

# ELIXIR-GR: Εκτίμηση κληρονομησιμότητας και γενετικής συσχέτισης από μελέτες συσχέτισης εύρους γονιδιώματος

Τσιφιντάρης Μαργαρίτης<sup>1</sup>, Τσέτσος Φώτιος<sup>1</sup>, Μαρκάκη Μαρία<sup>1</sup>, Γαλάνης Αλέξης<sup>1</sup>, Πάσχου Περιστέρα<sup>1</sup>

<sup>1</sup>Τμήμα Μοριακής Βιολογίας και Γενετικής, Δημοκρίτειο Πανεπιστήμιο Θράκης



ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ  
DEMOCRITUS UNIVERSITY OF THRACE

## Εισαγωγή

Τα εργαλεία βιοπληροφορικής χρησιμοποιούνται ολοένα και περισσότερο για τις δυνατότητες που προσφέρουν στην ταχύτητα της ανάλυσης και επεξεργασίας μεγάλου όγκου βιολογικών δεδομένων και την αποτελεσματική μείωση του συνολικού κόστους των αναλύσεων. Στην παρούσα εργασία παρουσιάζουμε ένα εργαλείο το οποίο αυτοματοποιεί τη χρήση της τεχνικής *Linkage Disequilibrium Score Regression* (LDSC). Η τεχνική αυτή ποσοτικοποιεί τη συνεισφορά της πολυγονιδιακής βάσης ενός φαινοτύπου στα αποτελέσματα, σε σύγκριση με άλλους παράγοντες, χρησιμοποιώντας τα αποτελέσματα (summary statistics) από Μελέτες Συσχέτισης Εύρους Γονιδιώματος (Genome-Wide Association Studies – GWAS). Η LDSC χρησιμοποιείται ευρέως και για την εκτίμηση της γενετικής συσχέτισης και της κληρονομησιμότητας μεταξύ διαφόρων ασθενειών (ή γενικώς δυαδικών φαινοτύπων).

Ο σκοπός αυτής της μελέτης είναι να κάνει διαθέσιμη προς χρήση αυτήν την μέθοδο σε οποιοδήποτε χρήστη χωρίς να απαιτούνται γνώσεις προγραμματισμού

## Ροή Εργασιών (Pipeline)

### Είσοδος Αρχείων

Summary Statistics

Πίνακας excel

Ως είσοδο χρησιμοποιεί αρχεία με summary statistics και ένα ακόμα αρχείο με έναν πίνακα που περιγράφει τα δεδομένα.

Αυτός ο πίνακας έχει τρεις στήλες και πρέπει να περιέχει τις εξής πληροφορίες: ονομασία μελέτης, τοποθεσία του αρχείου στον δίσκο και μέγεθος του δείγματος της μελέτης.

### Επεξεργασία Αρχείων

Manhattan Plots

Έλεγχος Ποιότητας

Ανάλυση LDSC

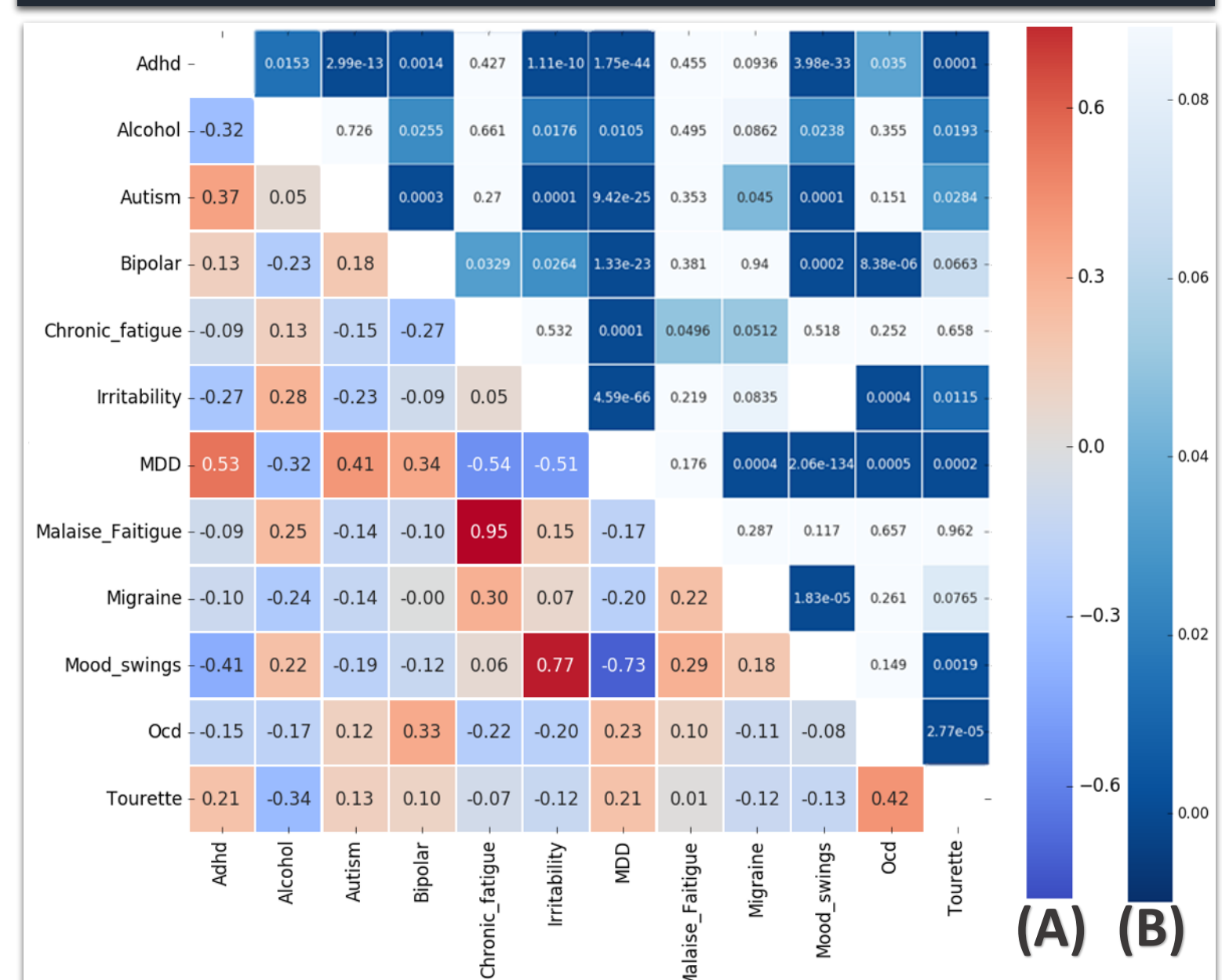
Τα αρχεία οργανώνονται σε φάκελους και για κάθε μια μελέτη δημιουργείται ένα Manhattan Plot. Στην συνέχεια, αναζητούνται στα αρχεία πληροφορίες για τις παραμέτρους: όνομα πολυμορφισμού, χρωμόσωμα, θέση (σε base pair), τα 2 αλληλόμορφα, Τιμή-P, και το effect (beta, or, z-score). Στο τέλος φιλτράρονται για διπλό-εγγραφές (double rsid), MAF μικρότερη του 1%, αφαιρούνται όσοι πολυμορφισμοί δεν είναι SNP (π.χ. INDELs), και από αυτά όσα είναι τρι-αλληλικά.

### Έξοδος Αρχείων

Γράφημα (heatmap)

Στο εξαγόμενο γράφημα (Εικόνα 1) απεικονίζονται τα αποτελέσματα μιας ανάλυσης LDSC. Η γενετική συσχέτιση παίρνει τιμές από -1 έως 1. Όσο μεγαλύτερη είναι η γενετική συσχέτιση (θετικές τιμές) μεταξύ 2 φαινοτύπων τόσο πιο κόκκινο γίνεται το τετράγωνο για το συγκεκριμένο ζεύγος. Από την άλλη πλευρά, αν αυτοί οι 2 φαινότυποι εμφανίζουν αντίθετη κατεύθυνση (αρνητικές τιμές) και όσο αυτό το φαινόμενο γίνεται εντονότερο τόσο πιο μπλε γίνεται το τετράγωνο. Η τιμή 0 δείχνει ότι δεν υπάρχει συσχέτιση.

## Αποτελέσματα



**Εικόνα 1.** Παράδειγμα απεικόνισης των αποτελεσμάτων της ανάλυσης LDSC για 11 διαφορετικούς φαινοτύπους που παρατηρείτε κλινική συσχέτιση με το σύνδρομο Tourette. Στη στήλη (A) και το κάτω μισό του τετραγώνου παρατηρούνται οι τιμές για την γενετική συσχέτιση ενώ στην στήλη (B) και το πάνω μισό του τετραγώνου οι τιμές-P (p-values)

Στο παραπάνω γράφημα παρατηρείται η γενετική συσχέτιση του Συνδρόμου Tourette με άλλες ασθένειες (η γενικότερα φαινοτύπους)

## Συμπεράσματα

Το εργαλείο (pipeline) που δημιουργήθηκε μπορεί να κάνει διαθέσιμη προς χρήση αυτήν την μέθοδο (LDSC) σε οποιοδήποτε χρήστη χωρίς να απαιτούνται γνώσεις προγραμματισμού. Ωστόσο, ο χρόνος που απαιτείται να γίνει μια ανάλυση σε ένα συμβατικό υπολογιστή μπορεί να φτάσει μέχρι και κάποιες ώρες. Για αυτό το σκοπό ο μελλοντικός στόχος είναι να δημιουργηθεί μια διαδικτυακή πλατφόρμα (π.χ. περιβάλλον Galaxy) η οποία θα τοποθετηθεί σε κάποιο server ώστε να κάνει ευκολότερη την χρήση και την προσβασιμότητα αυτού του εργαλείου καθώς και να μειώσει τον χρόνο των αναλύσεων.



Η εργασία αυτή υλοποιήθηκε στο πλαίσιο της Πράξης «ELIXIR-GR: Η Ελληνική Ερευνητική Υποδομή για Διαχείριση και Ανάλυση Δεδομένων στις Βιοεπιστήμες» (MIS 5002780) που εντάσσεται στη Δράση «Ενίσχυση των Υποδομών Έρευνας και Καινοτομίας» και χρηματοδοτείται από το Επιχειρησιακό Πρόγραμμα «Ανταγωνιστικότητα, Επιχειρηματικότητα και Καινοτομία» στο πλαίσιο του ΕΣΠΑ 2014-2020, με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης (Ευρωπαϊκό Ταμείο Περιφερειακής Ανάπτυξης).

Η εργασία αυτή υλοποιήθηκε επίσης στο πλαίσιο του έργου «Διεπιστημονικό Δίκτυο Εκπαίδευσης για το σύνδρομο Tourette. Διαμόρφωση Ευρωπαϊκών υποδομών εκπαίδευσης για τις νευροαναπτυξιακές διαταραχές (MATCHING FUNDS), KE-81523»



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ  
ΥΠΟΥΡΓΕΙΟ  
ΟΙΚΟΝΟΜΙΑΣ & ΑΝΑΠΤΥΞΗΣ  
ΕΙΔΙΚΗ ΓΡΑΜΜΑΤΕΙΑ ΕΠΠΑ & ΤΣ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ ΕΠΑΝΕΚ

ΕΠΑΝΕΚ 2014-2020  
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ  
ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑ  
ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΤΗΤΑ  
ΚΑΙΝΟΤΟΜΙΑ

ΕΣΠΑ  
2014-2020  
ανάπτυξη - εργασία - αλληλεγγύη

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης