



THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA

CREATE CHANGE

Post-GWAS: Gene-based test

Genetics and Genomics Winter School

Ang Li

20 June 2023



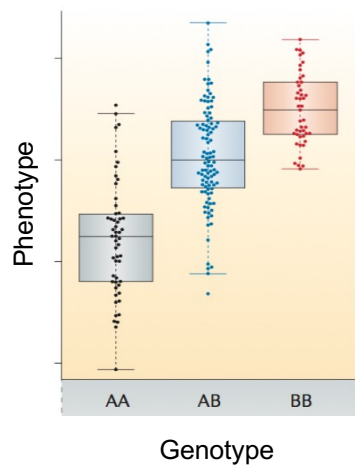
[@Ang_Li_19](https://twitter.com/Ang_Li_19)



Outline

1. Why is it important to conduct gene-based tests?
2. What are the state-of-the-art methods?
3. How to perform gene-based tests?

Let's do a genome-wide genetic association for a disease

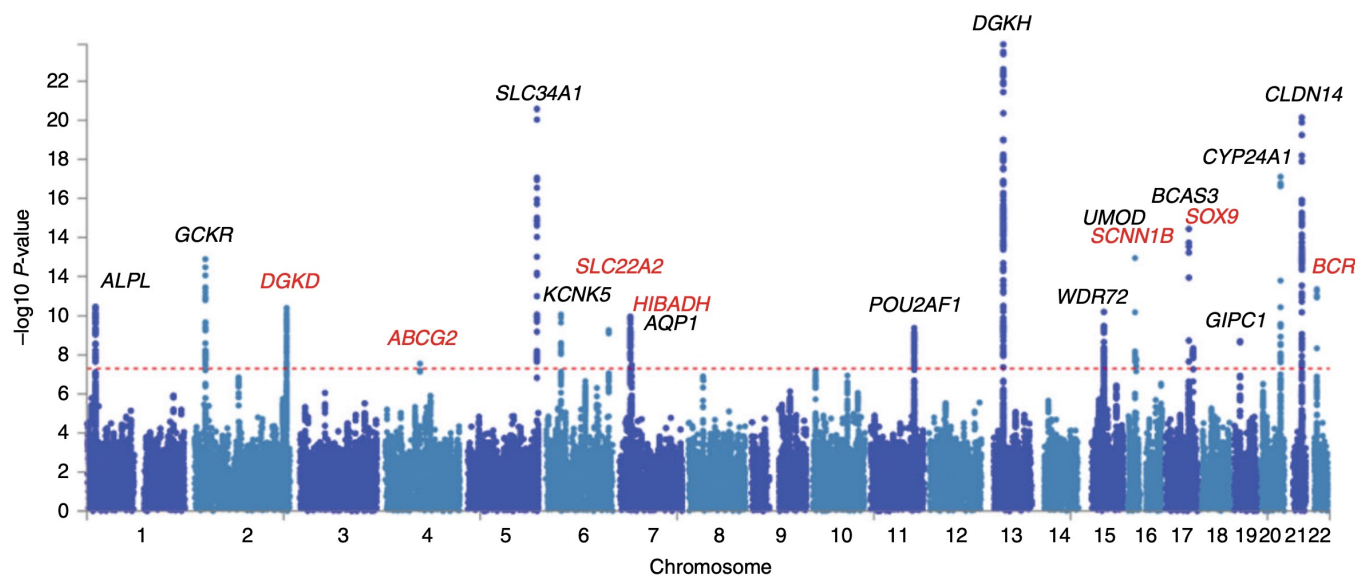


Single SNP-based test

Summary statistics

1. b , SNP effect estimate
2. p , the significance of association per SNP

Manhattan plot

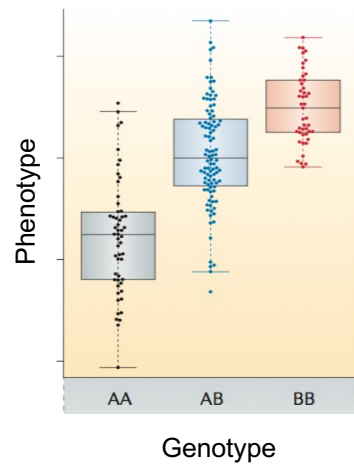


SNP	A1	A2	freq	BETA	SE	P	N
rs3949130	A	C	0.0761272	-8.62939	4.45554	0.0527723	34311
rs200167968	C	G	0.0753923	-7.64071	4.47458	0.0877138	34347
rs79847867	A	C	0.0780728	-7.96104	4.36084	0.0679142	34993
rs200058026	A	G	0.0692703	-9.70190	4.64611	0.0367816	34286
rs131523	A	G	0.3089400	-1.88391	2.54537	0.4592180	34442
rs185518626	T	C	0.0122513	-2.84866	10.82540	0.7924390	33425

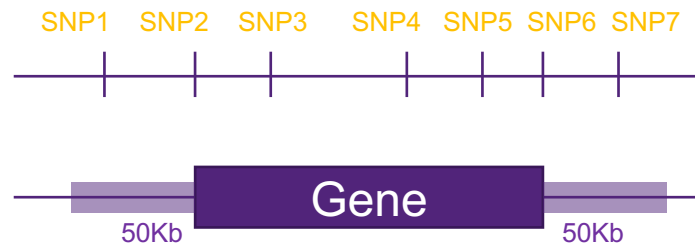
What's the gene underlying a "tower" of SNP association signals?

Why is it important to conduct gene-based testing?

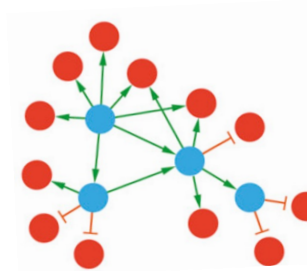
Single SNP-based test



Gene-based test



Reduce test burden



Assessing Functional Relevance

Integrating Additional Data

What gene-based tests that are currently available?

For common variants (minor allele frequency or MAF > 0.01):

VEGAS¹, MAGMA², fastBAT³, Pascal-sum⁴

1. Liu et al., 2010
2. de Leeuw et al., 2015
3. Bakshi et al., 2016
4. Lamperter et al., 2016

“Weighted-sum” strategy

- Compute the test statistic as the weighted sum of χ^2 variables
- P-value is not analytically available
- Permutation (VEGAS), simulation (MAGMA), approximation (fastBAT, Pascal-sum)

$$\text{Test statistic} = \sum_i z_i^2 = \sum_i \lambda_i \chi_1^2$$

(λ_i is the eigenvalues of \mathbf{R})

Limitation of the “weighted-sum” strategy

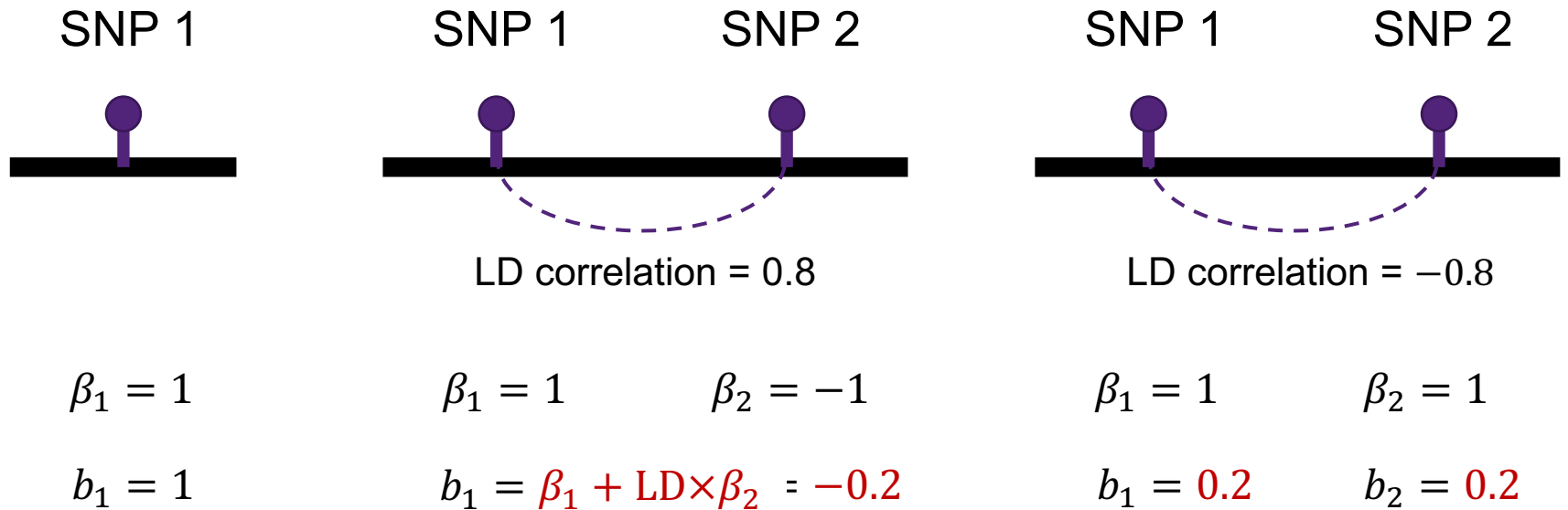
- Do not account for the signs of SNP effects and LD
- Loss of power when there exist masking effects



The larger the weight, the more influential the variant is in contributing to the overall trait or disease.

What is “masking effect”?

The marginal effect of a SNP can be masked by another SNP in linkage disequilibrium (LD) that has an opposite effect size.



The masking effect occurs when the product of the true effects and LD correlation is negative.

mBAT (Multivariate set-Based Association Test)

Tests whether the variance explained by the SNPs in a gene is zero.

$$\text{Variance explained} = \boldsymbol{\beta}'\mathbf{R}\boldsymbol{\beta} = \mathbf{b}'\mathbf{R}^{-1}\mathbf{b} \rightarrow \mathbf{z}'\mathbf{R}^{-1}\mathbf{z} = T_{\text{mBAT}} \sim \chi_m^2 \text{ under the null}$$

- Only requires GWAS summary statistics (z-scores) and LD correlation matrix \mathbf{R} .
- Account for the directions of both effect sizes and LD correlations between SNPs.

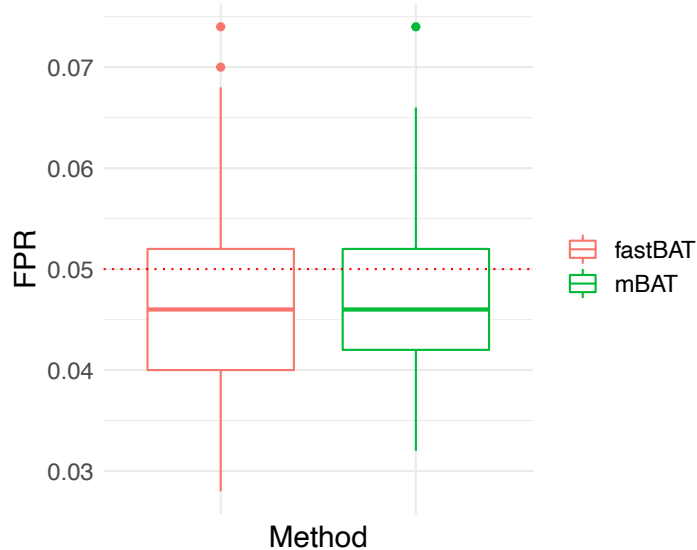
In contrast, methods using “weighted-sum” strategy (VEGAS, MAGMA, and fastBAT) have

$$\text{Test statistic} = \sum_i z_i^2 = \sum_i \lambda_i \chi_1^2 \quad (\lambda_i \text{ is the eigenvalues of } \mathbf{R})$$

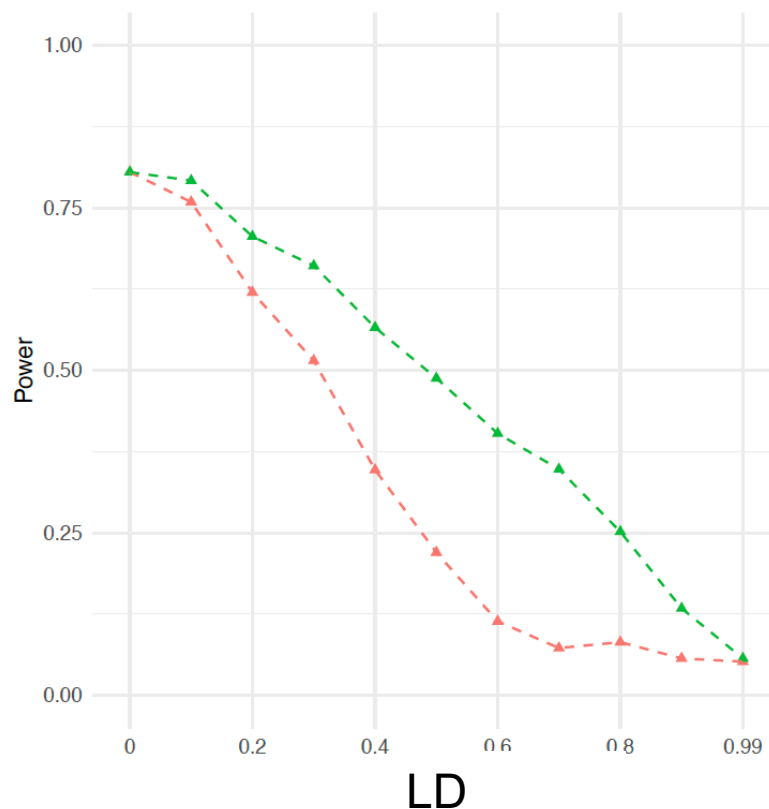
where the directions of effect sizes are ignored.

Proof-of-concept simulation

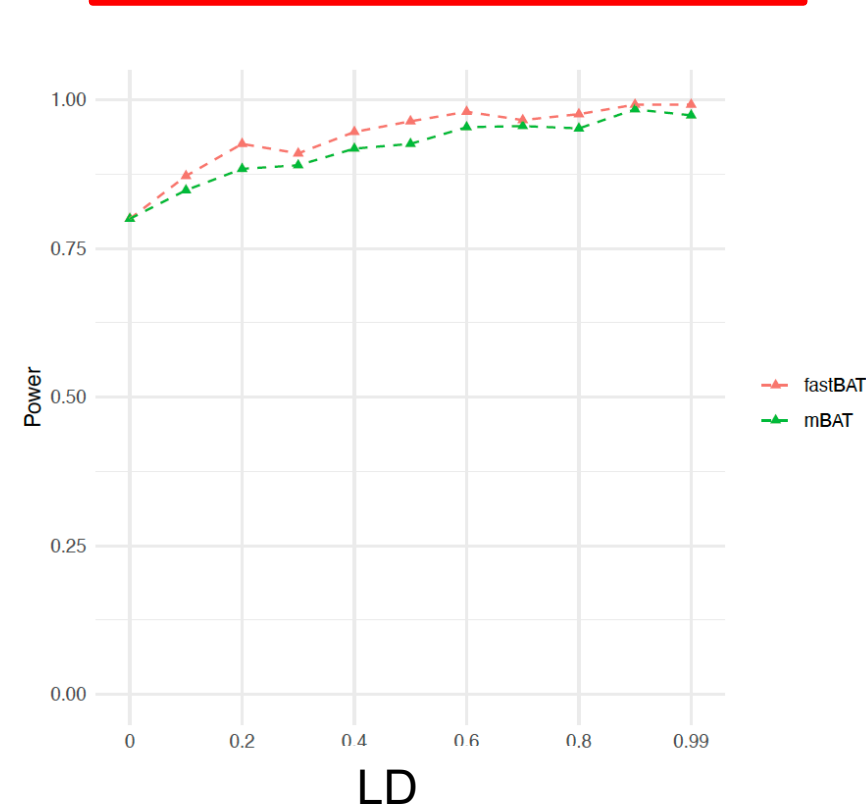
2 SNPs with simulated genotypes and phenotypes



$$(\beta_1 = 1, \beta_2 = -1)$$

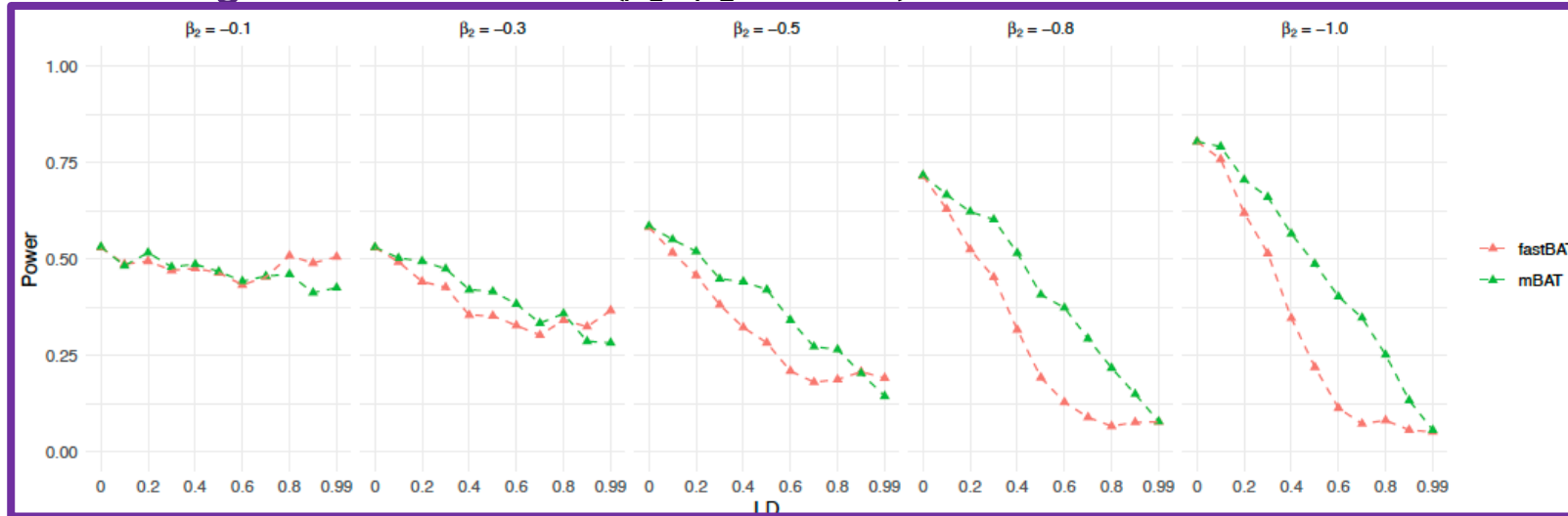


$$(\beta_1 = 1, \beta_2 = 1)$$

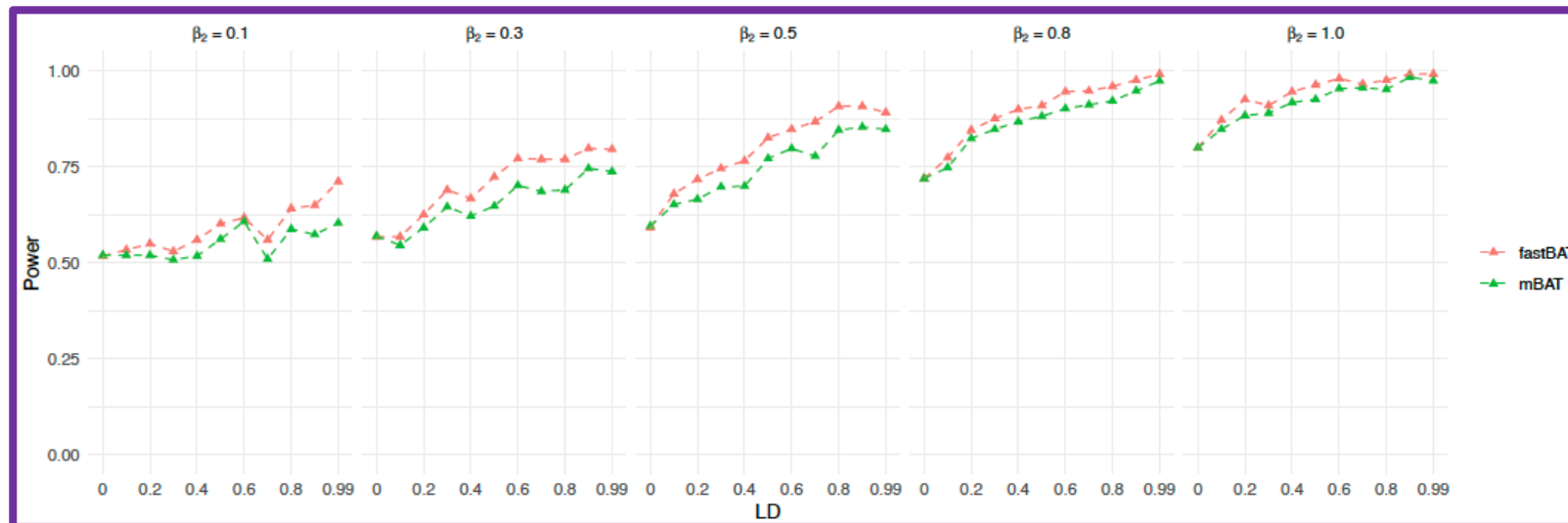


Proof-of-concept simulation

Masking effect scenario ($\beta_1 \times \beta_2 \times LD < 0$)

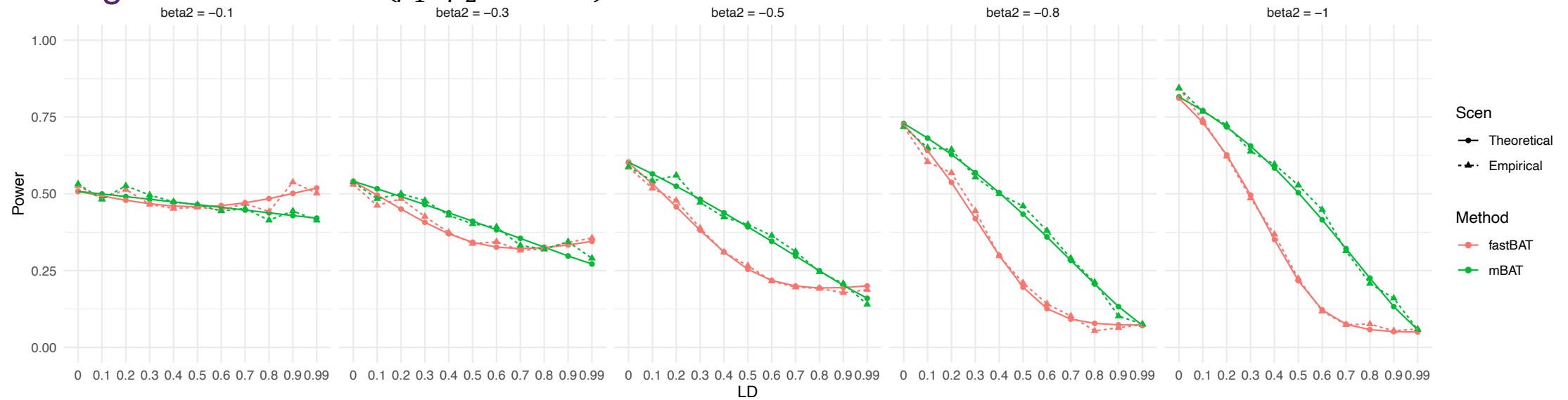


Non-masking effect scenario ($\beta_1 \times \beta_2 \times LD > 0$)

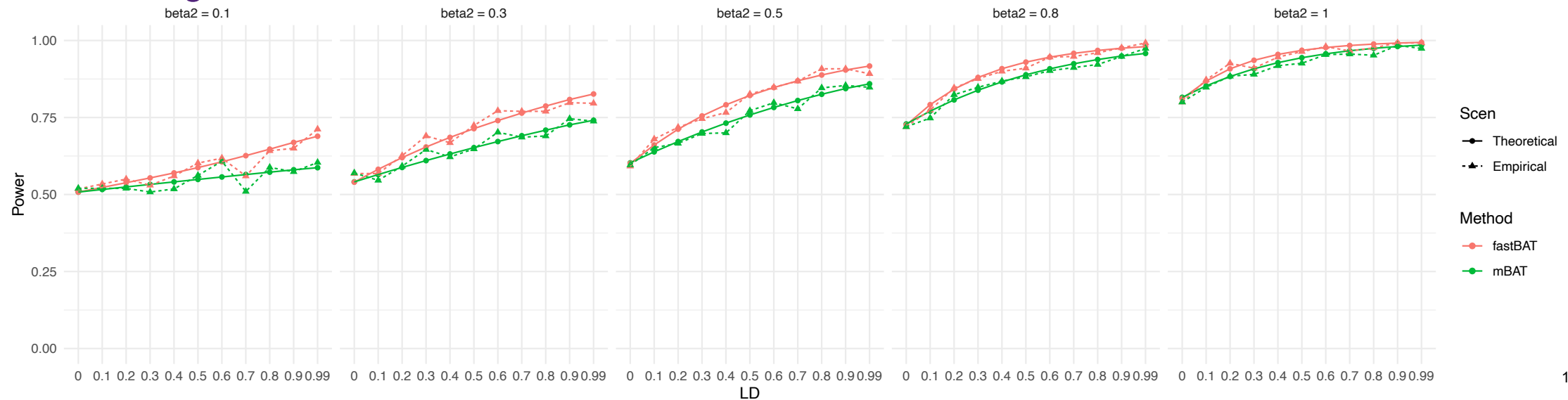


Proof-of-concept simulation

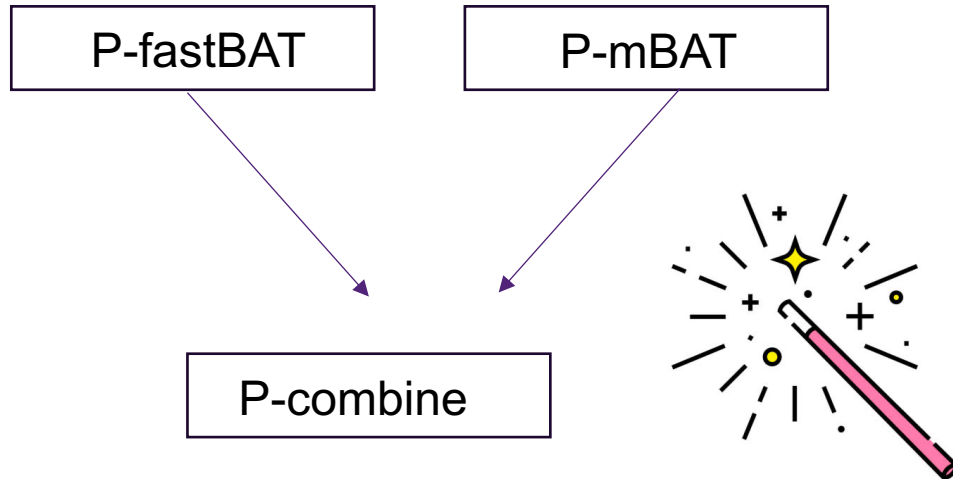
Masking effect scenario ($\beta_1 \times \beta_2 \times LD < 0$)



Non-masking effect scenario ($\beta_1 \times \beta_2 \times LD > 0$)



How to maximize the power combining two strategies?

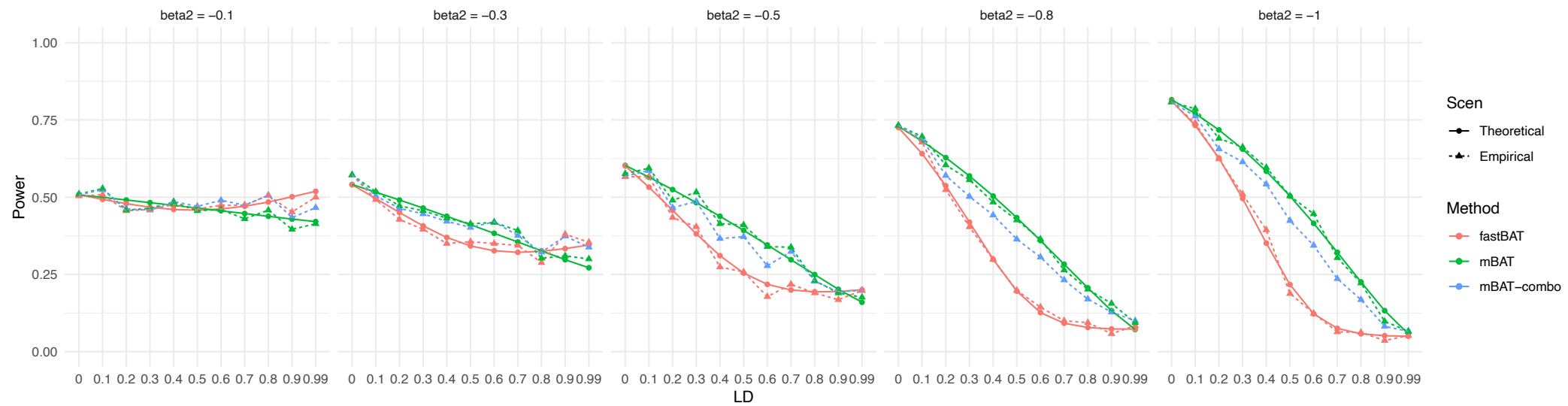


**Cauchy Combination Test:
A Powerful Test With Analytic p -Value Calculation
Under Arbitrary Dependency Structures**

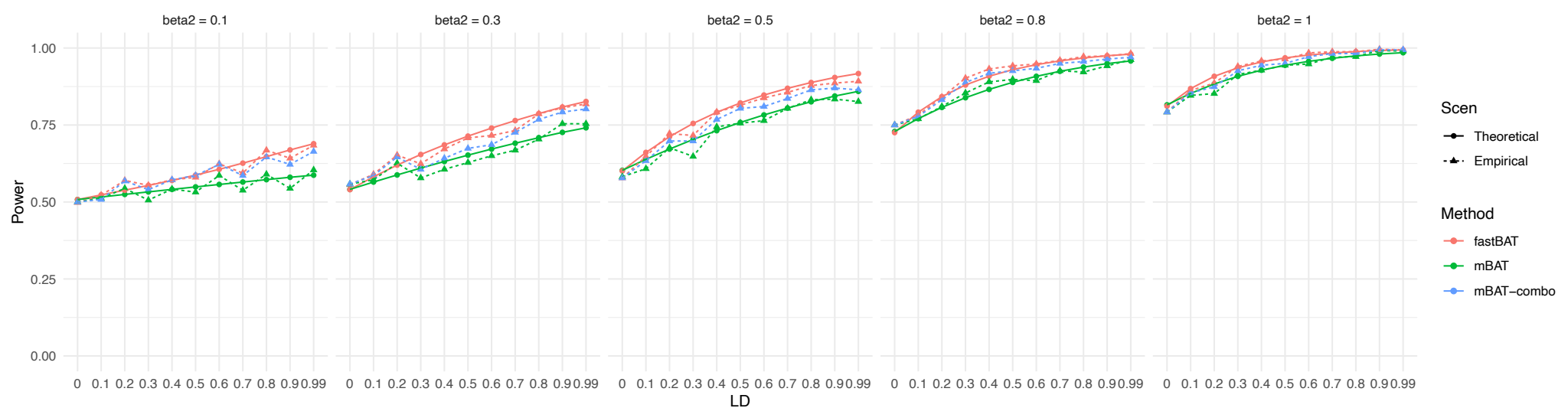
Liu, Y., & Xie, J. (2019).

Cauchy combination method produces a single P-value from multiple tests, no need to know the correlation structure between test statistics, owing to the statistical properties of Cauchy distribution.

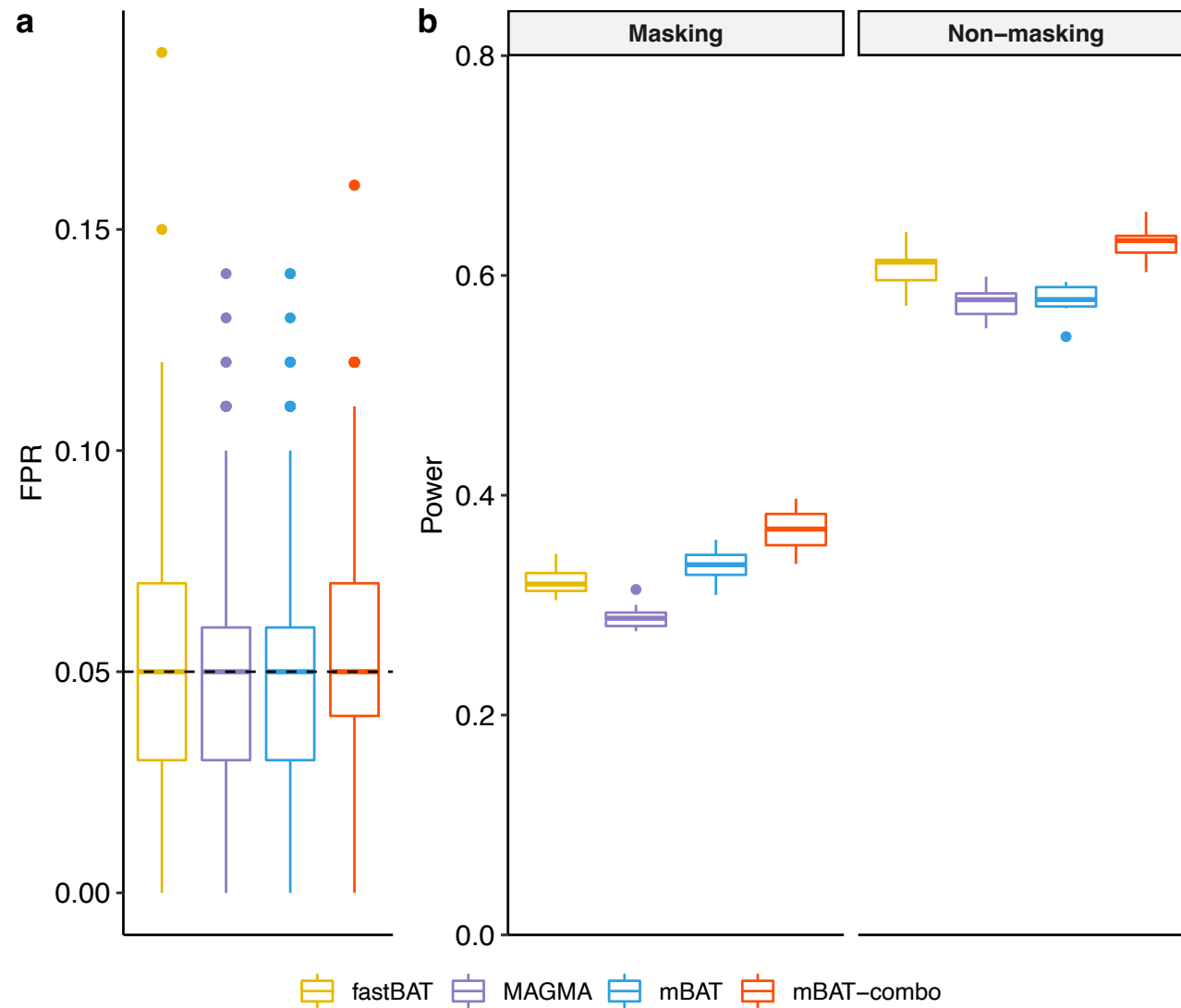
Masking effect scenario ($\beta_1 \times \beta_2 \times LD < 0$)



Non-masking effect scenario ($\beta_1 \times \beta_2 \times LD > 0$)

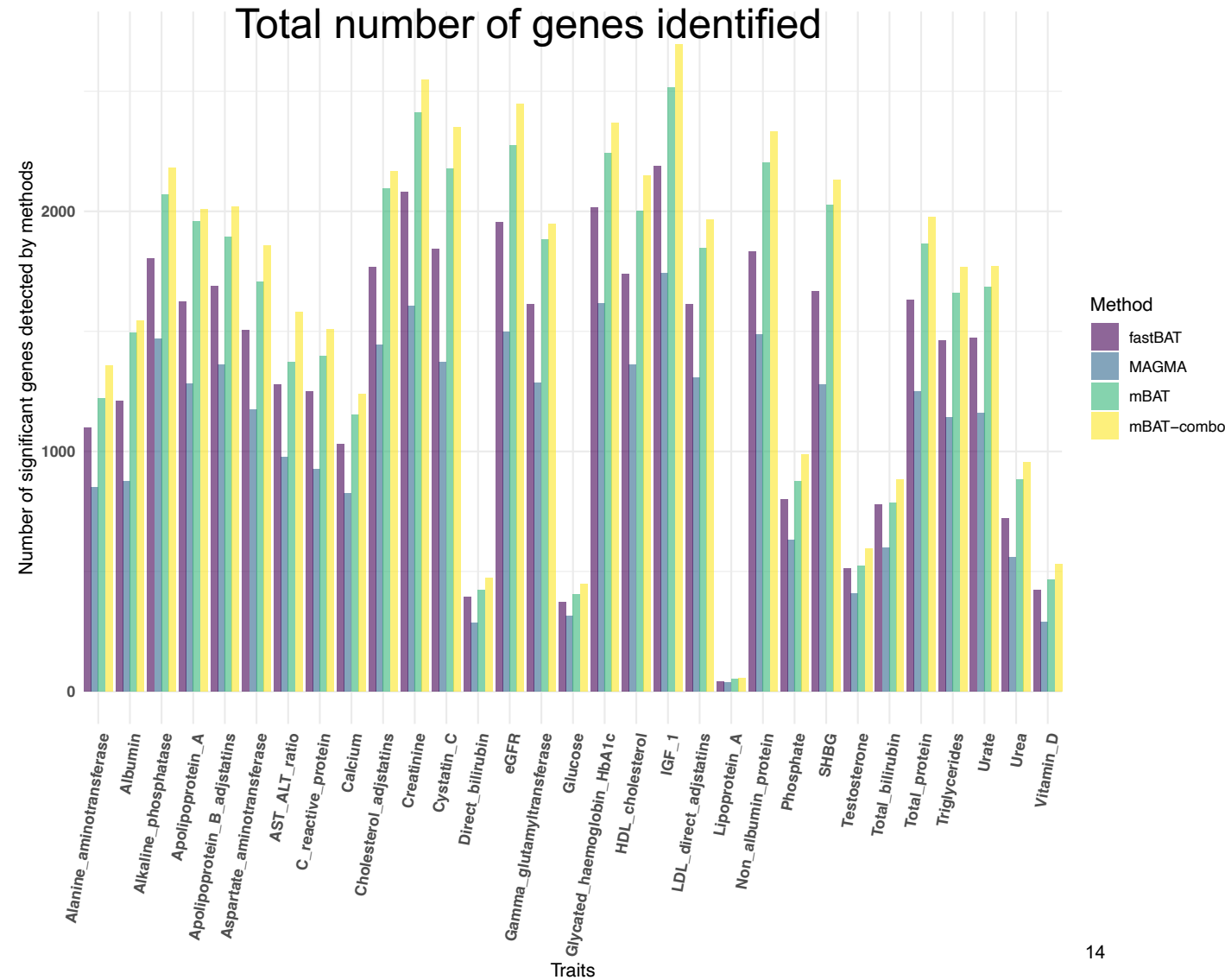


Simulation based on real genotypes



Application to 35 blood and urine metabolite traits

On average across traits, mBAT-combo (mBAT) identified 19.7% (11.5%) more genes than fastBAT and 56.9% (43.9%) more than MAGMA.

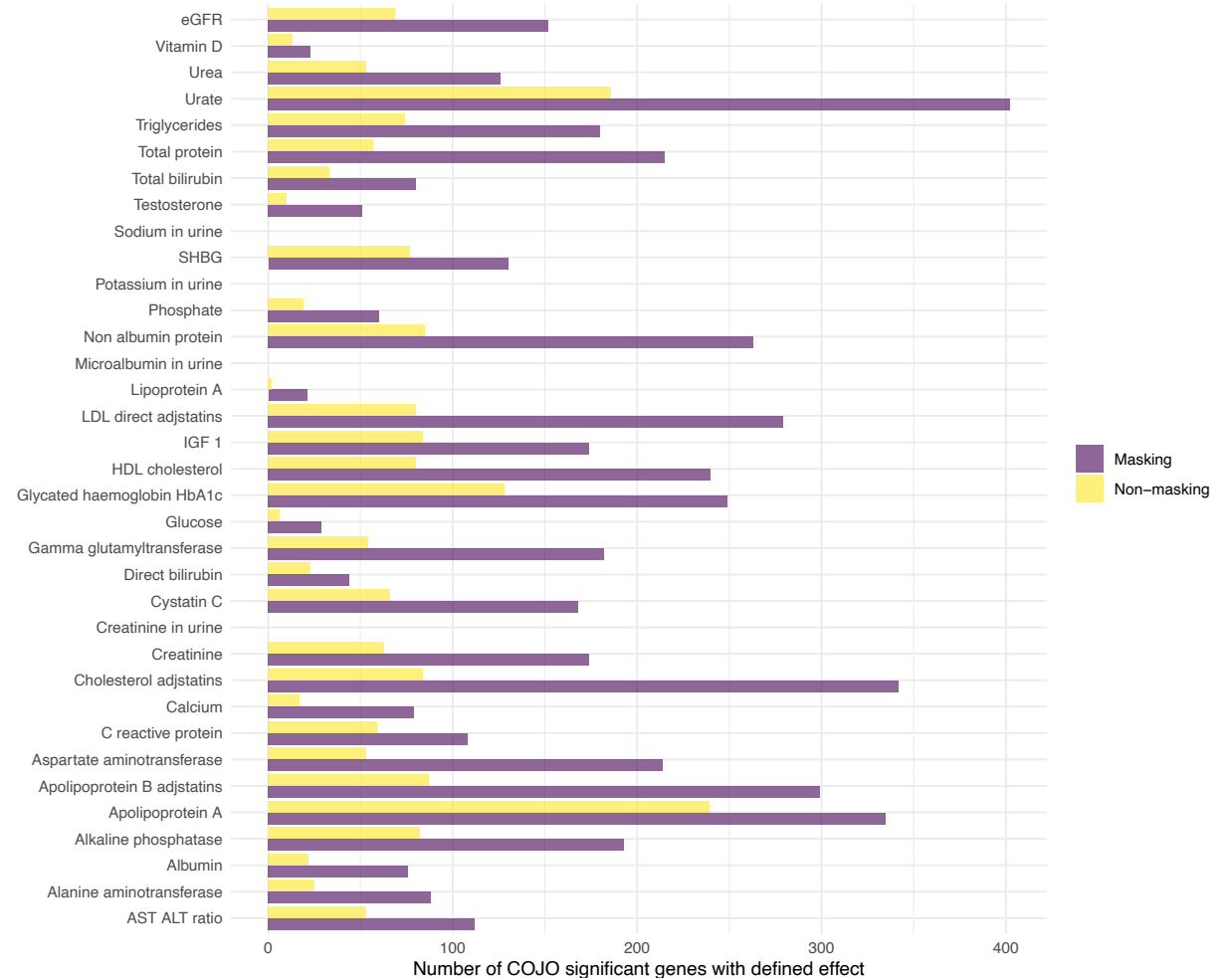
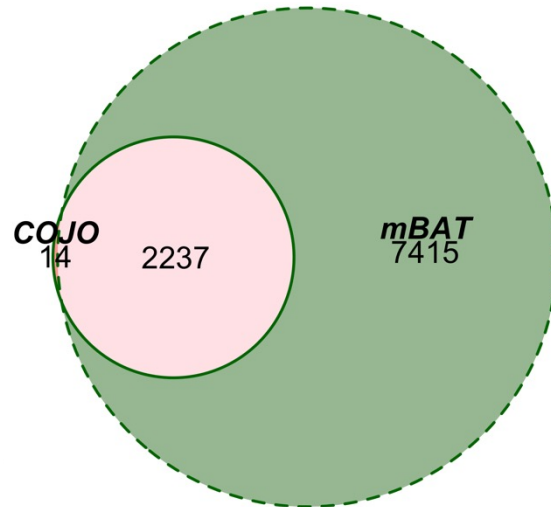


Masking is common if not ubiquitous

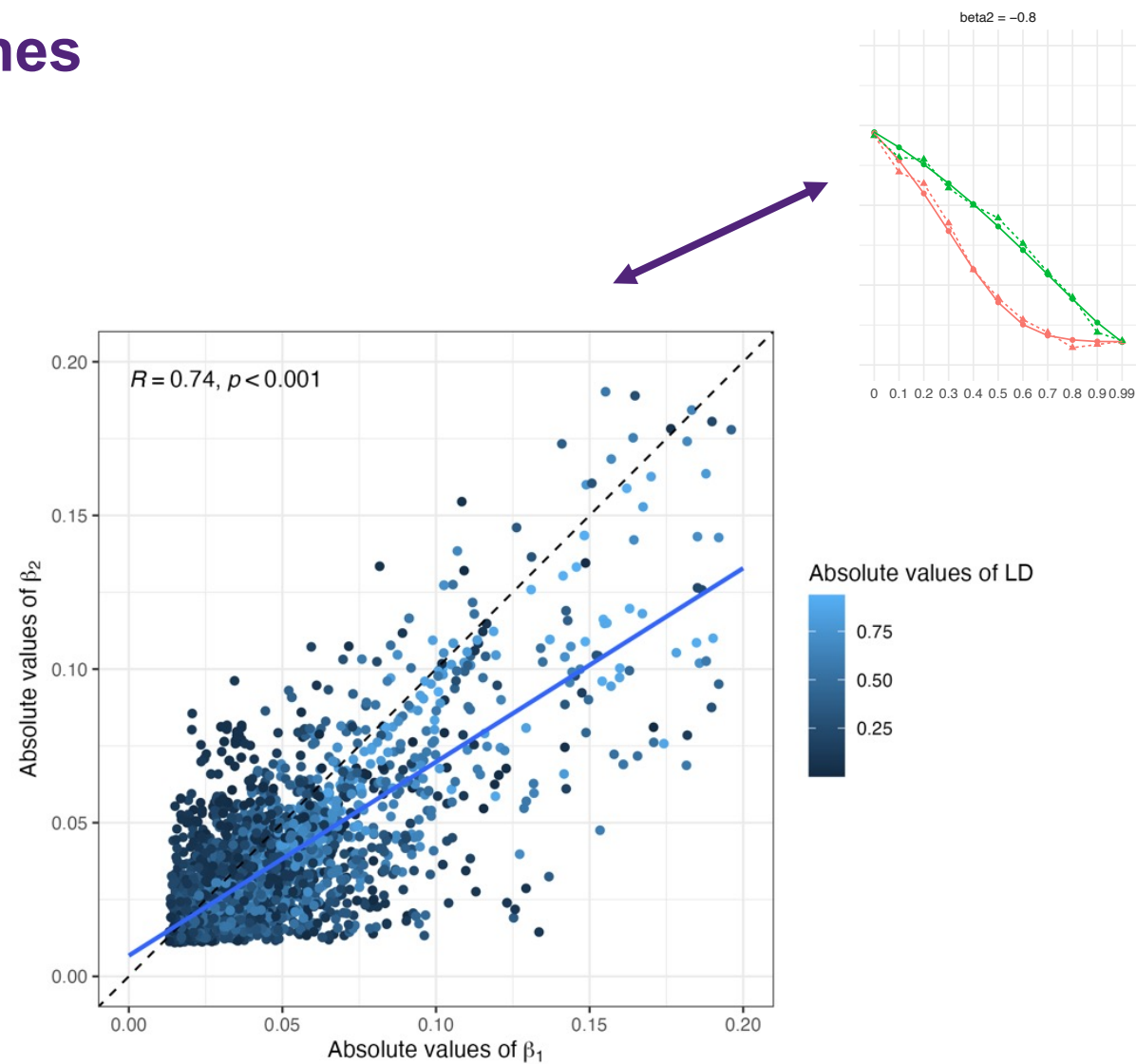
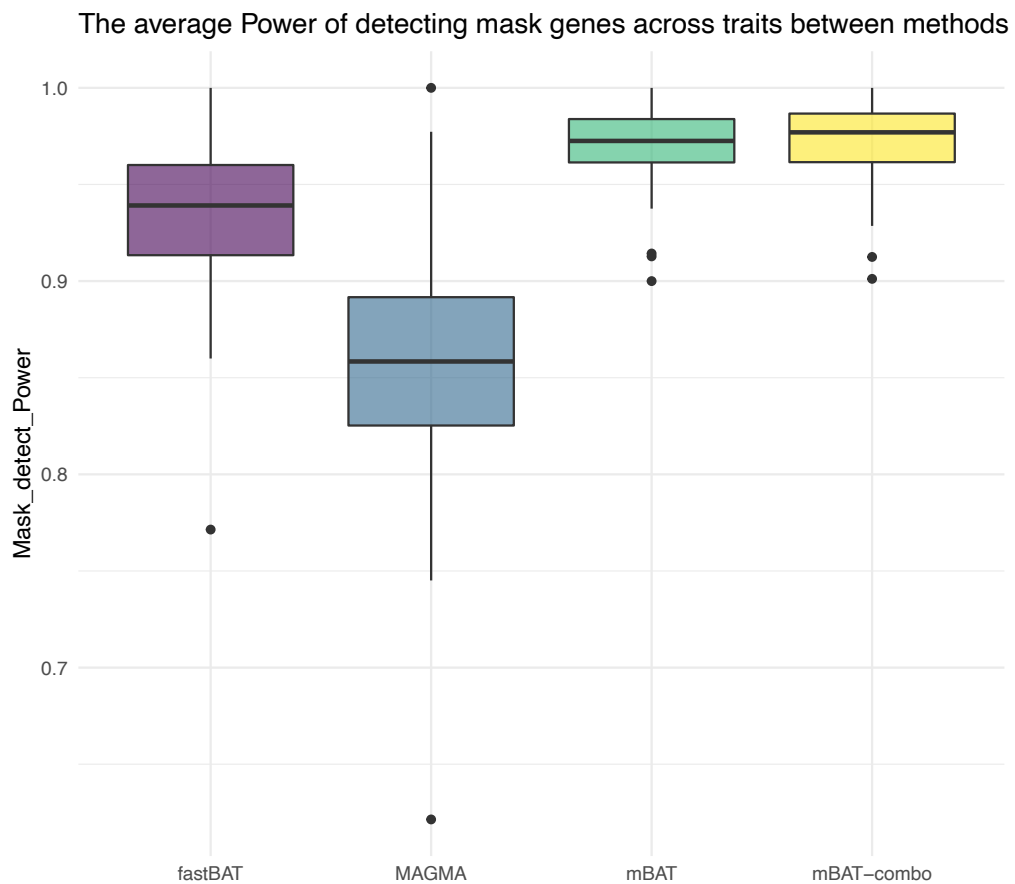
Definition of Masking Genes:

- At least 2 SNPs with joint-effect $P_{COJO} < 5 \times 10^{-6}$
- For the top 2 SNPs: $\hat{\beta}_1 \times LD \times \hat{\beta}_2 < 0$

99.4% COJO masking genes could be detected by mBAT.



High power for identifying masking genes



mBAT-significant genes with evidence of masking effects from COJO

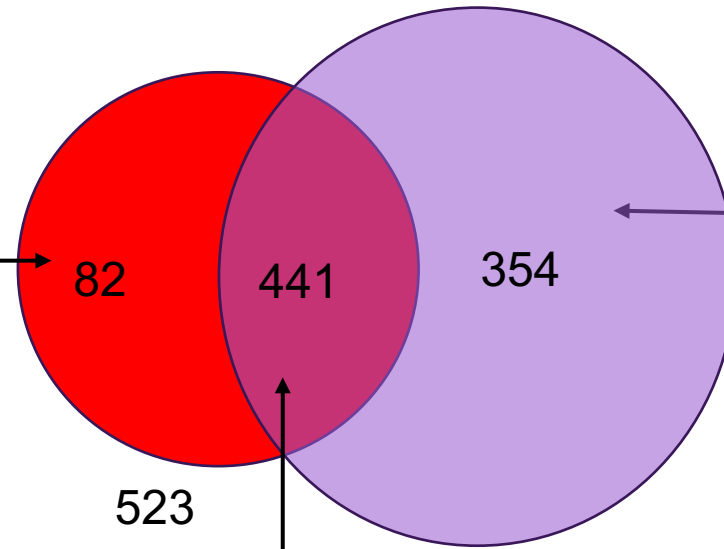
Validate mBAT-combo novel genes in larger GWAS

Schizophrenia (SCZ)

Novel genes detected by mBAT-combo in PGC-SCZ2 GWAS

Significant in mBAT-combo but no GWS SNP in the gene \pm 1Mb

$N_{\text{effective}} = 57\text{K}$



Significant genes in PGC-SCZ3 GWAS

No GWS hit in the gene \pm 1Mb in PGC2 & at least one GWS hit in the gene \pm 50kb

$N_{\text{effective}} = 73\text{K}$

87% mBAT-combo

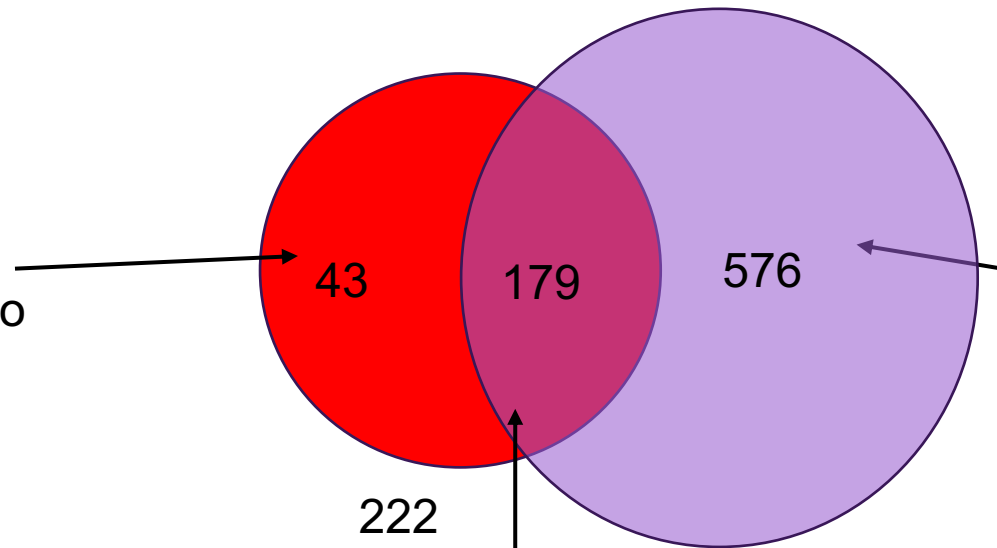
Genes successfully replicated

Validate mBAT-combo novel genes in larger GWAS

Schizophrenia (SCZ)

Novel genes detected by **mBAT-combo** missed by fastBAT/MAGMA in PGC-SCZ2
 Significant in mBAT-combo but no GWAS hit in the gene $\pm 1\text{Mb}$

$N_{\text{effective}} = 57\text{K}$



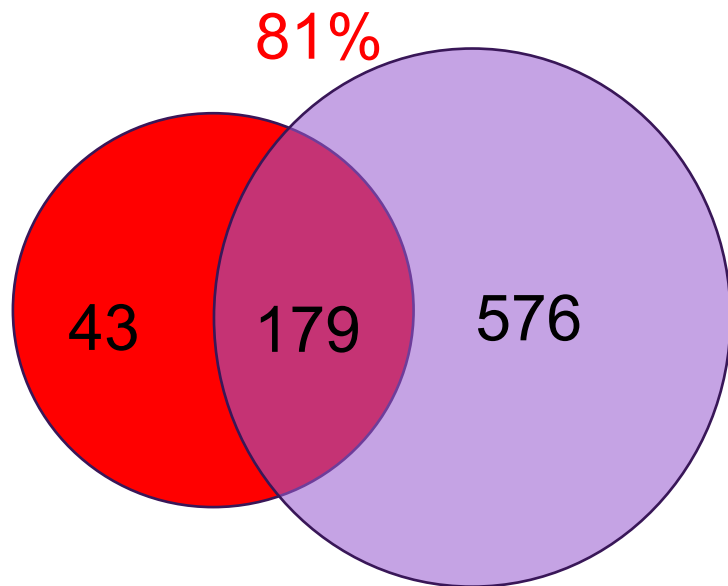
Significant genes in PGC-SCZ3 GWAS
 No GWS hit in the gene $\pm 1\text{Mb}$ in PGC2 & at least one GWS hit in the gene $\pm 50\text{kb}$ region

$N_{\text{effective}} = 73\text{K}$

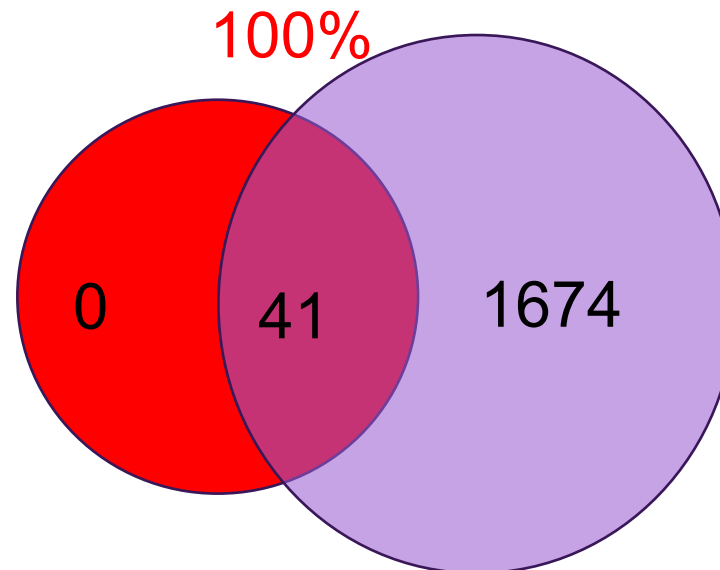
81% mBAT-combo
 Genes successfully replicated

Validate mBAT-combo novel genes in larger GWAS

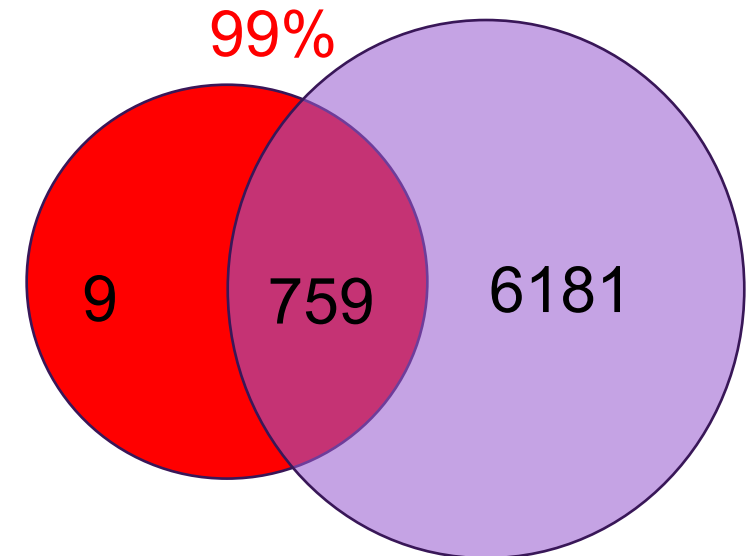
a) Schizophrenia:
N: 1.3-fold



b) BMI:
N: 1.9-fold



c) Height:
N: 2.0-fold



Novel genes detected by **mBAT-combo only** but missed by fastBAT/MAGMA in smaller GWAS

How can we perform gene-based tests?



AJHG



Volume 110, Issue 1, 5 January 2023, Pages 30-43

Article

mBAT-combo: A more powerful test to detect gene-trait associations from GWAS data

Ang Li¹, Shouye Liu¹, Andrew Bakshi², Longda Jiang³, Wenhan Chen⁴, Zhili Zheng¹, Patrick F. Sullivan^{5,6}, Peter M. Visscher¹, Naomi R. Wray^{1,7}, Jian Yang^{8,9}, Jian Zeng¹  

In Practical:

1. Run mBAT-combo in human height (Wood et al data)
2. Compare power to the standard GWAS approach
3. Detect gene with masking effects

GCTA

a tool for Genome-wide Complex Trait Analysis

GCTA
SMR
GSMR
OSCA

- [Overview](#)
- [Download](#)
- [FAQ](#)
- [Basic Options](#)
- [GREML](#)
- [GWAS Analysis](#)
- [GWAS Simulation](#)
- [Conditional Analysis](#)
- [Gene-based Test](#)
- fastBAT
- mBAT-combo
- fastGWA-BB
- ACAT-V
- [Mendelian Randomisation](#)
- [Genomic Risk Prediction](#)
- [Linkage Disequilibrium](#)
- [Population Genetics](#)
- [Data Resource](#)

mBAT-combo

GCTA-mBAT-combo: a powerful method to c association study (GWAS) summary data.

This method performs a set-based associati individual-level genotypes. Please see Li et al

Note: most other GCTA options are also valid

Examples

Gene-based test

```
# Output mBAT-combo p-values:
gcta64 --bfile test --mBAT-combo test.ma
```

mBAT-combo performs both mBAT and fastBAT effects, because mBAT test is more powerful

```
# Output mBAT-combo, mBAT and fastBAT p-values:
gcta64 --bfile test --mBAT-combo test.ma
```

The testing data can be downloaded [here](#).

Thanks

Dr. Jian Zeng

Prof. Naomi Wray

Prof. Jian Yang

All co-authors of mBAT-combo.

