IN TIME OF TEST, FAMILY IS BEST

Family-based GWAS: methods and applications to addiction phenotypes

CAMELIA C. MINICĂ
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-**W**ide **A**ssociation **S**tudies (GWAS) = test the statistical association between the GV and the phenotype in a regression model

Family-based GWAS

$$\mathbf{y}_{ij} = \mathbf{b}_0 + \mathbf{b}_1 * \mathbf{g}_{ij} + \boldsymbol{\varepsilon}_{ij}$$

where i is indicator of family and j is subjects within families.
y, **b**, **g** and **ε** are vectors (n = number of phenotypes within family)

$$X = \begin{pmatrix} 1 & g_1 \\ 1 & g_2 \\ \vdots & \vdots \\ 1 & g_N \end{pmatrix} \qquad b = \begin{pmatrix} b0 \\ b1 \end{pmatrix} \qquad y = \begin{pmatrix} ph_1 \\ ph_2 \\ \vdots \\ ph_N \end{pmatrix}$$

Statistical Power - paramount in GWAS for:

- small effect genes: $< 1\%$ explained variance
- up to 6 million tests $\rightarrow$ 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:

THE FAMILIAL COVARIANCE MATRIX V

$$\boldsymbol{\varepsilon} | X \sim N(0, V)$$

$$V(\Theta)$$

$$\Theta = [\sigma^2_A, \sigma^2_C, \sigma^2_E]$$

What model for V is most powerful *and* fast?

Methods: Use simulations to compare the standard and sandwich corrected Unweighted Least Squares (ULS) and Maximum Likelihood (ML) .

SIMULATION ACE trait 4-sib family

	ML standard ACE model (true)	Sandwich corrected ML (false: AE Model)	Sandwich corrected ML (false: CE model)	Sandwich corrected ULS (false: E model)
mean(b1)	-0.142	-0.142	-0.142	-0.142
mean (st.err.)	0.023	0.024	0.024	0.031
mean (t-value)	-6.03	-5.98	-5.98	-4.65
power	**75.7**	**74.2**	**74.2**	**25.1**

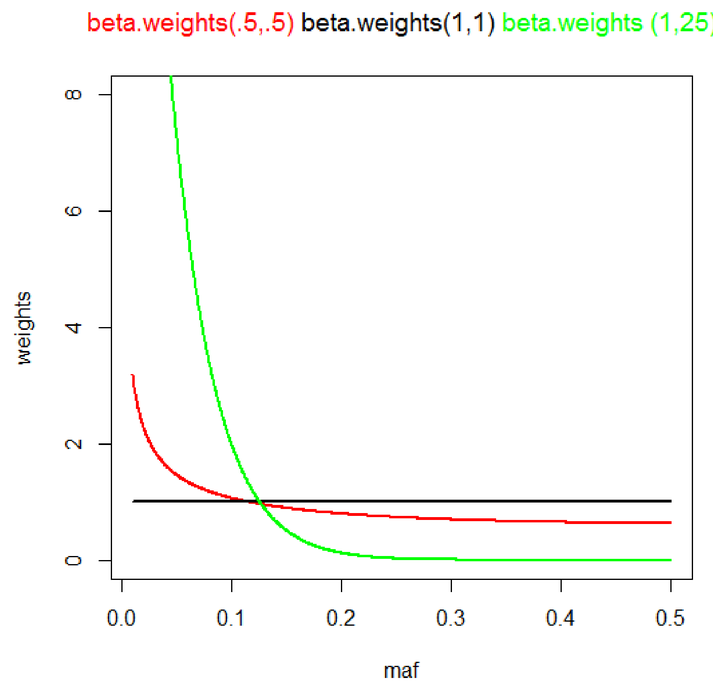
Conclusion: Model V as an AE or a CE & use ML with a SANDWICH!

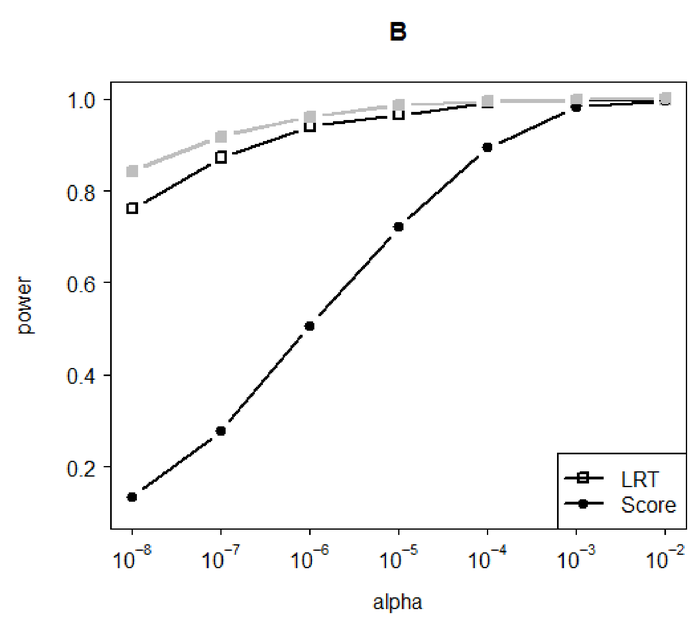
THE WEIGHTING IS THE HARDEST PART

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?

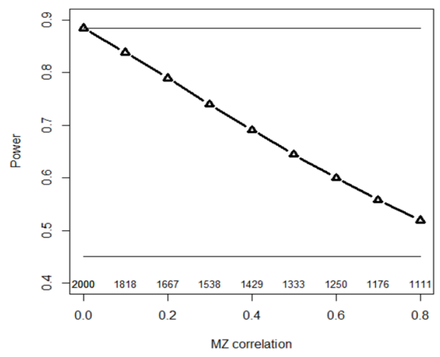
Compute effective sample size:

$N_E = (2*N) / (1+r)$

intraclass correlation

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

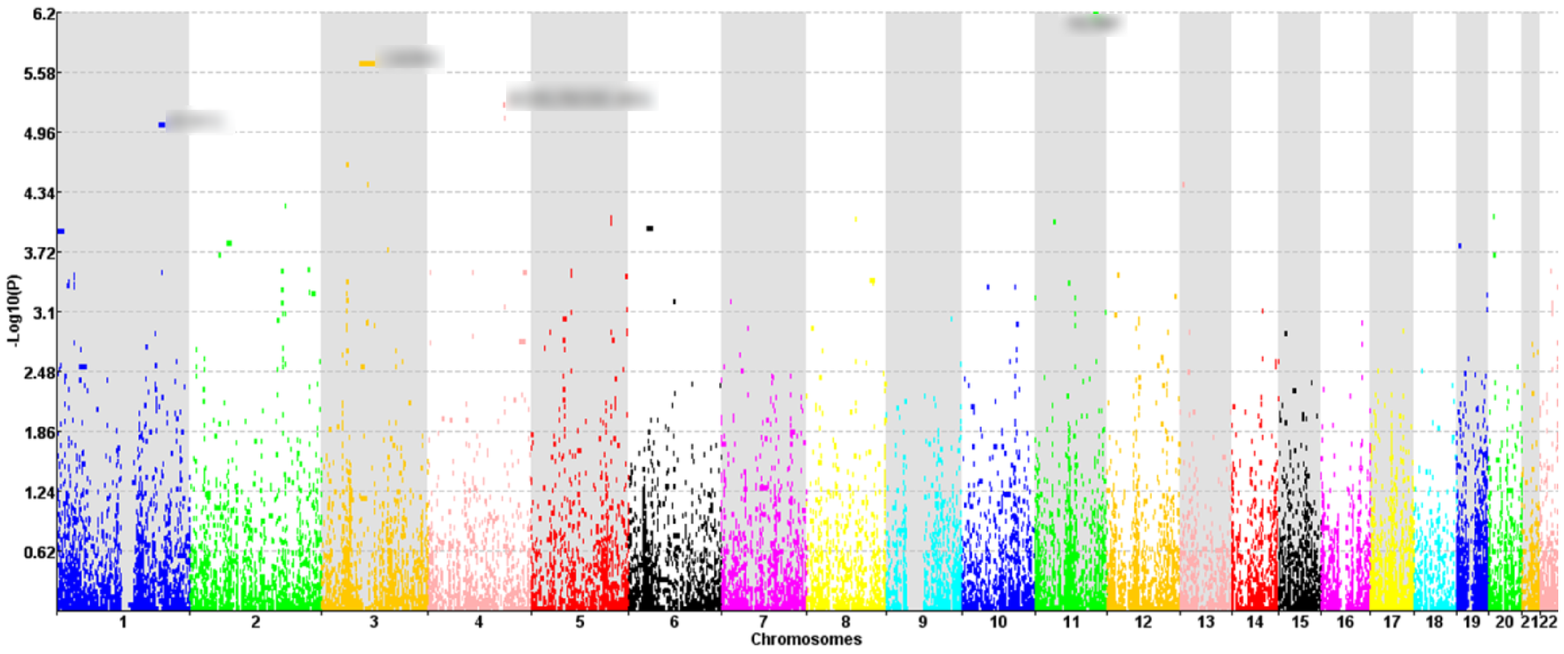
MZ pairs ~~or MZ singletons?~~



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.