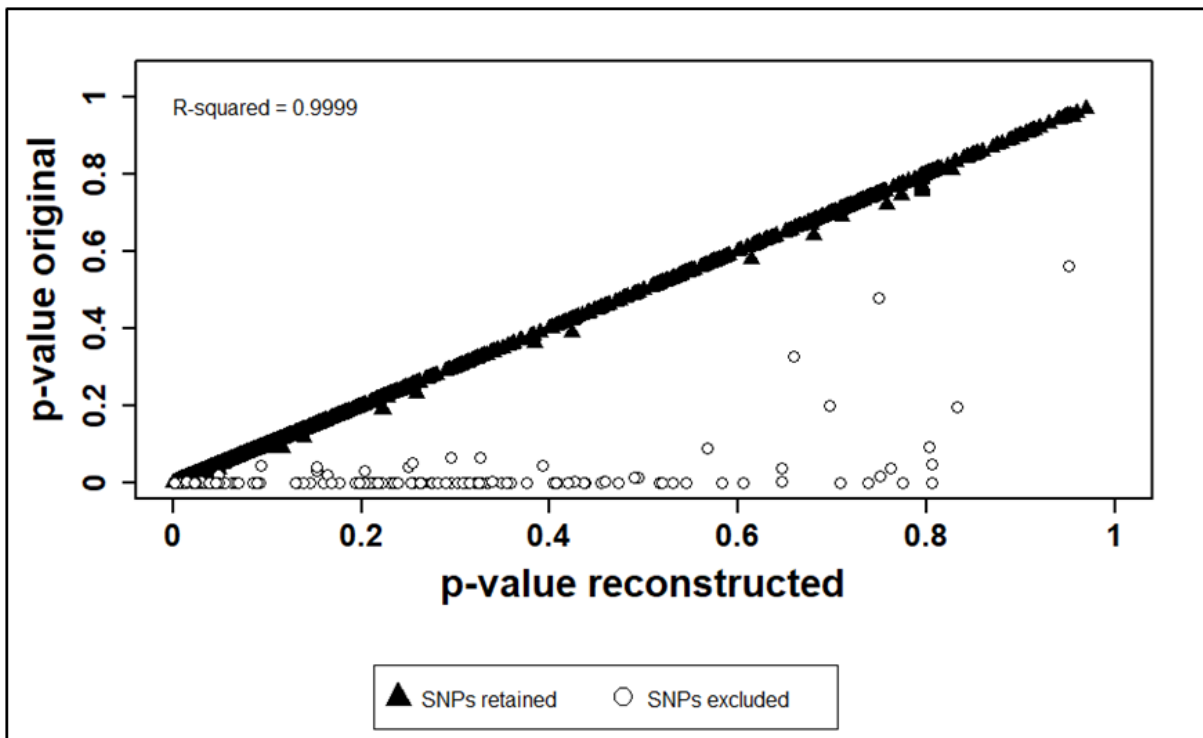
Εικόνα 7: Σύγκριση των 3 μεθόδων ανακατασκευής με το δείκτη Δ_{logp} αναφορικά με τις τιμές $R_{logpvalue}^2$ και το ποσοστό SNPs που κρατήθηκαν για κάθε κατώφλι.



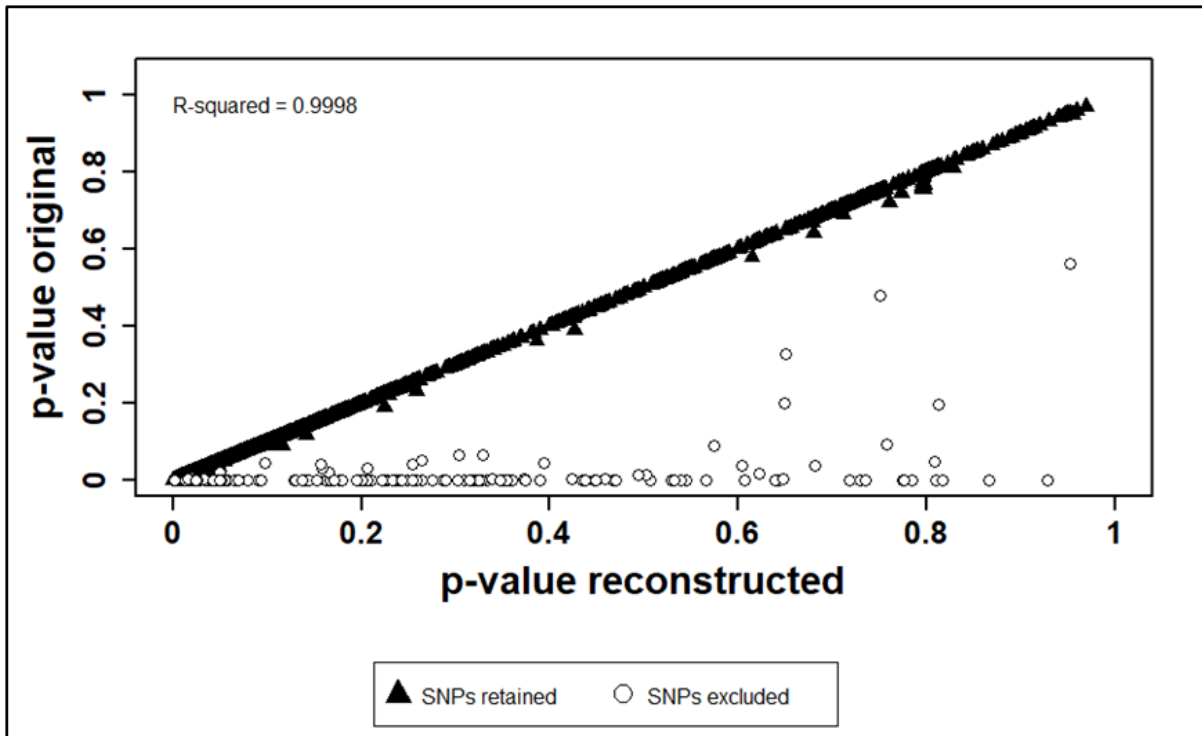
Εικόνα 8: Σύγκριση των 3 μεθόδων ανακατασκευής με το δείκτη Δ_{logp} αναφορικά με τις τιμές R_z^2 και το ποσοστό SNPs που κρατήθηκαν για κάθε κατώφλι.



Εικόνα 9: Συσχέτιση των παρατηρούμενων p-value και των ανακατασκευασμένων με τη μέθοδο Direct και το δείκτη $\Delta_{\log p}$ παρουσιάζοντας τα SNPs που κρατήθηκαν και τα SNPs που διαγράφηκαν για το κατώφλι 20%.



Εικόνα 10: Συσχέτιση των παρατηρούμενων p-value και των ανακατασκευασμένων με τη μέθοδο Di Pietrantonij και το δείκτη $\Delta_{\log p}$ παρουσιάζοντας τα SNPs που κρατήθηκαν και τα SNPs που διαγράφηκαν για το κατώφλι 20%.



Εικόνα 11: Συσχέτιση των παρατηρούμενων p-value και των ανακατασκευασμένων με τη μέθοδο Yang και το δείκτη Δ_{logR} παρουσιάζοντας τα SNPs που κρατήθηκαν και τα SNPs που διαγράφηκαν για το κατώφλι 20%.

Οι μέθοδοι των Di Pietrantonj και Yang είναι αρκετά παρόμοιες και λειτουργούν καλύτερα σε περιπτώσεις όπου τα συνοπτικά στατιστικά στοιχεία είναι ανεπαρκή. Ωστόσο, ενδέχεται να υποφέρουν όταν το διάστημα εμπιστοσύνης δίνεται με χαμηλή ακρίβεια. Αντίθετα, εάν χρησιμοποιηθεί το τυπικό σφάλμα ή η τιμή p-value, η ανακατασκευή είναι αρκετά ακριβής. Η μέθοδος Direct είναι λιγότερο ακριβής και χειροτερεύει εάν οι συχνότητες των αλληλομόρφων δεν αναφέρονται στη μελέτη, οπότε χρησιμοποιούνται εξωτερικές πληροφορίες από βάσεις δεδομένων. Επίσης, είναι προφανές ότι το κατώφλι αποκοπής 20% επιτυγχάνει βέλτιστα αποτελέσματα. Με βάση αυτά, παρουσιάζουμε ορισμένα απλά διαγνωστικά κριτήρια και πρακτικούς κανόνες που θα καθοδηγήσουν την ανακατασκευή σε προβλήματα του πραγματικού κόσμου. Όλες οι προαναφερθείσες μέθοδοι ανακατασκευής γονοτύπων έχουν υλοποιηθεί σε μια δημόσια διαθέσιμη διαδικτυακή υπηρεσία που φιλοξενείται σε αυτήν τη διεύθυνση: <https://compgen.dib.uth.gr/GWASreconstruction>. Η εργασία παρουσιάστηκε σε διεθνές συνέδριο το 2025, **Joint Conference of the Italian and Eastern Mediterranean Regions of the International Biometric Society (IBS-IR-EMR2025)** (Exarchou-Kouveli, et al. 2025).

7. Βιβλιογραφία

Chen, Y., Cai, Y., Hong, C., & Jackson, D. (2016). Inference for correlated effect sizes using multiple univariate meta-analyses. *Statistical Medicine*, 35(9), 1405–1422. <https://doi.org/10.1002/sim.6789>

Di Pietrantonj, C. (2006). Four-fold table cell frequencies imputation in meta analysis. *Statistics in medicine*, 25(13), 2299-2322.

Evangelou, E., & Ioannidis, J. P. A. (2013). Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics*, 14(6), 379–389.

Evangelou, E., Maraganore, D. M., & Ioannidis, J. P. A. (2007). Meta-analysis in genome-wide association datasets: strategies and application in Parkinson disease. *PLoS One*, 2(2), e196.

Exarchou-Kouveli, K.K., Stasi, E.A., Manios, G.A., Kontou, P.I., Bagos, P.G. (2025). GWAS genotype reconstruction from summary statistics: a comparison of methods. Joint Conference of the Italian and Eastern Mediterranean Regions of the International Biometric Society (IBS-IR-EMR2025)

Gibbs, R. A., Belmont, J. W., Hardenbol, P., Willis, T. D., Yu, F. L., Yang, H. M., Ch'ang, L.-Y., Huang, W., Liu, B., & Shen, Y. (2003). The international HapMap project.

Huang, L., Rosen, J. D., Sun, Q., Chen, J., Wheeler, M. M., Zhou, Y., Min, Y.-I., Kooperberg, C., Conomos, M. P., & Stilp, A. M. (2022). TOP-LD: A tool to explore linkage disequilibrium with TOPMed whole-genome sequence data. *The American Journal of Human Genetics*, 109(6), 1175–1181.

Julienne, H., Shi, H., Pasaniuc, B., & Aschard, H. (2019). RAISS: robust and accurate imputation from summary statistics. *Bioinformatics*, 35(22), 4837–4839.

Kontou, P. I., & Bagos, P. G. (2024). The goldmine of GWAS summary statistics: a systematic review of methods and tools. *BioData Mining*, 17(1), 31.

Kwan, J. S. H., Li, M.-X., Deng, J.-E., & Sham, P. C. (2016). FAPI: fast and accurate P-value imputation for genome-wide association study. *European Journal of Human Genetics*, 24(5), 761–766.

Lee, D., & Bacanu, S.-A. (2024). GAUSS: a summary-statistics-based R package for accurate estimation of linkage disequilibrium for variants, Gaussian imputation, and TWAS analysis of cosmopolitan cohorts. *Bioinformatics*, 40(4), btae203.

Lee, D., Bigdeli, T. B., Riley, B. P., Fanous, A. H., & Bacanu, S.-A. (2013). DIST: direct imputation of summary statistics for unmeasured SNPs. *Bioinformatics*, 29(22), 2925–2927. <https://doi.org/10.1093/bioinformatics/btt500>

Lee, D., Bigdeli, T. B., Williamson, V. S., Vladimirov, V. I., Riley, B. P., Fanous, A. H., & Bacanu, S.-A. (2015). DISTMIX: direct imputation of summary statistics for unmeasured SNPs from mixed ethnicity cohorts. *Bioinformatics*, 31(19), 3099–3104.

Mägi, R., & Morris, A. P. (2010). GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics*, 11, 1–6.

Manios, G. A., Kandylas, D., Kylonis, A., Bagos, P. G., & Kontou, P. I. (2025). fastMETA: a fast and efficient tool for multivariate meta-analysis of GWAS. *Frontiers in Genetics, Volume 16-2025*.

Manios, G. A., Michailidi, A., Kontou, P. I., & Bagos, P. G. (2025). PRED-LD: efficient imputation of GWAS summary statistics. *BMC bioinformatics*, 26(1), 107.

Manios, G. A., Nteli, S., Kontou, P. I., & Bagos, P. G. (2026). PYRAMA: An open-source tool for advanced meta-analysis of genome wide association studies. *Bioinformatics*.

Medina-Gomez, C. (2017). Bivariate genome-wide association meta-analysis of pediatric musculoskeletal traits reveals pleiotropic effects at the SREBF1/TOM1L2 locus. *Nature Communications*, 8(1), 121.

Oehlert, G. W. (1992). A note on the delta method. *The American Statistician*, 46(1), 27–29.

Pasaniuc, B., Zaitlen, N., Shi, H., Bhatia, G., Gusev, A., Pickrell, J., Hirschhorn, J., Strachan, D. P., Patterson, N., & Price, A. L. (2014). Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics*, 30(20), 2906–2914.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3), 559–575.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, 47(W1), W191–W198.

Rüeger, S., McDaid, A., & Kutalik, Z. (2018). Evaluation and application of summary statistic imputation to discover new height-associated loci. *PLoS Genetics*, 14(5), e1007371.

- Siva, N. (2008). 1000 Genomes project. *Nature Biotechnology*, 26(3), 256–257.
- Staley, J. R., Blackshaw, J., Kamat, M. A., Ellis, S., Surendran, P., Sun, B. B., Paul, D. S., Freitag, D., Burgess, S., & Danesh, J. (2016). PhenoScanner: a database of human genotype–phenotype associations. *Bioinformatics*, 32(20), 3207–3209.
- Uffelmann, E., Huang, Q. Q., Munung, N. S., De Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, 1(1), 59.
- Vennou, K. E., Piovani, D., Kontou, P. I., Bonovas, S., & Bagos, P. G. (2020). Multiple outcome meta-analysis of gene-expression data in inflammatory bowel disease. *Genomics*, 112(2), 1761–1767.
- Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T. J. C., van der Sluis, S., Andreassen, O. A., Neale, B. M., & Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics*, 51(9), 1339–1348.
- Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, 26(17), 2190–2191.
- Yang, Z., Paschou, P., & Drineas, P. (2022). Reconstructing SNP allele and genotype frequencies from GWAS summary statistics. *Scientific Reports*, 12(1), 8242.
- Ziegler, A., König, I. R., & Thompson, J. R. (2008). Biostatistical aspects of genome-wide association studies. *Biometrical Journal*, 50(1), 8–28.
- Zondervan, K. T., & Cardon, L. R. (2004). The complex interplay among factors that influence allelic association. *Nature Reviews Genetics*, 5(2), 89–100.