Multivariate MAPIT

Leveraging the genetic correlation between traits improves the detection of epistasis in genome-wide association studies

Julian Stamp
Center for Computational Molecular Biology
Brown University

CCMB

BROWN

Outline

Introduction

Methods

Simulated Data

Real Data

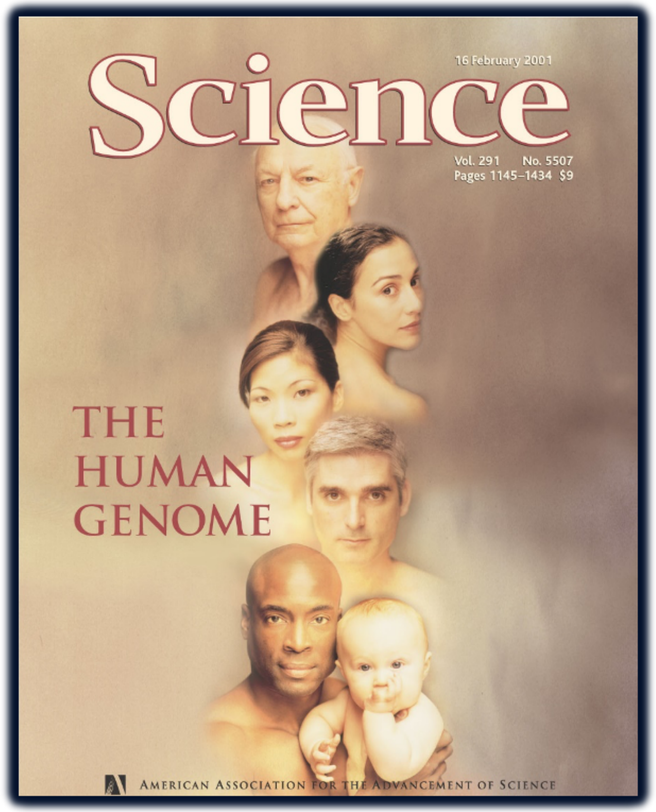
Summary

Genome Wide Association Studies

- Genotype large cohorts
- Map traits through statistical tests

20 years later & many open problems:

- Rare variants
- Inaccurate predictions
- Missing heritability



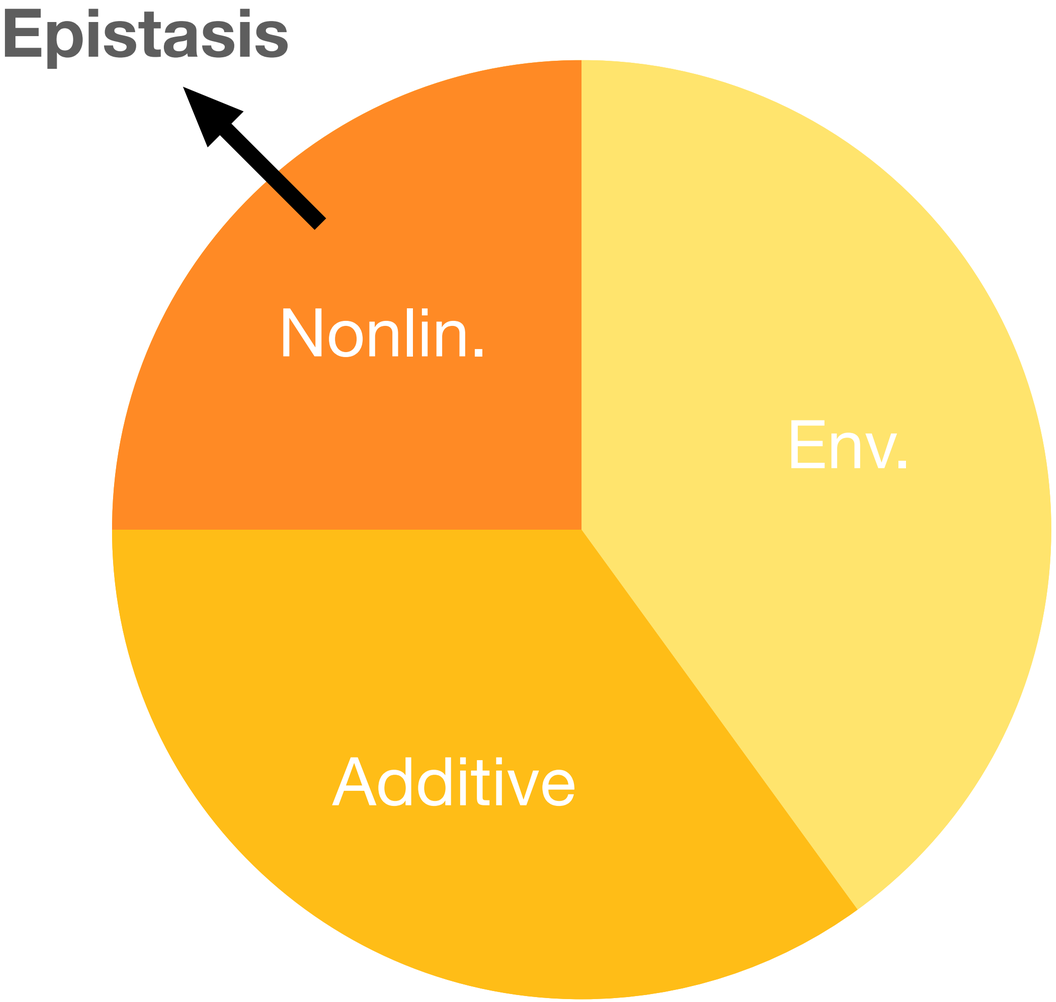
Phenotypic Variance

Genetic & Environmental Factors

$$P = G + E$$

Broad sense Heritability

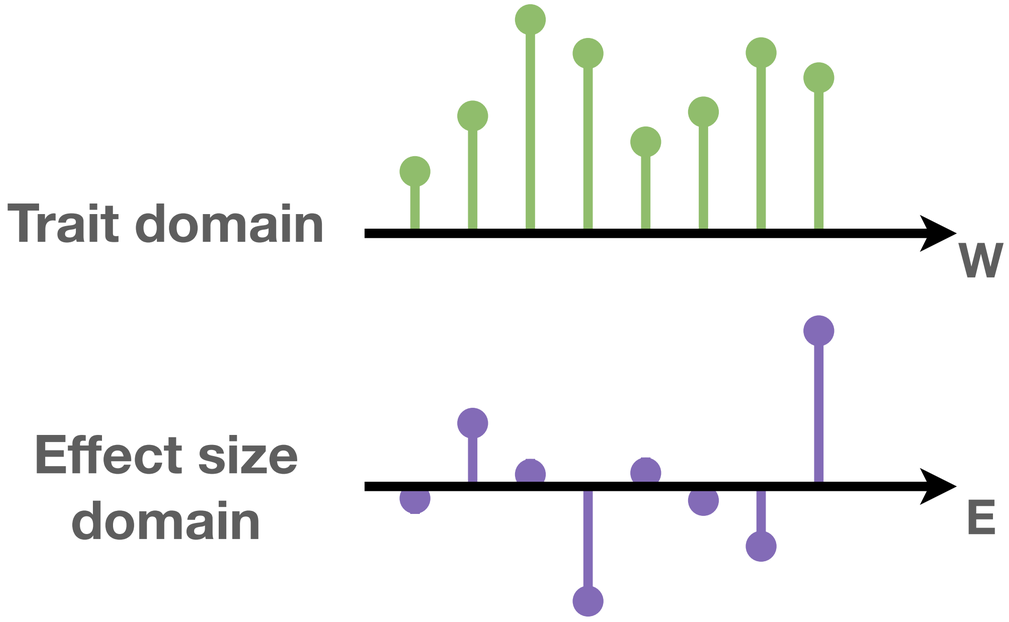
$$H^2 = \frac{\mathrm{Var}[G]}{\mathrm{Var}[P]}$$



Hadamard-Walsh Transform

- For combinatorially complete data transform linearly from trait to epistatic effects
- Rows are mapped to genotypes

$$\frac{1}{2^L}\Psi\vec{W} = \vec{E}_W$$



$$\frac{1}{2^L} \cdot \underbrace{\begin{pmatrix} +1 & +1 & +1 & \cdots & +1 \\ +1 & -1 & +1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ +1 & -1 & -1 & \cdots & -1 \end{pmatrix}}_{\text{Hadamard matrix}} \cdot \underbrace{\begin{pmatrix} 1.07 \\ 2.54 \\ \vdots \\ 0.41 \end{pmatrix}}_{\text{Trait values}} = \underbrace{\begin{pmatrix} 2.28 \\ 0.0857 \\ \vdots \\ -0.1051 \end{pmatrix}}_{\text{Effect size of interactions}}$$

Biobank Scale Data

- ~ 10^5 to 10^6 variants
- ~ 10^4 to 10^5 samples
- ~ 3^{100000} genotype combinations*

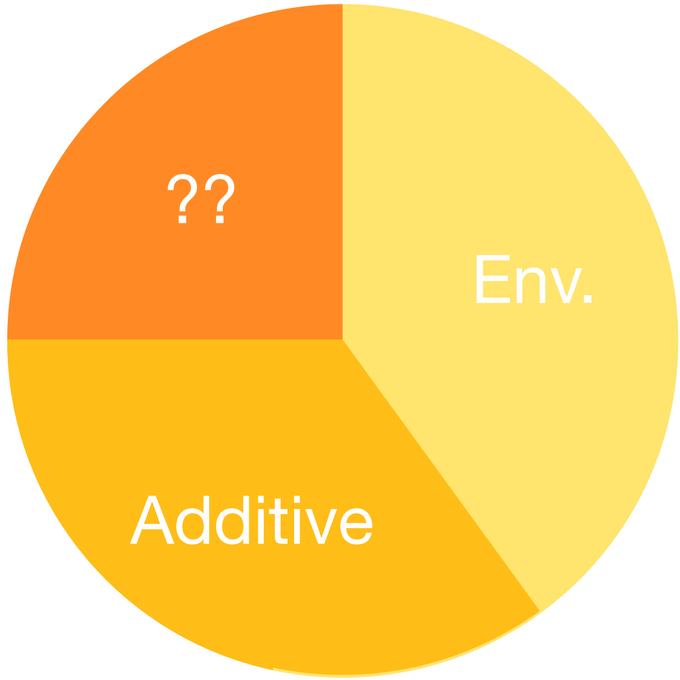
$\Longrightarrow$ underdetermined & combinatorially incomplete

*Number of atoms in the universe: ~ 10^{80}

Motivation

Multivariate approach to studying non-linear contributions in complex traits

- More than 11 million SNPs in human genome[1], ∼400k trait associations[2]
- Majority of the heritability of complex traits “missing”[3]
- Epistasis could explain missing heritability
- Computational methods to detect epistasis are underpowered or computationally resource intensive[4]

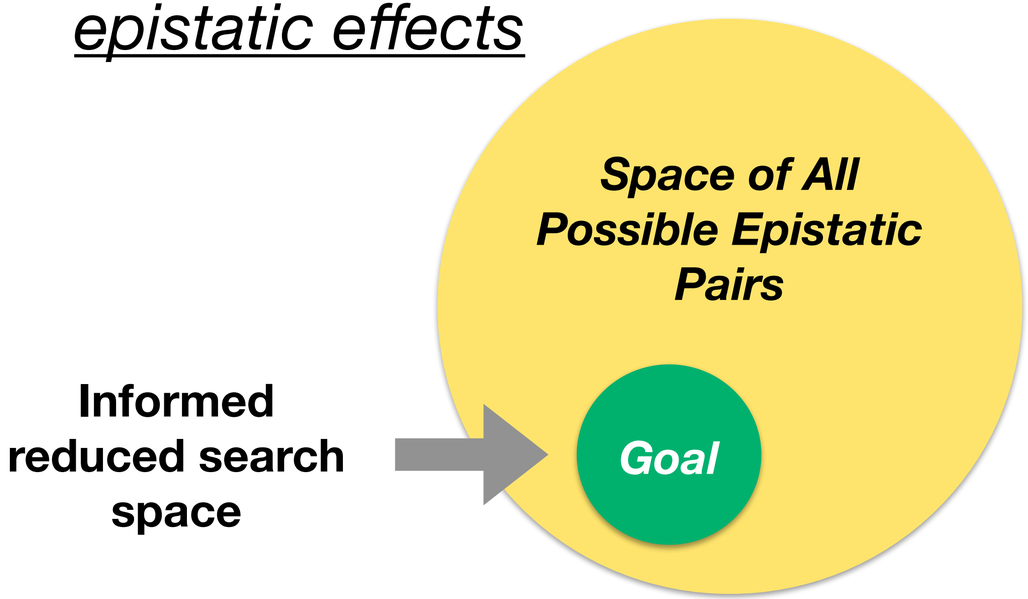


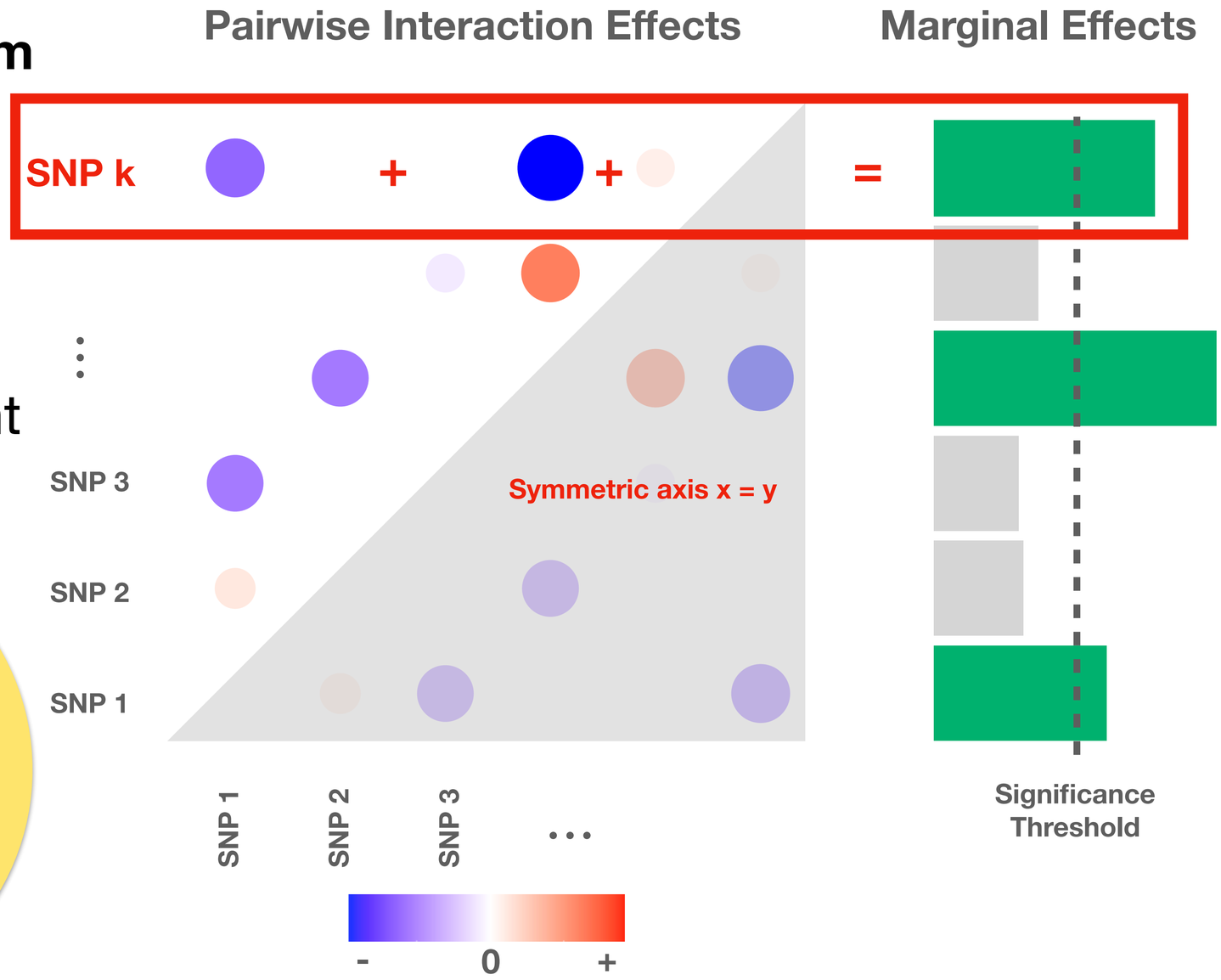
1 Madsen et al. (2007), *Genome Research*
2 Sollis et al. (2022), *Nucleic Acids Research*
3 Young (2019), *PLOS Genetics*
4 Crawford et al. (2017), *PLOS Genetics*

Explicit search space

Epistasis as combinatorial problem

- There are $p(p-1)/2$ possible interacting pairs for p SNPs
- **Idea:** Prioritize search for variant interactions using *marginal epistatic effects*

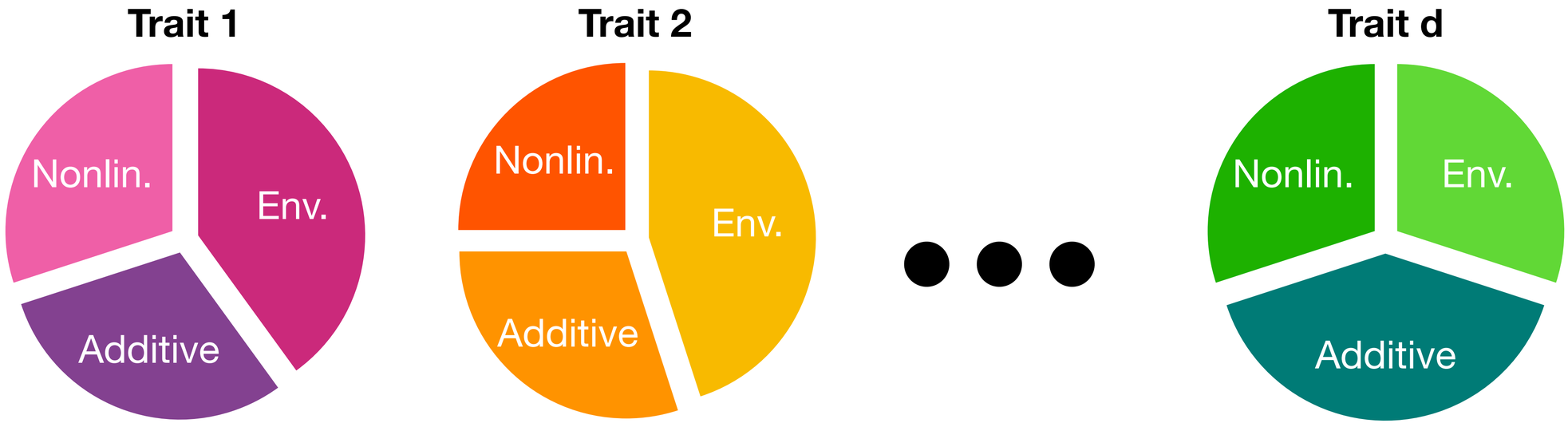




Multivariate LMM

- Genetic correlations between traits maintained by pleiotropy[1]
- Multivariate modelling improves GWAS[2]

⇒ Can we leverage **genetic correlations** to improve detection of epistasis?



1 Chebib and Guillaume (2021), *Genetics*
2 Zhou and Stephens (2014), *Nature*

Approach

Starting point: The Marginal Epistasis Test (MAPIT)

Population mean

Variant vectors

Error

$$\mathbf{y} = \mu + \beta_k \mathbf{x}_k + \sum_{l \neq k} \beta_l \mathbf{x}_l + \sum_{l \neq k} \alpha_l (\mathbf{x}_k \circ \mathbf{x}_l) + \varepsilon, \quad \varepsilon \sim \mathbf{MVN}\left(\mathbf{0}, \tau^2 \mathbf{I}\right)$$

Quantitative Trait

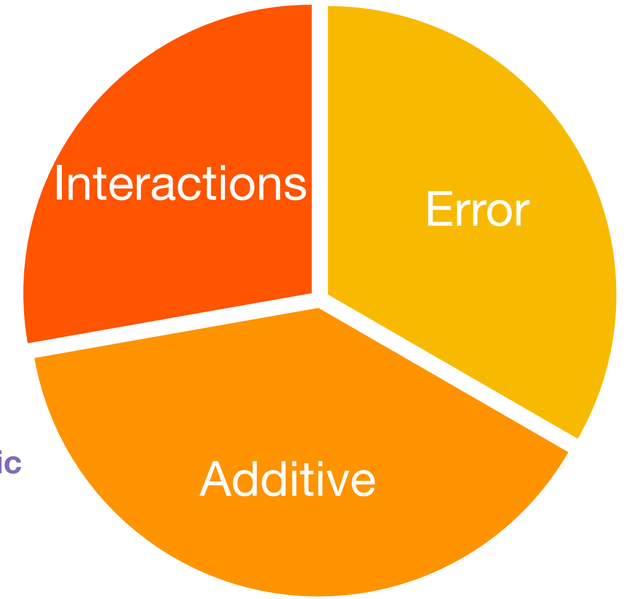
Linear Effect Sizes

Interaction Effect Size

$$= \mu + \beta_k \mathbf{x}_k + \mathbf{m}_k + \mathbf{g}_k + \varepsilon$$

Additive Variance Component

Pariwise Epistatic Variance Component



Crawford et al. (2017), *PLOS Gen*

Approach

Normal assumption for effect size trick for underdetermined data

$$\mathbf{y} = \mu + \beta_k \mathbf{x}_k + \mathbf{m}_k + \mathbf{g}_k + \varepsilon$$

- Genetic Relatedness Matrix
  $\mathbf{K} = \mathbf{X}_{-k}\mathbf{X}_{-k}^T$
- Covariance of the interaction of SNP *k* with it's background
  $\mathbf{G} = \mathbf{D}_k\mathbf{K}\mathbf{D}_k$ with
  $\mathbf{D}_k = \mathrm{diag}(\mathbf{x}_k)$
- Estimate variance parameters jointly using MQS

$$\mathbf{m}_k \sim \mathbf{MVN}\left(\mathbf{0}, \omega^2\mathbf{K}\right)$$

$$\mathbf{g}_k \sim \mathbf{MVN}\left(\mathbf{0}, \sigma^2\mathbf{G}\right)$$

$$\varepsilon \sim \mathbf{MVN}\left(\mathbf{0}, \tau^2\mathbf{I}\right)$$

Crawford et al. (2017), *PLOS Gen*

Approach

Multivariate extension of MAPIT (mvMAPIT)

MAPIT

- One trait $\mathbf{y} = (y_1, \ldots, y_n)^{\mathsf{T}}$
- Only covariance between samples $\mathbf{g}_k \sim \mathbf{MVN}\left(\mathbf{0}, \sigma^2\mathbf{G}\right)$
- Estimate variance components $\hat{\sigma}^2 = \mathbf{y}^T\mathbf{A}_k\mathbf{y}$

Crawford et al. (2017), *PLOS Gen*

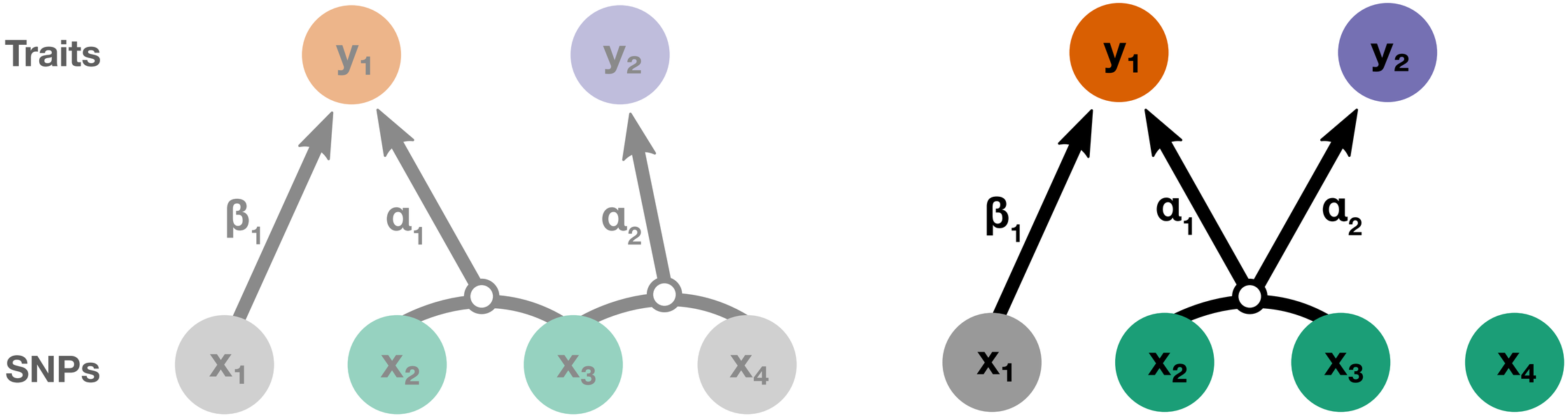
mvMAPIT

- Many traits $\mathbf{Y} = \begin{pmatrix} y_{11} & \cdots & y_{1d} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{nd} \end{pmatrix}$
- Covariance between samples and variance components $\mathbf{g}_k \sim \mathrm{MN}_{n\times d}\left(0, \mathbf{V}_G, \sigma^2\mathbf{G}\right)$
- Estimate d choose 2 variance and covariance components $\hat{\sigma}^2_{12} = \mathbf{y}_1^T\mathbf{A}_k\mathbf{y}_2$

Stamp et al. (2023), *G3*

mvMAPIT

Modelling cross-trait genetic correlations of interaction effects



Stamp et al. (2023), *G3*

MAPIT

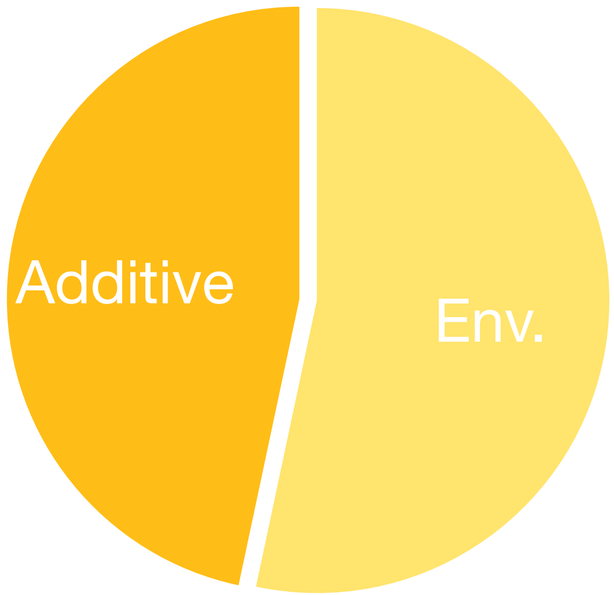
Simulations of complex traits

Scenarios

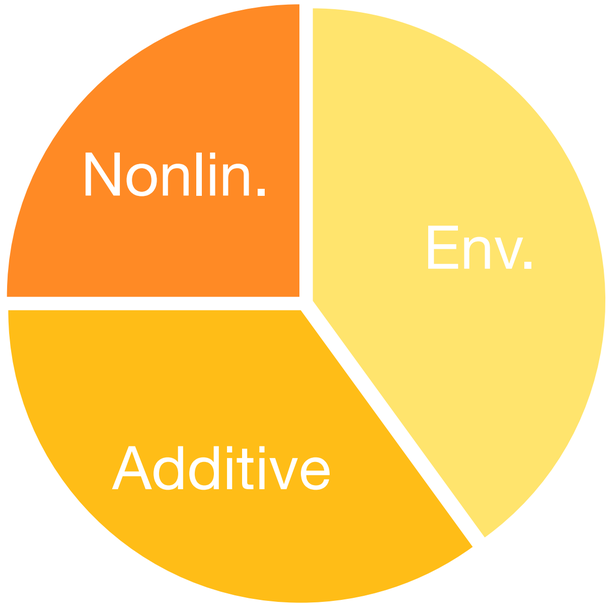
- Null Hypothesis true: no epistasis
- Epistasis with varying parameters

Parameters

- Broad sense heritability H^2
- Proportion of heritable variance due to epistasis $H^2(1-\rho)$



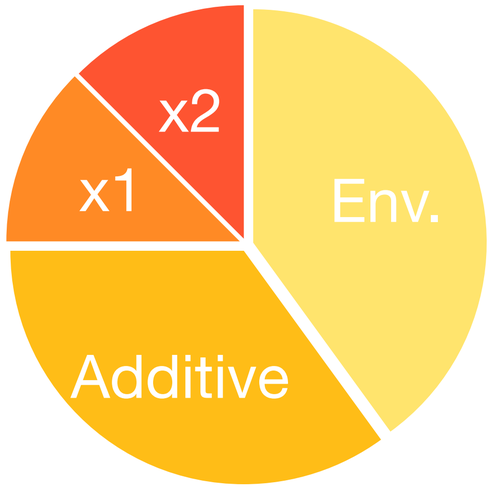
$H_0 : \sigma^2 = 0$



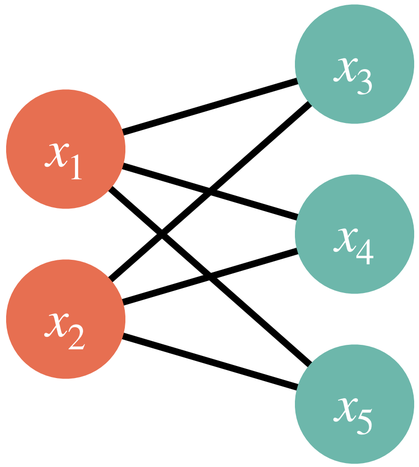
$H_1 : \sigma^2 \neq 0$

MAPIT

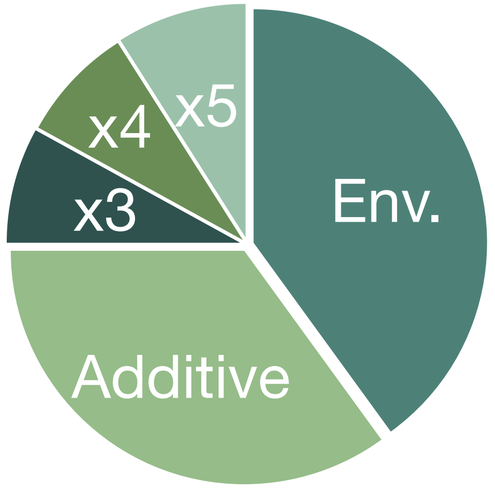
Simulations of complex traits



Group 1



Group 2



Additive SNPs

Epistatic Group 1

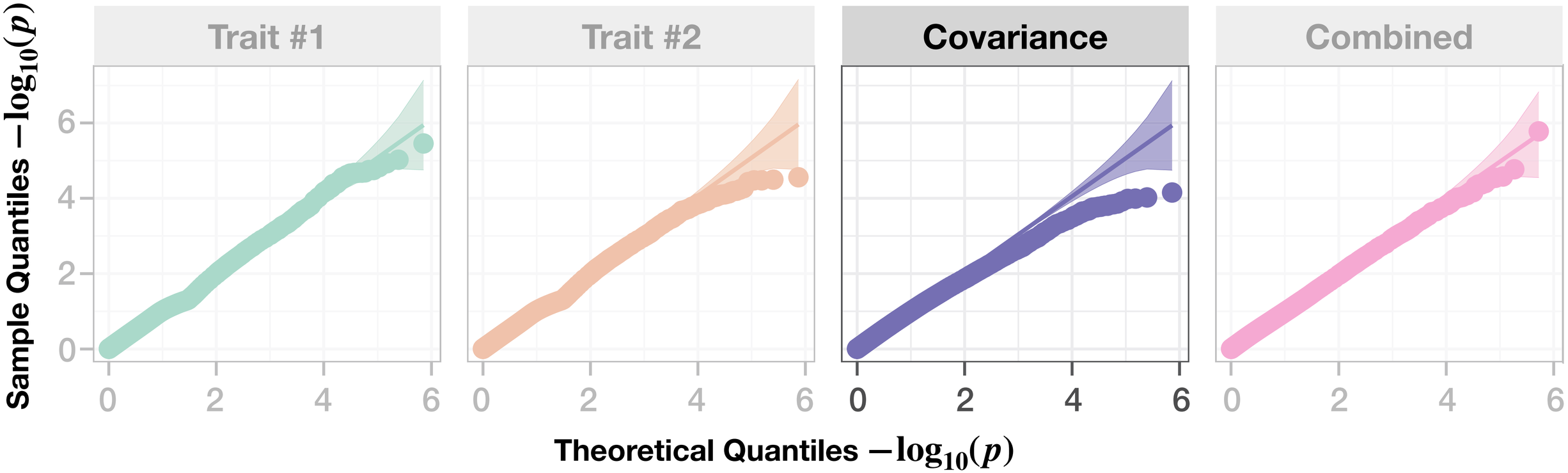
Epistatic Group 2

Marginal epistasis e.g.

- $\mathbf{g}_{x_1} = (\mathbf{x}_1 \circ \mathbf{x}_3) \cdot \alpha_{13} + (\mathbf{x}_1 \circ \mathbf{x}_4) \cdot \alpha_{14} + (\mathbf{x}_1 \circ \mathbf{x}_5) \cdot \alpha_{15}$
- $\mathbf{g}_{x_3} = (\mathbf{x}_1 \circ \mathbf{x}_3) \cdot \alpha_{13} + (\mathbf{x}_2 \circ \mathbf{x}_3) \cdot \alpha_{23}$

QQ-Plots*

mvMAPIT is well calibrated

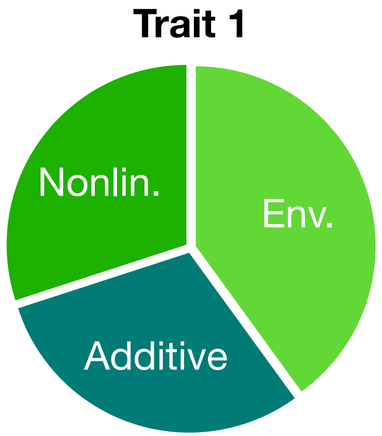


* Simulated data with null hypothesis true

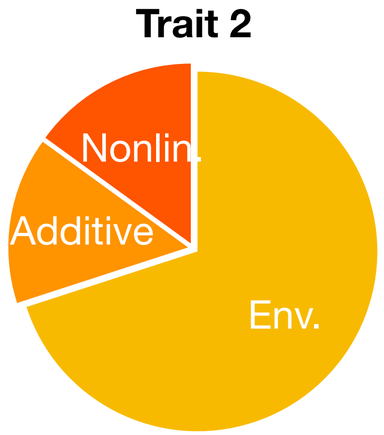
Stamp et al. (2023), *G3*

Empirical Power

Genetic correlations improve power of mvMAPIT

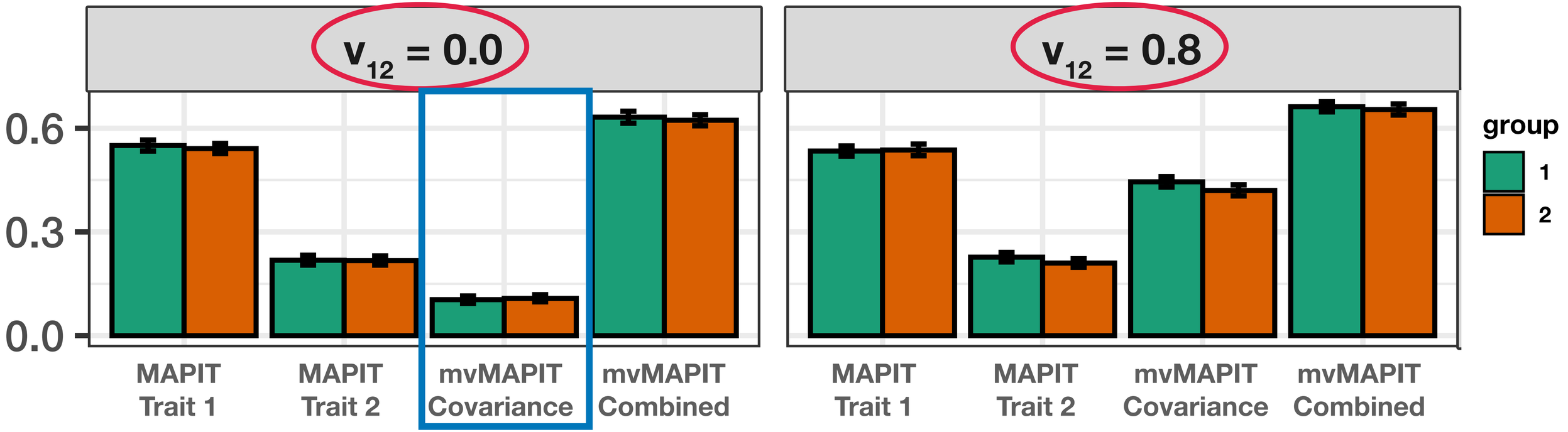


$H^2 = 0.6$



$H^2 = 0.3$

Correlation between epistatic effect sizes V_{12}



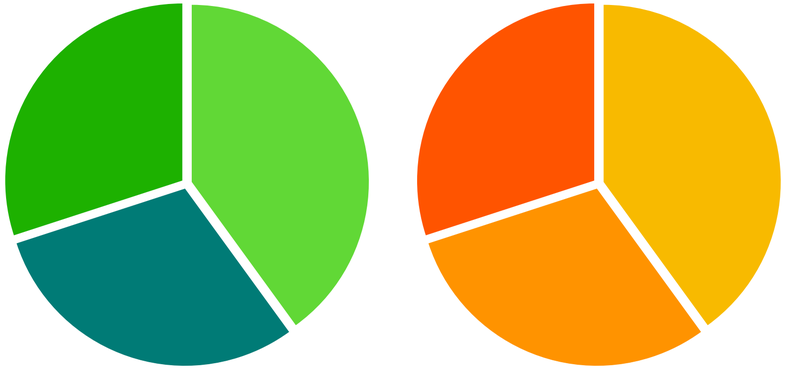
*significance threshold $\alpha = 0.05; \rho = 0.5$

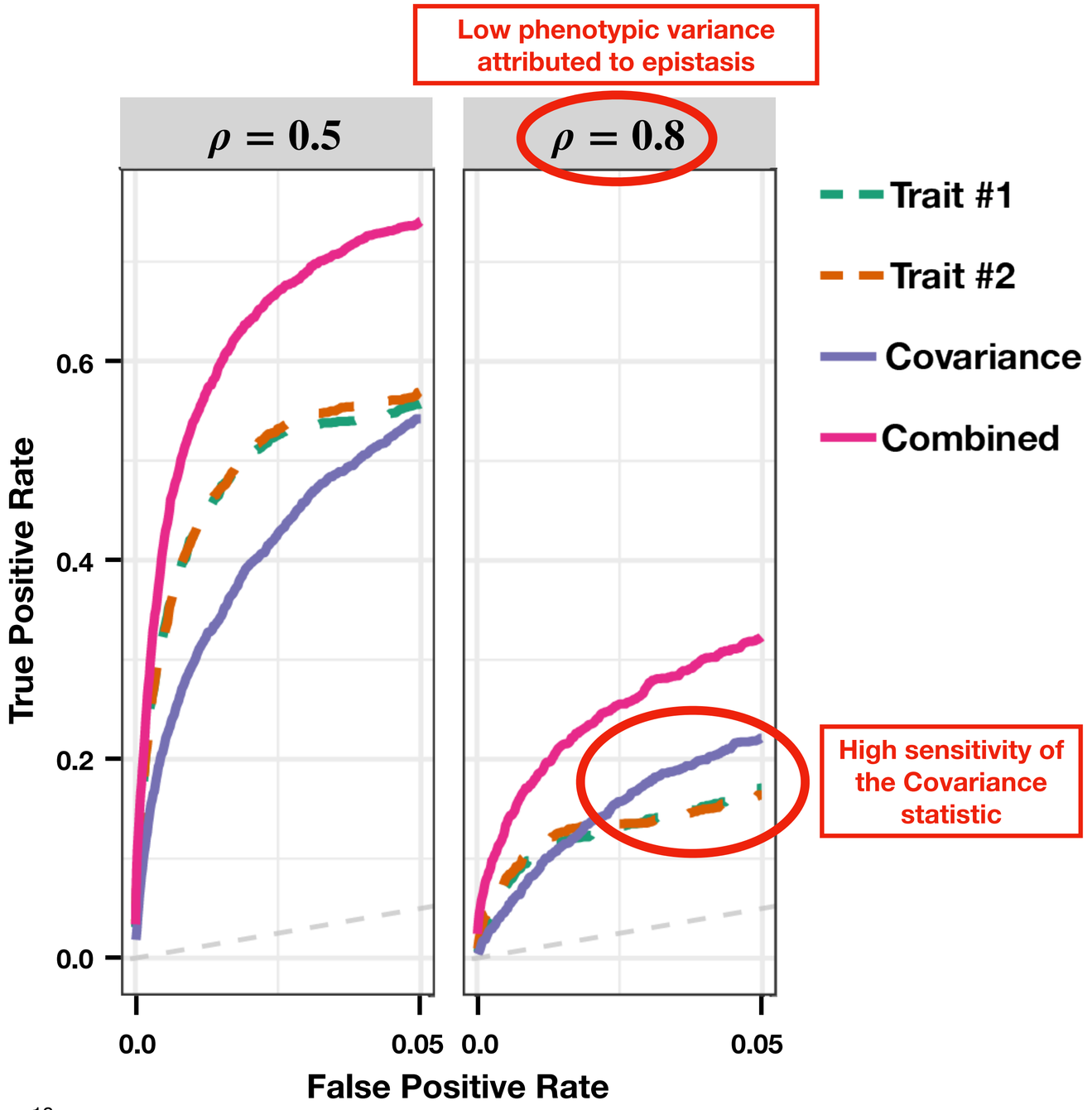
Stamp et al. (2023), *G3*

ROC Curves

Genetic correlations increase sensitivity of mvMAPIT

- Heritability $H^2 = 0.6$
- High correlation $v_{12} = 0.8$
- Steeper curves indicate higher sensitivity

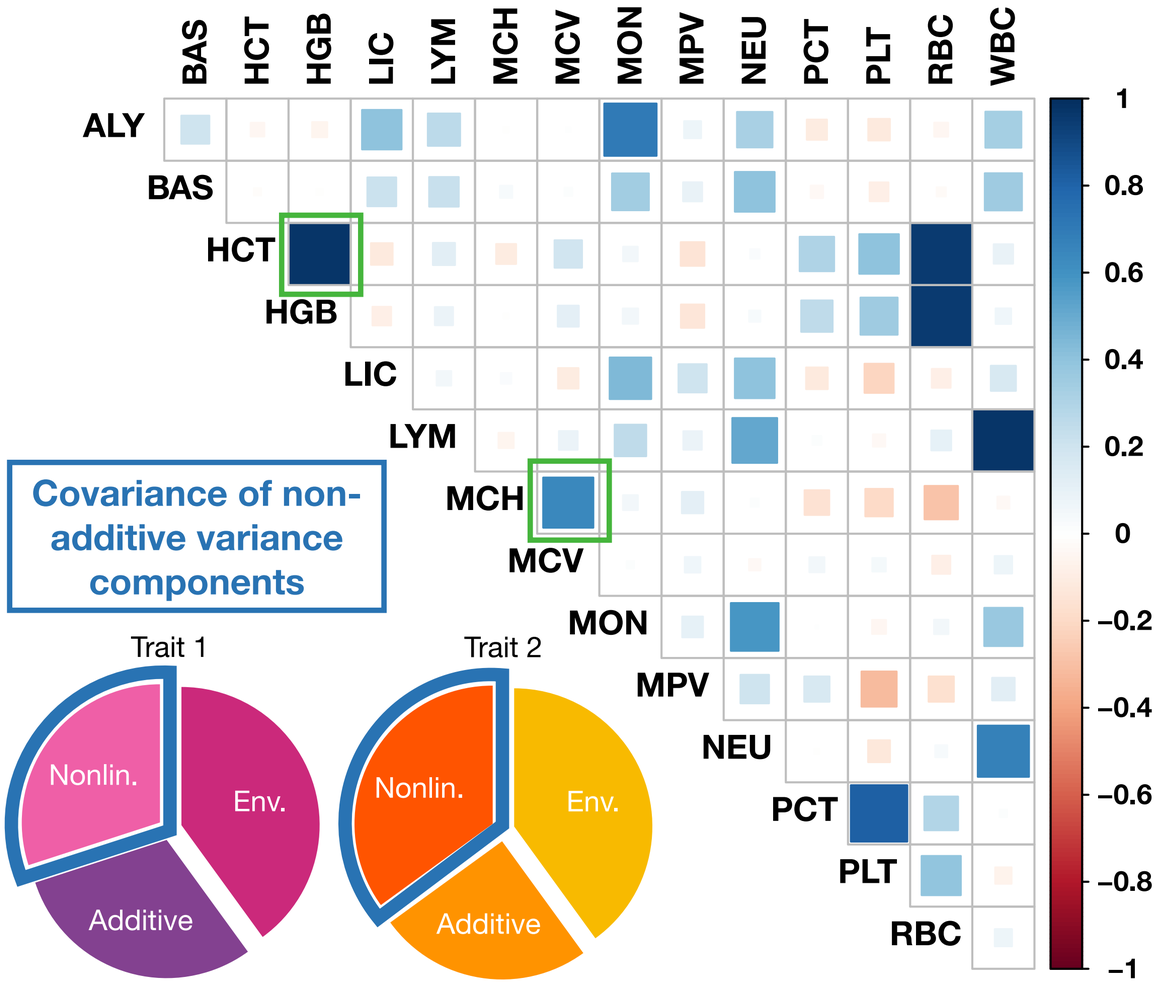




*Trait 1 & Trait 2 correspond to baseline (MAPIT)

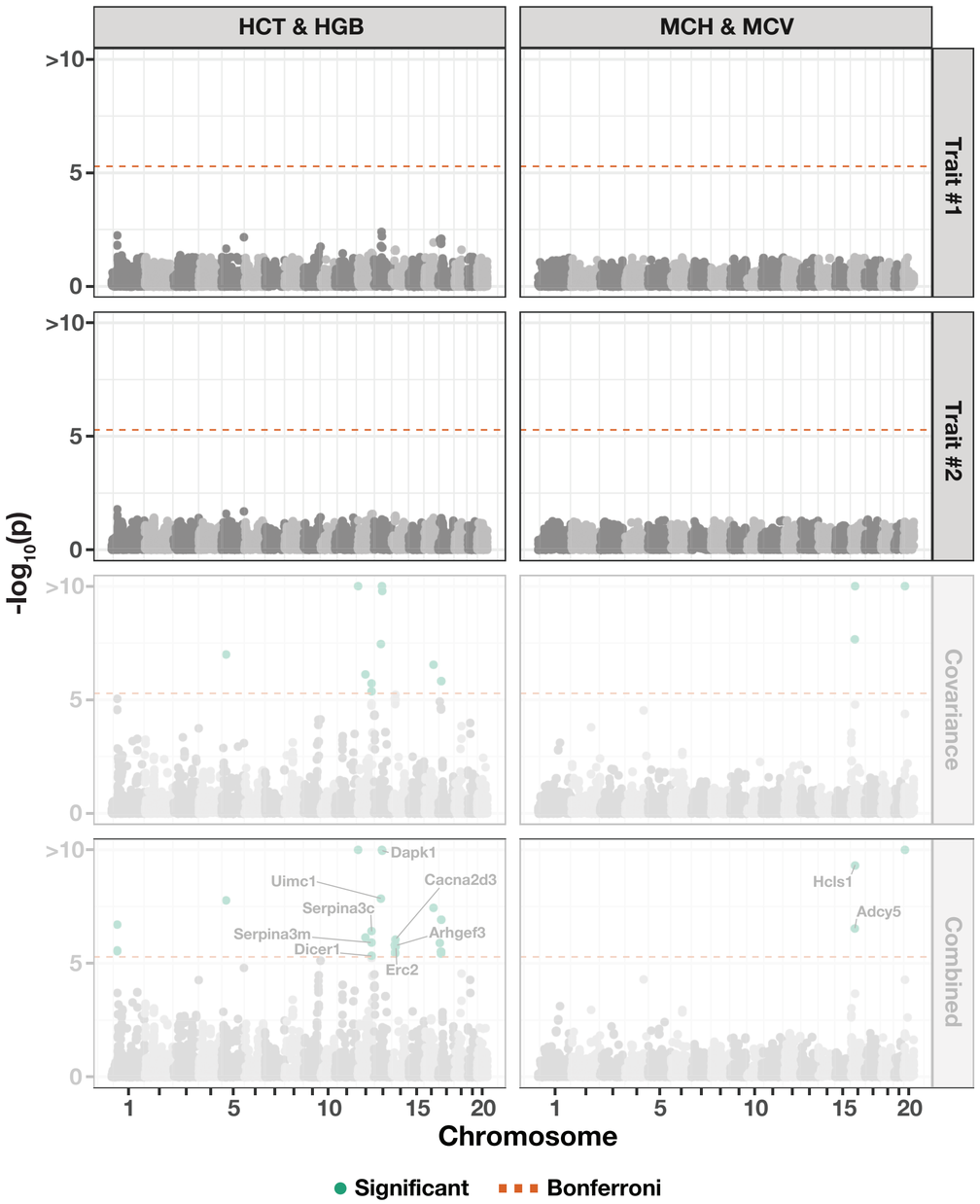
Real Data*

Genetic correlations reveal strong signal of epistasis



Covariance of non-additive variance components

* Hematology traits of WTCCC Mice



MAPIT
Trait 1

MAPIT
Trait 2

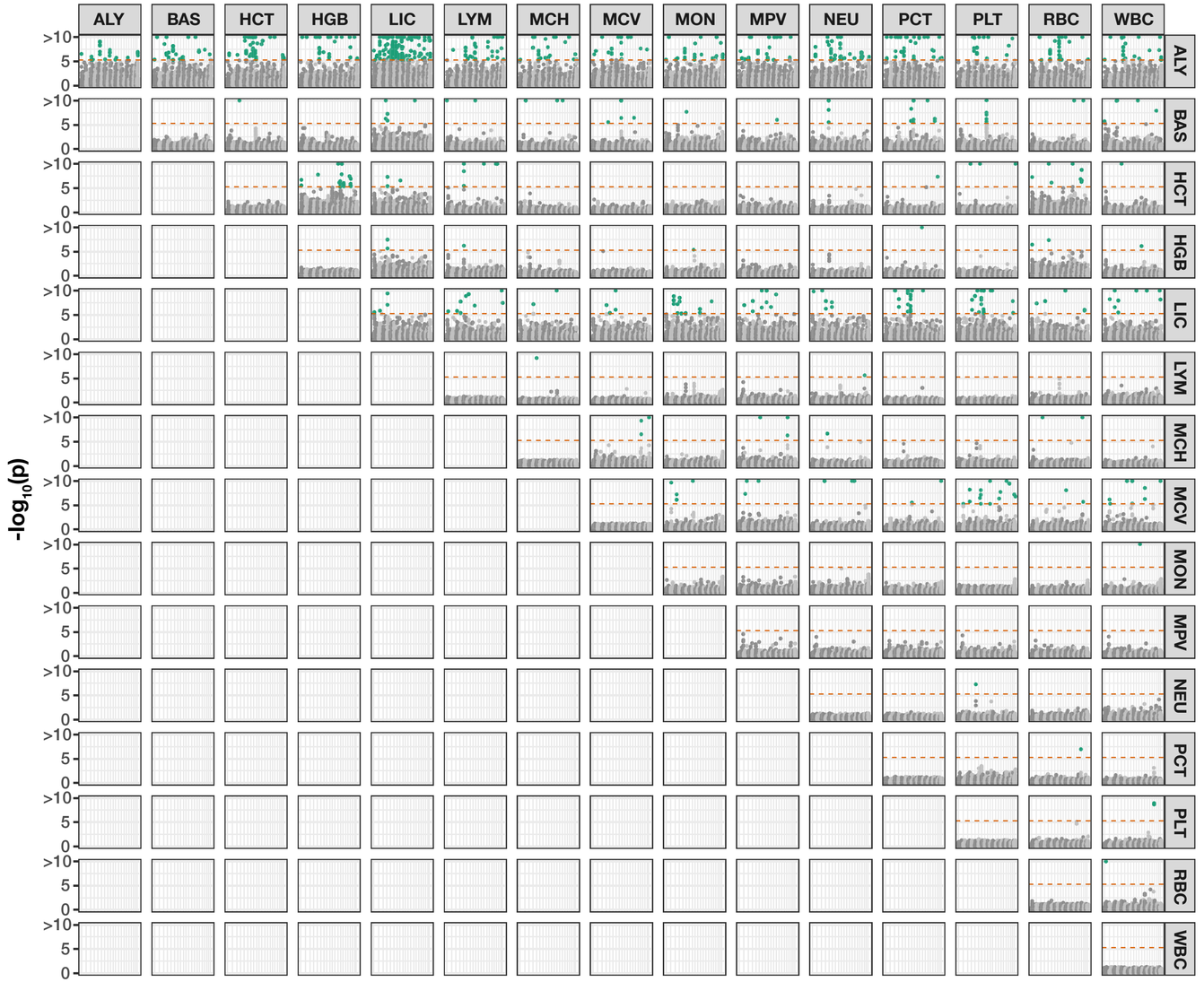
mvMAPIT
Covariance
Trait 1 & Trait 2

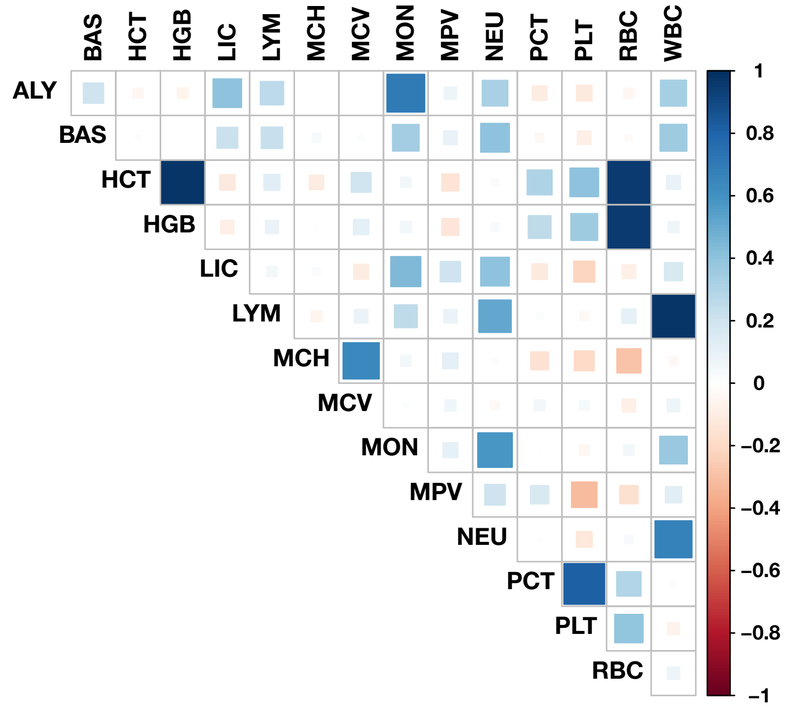
mvMAPIT
Combined
Trait 1 & Trait 2 & Covariance

Stamp et al. (2023), *G3*

Real Data*

New significant associations across many trait pairs



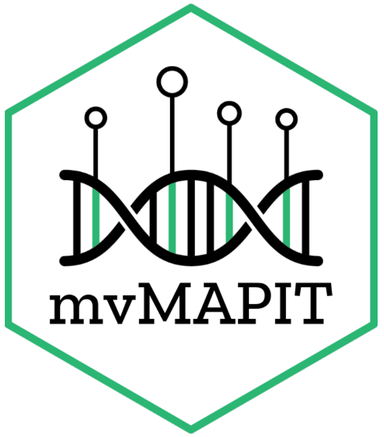


* Hematology traits of WTCCC Mice

Stamp et al. (2023), *G3*

Weaknesses

- Time complexity scaling with sample size
- Unknown interaction partner
- Meta analysis p-value interpretability



Strengths

- Analysis of shared genetic architecture of traits
- Correlation between effects improves sensitivity
- Marginalisation accumulates weak effects to strong signal
- Marginalisation reduces search space

Acknowledgements

Advisors

Lorin Crawford
Dan Weinreich

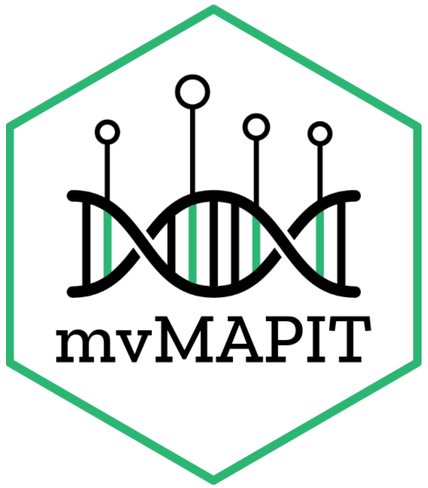
Crawford Lab and Dave

Chibuikem Nwizu
Dave Peede
Ria Vinod
Alex Wong
Emily Winn
Ashley Conard
Dana Edwin
Wai Shing Tang
Whitney Sloneker
Yu Zhong
Collin Small
Ryan Huang

mvMAPIT

- Code and documentation on GitHub: https://lcrawlab.github.io/mvMAPIT/
- R package published on CRAN: https://cran.r-project.org/package=mvMAPIT

```
install.packages('mvMAPIT')
```



Stamp et al. (2023), G3

Relevant References

Variance Component Estimation

- X. Zhou. "A unified framework for variance component estimation with summary statistics in genome-wide association studies." Ann. Appl. Stat. 11 (4) 2027 - 2051, December 2017. https://doi.org/10.1214/17-AOAS1052

Marginal Epistasis Detection

- L. Crawford, P. Zeng, S. Mukherjee, & X. Zhou, (2017). Detecting epistasis with the marginal epistasis test in genetic mapping studies of quantitative traits. PLOS Genetics, 13(7), e1006869. https://doi.org/10.1371/journal.pgen.1006869
- **J. Stamp**, A. DenAdel, D. Weinreich, & L. Crawford, (2023). Leveraging the Genetic Correlation between Traits Improves the Detection of Epistasis in Genome-wide Association Studies. G3 Genes|Genomes|Genetics, jkad118. https://doi.org/10.1093/g3journal/jkad118

Related Software/Source Code:

- mvMAPIT: https://lcrawlab.github.io/mvMAPIT/