IN TIME OF TEST, FAMILY IS BEST

Family-based GWAS: methods and applications to addiction phenotypes

AMELIA C. MINICA

Department of Biological Psychology, Vrije Universiteit Amsterdam, camelia.minica@gmail.com

Supervisors: Prof. Conor V. Dolan, Prof. Jacqueline M. Vink and Prof. Dorret I. Boomsma

PROJECT: WHY & WHAT FOR?

Genome-Wide Association Studies (GWAS) = test the statistical association between the GV and the phenotype in a regression model

Family-based GWAS

\[ y_i = \beta_0 + \beta_1 g_i + \ldots + \beta_p g_{pi} + e_i \]

where \( i \) is indicator of family and \( i \) is subjects within families.

\( y, g, e \) are vectors (n = number of phenotypes within family)

\[ X = \begin{pmatrix} 1 \\ g_{1i} \\ \vdots \\ g_{ni} \end{pmatrix}, \quad b = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \]

Statistical Power - paramount in GWAS for:

- small effect genes: < 1% explained variance
- up to 6 million tests → adapted \( \alpha = 10^{-8} \)

Aim: Increase power by refinement of statistical methodologies and meta-analyses

Retain computational speed

SANDWICH CORRECTED SE

Background: Relatives resemble each other because they share genes (A) and environment (C). Resemblance is expressed in:

\[ \mathbf{V} = \mathbf{A} + \mathbf{C} \]

Background: SKAT - important rare variants (RV) test based on a random effects model.

Weights assigned to capture the likelihood of a RV being functional.

Correct weighting increases power and yet correct weights are not known. What is the effect of weight misspecification in SKAT?

Methods: Compare LRT and score test under weight misspecification using simulations.

Figure: LEFT: Weights assigned based on frequency (maf)

RIGHT: Simulated weights: beta.weights(1,1), Fitted weights: beta.weights(5,5).

Conclusion: LRT is more robust and powerful than score under weight misspecification. This is a paramount result, as misspecified models are likely to be the rule rather than the exception.

MZ TWINS OR MZ SINGLETONS?

Background: Occasionally in family-based GWAS, including monozygotic (MZ) twins, the data from one MZ twin are dropped, thus reducing the MZ pairs to singletons.

Is this practice optimal?

Compute effective sample size:

\[ N_e = \frac{(2^rN)}{(1+r)} \]

ranges from N (r = 1) to 2^N (r = 0)

Conclusion: the presence of MZ twin pairs does not affect the type I error rate, and reducing MZ pairs to singletons reduces power.

5 GENES IMPLICATED IN CANNABIS USE: A META-ANALYSIS

Background: Regular cannabis use has been associated with health problems (mood and anxiety disorders) and predicts diminished educational and professional attainment.

Methods: Fixed effects meta-analysis in a sample >32,000 individuals.

Implications: One can start building a road map for developing drugs to treat cannabis dependence and abuse.