Partitioning the non-additive variation of complex traits

Center for Computational Molecular Biology
Brown University

Julian Stamp

Outline

Introduction

Marginal Epistasis

Multivariate Linear Mixed Models

i-LDSC regression

Conclusion

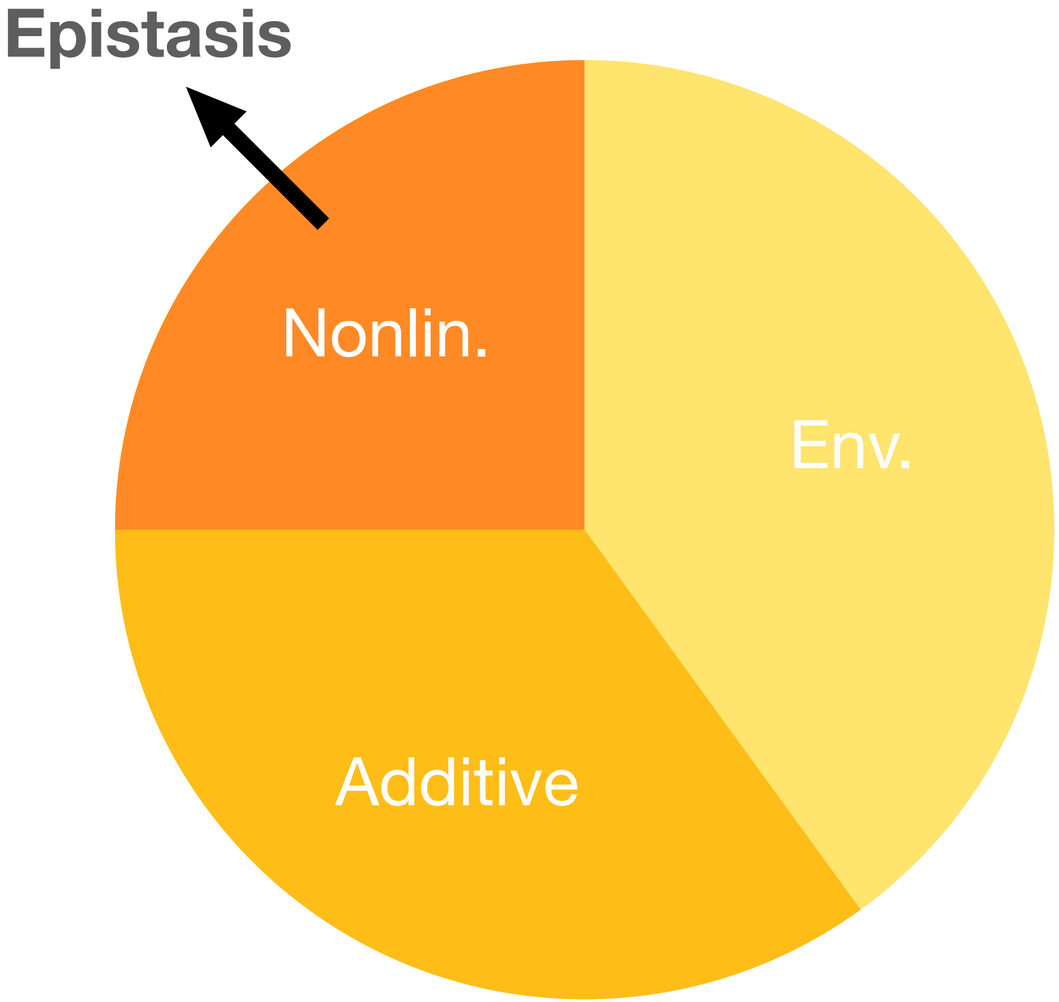
Phenotypic Variance

Genetic & Environmental Factors

$$P = G + E$$

Broad sense Heritability

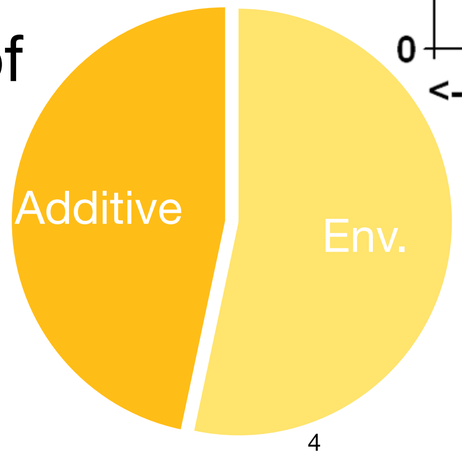
$$H^2 = \frac{\mathrm{Var}[G]}{\mathrm{Var}[P]}$$



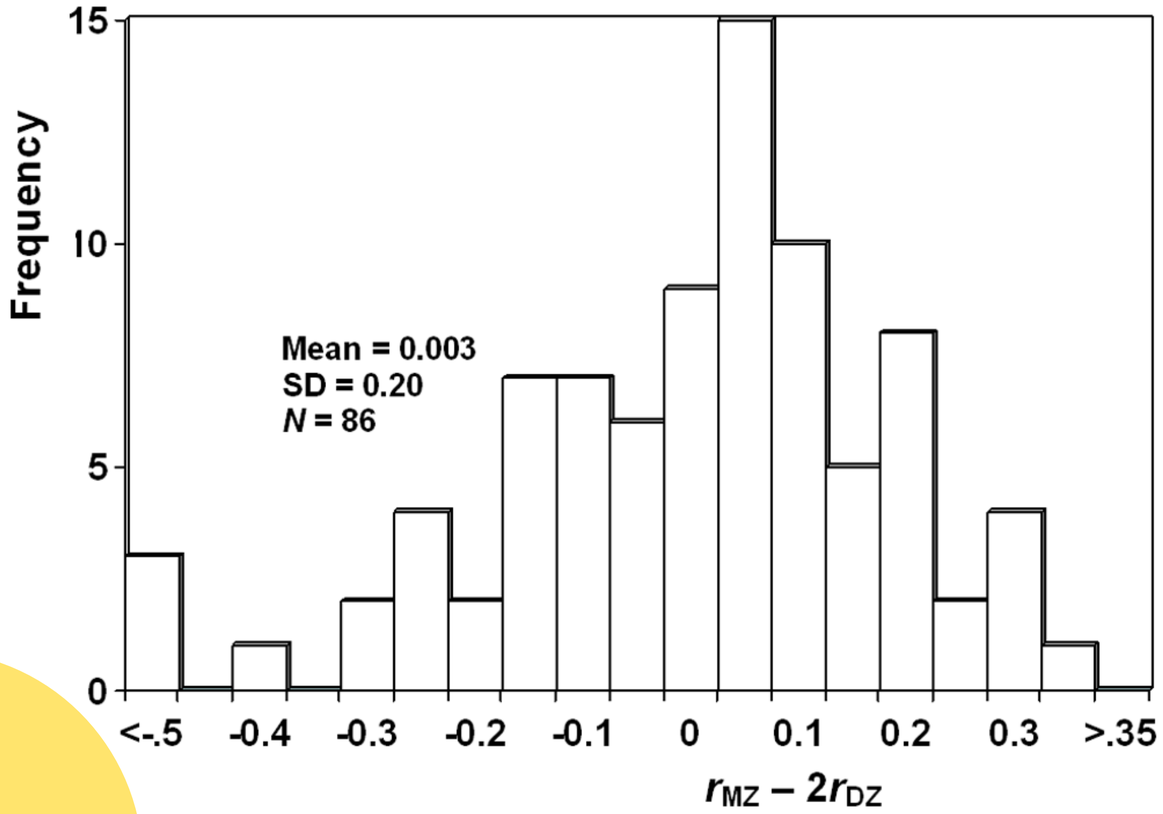
Non-additive variation in human traits

Relative importance of epistasis is controversial[1]

- Epistatic gene action is different from statistical epistasis
- Statistical epistatic trait variance depends on allele frequencies[2]
- Estimated “additive” effects are function of non-additive effects[2]
- Majority of the heritability of complex traits “missing”[3]



Hill, Goddard, and Visscher (2008). Distribution $r_{MZ} - 2r_{DZ}$ for all traits on human twins.



r_{MZ} - trait correlation monozygous twins

r_{DZ} - trait correlation dizygous twins

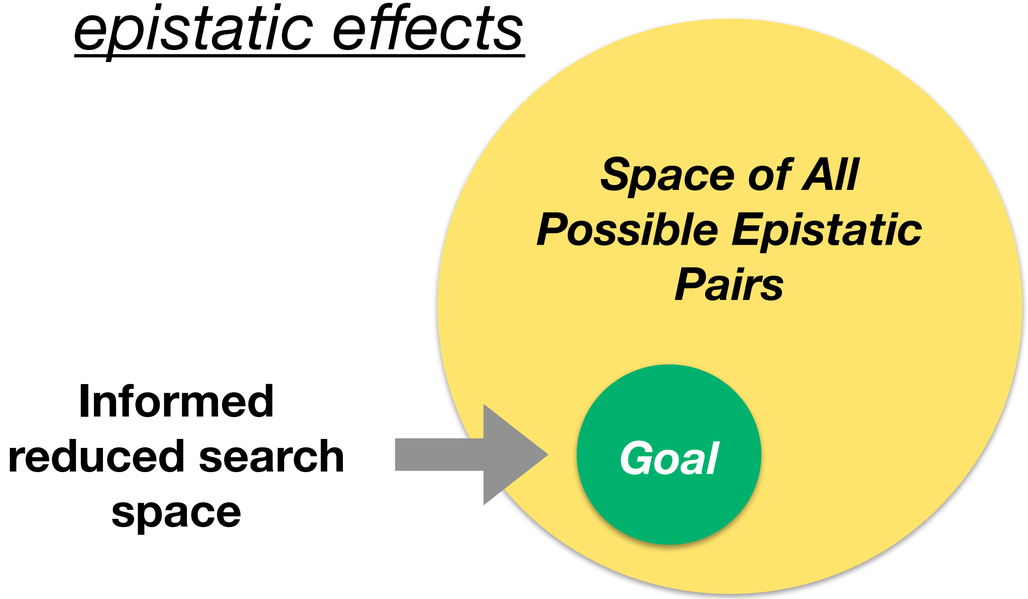
$r_{MZ} - 2r_{DZ} > 0$ implies **nonlinear** contributions

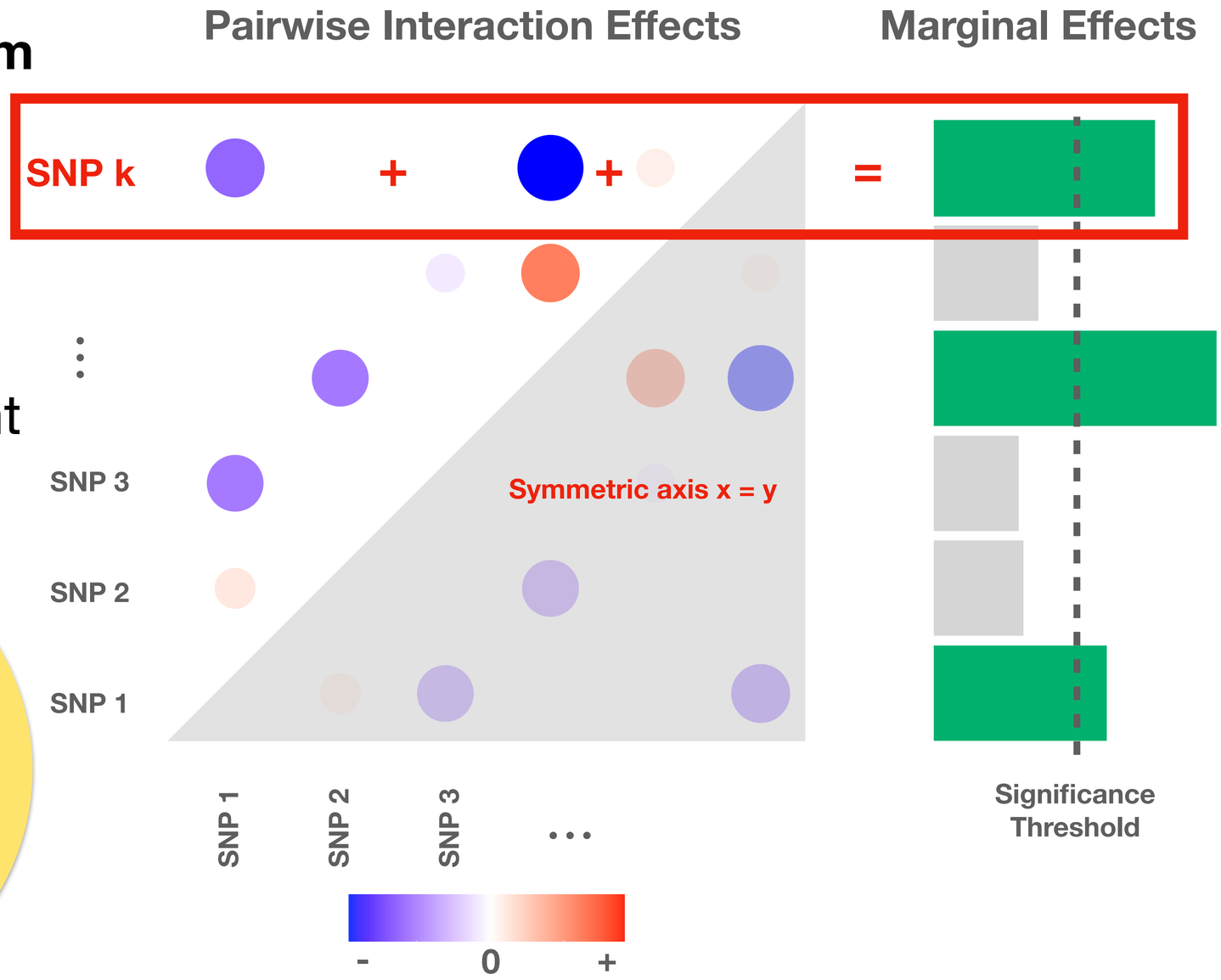
1 Hill et al. (2008), *PLOS Gen*
2 Hivert et al. (2021), *PLOS Gen*
3 Young (2019), *PLOS Gen*

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*





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

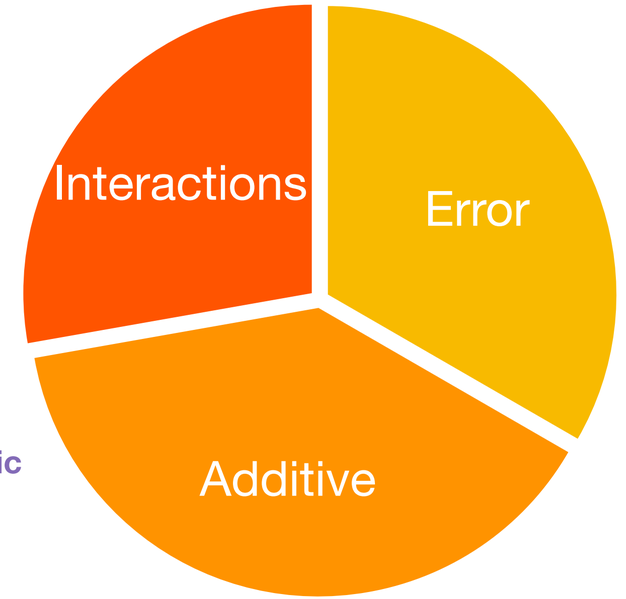
Linear Effect Sizes

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

- 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

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

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

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

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

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

MAPIT

Null Hypothesis and Test

- We want to test for marginal epistatic effects
- Use MQS[1] to estimate variance components
- Under the null hypothesis assume mixture of chi-squared[2]

$$H_0 : \mathbf{g}_k = 0 \quad \Leftrightarrow \quad H_0 : \sigma^2 = 0$$

$$\hat{\sigma}^2 = \mathbf{y}^T \mathbf{A}_k \mathbf{y}$$

$$\sigma^2 \sim \sum_{i=1}^{n} \lambda_i \chi^2_{1,i}$$

1 Zhou (2017), *AOAS*
2 Crawford et al. (2017), *PLOS Gen*

MAPIT

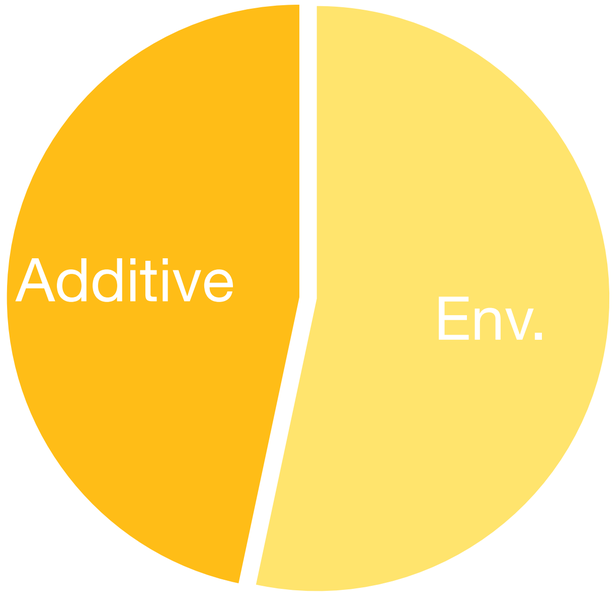
Simulations of complex traits

Scenarios

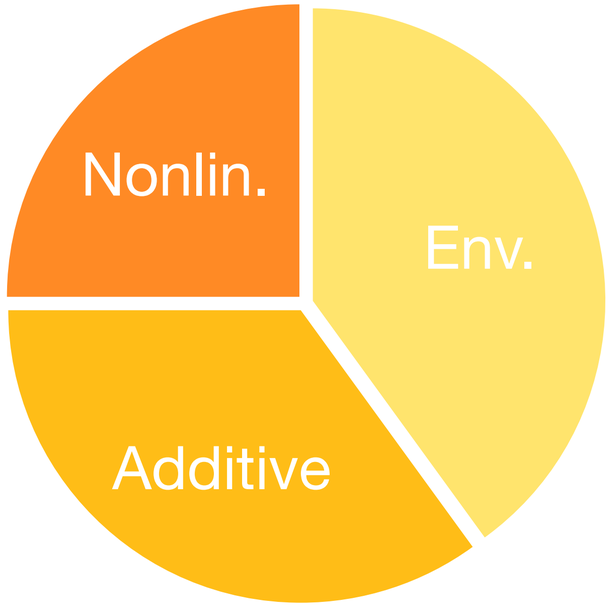
- Null Hypothesis true: no epistasis
- Epistasis with varying parameters

Parameters

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



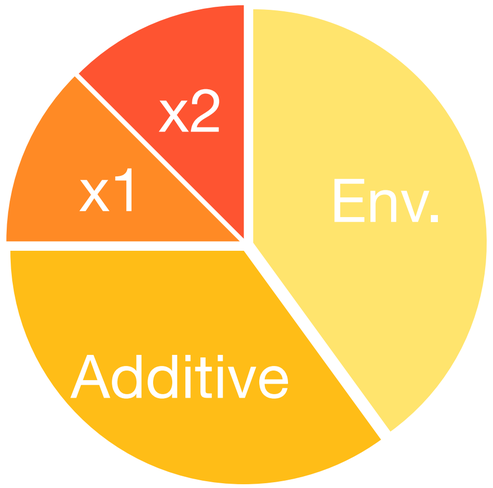
$H_0 : \sigma^2 = 0$



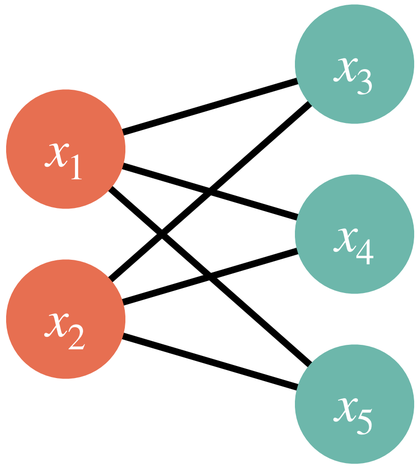
$H_1 : \sigma^2 \neq 0$

MAPIT

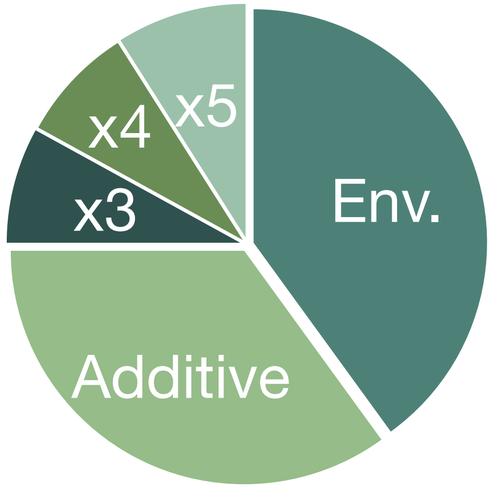
Simulations of complex traits



Group 1



Group 2

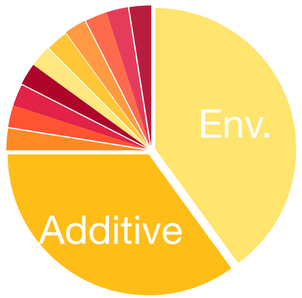


- Additive SNPs
- Epistatic Group 1
- Epistatic Group 2

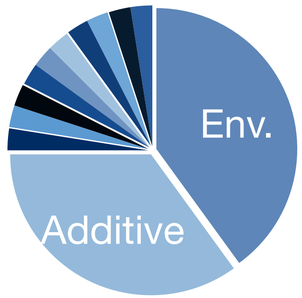
Marginal epistasis e.g.

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

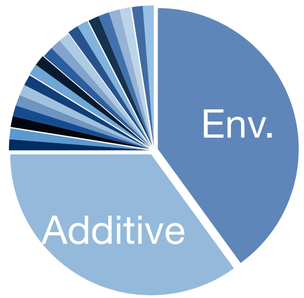
Simulations
Estimating PVE



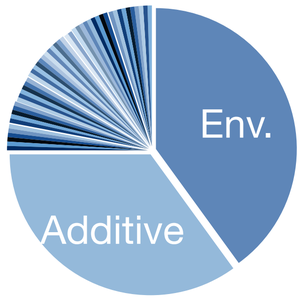
Group 1



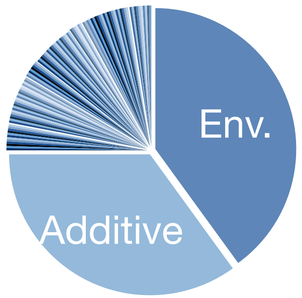
Group 2 - I



Group 2 - II

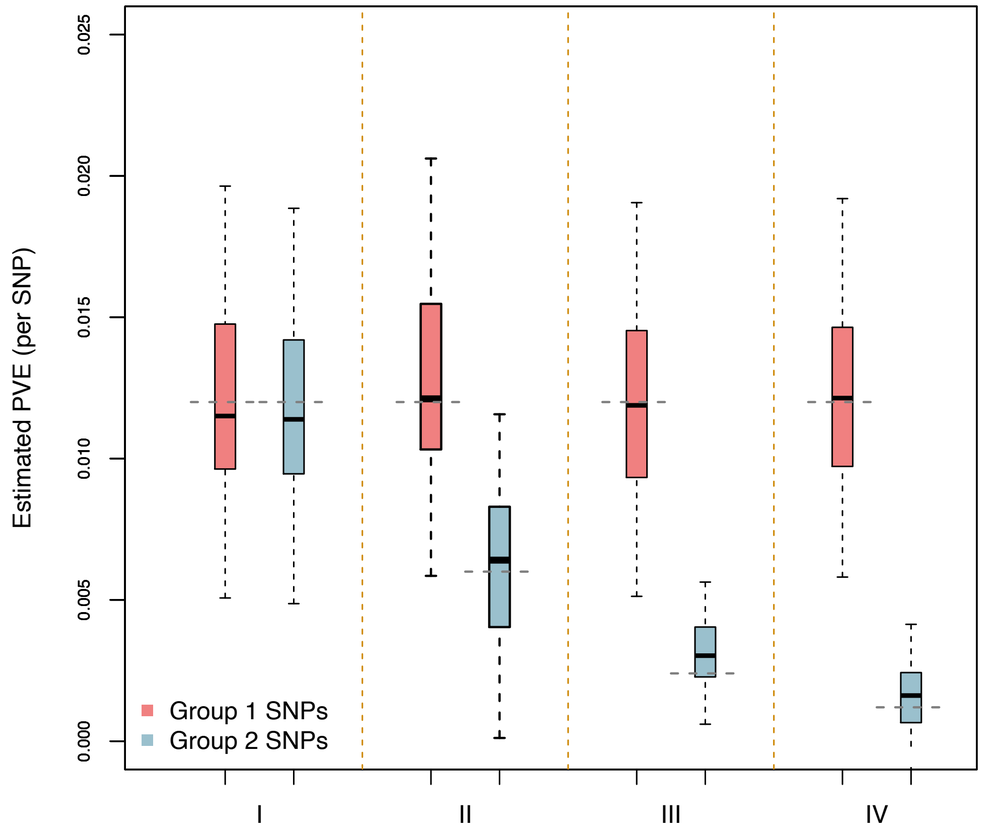


Group 2 - III



Group 2 - IV

- 10 Causal SNPs in Group 1
- **Scenario I**: 10 SNPs in Group 2
- **Scenario II**: 20 SNPs in Group 2
- **Scenario III**: 50 SNPs in Group 2
- **Scenario IV**: 100 SNPs in Group 2

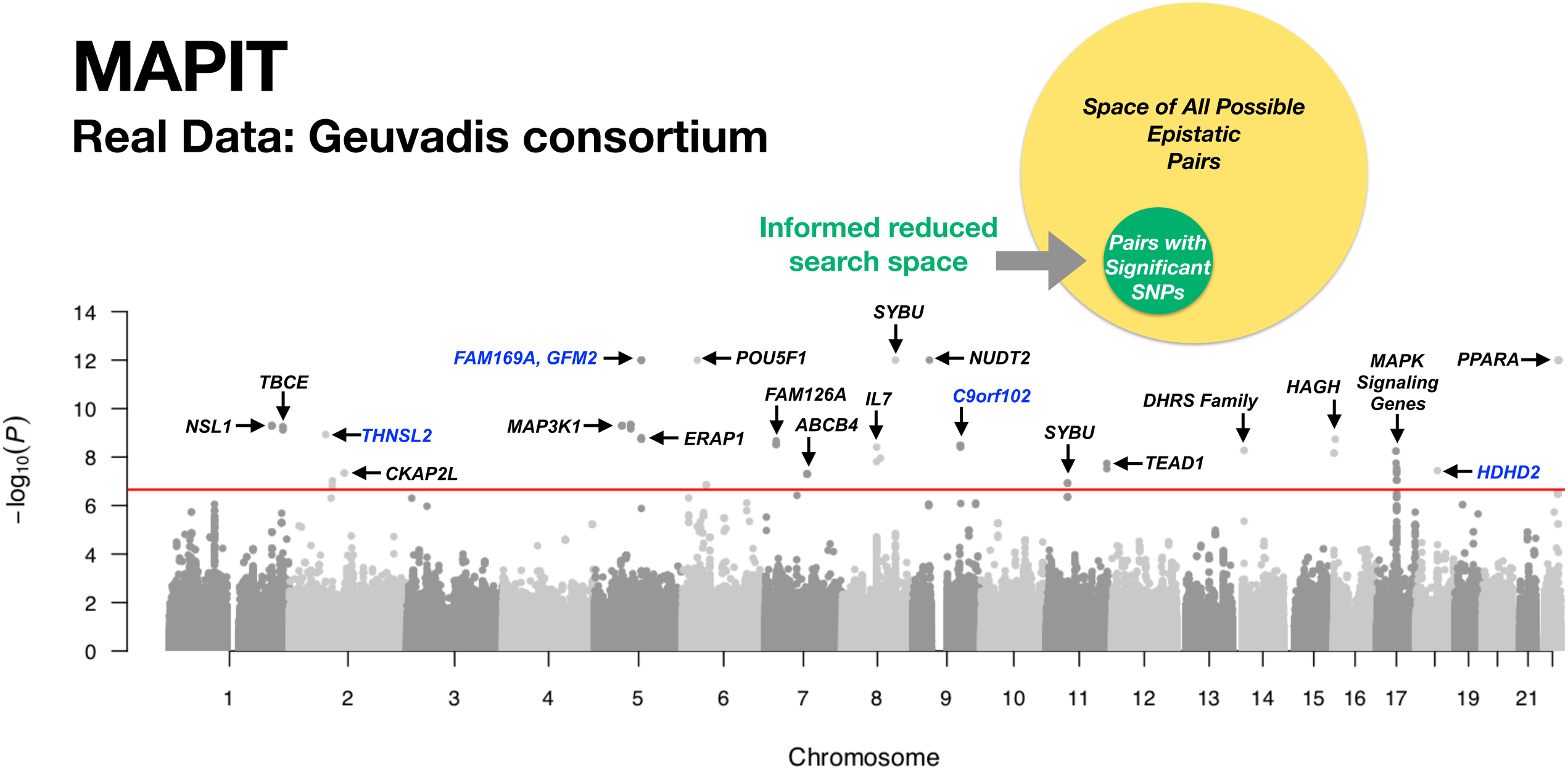


* Epistasis is 20% of PVE

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

MAPIT

Real Data: Geuvadis consortium



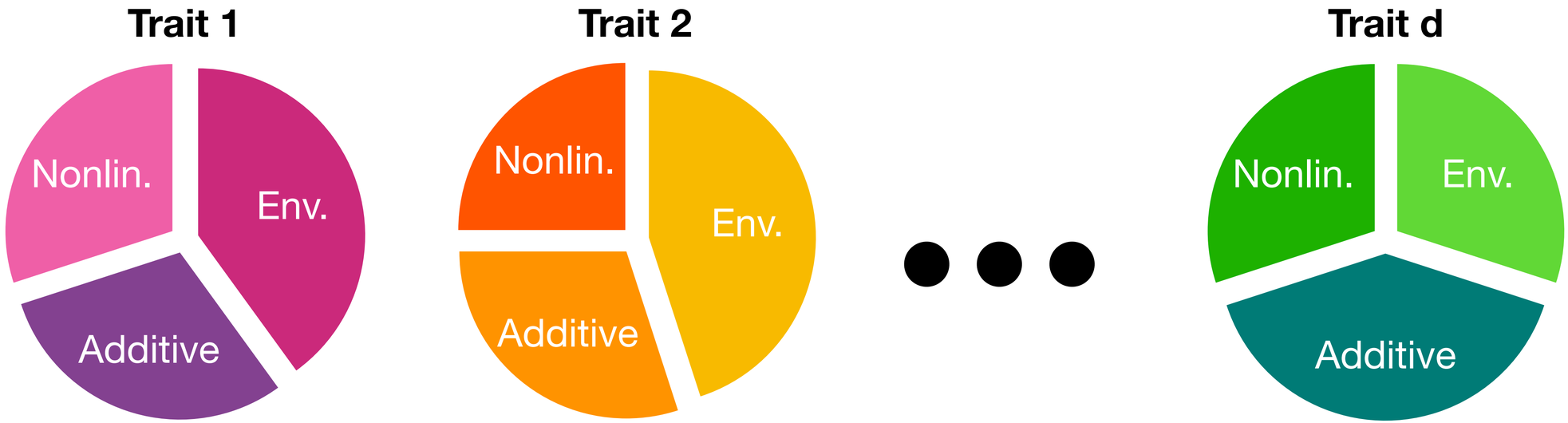
Red Line: Genome-wide significance threshold

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

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

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*

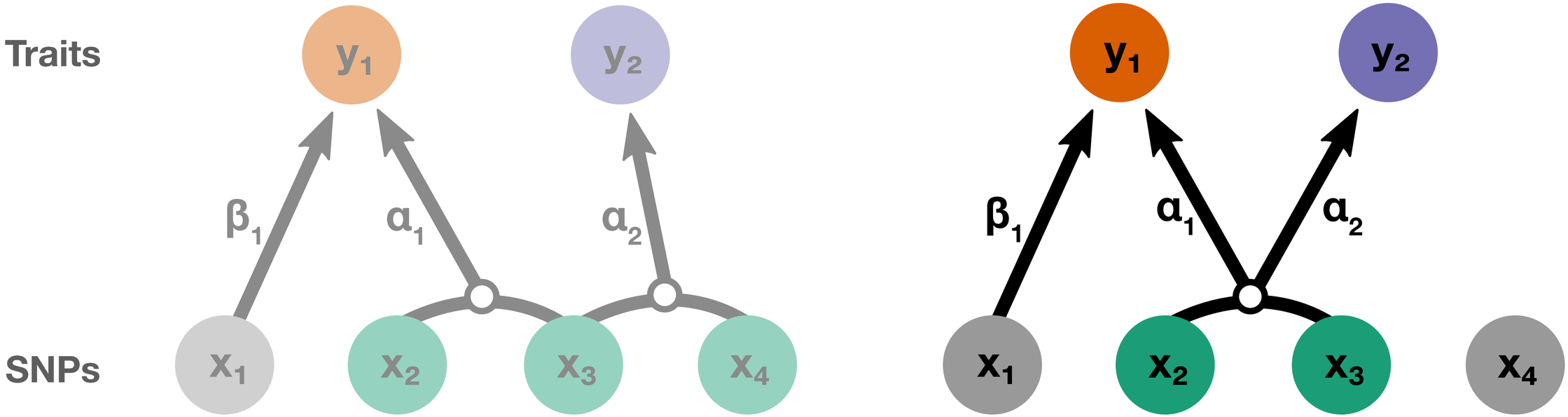
mvMAPIT

- 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

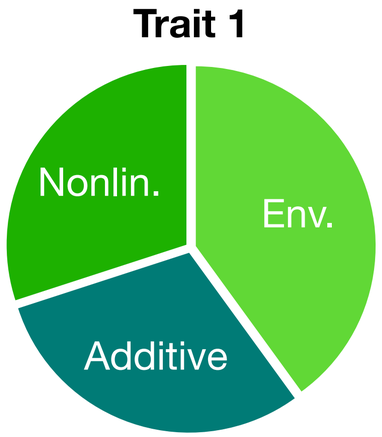
Modelling cross-trait genetic correlations of interaction effects



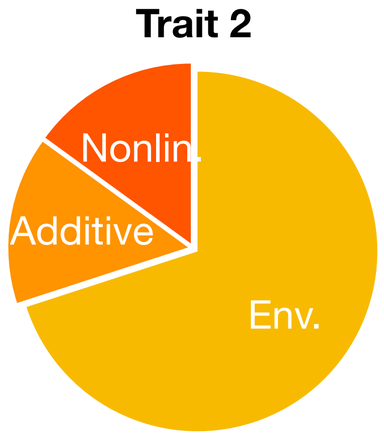
Stamp et al. (2023), *G3*

Empirical Power

Genetic correlations improve power of mvMAPIT

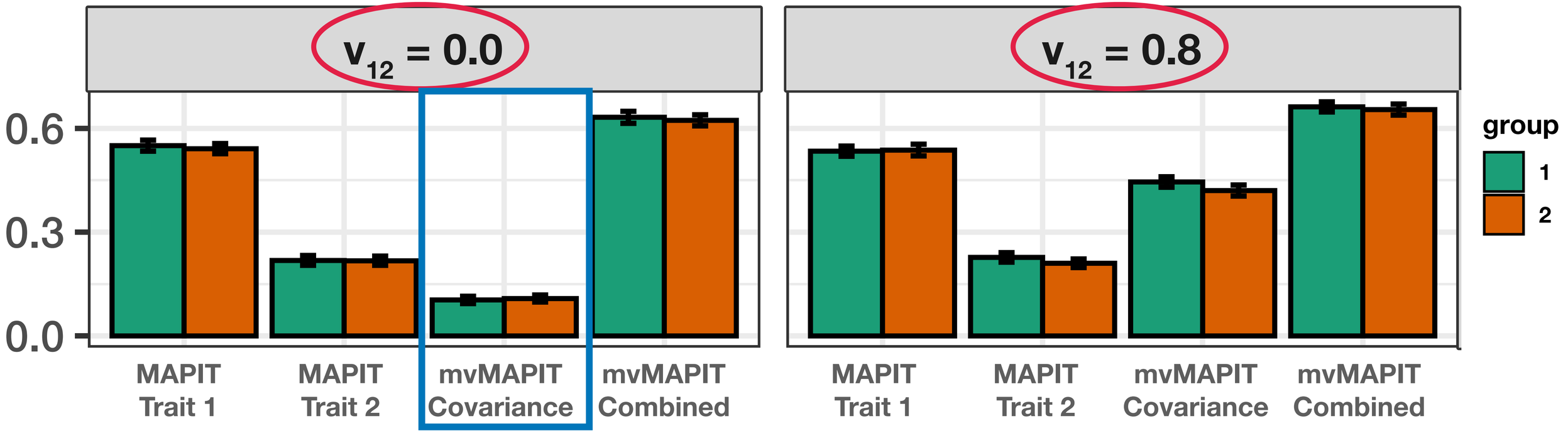


$H^2 = 0.6$



$H^2 = 0.3$

Correlation between epistatic effect sizes V_{12}

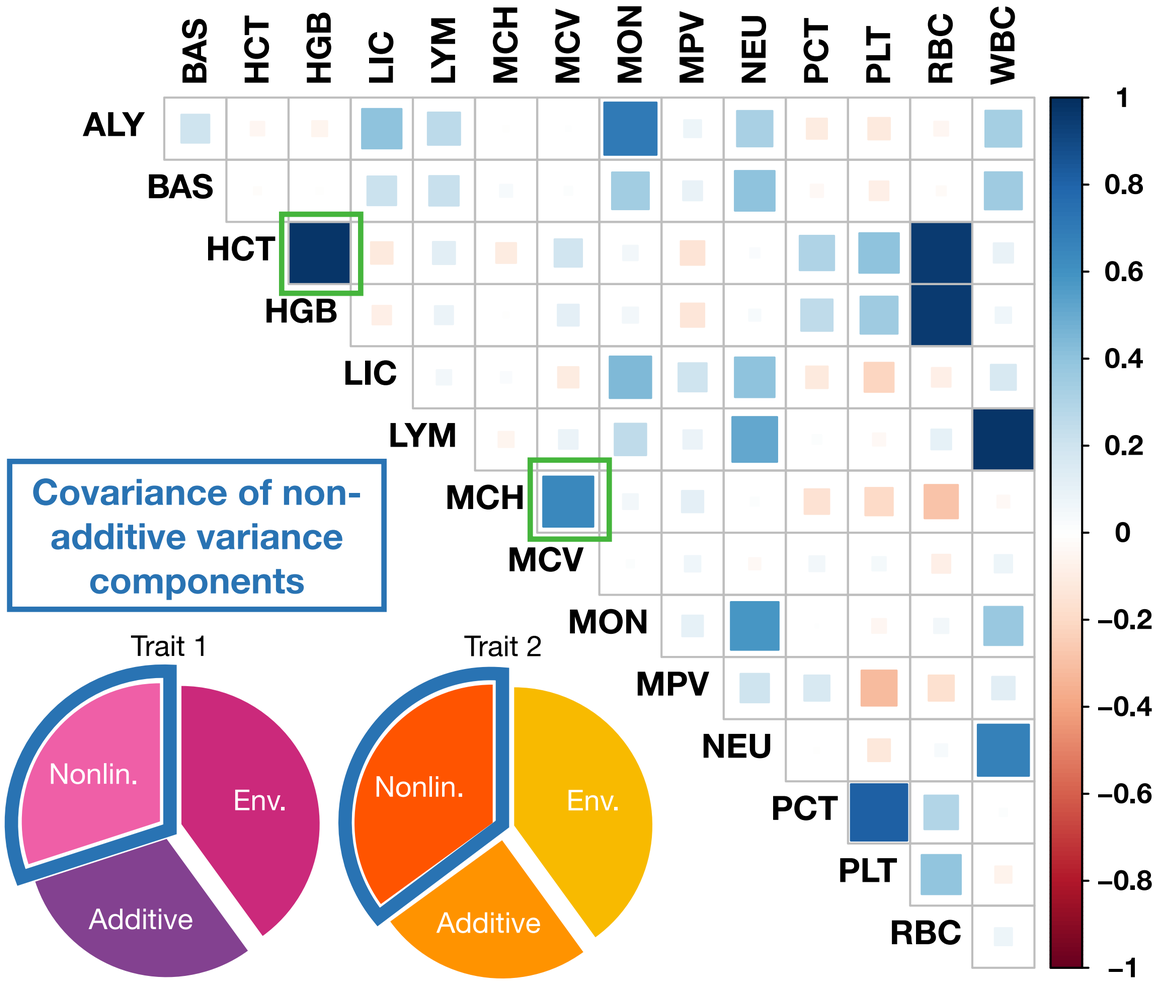


*significance threshold $\alpha = 0.05; \rho = 0.5$

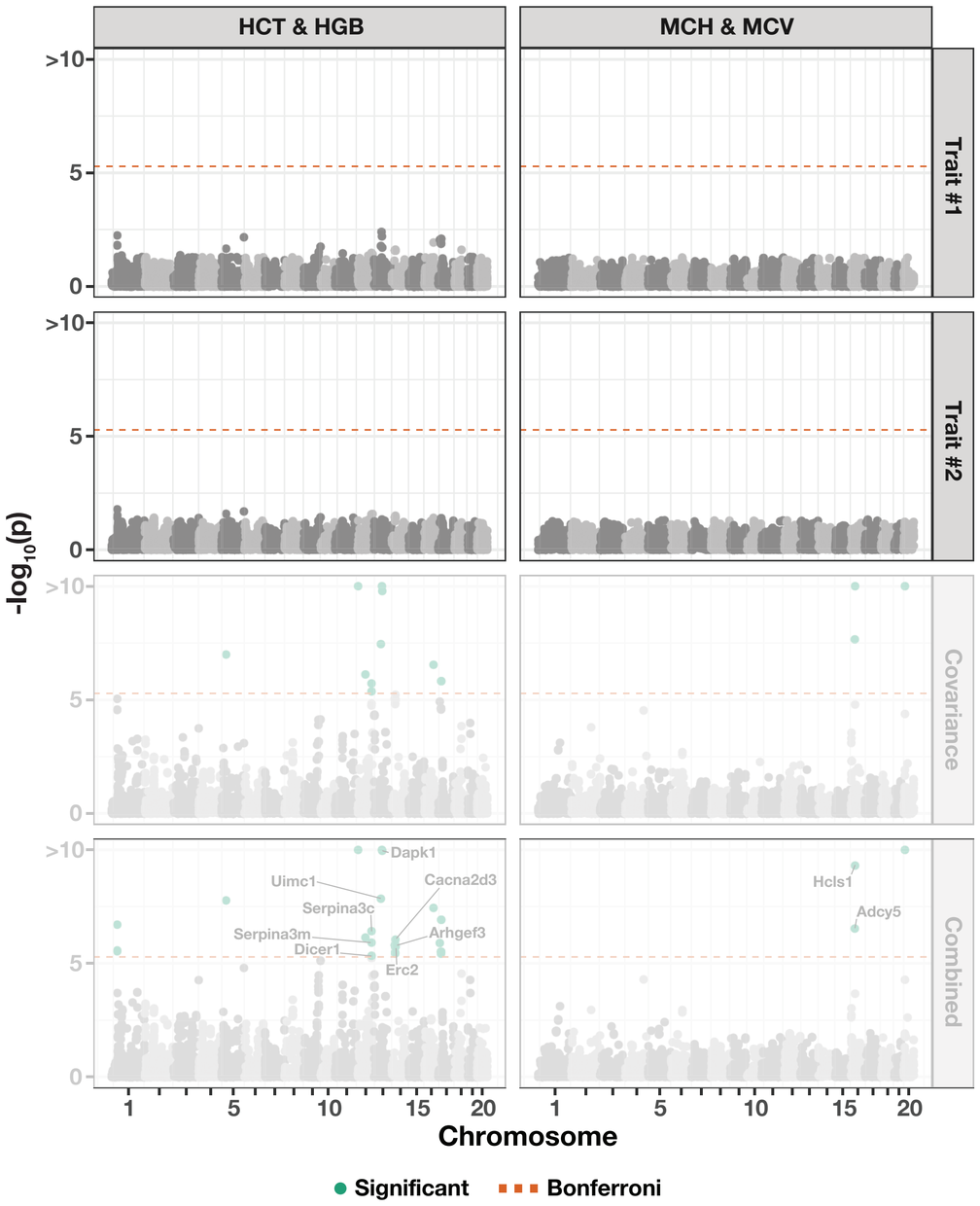
Stamp et al. (2023), *G3*

Real Data*

Genetic correlations reveal strong signal of epistasis



Covariance of non-additive variance components



MAPIT
Trait 1

MAPIT
Trait 2

mvMAPIT
Covariance
Trait 1 & Trait 2

mvMAPIT
Combined
Trait 1 & Trait 2 & Covariance

* Hematology traits of WTCCC Mice

Stamp et al. (2023), *G3*

i-LDSC regression

Non-additive effects in complex human traits

- Including epistasis improves heritability estimates in GWAS
- Epistasis is more pervasive in human traits than previously reported

New Results

Follow this preprint

Accounting for statistical non-additive interactions enables the recovery of missing heritability from GWAS summary statistics

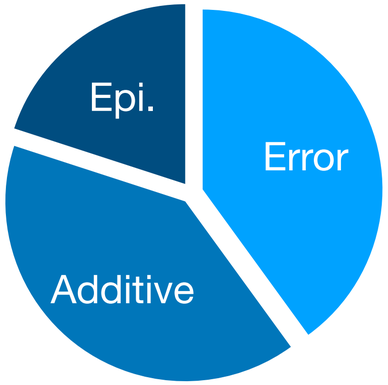
Samuel Pattillo Smith, Gregory Darnell, Dana Udwin, Arbel Harpak, Sohini Ramachandran, Lorin Crawford

doi: https://doi.org/10.1101/2022.07.21.501001

Generative Model

Polygenic trait architecture

- $\mathbb{V}[\mathbf{X}\boldsymbol{\beta}] + \mathbb{V}[\mathbf{W}\boldsymbol{\theta}] = H^2$ is the broad-sense heritability
- $\mathbb{V}[\mathbf{X}\boldsymbol{\beta}] = h^2 = \rho H^2$ is the narrow-sense heritability
- $\mathbb{V}[\mathbf{W}\boldsymbol{\theta}] = (1-\rho)H^2$ makes up the remaining variation
- ρ measures the proportion of variance that is explained by additivity.



Genotype Matrix

Pairwise cis-Acting Interaction Matrix

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

Linear Effects

Interaction Effects

$$\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, (1-H^2)\mathbf{I})$$

Smith, Darnell et al., *bioRxiv*

i-LDSC regression

Extending the LD Score Regression Framework

LD Score Regression

- Taking the expectation of GWA test statistics $\chi^2 = N\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^\intercal$ yields:

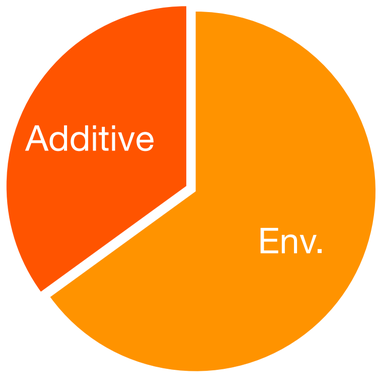
$$\mathbb{E}[\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^\intercal] = \lambda\mathbf{R} + \left(\frac{\rho H^2}{J}\right)\mathbf{R}^2$$

- A model to estimate heritability:

$$\mathbb{E}[\chi^2] \propto \mathbf{1} + \boldsymbol{\ell}\tau$$

LD Scores are given by:

$$\boldsymbol{\ell}_j = \sum_k r_{jk}^2$$



Bulik-Sullivan et al. (2015), *Nature Gen*

Interaction-LD Score

- Taking the expectation of GWA test statistics $\chi^2 = N\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^\intercal$ yields:

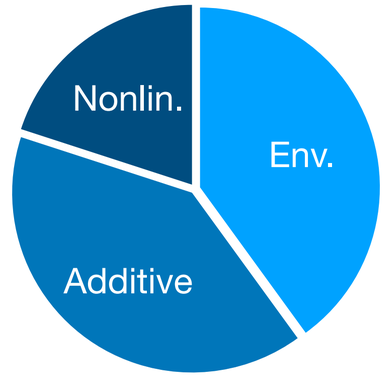
$$\mathbb{E}[\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^\intercal] = \lambda\mathbf{R} + \left(\frac{\rho H^2}{J}\right)\mathbf{R}^2 + \left(\frac{(1-\rho)H^2}{M}\right)\mathbf{V}^2$$

- A model to estimate heritability:

$$\mathbb{E}[\chi^2] \propto \mathbf{1} + \boldsymbol{\ell}\tau + \boldsymbol{f}\sigma^2$$

- LD and i-LD Scores are given by:

$$\boldsymbol{\ell}_j = \sum_k r_{jk}^2, \qquad \boldsymbol{f}_j = \sum_m v_{jm}^2$$

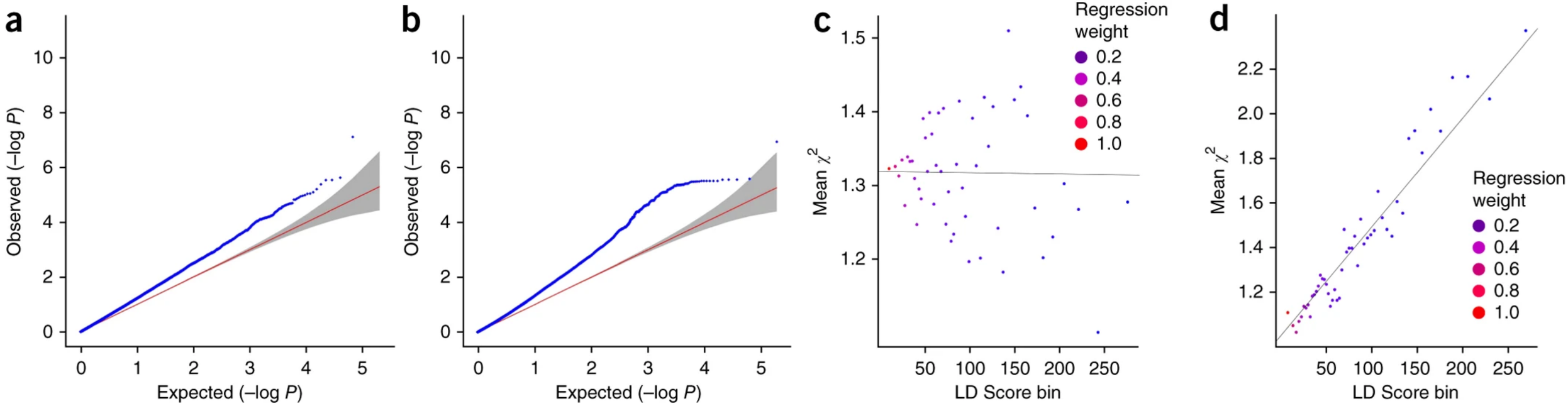


Smith, Darnell et al., *bioRxiv*

LDSC regression

Estimating narrow sense heritability from GWA summary statistics

- Regress on $\mathbb{E}[\chi^2] \propto \mathbf{1} + \ell\tau$

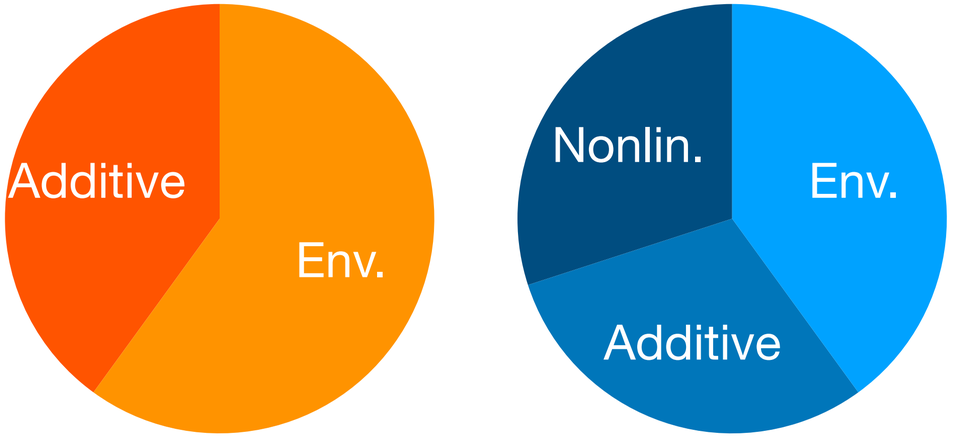


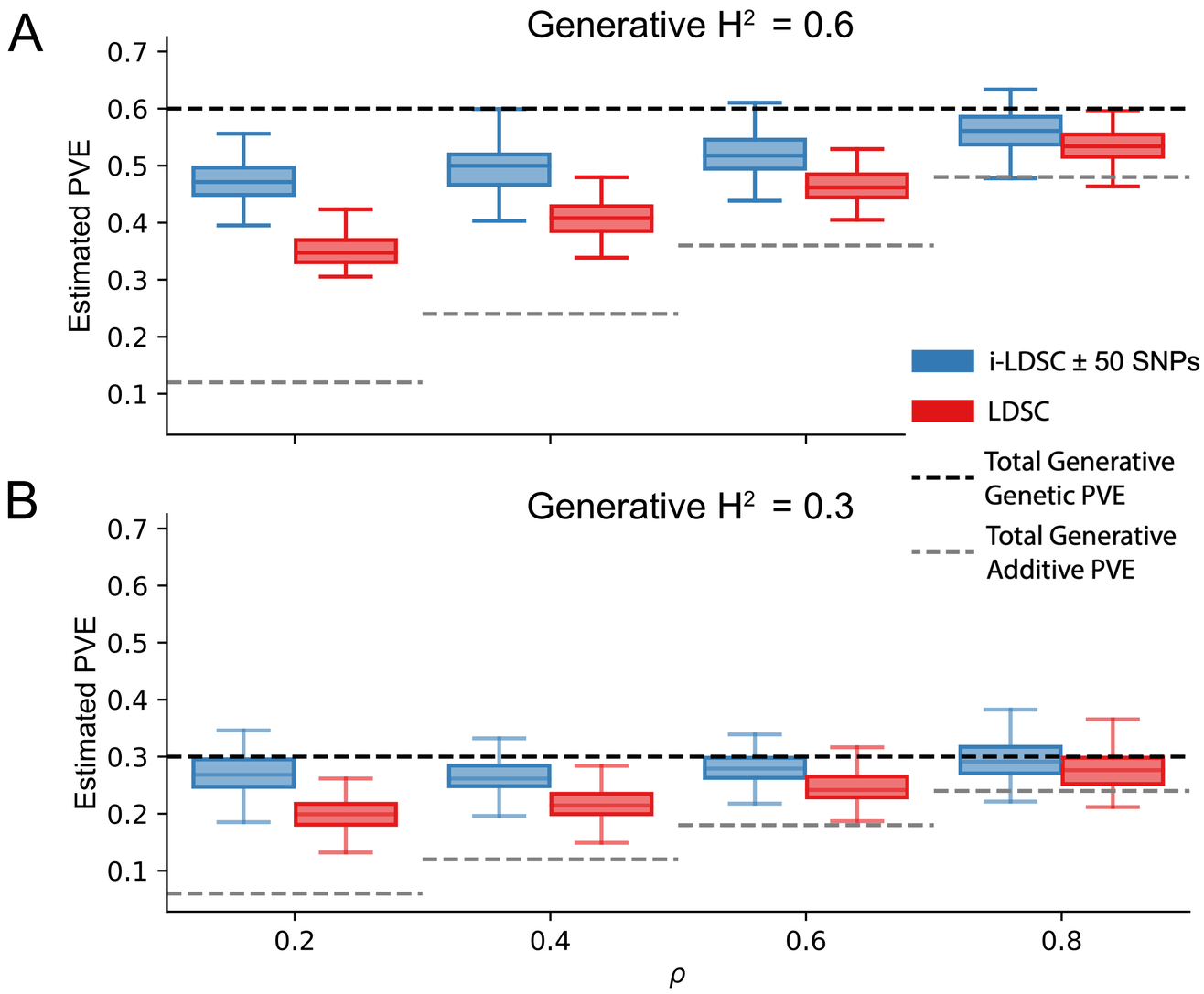
Bulik-Sullivan et al. (2015), *Nature Gen*

i-LDSC regression

Epistatic LD score improves estimate of narrow sense heritability

- Include marginal epistatic LD score f
- Regress on $\mathrm{E}(\chi^2) \sim \mathbf{1} + \ell\tau + f\sigma$

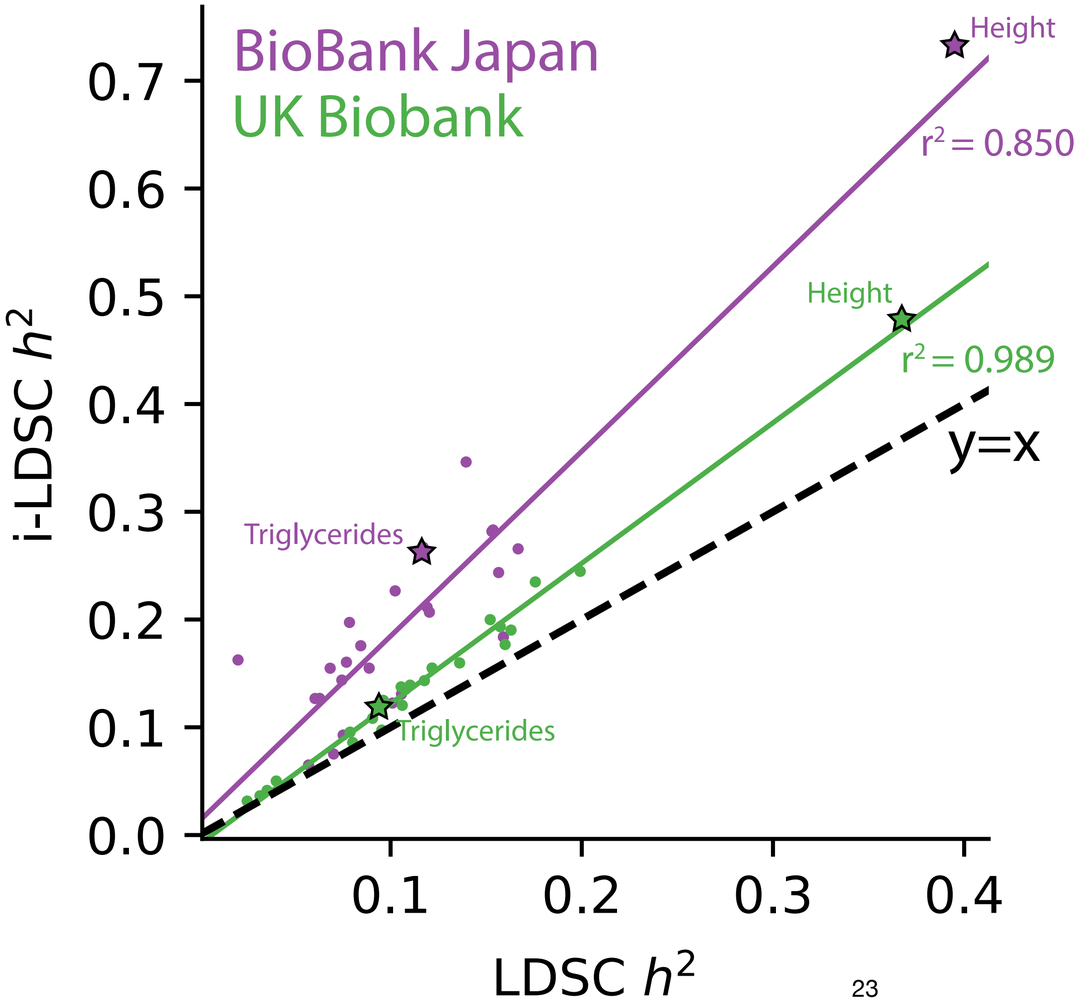




Smith, Darnell et al., *bioRxiv*

i-LDSC regression

Evidence of non-additive effects in human traits



LDSC = LD Score Regression

i-LDSC = Interaction-LD Score Regression

G. Darnell

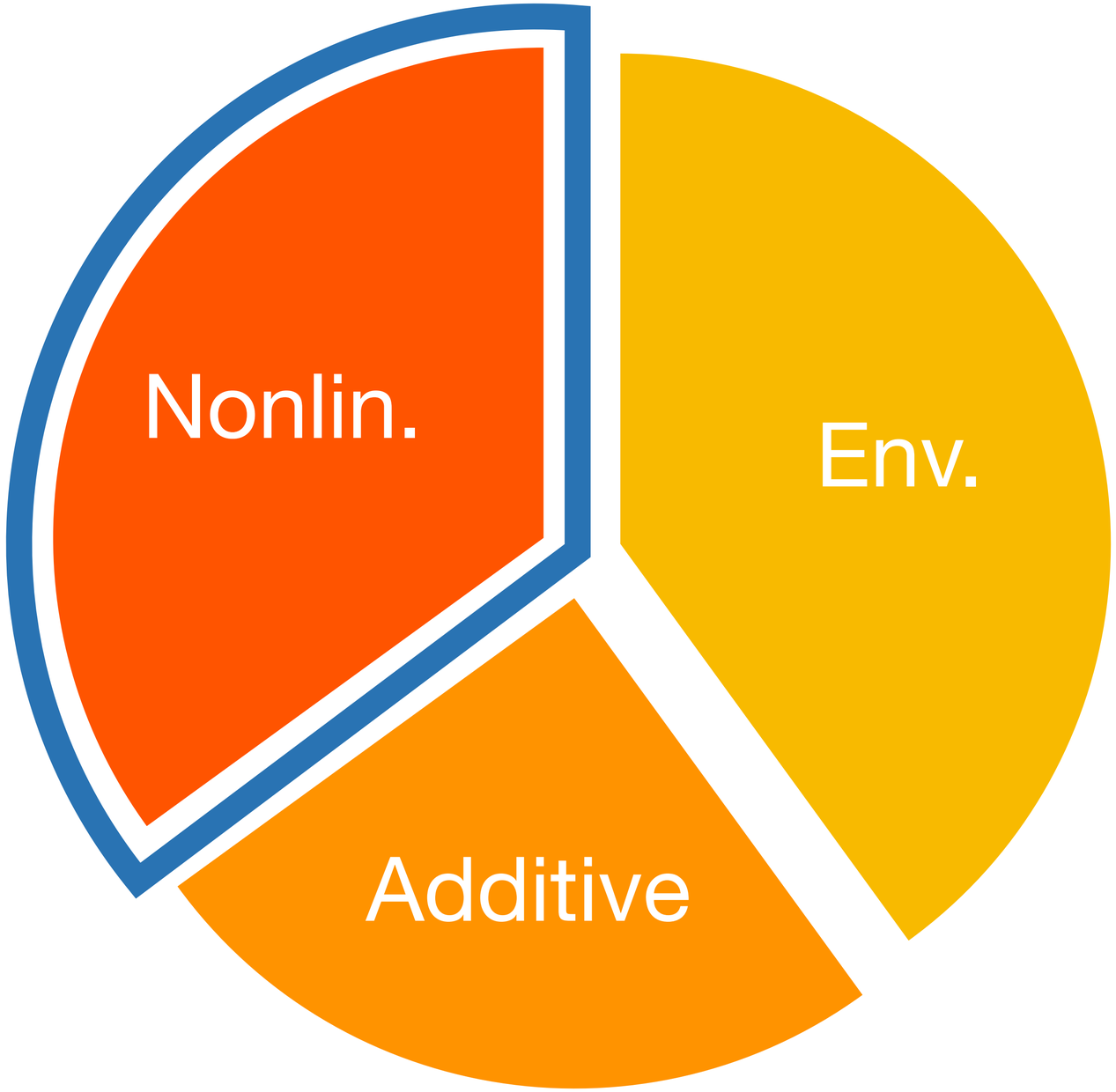
S. Smith

Smith, Darnell et al., *bioRxiv*

Non-additive variation of complex traits

Variance component partitioning improves detection of epistasis.

- Marginal epistasis addresses search space and small effect problem
- Modeling genetic correlations reveals pleiotropic trait architecture and improves sensitivity
- interaction-LD score regression reveals non-additive variation in human traits



Acknowledgements

Advisors

Lorin Crawford
Dan Weinreich

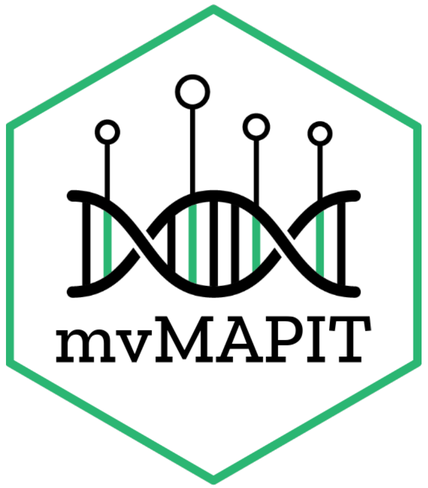
Crawford Lab and Dave

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) 2017 - 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

Interaction-LD Score Regression:

- G. Darnell*, S.P. Smith*, D. Udwin, S. Ramachandran, and L. Crawford. Partitioning tagged non-additive genetic effects in summary statistics provides evidence of pervasive epistasis in complex traits. bioRxiv. 2022.07.21.501001.

Related Software/Source Code:

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