G3, 2023, 13(8), jkad118
https://doi.org/10.1093/g3journal/jkad118
Advance Access Publication Date: 27 May 2023
Investigation

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

Julian Stamp,[1,*] Alan DenAdel,[1] Daniel Weinreich,[1,2] Lorin Crawford[1,3,4,*]

[1]Center for Computational Molecular Biology, Brown University, Providence, RI 02906, USA
[2]Department of Ecology, Evolution, and Organismal Biology, Brown University, Providence, RI 02906, USA
[3]Department of Biostatistics, Brown University, Providence, RI 02903, USA
[4]Microsoft Research New England, Cambridge, MA 02142, USA

*Corresponding author. Center for Computational Molecular Biology, Brown University, 164 Angell St, Box 1903, Providence, RI 02906, USA. Email: julian_stamp@brown.edu;
*Corresponding author. Center for Computational Molecular Biology, Brown University, 164 Angell St, Box 1903, Providence, RI 02906, USA; Microsoft Research New England, 1 Memorial Dr, Cambridge, MA 02142, USA. Email: lcrawford@microsoft.com

## Abstract

Epistasis, commonly defined as the interaction between genetic loci, is known to play an important role in the phenotypic variation of complex traits. As a result, many statistical methods have been developed to identify genetic variants that are involved in epistasis, and nearly all of these approaches carry out this task by focusing on analyzing one trait at a time. Previous studies have shown that jointly modeling multiple phenotypes can often dramatically increase statistical power for association mapping. In this study, we present the "multivariate MArginal ePIstasis Test" (mvMAPIT)—a multioutcome generalization of a recently proposed epistatic detection method which seeks to detect *marginal epistasis* or the combined pairwise interaction effects between a given variant and all other variants. By searching for marginal epistatic effects, one can identify genetic variants that are involved in epistasis without the need to identify the exact partners with which the variants interact—thus, potentially alleviating much of the statistical and computational burden associated with conventional explicit search-based methods. Our proposed mvMAPIT builds upon this strategy by taking advantage of correlation structure between traits to improve the identification of variants involved in epistasis. We formulate mvMAPIT as a multivariate linear mixed model and develop a multitrait variance component estimation algorithm for efficient parameter inference and *P*-value computation. Together with reasonable model approximations, our proposed approach is scalable to moderately sized genome-wide association studies. With simulations, we illustrate the benefits of mvMAPIT over univariate (or single-trait) epistatic mapping strategies. We also apply mvMAPIT framework to protein sequence data from two broadly neutralizing anti-influenza antibodies and approximately 2,000 heterogeneous stock of mice from the Wellcome Trust Centre for Human Genetics. The mvMAPIT `R` package can be downloaded at https://github.com/lcrawlab/mvMAPIT.

**Keywords:** epistasis, complex traits, covariance component analysis, multivariate linear mixed models

## Introduction

Genome-wide association (GWA) studies have contributed substantially in the discovery of genetic markers associated with the architecture of disease phenotypes (Ripke *et al.* 2014; Ellinghaus *et al.* 2016; Fuchsberger *et al.* 2016; Visscher *et al.* 2012, 2017; Loos 2020). Epistasis, commonly defined as the interaction between genetic loci, has long been thought to play a key role in defining the genetic architecture underlying many complex traits and common diseases (Tao *et al.* 2012; Wei *et al.* 2012; Kirino *et al.* 2013; Chen *et al.* 2017; Gusareva *et al.* 2018). Indeed, previous studies have detected pervasive epistasis in many model organisms (Peripato *et al.* 2004; Tong *et al.* 2004; Brem *et al.* 2005; Deutschbauer and Davis 2005; Kroymann and Mitchell-Olds 2005; Collins *et al.* 2006; Lehner *et al.* 2006; Onge *et al.* 2007; Wentzell *et al.* 2007; Shao *et al.* 2008; Flint and Mackay 2009; Gerke *et al.* 2009; Costanzo *et al.* 2010; He *et al.* 2010; Horn *et al.* 2011; Jarvis and Cheverud 2011; Leamy *et al.* 2011; Pettersson *et al.* 2011; Szappanos *et al.* 2011; Gaertner *et al.* 2012; Bloom *et al.* 2013; Chari and Dworkin 2013; Monnahan and Kelly 2015; Darnell *et al.* 2022). Substantial contributions of epistasis to phenotypic variance have been revealed for many complex traits (Huang and Mackay 2016; Crawford *et al.* 2018) and have been suggested to constitute an important component of evolution (Weinreich *et al.* 2006). Furthermore, modeling epistasis in addition to additive and dominant effects has been argued to increase phenotypic prediction accuracy in model organisms (Forsberg *et al.* 2017; Zhou *et al.* 2022) and facilitate genomic selection in breeding programs (Muñoz *et al.* 2014; Jiang and Reif 2015; Runcie *et al.* 2021). Despite a longstanding and currently ongoing debate about the contribution of nonadditive effects on the architecture of human complex traits (Long and Langley 1999; Marchini

**Received:** January 11, 2023. **Accepted:** May 23, 2023

et al. 2005; Hill et al. 2008; Shao et al. 2008; Crow 2010; Zuk et al. 2012; Hivert et al. 2021; Pazokitoroudi et al. 2021; Wainschtein et al. 2022), recent genetic mapping studies have also identified evidence of epistatic interactions that significantly contribute to quantitative traits and diseases (Brown et al. 2014; Fang et al. 2019; van de Haar et al. 2019; Sheppard et al. 2021), and some have recently argued that gene-by-gene interactions can drive heterogeneity of causal variant effect sizes across diverse human populations (Patel et al. 2022). Importantly, epistasis is often proposed as a key contributor to missing heritability—the proportion of heritability not explained by the top associated variants in GWA studies (Wanstrat and Wakeland 2001; Eichler et al. 2010; Gusareva et al. 2014; Gusareva and Van Steen 2014; Gusareva et al. 2018).

Many statistical methods have been developed to facilitate the identification of epistasis in complex traits and diseases. Generally, these existing tools can be classified into two frameworks. In the first framework, explicit searches are performed to detect significant pairwise or higher order interactions. More specifically, they use various strategies including exhaustive search (Purcell et al. 2007; Schüpbach et al. 2010; Ma et al. 2013), probabilistic search (Prabhu and Pe'er 2012), or prioritization based on a predefined set of biological annotations of signaling pathways or genomic regulatory units (Lewinger et al. 2013; Guo et al. 2021). Different statistical paradigms have been implemented for these explicit search-based approaches including various frequentist tests (Purcell et al. 2007; Wan et al. 2010; Lishout et al. 2013), Bayesian inference (Zhang and Liu 2007; Tang et al. 2009; Zhang et al. 2011; Guo et al. 2019), and, most recently, detecting epistasis using machine learning (Chang et al. 2020; Fergus et al. 2020). Indeed, the explosion of large-scale genomic data sets has provided the unique opportunity to integrate many of these techniques as standard statistical tools within GWA analyses. Many modern GWA applications have data sets that can include hundreds of thousands of individuals genotyped at millions of markers and phenotyped for thousands of traits (Nagai et al. 2017; Bycroft et al. 2018). Due to the potentially large space of genetic interactions (e.g. $J(J-1)/2$ possible pairwise combinations for $J$ variants in a study), explicit search-based methods often suffer from heavy computational burden. Even with various efficient computational improvements (Wan et al. 2010; Hemani et al. 2011; Prabhu and Pe'er 2012; Zhu and Fang 2018; Bayat et al. 2021), exploring over a large combinatorial domain remains a daunting task for many epistatic mapping studies. More importantly, because of a lack of a priori knowledge about epistatic loci, exploring all possible combinations of genetic variants can result in low statistical power after correcting for multiple hypothesis tests.

As a departure from the explicit search strategy, the second category of epistatic mapping methods attempts to address the previously mentioned challenges by detecting *marginal* epistasis. Specifically, instead of directly identifying individual pairwise or higher order interactions, these approaches focus on identifying variants that have a nonzero interaction effect with any other variant in the data set. For example, the "MArginal ePIstasis Test" (MAPIT) (Crawford et al. 2017) assesses each variant (in turn) and identifies candidate markers that are involved in epistasis without the need to identify the exact partners with which the variants interact—thus, alleviating much of the statistical power concerns and heavy computational burdens associated with explicit search-based methods. As a framework, the marginal epistatic strategy has been implemented in both linear mixed models and machine learning and has been used for case–control studies (Crawford and Zhou 2018), pathway enrichment applications (Turchin et al. 2020), heritability estimation (Darnell et al. 2022), and even extended to explore different sources of nonadditive genetic variation (e.g. gene-by-environment interactions) (Moore et al. 2019; Kerin and Marchini 2020b). However, despite its wide adoption, this approach can still be underpowered for traits with low heritability or "polygenic" traits which are generated by many mutations of small effect (Crawford et al. 2017).

To date, both the explicit search and marginal epistasis detection methodologies have only focused on analyzing one phenotype at a time. However, many previous genetic association studies have extensively shown that jointly modeling multiple phenotypes can often dramatically increase power for variant detection (Zhou and Stephens 2014). In this work, we present the "multivariate MArginal ePIstasis Test" (mvMAPIT)—a multioutcome generalization of the MAPIT model which aims to take advantage of the relationship between traits to improve the identification of variants involved in epistasis. We formulate mvMAPIT as a multivariate linear mixed model (mvLMM) and extend a previously developed variance component estimation algorithm for efficient parameter inference and $P$-value computation in the multitrait setting (Zhou 2017). Together with reasonable model approximations, our proposed approach is scalable to moderately sized GWA studies. With detailed simulations, we illustrate the benefits of mvMAPIT in terms of providing effective type I error control and compare its power to the univariate (or single-trait) mapping strategy used in the original MAPIT model. Here, part of our main contribution is the demonstration that the power in our proposed multivariate approach is driven by the correlations between the effects of pairwise interactions on multiple traits. To close, we also apply the mvMAPIT framework to protein sequence data from a nearly combinatorially complete library of two broadly neutralizing anti-influenza antibodies (Phillips et al. 2021) and to 15 quantitative hematology traits assayed in a heterogeneous stock of mice from Wellcome Trust Centre for Human Genetics (Valdar, Flint, et al. 2006; Valdar, Solberg, et al. 2006).

## Materials and methods

### The marginal epistasis test for single traits

The original motivation behind the original "MArginal ePIstasis Test" (MAPIT) was to identify variants that are involved in epistasis while avoiding the need to explicitly conduct an exhaustive search over all possible pairwise interactions (Crawford et al. 2017). In this section, we give a statistical overview of the univariate version of MAPIT where the objective is to search for marginal epistatic effects (i.e. the combined pairwise interaction effects between a given variant and all other variants) that drive the genetic architecture of single traits. To begin, consider a GWA study with $N$ individuals who have been genotyped for $J$ single nucleotide polymorphisms (SNPs) encoded as {0, 1, 2} copies of a reference allele at each locus. In the MAPIT framework, we examine one SNP at a time (indexed by $j$) and consider the following linear model:

$$\mathbf{y} = \mu + \mathbf{x}_j\beta_j + \sum_{l\neq j}\mathbf{x}_l\beta_l + \sum_{l\neq j}(\mathbf{x}_j \circ \mathbf{x}_l)\alpha_l + \varepsilon, \quad \varepsilon \sim \mathcal{N}(\mathbf{0}, \tau^2\mathbf{I}), \qquad (1)$$

where $\mathbf{y}$ is an $N$-dimensional vector of phenotypic states for a quantitative trait of interest measured in the $N$ individuals; $\mu$ is an intercept term; $\mathbf{X}$ denotes an $N \times J$ matrix of genotypes with $\mathbf{x}_j$ and $\mathbf{x}_l$ representing $N$-dimensional vectors for the $j$th and $l$th SNPs; $\beta_j$ and $\beta_l$ are the respective additive effects; $\mathbf{x}_j \circ \mathbf{x}_l$ denotes the Hadamard (elementwise) product between two genotypic

vectors with corresponding interaction effect size $\alpha_l$; $\varepsilon$ is a normally distributed error term with mean zero and scale variance term $\tau^2$; and $\mathbf{I}$ denotes an $N \times N$ identity matrix. For convenience, we will assume that both the genotype matrix (columnwise) and trait of interest have been mean-centered and standardized. It is also worth noting that, while we limit the above to the task of identifying second-order (i.e. pairwise) interactions between genetic variants, extensions of MAPIT to higher order epistatic and gene-by-environmental effects have been shown to be straightforward to implement (Moore *et al.* 2019; Kerin and Marchini 2020a, 2020b; Zhu *et al.* 2022).

### *Variance component model formulation*

Since many modern GWA applications present scenarios that would make equation (1) an underdetermined linear system (e.g. in modern biobank scale studies where genotyped markers $J > N$ individuals), the MAPIT framework follows other standard approaches (Yang *et al.* 2010; Wu *et al.* 2011; Zhou *et al.* 2013; Bulik-Sullivan *et al.* 2015; Crawford *et al.* 2017) to ensure model identifiability by assuming that the additive and interaction effect sizes follow univariate normal distributions where $\beta_l \sim \mathcal{N}(0, \omega^2/(J-1))$ and $\alpha_l \sim \mathcal{N}(0, \sigma^2/(J-1))$ for $l \neq j$, respectively. This key normal assumption on the regression coefficients allows for equation (1) to be equivalently represented as the following variance component model:

$$\mathbf{y} = \mu + \mathbf{x}_j\beta_j + \mathbf{m}_j + \mathbf{z}_j + \varepsilon, \quad \varepsilon \sim \mathcal{N}(\mathbf{0}, \tau^2\mathbf{I}), \tag{2}$$

where, in addition to previous notation, $\mathbf{m}_j = \sum_{l \neq j} \mathbf{x}_l\beta_l$ is the combined additive effects from all variants other than the $j$th and $\mathbf{z}_j = \sum_{l \neq j} (\mathbf{x}_j \circ \mathbf{x}_l)\alpha_l$ denote the summation of all pairwise interaction effects between the $j$th variant and all other variants. Under the variance component formulation in equation (2), the two random effects can also be expressed probabilistically as $\mathbf{m}_j \sim \mathcal{N}(\mathbf{0}, \omega^2\mathbf{K}_j)$ where $\mathbf{K}_j = \mathbf{X}_{-j}\mathbf{X}_{-j}^{\mathsf{T}}/(J-1)$ is an additive genetic relatedness matrix that is computed using all genotypes other than the $j$th SNP, and $\mathbf{z}_j \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{G}_j)$ where $\mathbf{G}_j = \mathbf{D}_j\mathbf{K}_j\mathbf{D}_j$ is a nonadditive relatedness matrix computed based on all pairwise interaction terms involving the $j$th SNP. Here, we let $\mathbf{D}_j = \mathrm{diag}(\mathbf{x}_j)$ denote an $N \times N$ diagonal matrix with the $j$th genotype as its only nonzero elements. It is also important to note that both $\mathbf{K}_j$ and $\mathbf{G}_j$ change with every new $j$th marker that is tested.

### *Univariate point estimates*

The key takeaway from the variance component model formulation in equation (2) is that $\sigma^2$ represents a measure on the marginal epistatic effect for each variant in the data. Therefore, in order to identify variants that have significant nonzero interaction effects, we must assess the null hypothesis $H_0 : \sigma^2 = 0$ for each variant in the data set. The original MAPIT framework uses a computationally efficient method of moments algorithm called MQS (Zhou 2017) to estimate model parameters and to carry out calibrated statistical tests. Briefly, MQS produces point estimates that are mathematically identical to the Haseman-Elston (HE) cross-product regression (Lee *et al.* 2011; Zhou 2017; Zhu and Zhou 2020). To implement this algorithm, we first specify a two-dimensional matrix $\mathbf{b}_j = [\mathbf{1}, \mathbf{x}_j]$ with $\mathbf{1}$ being an $N$-dimensional vector of ones. Next, we then multiply both sides of equation (2) by a variant-specific projection $\mathbf{P}_j = \mathbf{I} - \mathbf{b}_j(\mathbf{b}_j^{\mathsf{T}}\mathbf{b}_j)^{-1}\mathbf{b}_j^{\mathsf{T}}$ which maps the model onto a column space that is orthogonal to both the intercept and the genotypic vector $\mathbf{x}_j$. This process simplifies the model specification of MAPIT in equation (2) to the following

$$\begin{aligned} &\mathbf{y}_j^* = \mathbf{m}_j^* + \mathbf{z}_j^* + \varepsilon_j^*, \quad \mathbf{m}_j^* \sim \mathcal{N}(\mathbf{0}, \omega^2\mathbf{K}_j^*), \quad \mathbf{z}_j^* \sim \mathcal{N}(\mathbf{0}, \sigma^2, \mathbf{G}_j^*), \\ &\varepsilon_j^* \sim \mathcal{N}(\mathbf{0}, \tau^2\mathbf{P}_j), \end{aligned} \tag{3}$$

where we denote $\mathbf{y}_j^* = \mathbf{P}_j\mathbf{y}$; $\mathbf{m}_j^* = \mathbf{P}_j\mathbf{m}_j$; $\mathbf{K}_j^* = \mathbf{P}_j\mathbf{K}_j\mathbf{P}_j$; $\mathbf{z}_j^* = \mathbf{P}_j^*\mathbf{z}_j$; $\mathbf{G}_j^* = \mathbf{P}_j\mathbf{G}_j\mathbf{P}_j$; and $\varepsilon_j^* = \mathbf{P}_j\varepsilon$, respectively. The estimators for the variance components in equation (3) are naturally based on the second moment matching equations where, in expectation, we have

$$\mathbb{E}[\mathbf{y}_j^{*\mathsf{T}}\mathbf{H}\mathbf{y}_j] = \sum_{k=1}^{3} \mathrm{tr}(\mathbf{H}\boldsymbol{\Sigma}_{jk})\delta_k \tag{4}$$

with $\mathbf{H}$ being a symmetric and nonnegative definite matrix used to create weighted second moments, $\mathrm{tr}(\bullet)$ denotes the trace of a matrix, and we use shorthand to represent $[\boldsymbol{\Sigma}_{j1}; \boldsymbol{\Sigma}_{j2}; \boldsymbol{\Sigma}_{j3}] = [\mathbf{K}_j^*; \mathbf{G}_j^*; \mathbf{P}_j]$ and $\boldsymbol{\delta} = (\omega^2, \sigma^2, \tau^2)$, respectively. In practice, we replace the left hand side of equation (4) with the realized value $\mathbf{y}_j^{*\mathsf{T}}\mathbf{H}\mathbf{y}_j$. Note that many choices of $\mathbf{H}$ will yield unbiased estimates for $(\omega^2, \sigma^2, \tau^2)$, but different choices of $\mathbf{H}$ can affect statistical efficiency of the estimates. The set of moment matching equations in MQS is generated by using the covariance matrices corresponding to the variance components in place of the arbitrary $\mathbf{H}$. This system of equations can then be rewritten as the following matrix multiplication

$$\boldsymbol{\delta} = \mathbf{S}^{-1}\mathbf{q}, \quad q_k = \mathbf{y}_j^{*\mathsf{T}}\boldsymbol{\Sigma}_{jk}\mathbf{y}_j, \quad \mathbf{S}_{rs} = \mathrm{tr}(\boldsymbol{\Sigma}_{jr}\boldsymbol{\Sigma}_{js}), \tag{5}$$

where $\mathbf{q}$ is a three-dimensional vector and $\mathbf{S}$ is a $3 \times 3$-dimensional matrix with $k, r, s \in \{1, 2, 3\}$ being indices to represent the different variance components. If we subset just to compute an estimate for the marginal epistatic variance component (i.e. for the second index), then equation (5) reduces to the following formula

$$\widehat{\sigma}_j^2 = \mathbf{y}_j^{*\mathsf{T}}\mathbf{H}_j\mathbf{y}_j^*, \tag{6}$$

where the variant-specific matrix $\mathbf{H}_j = (\mathbf{S}^{-1})_{21}\mathbf{K}_j^* + (\mathbf{S}^{-1})_{22}\mathbf{G}_j^* + (\mathbf{S}^{-1})_{23}\mathbf{P}_j$ is now used in place of the arbitrary $\mathbf{H}$.

### *Univariate hypothesis testing*

In general, there are two ways to compute $P$-values in the MAPIT framework (Crawford *et al.* 2017). The first option uses a two-sided z-score or normal test. This particular test only requires the variance component estimate $\widehat{\sigma}_j^2$ from equation (6) and its corresponding standard error, which is approximated in the MQS approach as

$$\mathbb{V}[\widehat{\sigma}_j^2] \approx 2\mathbf{y}_j^{*\mathsf{T}}\mathbf{H}_j^{\mathsf{T}}\mathbf{V}_j\mathbf{H}_j\mathbf{y}_j^*, \tag{7}$$

where $\mathbf{V}_j = \widehat{\omega}_j^2\mathbf{K}_j^* + \widehat{\sigma}_j^2\mathbf{G}_j^* + \widehat{\tau}_j^2\mathbf{P}_j$. The second option for deriving $P$-values in the MAPIT framework uses an exact test which is based on the fact that the MQS variance component estimate follows a mixture of chi-square distributions under the null hypothesis. This is derived from both the normality assumption on $\mathbf{y}^*$ and the quadratic form of the statistic in equation (6). Namely, $\widehat{\sigma}_j^2 \sim \sum_{i=1}^{N} \lambda_i\chi_{1,i}^2$ where $\chi_1^2$ are chi-square random variables with one degree of freedom and $(\lambda_1, \ldots, \lambda_N)$ are the eigenvalues of the matrix

$$(\widehat{\omega}_0^2\mathbf{K}_j^* + \widehat{\tau}_0^2\mathbf{P}_j)^{1/2}\mathbf{H}_j(\widehat{\omega}_0^2\mathbf{K}_j^* + \widehat{\tau}_0^2\mathbf{P}_j)^{1/2} \tag{8}$$

with $(\widehat{\omega}_0^2, \widehat{\tau}_0^2)$ being the MQS estimates of $(\widehat{\omega}^2, \widehat{\tau}^2)$ under the null hypothesis $H_0 : \sigma^2 = 0$. Several approaches have been proposed to obtain $P$-values under a mixture of chi-square distributions, including the Davies method (Davies 1980) (see Data and Software Availability). In practice, while the Davies method is an exact test and is expected to produce calibrated $P$-values, it can become computationally intensive since it scales cubically in the number of individuals $N$. On the other hand, while the normal test only scales quadratically in $N$ because of the standard error approximation in equation (7), it has been shown to lead to miscalibrated $P$-values for data sets with small sample sizes. As a result, MAPIT uses a hybrid procedure which implements the normal test by default, and then applies the Davies method when the $P$-value from the normal test is below the nominal threshold of 0.05 (Crawford *et al.* 2017).

## Derivation of the multivariate marginal epistasis test

The "multivariate MArginal ePIstasis Test" (mvMAPIT) is a multioutcome extension of the statistical framework MAPIT which aims to identify variants that are involved in epistatic interactions by leveraging the covariance structure of nonadditive genetic variation that is shared between multiple traits. Once again, consider a GWA study with $N$ individuals—however, this time, assume that each observation has been measured for $D$ different phenotypes. We will denote these sets of outcomes via a $D \times N$-dimensional matrix $\mathbf{Y} = [\mathbf{y}_1^\intercal, \ldots, \mathbf{y}_D^\intercal]$ with $\mathbf{y}_d$ denoting an $N$-dimensional phenotypic vector for the $d$th trait. Given the $j$th variant of interest, we specify the mvMAPIT approach as the following multivariate linear mixed model (mvLMM) (Zhou and Stephens 2014)

$$\mathbf{Y} = \mathbf{U} + \boldsymbol{\beta}_j \mathbf{x}_j^\intercal + \sum_{l \neq j} \boldsymbol{\beta}_l \mathbf{x}_l^\intercal + \sum_{l \neq j} \boldsymbol{\alpha}_l (\mathbf{x}_j \circ \mathbf{x}_l)^\intercal + \mathbf{E} \quad \mathbf{E} \sim \mathcal{MN}(\mathbf{0}, \mathbf{V}_\varepsilon, \mathbf{I}), \quad (9)$$

where, in addition to previous notation, $\mathbf{U}$ is a $D \times N$-dimensional matrix which contains population-level intercepts that are the same for all individuals within each trait; $\boldsymbol{\beta}_j$ and $\boldsymbol{\beta}_l$ are $D$-dimensional vectors of additive effects for the $j$th and $l$th genotypic vectors; $\boldsymbol{\alpha}_l$ is a $D$-dimensional vector of coefficients for the interaction effects between the $j$th and $l$th SNPs spanning all traits; and $\mathbf{E}$ denotes an $D \times N$ matrix of residual errors that is assumed to follow a matrix-variate normal distribution with mean $\mathbf{0}$, within column covariance $\mathbf{V}_\varepsilon$ among the $D$ traits, and independent within row covariance among the $N$ individuals in the study.

Similar to the univariate setting, we need to make additional probabilistic assumptions to ensure model identifiability when equation (9) is an underdetermined linear system. To that end, let $\mathbf{B} = [\boldsymbol{\beta}_l]_{l \neq j}$ and $\mathbf{A} = [\boldsymbol{\alpha}_l]_{l \neq j}$ denote the collection of coefficients not involving the $j$th variant of interest. Here, we will assume that these $D \times (J-1)$ effect size matrices also follow matrix-variate normal distributions where $\mathbf{B} \sim \mathcal{MN}(\mathbf{0}, \mathbf{V}_\beta, \mathbf{I})$ and $\mathbf{A} \sim \mathcal{MN}(\mathbf{0}, \mathbf{V}_\alpha, \mathbf{I})$, respectively. Note that this formulation is largely similar to the univariate case except with the additional property that the phenotypes being studied share some genetic covariance through $\mathbf{V}_\beta$ and $\mathbf{V}_\alpha$. This assumption, coupled with the affine transformation property of matrix normal distributions, allows for us to equivalently represent the mvMAPIT model in equation (9) as the following multivariate variance component model:

$$\mathbf{Y} + \mathbf{U} + \boldsymbol{\beta}_j \mathbf{x}_j^\intercal + \mathbf{M}_j + \mathbf{Z}_j + \mathbf{E}, \quad \mathbf{E} \sim \mathcal{MN}(\mathbf{0}, \mathbf{V}_\varepsilon, \mathbf{I}), \quad (10)$$

where $\mathbf{M}_j = \sum_{l \neq j} \boldsymbol{\beta}_l \mathbf{x}_l^\intercal$ with $\mathbf{M}_j \sim \mathcal{MN}(\mathbf{0}, \mathbf{V}_\beta, \mathbf{K}_j)$ represents the combined additive effects from all other variants across the $D$ traits and $\mathbf{Z}_j = \sum_{l \neq j} \boldsymbol{\alpha}_l (\mathbf{x}_j \circ \mathbf{x}_l)^\intercal$ with $\mathbf{Z}_j \sim \mathcal{MN}(\mathbf{0}, \mathbf{V}_\alpha, \mathbf{G}_j)$ encodes all pairwise interaction terms involving the $j$th SNP across the $D$ traits. Here, the term $\mathbf{Z}_j$ becomes the main focus of model inference.

As we will show below, mvMAPIT works by analyzing pairs of traits at a time; this allows the framework to be applied to any number of measured phenotypes. Analyzing more traits simply requires more computational resources both in terms of wall-clock time and computer memory. For each point estimate, mvMAPIT performs matrix operations that scale quadratically with sample size. The software also needs to store covariance matrices corresponding to the number of random effects in the model. Both these added costs scale as $D(D+1)/2$ for $D$ traits. If one were to also consider higher order interactions, an additional resource burden would come from requiring additional covariance matrices to be stored as well as projecting these covariance matrices onto a space that is orthogonal to the variant of interest and the population intercept. In the multivariate setting, the time complexity of the projection step scales on the order of $DN^2$ with again $N$ being the number of samples in the data.

## Hypothesis testing in the mvMAPIT framework

The goal of the mvMAPIT framework still comes down to assessing the null hypothesis that tests for nonzero marginal epistatic effects. However, parameter estimation in mvLMMs can present substantial computational challenges. For example, one common way in the literature to rewrite the model specified in equation (10) is to vectorize (or stack) the columns of each matrix in the regression such that $\mathbf{y} = \text{vec}(\mathbf{Y})$, $\boldsymbol{\mu} = \text{vec}(\mathbf{U})$, $\mathbf{m}_j = \text{vec}(\mathbf{M}_j)$, $\mathbf{z}_j = \text{vec}(\mathbf{Z}_j)$, and $\varepsilon = \text{vec}(\mathbf{E})$. Under this reformulation, we could simply follow the procedures in equations (3)–(8) to find significant variance components; but since $\mathbb{V}[\mathbf{m}_j] = \mathbf{K}_j \otimes \mathbf{V}_\beta$ and $\mathbb{V}[\mathbf{z}_j] = \mathbf{G}_j \otimes \mathbf{V}_\alpha$ are each $ND \times ND$ dimensions (via the Kronecker product $\otimes$), the periterative computation time for performing hypothesis testing on each $j$th SNP would now increase both with the number of individuals ($N$) and with the number of phenotypes ($D$). This could make model fitting infeasible for large biobanks even with only two traits. As an alternative, we present a combinatorial approach which first fits univariate MAPIT models and then combines the resulting $P$-values with those stemming from a "covariance statistic" which looks for shared marginal epistatic effects between all pairwise combinations of the $D$ traits. Importantly, our combinatorial approach does not make assumptions about the covariance structure between traits, which would need to be known (or assumed) in the Kronecker formulation.

To implement the multivariate marginal epistasis test, we follow a similar strategy used in the univariate MAPIT model and right multiply equation (10) by a variant-specific projection $\mathbf{P}_j = \mathbf{I} - \mathbf{b}_j(\mathbf{b}_j^\intercal \mathbf{b}_j)^{-1}\mathbf{b}_j^\intercal$ which maps the model onto a column space that is orthogonal to the population-level intercepts and the genotypic vector $\mathbf{x}_j$. This results in a simplified mvLMM of the following form:

$$\mathbf{Y}_j^* = \mathbf{M}_j^* + \mathbf{Z}_j^* + \mathbf{E}_j^*, \quad \mathbf{E}_j^* \sim \mathcal{MN}(\mathbf{0}, \mathbf{V}_\varepsilon, \mathbf{P}_j), \quad (11)$$

where, in addition to previous notation, $\mathbf{Y}_j^* = \mathbf{Y}\mathbf{P}_j$; $\mathbf{M}_j^* = \mathbf{M}_j\mathbf{P}_j$; $\mathbf{Z}_j^* = \mathbf{Z}_j\mathbf{P}_j$; and $\mathbf{E}_j^* = \mathbf{E}_j\mathbf{P}_j$, respectively. Probabilistically, this transformation assumes $\mathbf{M}_j^* \sim \mathcal{MN}(\mathbf{0}, \mathbf{V}_\beta, \mathbf{K}_j^*)$ with $\mathbf{K}_j^* = \mathbf{P}_j\mathbf{K}_j\mathbf{P}_j$; and $\mathbf{Z}_j^* \sim \mathcal{MN}(\mathbf{0}, \mathbf{V}_\alpha, \mathbf{G}_j^*)$ with $\mathbf{G}_j^* = \mathbf{P}_j\mathbf{G}_j\mathbf{P}_j$. The joint analysis of multiple outcomes requires a generalization of the MQS algorithm to also include moment estimates for the covariance components

between traits. Let $\mathbf{y}_c^*$ and $\mathbf{y}_d^*$ be the $c$th and $d$th rows of the adjusted phenotypic matrix $\mathbf{Y}_j^*$, respectively. The general MQS estimates for the marginal epistatic effect is a generalization of equation (6) which is given in the following quadratic form:

$$\widehat{\sigma}^2_{j,(cd)} = \mathbf{y}_c^{*\intercal}\mathbf{H}_j\mathbf{y}_d^*, \tag{12}$$

where $\mathbf{H}_j$ is as previously defined in the univariate MAPIT case and the trait-specific indices span between the $c, d \in 1, \ldots, D$ phenotypes. Here, when $c = d$, the above is exactly equal to equation (6) and the variance component point estimate is computed using only one trait row in $\mathbf{Y}_j^*$. On the other hand, when $c \neq d$, equation (12) takes on a bilinear form where $\mathbb{E}[\mathbf{y}_c^{*\intercal}\mathbf{H}_j\mathbf{y}_d^*] = \mathrm{tr}(\mathbf{H}_j\mathbf{V}_{j,(cd)})$ with $\mathbf{V}_{j,(cd)} = \mathbb{V}[\mathbf{y}_c^*, \mathbf{y}_d^*]$ being the covariance between any two traits of interest. The corresponding standard error of the bilinear covariance component can then be estimated via the following approximation (Searle 1979):

$$\mathbb{V}[\widehat{\sigma}^2_{j,(cd)}] \approx \mathbf{y}_c^{*\intercal}\mathbf{H}_j^\intercal\mathbf{V}_{j,(cd)}\mathbf{H}_j\mathbf{y}_d^* + \mathbf{y}_c^{*\intercal}\mathbf{H}_j^\intercal\mathbf{V}_{j,(dd)}\mathbf{H}_j\mathbf{y}_c^*. \tag{13}$$

Once again, notice that when $c = d$, the term $\mathbf{V}_{j,(cd)} = \mathbf{V}_{j,(dd)}$ and the above approximation in equation (13) is equal to equation (7).

The combinatorial hypothesis testing procedure that is used in mvMAPIT occurs in three key steps:

1) In the first step, the model fits univariate models for all $D$ traits of interests (i.e. using equations (3)–(8) from the MAPIT model or equivalently equations (12) and (13) with $c = d$). Here, we use the proposed hybrid testing approach where we first implement a normal test by default, and then apply the exact Davies method when the $P$-value from the normal test is below the nominal significance threshold of 0.05 (Crawford *et al.* 2017).
2) In the second step, we derive $P$-values for the covariance components (i.e. using equations (12) and (13) when $c \neq d$) with a normal test. As we will show below, the $P$-values derived for the covariance components with the asymptotic normal approximation tend to be slightly deflated under the null hypothesis. While this leads to generally conservative behavior with respect to type I error control, the downside is that the test may result in reduced power under the alternative, especially after multiple correction for data sets with small sample sizes or for traits that have low genetic correlation. In these cases, deriving an exact test to obtain more calibrated $P$-values could be done; however, we do not explore this line of work here.
3) In the third and final step, mvMAPIT combines the $P$-values from the first two steps into an overall marginal epistatic $P$-value. Assume that we have $T = 3$ sets of $P$-values (two sets corresponding to marginal effects and one covariance set). The mvMAPIT software carries out the $P$-value combining procedure in three different ways. The first assumes that each of the $t = 1, \ldots, T$ tests are (effectively) independent and implements Fisher's method (Fisher 1925) which combines $P$-values into a single chi-square test statistic using the formula $\chi^2_{2T} \sim -2\sum_{t=1}^{T}\log(p_t)$, where $p_t$ denotes the $P$-value from the $t$th test. In Fisher's method, the $\chi^2$ test statistic will be large when $P$-values tend to be small (i.e. when the null hypothesis is not true for every test). The second approach assumes an unknown dependency structure between each of the $T$ tests and computes a harmonic mean (Wilson 2019a) $P$-value, where $\mathring{p} = \sum_t w_t / \sum_t w_t/p_t$. Here, $\sum_t w_t = 1$ are weights which we uniformly set to be $w_t = 1/T$ for all $P$-values. The last approach implements the Cauchy combination test which produces analytic $P$-values for any arbitrary dependency structure (Liu and Xie 2020). Under this method, the Cauchy combination test statistic is defined as $CCT = \sum_t w_t \tan\{(0.5 - p_t)\pi\}$, where $\tan(\bullet)$ is the tangent function and again we set $w_t = 1/T$ to be uniform for all $P$-values. The test statistic follows a standard Cauchy distribution $CCT \sim C(0, 1)$.

In practice, epistatic effects are assumed to make small contributions to the overall broad-sense heritability of complex traits (Hivert *et al.* 2021; Pazokitoroudi *et al.* 2021; Wainschtein *et al.* 2022). As a result, detecting associated variants that significantly contribute to nonadditive variation can be difficult. Intuitively, this combinatorial approach is meant to aggregate over the signal identified in both the marginal and covariance tests to improve power. In our results below, we show that Fisher's method, the harmonic mean, and the Cauchy combination test approach are well calibrated under the null hypothesis (i.e. only additive effects for all traits analyzed) and increase the ability to detect marginal epistatic variants under the alternative in both simulations and real data.

### *Note on settings where mvMAPIT is designed to be most powered*

The formulation of the general estimates in equations (12) and (13) highlight an important takeaway in that the mvMAPIT covariance statistic models epistatic pairs that together affect the architecture of multiple traits. It is not meant to identify individual SNPs that are involved in epistasis for multiple traits while being a member of different interacting pairs. To clarify this, consider two simple scenarios in Fig. 1 where we have two phenotypes ($\mathbf{y}_1$ and $\mathbf{y}_2$) that are generated by a combination of four SNPs ($\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$). In the first scenario, we say that (in expectation) $\mathbb{E}[\mathbf{y}_1] = \mathbf{x}_1\beta_1 + (\mathbf{x}_2 \circ \mathbf{x}_3)\alpha_1$ and $\mathbb{E}[\mathbf{y}_2] = (\mathbf{x}_2 \circ \mathbf{x}_3)\alpha_2$ (Fig. 1a), while in the second scenario, $\mathbb{E}[\mathbf{y}_1] = \mathbf{x}_1\beta_1 + (\mathbf{x}_2 \circ \mathbf{x}_3)\alpha_1$ and $\mathbb{E}[\mathbf{y}_2] = (\mathbf{x}_3 \circ \mathbf{x}_4)\alpha_2$ (Fig. 1b). The key to power in the mvMAPIT framework is that, in the first scenario, the interaction between $\mathbf{x}_2$ and $\mathbf{x}_3$ appears in both traits with nonzero correlation between the effect sizes $\alpha_1$ and $\alpha_2$. This is in contrast to the second scenario where there is a common variant involved in epistasis but it is a member of two different sets of interactions that affect each trait. The mvMAPIT covariance statistic captures the situation illustrated in the first scenario (Fig. 1a) but not in the second (Fig. 1b).

## Simulation study design

To test the utility of the mvMAPIT framework, we modified a frequently used simulation scheme (Crawford *et al.* 2017; Darnell *et al.* 2022) to generate collections of synthetic quantitative traits under multiple genetic architectures using real genotypes from chromosome 22 of the control samples in the Wellcome Trust Case Control Consortium (WTCCC) 1 study. After preprocessing, considering this particular group of individuals and SNPs resulted in a data set consisting of $N = 2,938$ individuals and $J = 5,747$ markers. In these simulations, we randomly choose 1,000 causal SNPs to directly affect $D = 2$ phenotypes. We generate these synthetic traits via the following general multivariate linear model:

$$\mathbf{Y} = \sum_{g \in \mathcal{G}} \boldsymbol{\beta}_g \mathbf{x}_g^\intercal + \mathbf{A}\mathbf{W}^\intercal + \mathbf{E}, \quad \mathbf{E} \sim \mathcal{MN}(\mathbf{0}, \mathbf{I}, \mathbf{I}), \tag{14}$$

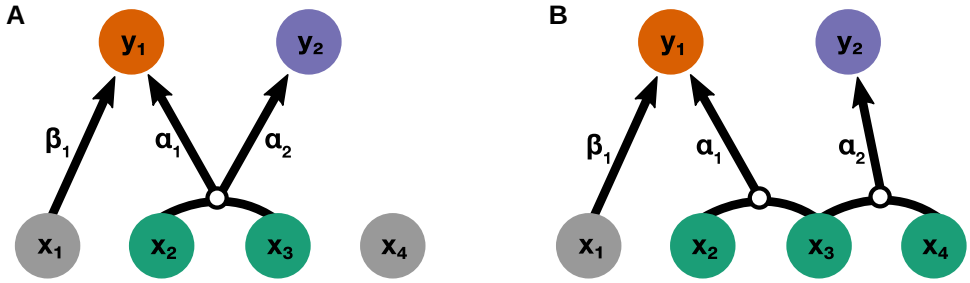


**Fig. 1.** Schematic of the types of shared interactions modeled by the multivariate marginal epistasis test. Consider two simple, proof-of-concept simulation scenarios where two traits ($\mathbf{y}_1$, $\mathbf{y}_2$) are generated by a combination of four SNPs ($\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{x}_4$). a) First scenario where (in expectation) $\mathbb{E}[\mathbf{y}_1] = \mathbf{x}_1\beta_1 + (\mathbf{x}_2 \circ \mathbf{x}_3)\alpha_1$ and $\mathbb{E}[\mathbf{y}_2] = (\mathbf{x}_2 \circ \mathbf{x}_3)\alpha_2$. b) Second scenario where $\mathbb{E}[\mathbf{y}_1] = \mathbf{x}_1\beta_1 + (\mathbf{x}_2 \circ \mathbf{x}_3)\alpha_1$ and $\mathbb{E}[\mathbf{y}_2] = (\mathbf{x}_3 \circ \mathbf{x}_4)\alpha_2$. In both panels, variant $\mathbf{x}_1$ only has an additive effect $\beta_1$ on trait $\mathbf{y}_1$. The mvMAPIT approach models correlations between the effects of a given interaction on multiple traits. Therefore, mvMAPIT is designed to identify SNPs involved in the first scenario where the interaction between variants $\mathbf{x}_2$ and $\mathbf{x}_3$ is shared between traits with nonzero correlated effect sizes $\alpha_1$ and $\alpha_2$. This is in contrast to the second case, where variant $\mathbf{x}_3$ is important to both traits but through distinct interactions with variants $\mathbf{x}_2$ and $\mathbf{x}_4$, respectively.

where $\mathbf{Y}$ is an $D \times N$ matrix containing all the phenotypes; $\mathcal{G}$ represents the set of 1,000 causal SNPs; $\mathbf{x}_g$ is the genotype for the $g$th causal SNP encoded as 0, 1, or 2 copies of a reference allele; $\boldsymbol{\beta}_g$ is a $D$-dimensional vector and represent the additive effect sizes for the $g$th SNP in the $D$ traits; $\mathbf{W}$ is an $N \times M$ matrix which holds pairwise interactions (i.e. Hadamard products) between some subset of causal SNPs; $\mathbf{A} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M]$ is a $D \times M$ matrix of interaction effect sizes with $\boldsymbol{\alpha}_m$ being $D$-dimensional epistatic coefficients for the $m$th interaction in the $d$th trait; and $\mathbf{E}$ is an $D \times N$ matrix of normally distributed environmental noise.

In these studies, we assume that the total phenotypic variances for both traits in $\mathbf{Y}$ are set to be 1. The additive and interaction effect sizes for causal SNPs are randomly drawn from matrix normal distributions where we control the correlation of effects between traits. This simplifies to us drawing coefficients as

$$\boldsymbol{\beta}_g \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_\beta), \quad \boldsymbol{\alpha}_m \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_\alpha), \tag{15}$$

where $\mathbf{V}_\beta$ and $\mathbf{V}_\alpha$ are $D \times D$ covariance matrices for additive effects and pairwise interactions between the phenotypes. Once these coefficients are sampled, we rescale them so that they explain a fixed proportion of the broad-sense heritability $H^2$. Similarly, the environmental noise matrix is rescaled such that it explains $1 - H^2$. When generating synthetic traits, we assume that the additive effects make up $\rho$% of the broad-sense heritability, while the pairwise interactions make up the remaining $(1-\rho)$%. Alternatively, we say that the proportion of the heritability explained by additivity is $\rho H^2$, while the proportion of phenotypic variance explained by pairwise interactions is $(1-\rho)H^2$. Setting $\rho = 1$ represents the null model where the variation of a trait is driven by solely additive effects. Here, we use the same simulation strategy used in previous studies (Crawford *et al.* 2017; Darnell *et al.* 2022) where we divide the causal variants into three groups where:

- $\mathcal{G}_1$ is a small number of SNPs with both additive and epistatic effects;
- $\mathcal{G}_2$ is a larger number of SNPs with both additive and epistatic effects;
- $\mathcal{G}_3$ is a large number of SNPs with only additive effects.

Here, the epistatic causal SNPs interact between sets. All SNPs in $\mathcal{G}_1$ interact with all SNPs in the $\mathcal{G}_2$, but do not interact with variants in their own group (with the same rule applies to the second group). Again, each interaction between SNPs in the two groups has an epistatic effect size $\alpha_d$ for the $d$th trait (see Fig. 1a). The correlation of epistatic effects between traits is determined via the correlation coefficient $\upsilon_{\alpha,12}$ (i.e. the off-diagonal elements of $\mathbf{V}_\alpha$). With this set up, one can think of the SNPs assigned to $\mathcal{G}_1$ as being the "hub nodes" in an interaction network. Note that we use this setup because it has been shown that the ability to detect two interacting variants depends on the proportion of phenotypic variance that they marginally explain. For example, in our case, this means that power is expected to depend on $\mathbb{V}[\mathbf{W}\boldsymbol{\alpha}]/|\mathcal{G}_1|$ and $\mathbb{V}[\mathbf{W}\boldsymbol{\alpha}]/|\mathcal{G}_2|$ for groups 1 and 2, respectively, where $|\mathcal{G}|$ denotes the cardinality of the set. Given different parameters for the generative model in equation (14), we simulate data mirroring a wide range of genetic architectures by varying the following parameters:

- broad-sense heritability: $H^2 = 0.3$ and 0.6;
- proportion of phenotypic variation that is explained by additive effects: $\rho = 0.5$, 0.8, and 1;
- causal SNPs in each of the three groups: $\{|\mathcal{G}_1|, |\mathcal{G}_2|, |\mathcal{G}_3|\} = \{10, 10, 980\}$ and $\{10, 20, 970\}$;
- correlation between additive effects: $\upsilon_{\beta,12} = 0$, 0.8, and 1;
- correlation between epistatic effects: $\upsilon_{\alpha,12} = 0$ and 0.8.

All figures and tables show the mean performances (and standard errors) for each parameter combination across 100 simulated replicates.

## Preprocessing of the heterogeneous stock of mice data set

As part of the analyses, this work makes use of GWA data from the Wellcome Trust Centre for Human Genetics (Valdar, Flint, *et al.* 2006; Valdar, Solberg, *et al.* 2006) (http://mtweb.cs.ucl.ac.uk/mus/www/mouse/index.shtml). The genotypes from this study were downloaded directly using the BGLR-R package (Perez and de los Campos 2014). This study contains $N = 1,814$ heterogeneous stock of mice from 85 families (all descending from eight inbred progenitor strains) (Valdar, Flint, *et al.* 2006; Valdar, Solberg, *et al.* 2006), and 131 quantitative traits that are classified into 6 broad categories including behavior, diabetes, asthma, immunology, hematology, and biochemistry. Phenotypic measurements for these mice can be found freely available online to download (details can be found at http://mtweb.cs.ucl.ac.uk/mus/www/mouse/HS/index.shtml and https://github.com/lcrawlab/mvMAPIT). In the analyses below, we focused on 15 hematological phenotypes including:

atypical lymphocytes (ALY; `Haem.ALYabs`), basophils (BAS; `Haem.BASabs`), hematocrit (HCT; `Haem.HCT`), hemoglobin (HGB; `Haem.HGB`), large immature cells (LIC; `Haem.LICabs`), lymphocytes (LYM; `Haem.LYMabs`), mean corpuscular hemoglobin (MCH; `Haem.MCH`), mean corpuscular volume (MCV; `Haem.MCV`), monocytes (MON; `Haem.MONabs`), mean platelet volume (MPV; `Haem.MPV`), neutrophils (NEU; `Haem.NEUabs`), plateletcrit (PCT; `Haem.PCT`), platelets (PLT; `Haem.PLT`), red blood cell count (RBC; `Haem.RBC`), red cell distribution width (RDW; `Haem.RDW`), and white blood cell count (WBC; `Haem.WBC`). All phenotypes were previously corrected for sex, age, body weight, season, year, and cage effects (Valdar, Flint, *et al.* 2006; Valdar, Solberg, *et al.* 2006). For individuals with missing genotypes, we imputed values by the mean genotype of that SNP in their corresponding family. Only polymorphic SNPs with minor allele frequency above 5% were kept for the analyses. This left a total of $J = 10,227$ autosomal SNPs that were available for all mice.

## Results

### mvMAPIT produces calibrated *P*-values and conservative type I error rates

In this section, we make use of a previously described simulation scheme (Crawford *et al.* 2017; Darnell *et al.* 2022) in order to investigate whether mvMAPIT and its combinatorial inference approach preserves the desired type I error rate and produces well-calibrated *P*-values under the null hypothesis. Here, we generate synthetic phenotypes using real genotypes from the 22nd chromosome of the control samples in the WTCCC 1 study (Burton *et al.* 2007). Altogether, these data consist of $N = 2,938$ individuals and $J = 5,747$ SNPs. Since the goal of mvMAPIT is to search for variants involved in epistatic interactions, we consider the null model to be satisfied when the phenotypic variation of the synthetic traits are solely driven by additive effects. Here, we first subsample the genotypes for $N = 1,000$, 1,750, and 2,500 observations. Next, we randomly select 1,000 causal SNPs and simulate continuous phenotypes by using the linear model $\mathbf{Y} = \mathbf{B}\mathbf{X}^{\intercal} + \mathbf{E}$. The additive effect sizes for each causal SNP are drawn as $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_{\beta})$ across traits, and then we scale all terms to ensure a narrow-sense heritability of 60%. In these simulations, we vary the correlation of the additive genetic effects such that we have traits with independent additive effects ($\upsilon_{\beta,12} = 0$), traits with highly correlated additive effects ($\upsilon_{\beta,12} = 0.8$), and traits with perfectly correlated additive effects ($\upsilon_{\beta,12} = 1$). We assess the calibration of the *P*-values that are produced by mvMAPIT during each of the three key steps in its combinatorial hypothesis testing procedure. That is, we evaluate (1) the *P*-values resulting from the univariate test on each trait, (2) the *P*-values derived from the covariance test, and (3) the final overall *P*-value that is computed by combining the first two sets of *P*-values via Fisher's method, the harmonic

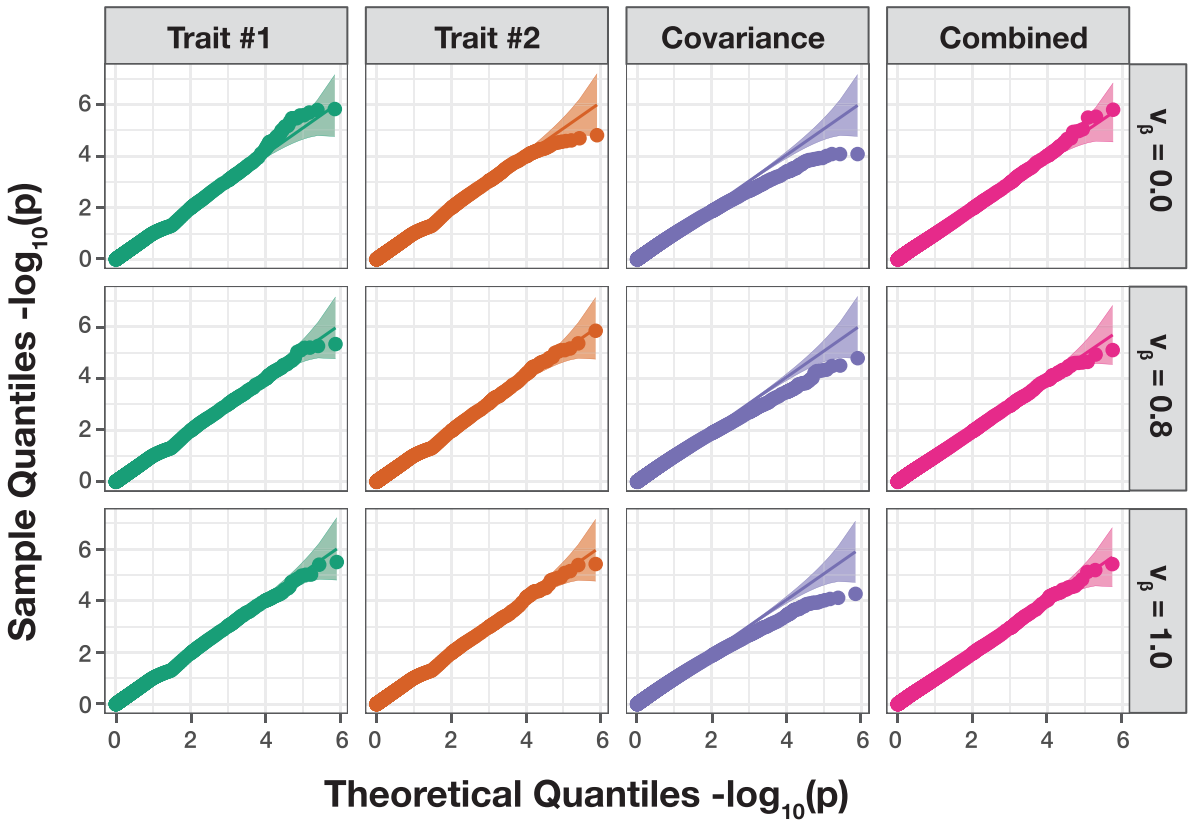


**Fig. 2.** The mvMAPIT framework using Fisher's method produces well-calibrated *P*-values when traits are generated by only additive effects (sample size $N = 2,500$ individuals). In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row of QQ plots corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have independent traits ($\upsilon_\beta = 0$), highly correlated traits ($\upsilon_\beta = 0.8$), and perfectly correlated traits ($\upsilon_\beta = 1$). The first two columns show *P*-values resulting from the univariate MAPIT test on "trait #1" and "trait #2," respectively. The third column depicts the "covariance" *P*-values which corresponds to assessing the pairwise interactions affecting both traits. Lastly, the fourth column shows the final "combined" *P*-value which combines the *P*-values from the first three columns using Fisher's method. The 95% confidence interval for the null hypothesis of no marginal epistatic effects is also shown as the shaded band. Each plot combines results from 100 simulated replicates.

mean, or the Cauchy combination procedure. Note that we expect the $P$-values from the first univariate test to be well-calibrated since it is equivalent to the MAPIT model. Figure 2 and Supplementary Figs. S1–S2 show the quantile–quantile (QQ) plots based on $P$-values combined using Fisher's method, while Supplementary Figs. S3–S8 depict results while using the harmonic mean and the Cauchy combination test. Similarly, Table 1 and Supplementary Tables S1–S8 show the empirical type I error rates estimated for mvMAPIT at significance levels $P = 0.05$, 0.01, and 0.001, respectively.

Overall, mvMAPIT conservatively controls type 1 error rate, both in the presence of nonzero correlation between additive effects on the two traits and even with small sample sizes in the data. This result holds regardless of how $P$-values are combined in the model. The QQ plots of the $P$-values for all three components in mvMAPIT follow the expected uniform distribution for the univariate and combined analysis. Notably, because of the approximations used to compute the standard error of the test statistic in equation (13), the multivariate extension of the MQS-based testing procedure in mvMAPIT can result in conservative $P$-values for the covariance components under the null.

## Improved detection of epistatic variants using mvMAPIT in simulations

We test the power of mvMAPIT across different genetic trait architectures via an extensive simulation study (see *Materials and Methods*). Once again, we generate synthetic phenotypes using real genotypes from the 22nd chromosome of the control samples in the WTCCC 1 study (Burton *et al.* 2007). As a reminder, these data consist of $N = 2,938$ individuals and $J = 5,747$ SNPs. In these simulations, we randomly choose 1,000 causal variants to directly affect the genetic architecture of $D = 2$ phenotypes. All causal SNPs are assumed to have a nonzero additive effect on both traits. Next, we randomly select a set of epistatic variants from the 1,000 causal SNPs and divide them into two interacting groups (again see Materials and Methods). We will denote these groups #1 and #2 as $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively, with $|\mathcal{G}|$ denoting the cardinality of the group. One may interpret the epistatic SNPs in $\mathcal{G}_1$ as being the "hub nodes" in an interaction network where each of these variants interact with all of the SNPs assigned to $\mathcal{G}_2$. We generate synthetic traits by using the multivariate linear model $\mathbf{Y} = \mathbf{B}\mathbf{X}^{\intercal} + \mathbf{A}\mathbf{W}^{\intercal} + \mathbf{E}$ where, in addition to previous notation, $\mathbf{W}$ is matrix of interactions between the SNPs assigned to the groups $\mathcal{G}_1$ and $\mathcal{G}_2$. The additive and interaction coefficients for causal SNP effects across traits are drawn as $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_\beta)$ and $\boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_\alpha)$, respectively. As a final step, we scale all terms to ensure that all genetic effects explain a fixed proportion of the total phenotypic variation. We assume a wide range of simulation scenarios by varying the following parameters:

- broad-sense heritability: $H^2 = 0.3$ and 0.6;
- proportion of phenotypic variation that is explained by additive effects: $\rho = 0.5$ and 0.8;
- number of causal SNPs assigned to the interaction groups: $\{|\mathcal{G}_1|, |\mathcal{G}_2|\} = \{10, 10\}$ and $\{10, 20\}$;
- correlation between epistatic effects: $\upsilon_{\alpha,12} = 0$ and 0.8.

All results presented in this section are based on 100 different simulated phenotypes for each parameter combination.

The main point of these simulations is to highlight the potential power gained from taking a multivariate approach to epistatic detection. To that end, in each of the simulation scenarios, we assess (i) the power of running the univariate MAPIT model on each trait individually, (ii) the marginal epistatic effects detected by the covariance test, (iii) the power from the overall association identified by mvMAPIT, and (iv) a baseline meta-analytic approach which combines only the $P$-values from the univariate MAPIT models on each trait. We will hereafter refer to that last model as "meta-MAPIT" for simplicity. Figure 3 and Supplementary Figs. S9–S11 show the empirical power of the univariate MAPIT model, the covariance test, mvMAPIT, and meta-MAPIT while using Fisher's method at various multiple hypothesis testing correction thresholds. Supplementary Figs. S12–S19 depict the same information but with mvMAPIT and meta-MAPIT using the harmonic mean and Cauchy combination test to combine $P$-values. We also compare each method's ability to rank true positives over false positives via receiver operating characteristic (ROC) and precision-recall curves (Fig. 4 and Supplementary Figs. S20–S24). There are several key takeaways from these simulation results. Overall, the ability of the univariate MAPIT framework to detect group #1 and #2 causal variants depends on the proportion of non-additive phenotypic variation that they explain. This has been shown in previous demonstrations of the method (Crawford *et al.* 2017). For example, when there are $|\mathcal{G}_2| = 10$ causal SNPs in group #2, each variant in the set is expected to explain $(1 - \rho)H^2/10\%$ of the genetic variance. As we increase that number of causal SNPs in group #2 to $|\mathcal{G}_2| = 20$, this proportion of variance explained by SNPs in group #2 will decrease which will make it more difficult to prioritize markers involved in interactions. Importantly, it is worth noting that the single-phenotypic test in MAPIT depends on the total interaction effects, rather than individual pairwise effects or the number of interacting pairs. An

**Table 1.** The mvMAPIT framework using Fisher's method preserves type I error rates under the null model when traits are generated by only additive effects (sample size $\mathbf{N} = 2,500$ individuals).

| | Add. effect corr. | $P = 0.05$ | $P = 0.01$ | $P = 0.001$ |
|---|---|---|---|---|
| Univariate | $\upsilon_\beta = 0.0$ | 0.030 $(1 \times 10^{-2})$ | 0.009 $(2 \times 10^{-3})$ | 0.0010 $(9 \times 10^{-4})$ |
| | $\upsilon_\beta = 0.8$ | 0.030 $(1 \times 10^{-2})$ | 0.009 $(2 \times 10^{-3})$ | 0.0010 $(7 \times 10^{-4})$ |
| | $\upsilon_\beta = 1.0$ | 0.030 $(1 \times 10^{-2})$ | 0.009 $(3 \times 10^{-3})$ | 0.0009 $(7 \times 10^{-4})$ |
| Covariance | $\upsilon_\beta = 0.0$ | 0.040 $(1 \times 10^{-2})$ | 0.006 $(2 \times 10^{-3})$ | 0.0003 $(3 \times 10^{-4})$ |
| | $\upsilon_\beta = 0.8$ | 0.040 $(1 \times 10^{-2})$ | 0.007 $(2 \times 10^{-3})$ | 0.0004 $(5 \times 10^{-4})$ |
| | $\upsilon_\beta = 1.0$ | 0.040 $(1 \times 10^{-2})$ | 0.006 $(2 \times 10^{-3})$ | 0.0003 $(4 \times 10^{-4})$ |
| Combined | $\upsilon_\beta = 0.0$ | 0.040 $(1 \times 10^{-2})$ | 0.006 $(2 \times 10^{-3})$ | 0.0003 $(3 \times 10^{-4})$ |
| | $\upsilon_\beta = 0.8$ | 0.040 $(1 \times 10^{-2})$ | 0.007 $(2 \times 10^{-3})$ | 0.0004 $(5 \times 10^{-4})$ |
| | $\upsilon_\beta = 1.0$ | 0.040 $(1 \times 10^{-2})$ | 0.006 $(2 \times 10^{-3})$ | 0.0003 $(4 \times 10^{-4})$ |

In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent additive effects ($\upsilon_\beta = 0$), traits with highly correlated additive effects ($\upsilon_\beta = 0.8$), and traits with perfectly correlated additive effects ($\upsilon_\beta = 1$). We assess the calibration of the $P$-values that are produced by mvMAPIT during each of the three key steps in its combinatorial hypothesis testing procedure (see *Materials and Methods*). We show type I error rates resulting from $P$-values taken from the "univariate" test on each trait independently, the "covariance" $P$-values which corresponds to assessing the pairwise interactions affecting both traits, and the final "combined" $P$-value. Results are summarized over 100 simulated replicates. Values in the parentheses are the standard deviations across replicates.

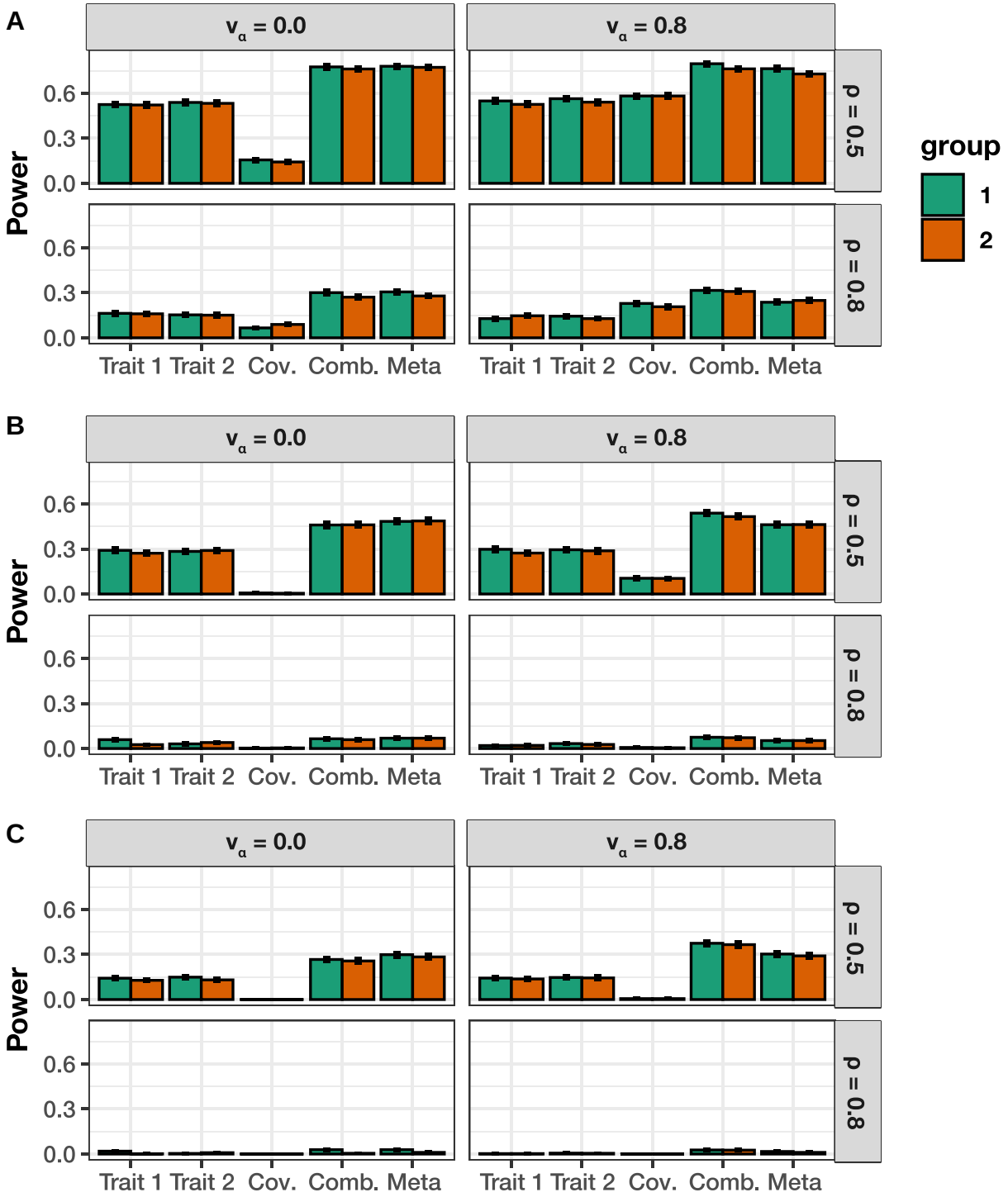


**Fig. 3.** Empirical power of mvMAPIT with Fisher's method to detect group #1 (10) and group #2 (10) epistatic variants across complex traits with moderate broad-sense heritability. In these simulations, both quantitative traits are simulated to have broad-sense heritability $H^2 = 0.6$ with architectures made up of both additive and epistatic effects. The parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to a setting where the epistatic effects for interactive pairs have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent epistatic effects ($v_\alpha = 0$) and traits with highly correlated epistatic effects ($v_\alpha = 0.8$). This plot shows the empirical power of mvMAPIT at significance levels a) $P = 5 \times 10^{-2}$, b) $P = 5 \times 10^{-4}$, and c) $P = 1 \times 10^{-5}$, respectively. Results for the group #1 and #2 causal markers are shown side-by-side, respectively. For comparison, the "trait #1" and "trait #2" bars correspond to the univariate MAPIT model, the "cov" bars corresponds to power contributed by the covariance test, and "comb" details power from the overall association identified by mvMAPIT in the combination approach. Lastly, the "meta" bars shows power from the association identified by combining only the *P*-values of the two univariate cases "trait #1" and "trait #2". Results are based on 100 simulations per parameter combination and the horizontal bars represent standard errors.

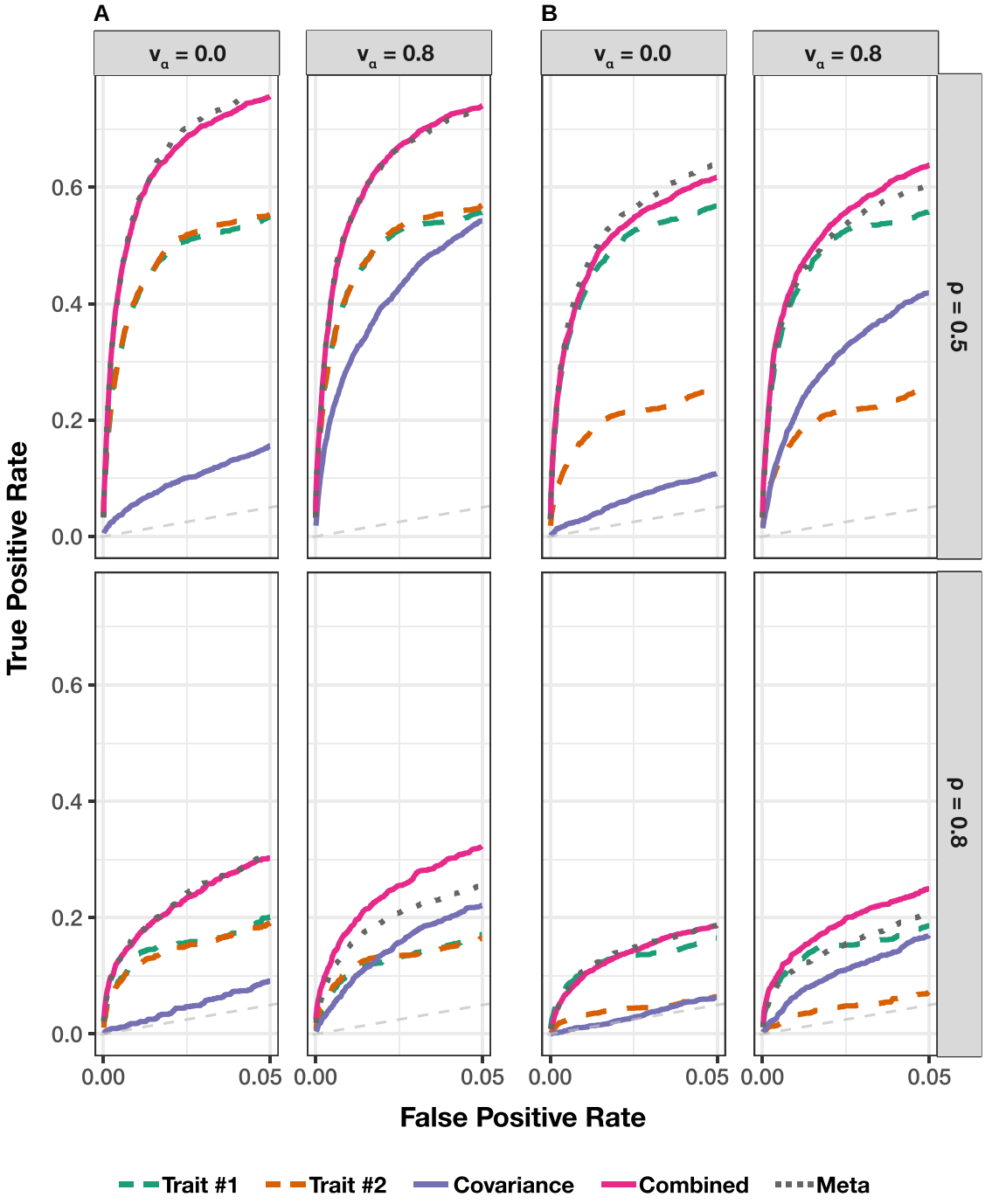


**Fig. 4.** Receiver operating characteristic (ROC) curves comparing the ability of mvMAPIT with Fisher's method to the univariate MAPIT model in detecting group #1 (10) and group #2 (10) epistatic variants across complex traits. In panel a) both traits have broad-sense heritability $H^2 = 0.6$, while in panel b) one of traits has broad-sense heritability $H^2 = 0.6$ and the other has heritability $H^2 = 0.3$. Across the rows, the parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to settings where the epistatic effects across traits are independent ($\upsilon_\alpha = 0$) or highly correlated ($\upsilon_\alpha = 0.8$). For comparison, the "trait #1" and "trait #2" dashed lines correspond to the univariate MAPIT model. The "covariance" solid line corresponds to power contributed by the covariance test. The "combined" line shows power from the overall association identified by mvMAPIT in the multivariate approach. Lastly, the "meta" dotted line shows power from the association identified by combining only the *P*-values of the two univariate cases "trait #1" and "trait #2". Note that the upper limit of the x-axis (i.e. false positive rate) has been truncated at 0.05. All results are based on 100 simulated replicates.

example of this can be seen by comparing Fig. 3a to Supplementary Fig. S9a where the ability to group #1 variants is independent of the number of variants in group #2.

Intuitively, the joint modeling approach of mvMAPIT provides a viable strategy for identifying SNPs contributing to nonadditive variation that would have otherwise gone undetected by univariate methods. The scenario where mvMAPIT shows significant gains over the univariate MAPIT modeling approach is when there is non-zero correlation between the effects of the epistatic interactions shared between traits (e.g. when $\upsilon_{\alpha,12} = 0.8$). The sensitivity of the covariance hypothesis test depends on the strength of this correlation which can help increase power when combining over

*P*-values in the final step of mvMAPIT. This becomes increasingly relevant in the low heritability cases. When there is no correlation of interaction effects shared between pairs of traits, combining only the univariate *P*-values in the meta-MAPIT model performs equally well as mvMAPIT which includes contributions from the covariance statistic. Figure 4 and Supplementary Figs. S20–S24 demonstrate that the sensitivity of the covariance statistic is comparable to the univariate statistic for highly correlated epistatic effects ($\upsilon_a = 0.8$) despite genetic variance being predominantly explained by additivity ($\rho = 0.8$). In this case, the multivariate approach performs better than simply doing a meta-analysis on only the univariate *P*-values. In Fig. 4, Supplementary Figs. S10, S11, S14, S15, and S18–S24, we simulated synthetic traits such that one has a moderate broad-sense heritability $H^2 = 0.6$ and the other has heritability $H^2 = 0.3$. In these scenarios, detecting variants involved in interactions increased for the trait with low heritability. In particular, the covariance component analysis is shown to play an important role in this improved detection (e.g. see Fig. 4b).

## Synergistic epistasis in binding affinity landscapes for neutralizing antibodies

We apply the mvMAPIT framework to protein sequence data from Phillips *et al.* (2021) who generated a nearly combinatorially complete library for two broadly neutralizing anti-influenza antibodies (bnAbs), CR6261, and CR9114. This data set includes almost all combinations of one-off mutations that distinguish between germline and somatic sequences which total to $J = 11$ heavy-chain mutations for CR6261 and $J = 16$ heavy-chain mutations for CR9114. Theoretically, a combinatorially complete data set for 11 and 16 mutations will have 2,048 and 65,536 samples, respectively. In this particular study, we have have access to $N = 1,812$ complete observations for CR6261 and $N = 65,091$ complete measurements for CR9114. For our analysis with mvMAPIT, residue sequence information was encoded as a binary matrix with the germline sequence residues marked by zeros and the somatic mutations represented as ones. As quantitative traits, Phillips *et al.* (2021) measure the binding affinity of the two antibodies to different influenza strains. Here, we assess the contribution of epistatic effects when binding to $H_1$ and $H_9$ for CR6261, and $H_1$ and $H_3$ for CR9114.

Once again, we report results after running mvMAPIT with Fisher's method, the harmonic mean, and the Cauchy combination approach (Supplementary Table S9). Figure 5a, Supplementary Figs. S25a and S26a show Manhattan plots for *P*-values corresponding to the trait-specific marginal epistatic tests (i.e. the univariate MAPIT model), the covariance test, and the mvMAPIT approach. Here, green colored dots are positions that have significant marginal epistatic effects beyond a Bonferroni corrected threshold for multiple testing ($P = 0.05/11 = 4.55 \times 10^{-3}$ for CR6261 and $P = 0.05/16 = 3.13 \times 10^{-3}$ for CR9114, respectively). Interestingly, while the univariate MAPIT approach was able to identify significant marginal epistatic effects for CR6261, it lacked the power to identify significant positions driving nonadditive variation in binding affinity for CR9114. Overall, the combined trait approach in mvMAPIT revealed marginal epistatic effects for positions 29, 35, 82, 83, and 84 in CR6261, and positions 30, 36, 57, 64, 65, 66, 82, and 83 for CR9114. Most notably, these same positions were also identified as contributing to pairwise epistasis by Phillips *et al.* (2021). In the original study, the authors first ran an exhaustive search to statistically detect significant interactions and then conducted downstream analyses to find that these positions are likely responsible for the antibodies binding to the influenza surface protein hemagglutinin. The regression coefficients from the exhaustive search, as reported by Phillips *et al.* (2021), are illustrated in panels b and c of Fig. 5, Supplementary Figs. S25 and S26. Panel b illustrates interaction coefficients when assessing binding of CR6261with $H_1$ (upper right triangle) and $H_9$ (lower left triangle). Panel c shows the same information when assessing binding of CR9114 with $H_1$ (upper right triangle) and $H_9$ (lower left triangle). Our results show that mvMAPIT identifies all required mutations in these systems as well as most positions involved in at least one epistatic pair.

## Joint modeling of hematology traits yields epistatic signal in stock of mice

In this section, we apply mvMAPIT to individual-level genotypes and 15 hematology traits in a heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics (Valdar, Flint, *et al.* 2006; Valdar, Solberg, *et al.* 2006). This collection of data contains approximately $N = 2,000$ individuals depending on the phenotype (see *Materials and Methods*), and each mouse has been genotyped at $J = 10,346$ SNPs. As noted by previous studies, these data represent a realistic mixture of the simulation scenarios we detailed in the previous sections (i.e. varying different values of the parameter $\rho$). Specifically, this stock of mice is known to be genetically related with population structure and the genetic architectures of these particular traits have been shown to have different levels of broad-sense heritability with varying contributions from nonadditive genetic effects.

For each pairwise trait analysis, we provide a summary table which lists the combined *P*-values after running mvMAPIT with Fisher's method and the harmonic mean (Supplementary Table S10). We also include results corresponding to the univariate MAPIT model and the covariance test for comparison. Overall, the single-trait marginal epistatic test only identifies significant variants for the large immature cells (LIC) after Bonferroni correction ($P = 4.83 \times 10^{-6}$). A complete picture of this can be seen in Supplementary Figs. S27–S29 which depict Manhattan plots of our genome-wide interaction study for all combinations of trait pairs. Here, we can see that most of the signal in the combined *P*-values from mvMAPIT likely stems from the covariance component portion of the model. This hypothesis holds true for the joint pairwise analysis of (i) hematocrit (HCT) and hemoglobin (HGB) and (ii) mean corpuscular hemoglobin (MCH) and mean corpuscular volume (MCV) (e.g. see the third and fourth rows of Fig. 6a, Supplementary Figs. S30a and S31a). One explanation for observing more signal in the covariance components over the univariate test could be derived from the traits having low heritability but high correlation between epistatic interaction effects. Recall that our simulation studies showed that the sensitivity of the covariance statistic increased for these cases. As a direct comparison, Fig. 6b, Supplementary Figs. S30b, S31b, and S32–S34 give examples of analyses that do not identify significant marginal epistatic effects even after using the mvMAPIT approach. This further supports the claim that the signal provided by the covariance statistic is not likely due to inflation.

Notably, the nonadditive signal identified by the covariance test is not totally dependent on the empirical correlation between traits (see Supplementary Fig. S35). Instead, as previously shown in our simulation study, the power of mvMAPIT over the univariate approach occurs when there is correlation between the effects of epistatic interactions shared between two traits. Importantly, many of the candidate SNPs selected by the mvMAPIT framework have been previously discovered by past publications as having some functional nonlinear relationship with the traits of interest. For example, the multivariate analysis with traits MCH and MCV show a significant SNP rs4173870 ($P = 4.89 \times 10^{-10}$) in the gene hematopoietic cell-specific Lyn substrate 1 (*Hcls1*) on chromosome 16 which has been shown to play a role in

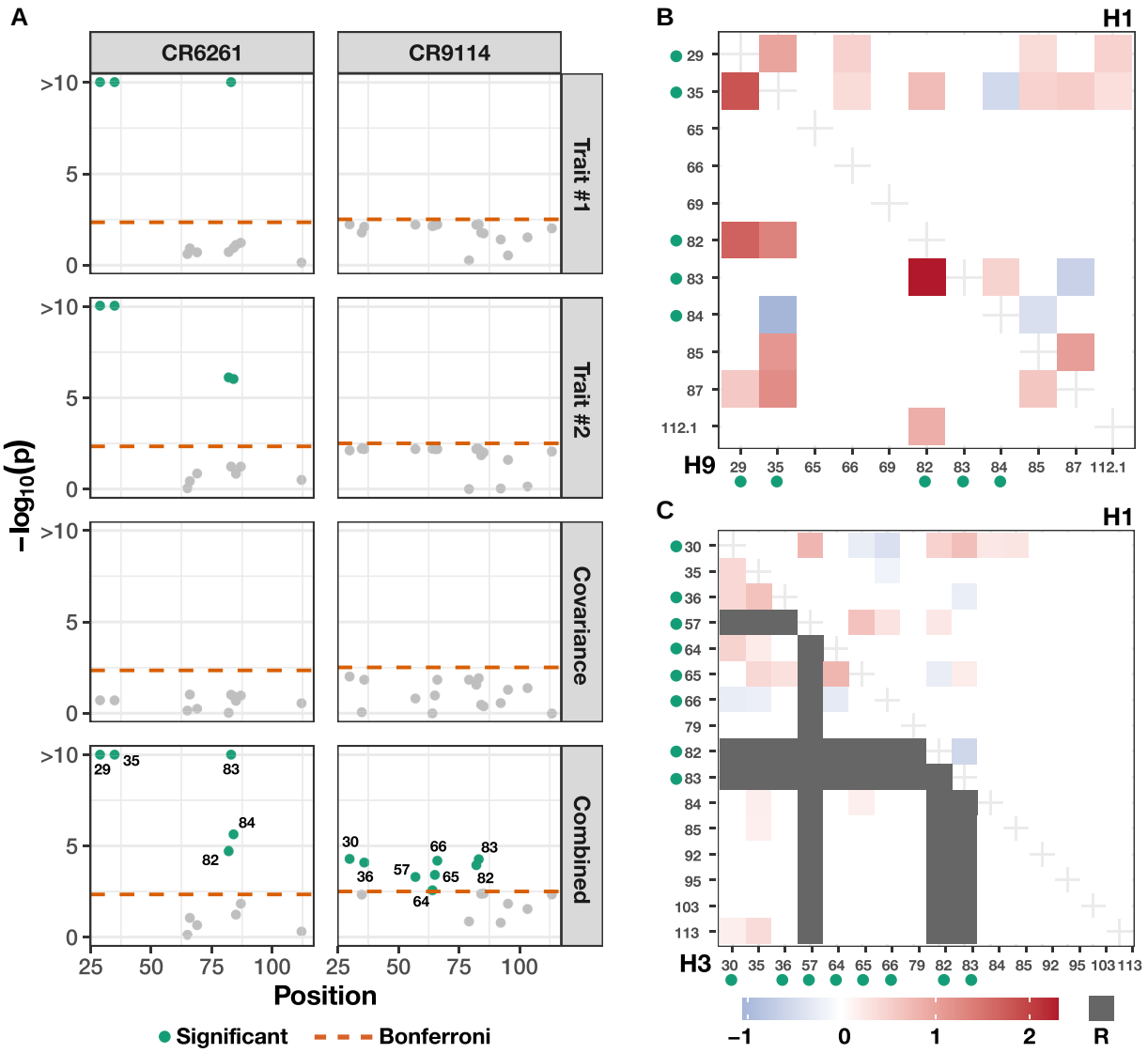


**Fig. 5.** Applying mvMAPIT with Fisher's method to broadly neutralizing antibodies recovers heavy-chain mutations known to be involved in epistatic interactions that affect binding against two influenza strains. These results are based on protein sequence data from Phillips *et al.* (2021) who generated a nearly combinatorially complete library for two broadly neutralizing anti-influenza antibodies (bnAbs), CR6261 and CR9114. For each antibody, we assess binding affinity to different influenza strains. For CR6261, traits #1 and #2 are binding measurements to the antigens $H_1$ and $H_9$, while for CR9114, we assess the same measurement for $H_1$ and $H_3$. a) Manhattan plots for the different sets of *P*-values computed during the mvMAPIT analysis. The dashed horizontal lines indicate a chain-wide Bonferroni corrected significance threshold ($P = 4.55 \times 10^{-3}$ for CR6261 and $P = 3.13 \times 10^{-3}$ for CR9114, respectively). The dots above the dashed horizontal lines are positions that have significant marginal epistatic effects after multiple correction. b, c) Exhaustive search results originally reported by Phillips *et al.* (2021). The dots next to the mutation labels on the axes are the residues that are significant in the multivariate MAPIT analysis and correspond to panel a). The shaded regions in panel b) are the regression coefficients for pairwise interactions between positions when assessing binding of CR6261with $H_1$ (upper right triangle) and $H_9$ (lower left triangle). Similarly, panel c) shows the same information when assessing binding of CR9114 with $H_1$ (upper right triangle) and $H_3$ (lower left triangle). Required mutations (indicated by R) are left out of the analysis (Phillips *et al.* 2021).

differentiation of erythrocytes (Castro-Ochoa *et al.* 2019). Similarly, the joint analysis of HGB and HCT shows hits in multiple coding regions. One example here are the SNPs rs3692165 ($P = 1.82 \times 10^{-6}$) and rs13482117 ($P = 8.94 \times 10^{-7}$) in the gene calcium voltage-gated channel auxiliary subunit alpha2delta 3 (*Cacna2d3*) on chromosome 14, which has been associated with decreased circulating glucose levels (IMPC 2014), and SNP rs3724260 ($P = 4.58 \times 10^{-6}$) in the gene *Dicer1* on chromosome 12 which has been annotated for anemia both in humans and mice (Raaijmakers *et al.* 2010).

Table 2 lists a select subset of SNPs in coding regions of genes that have been associated with phenotypes related to the hematopoietic system, immune system, or homeostasis and metabolism. Each of these are significant (after correction for multiple hypothesis testing) in the mvMAPIT analysis of related hematology traits. Some of these phenotypes have been reported as having large broad-sense heritability, which improves the ability of mvMAPIT to detect the signal. For example, the genes *Arf2* and *Cacna2d3* are associated with phenotypes related to glucose homeostasis, which has been reported to have a large heritable component (estimated $H^2 = 0.3$ for insulin sensitivity, Norris and Rich (2012)). Similarly, the genes *App* and *Pex1* are associated with thrombosis where (an estimated) more than half of phenotypic variation has been attributed to genetic effects (estimated

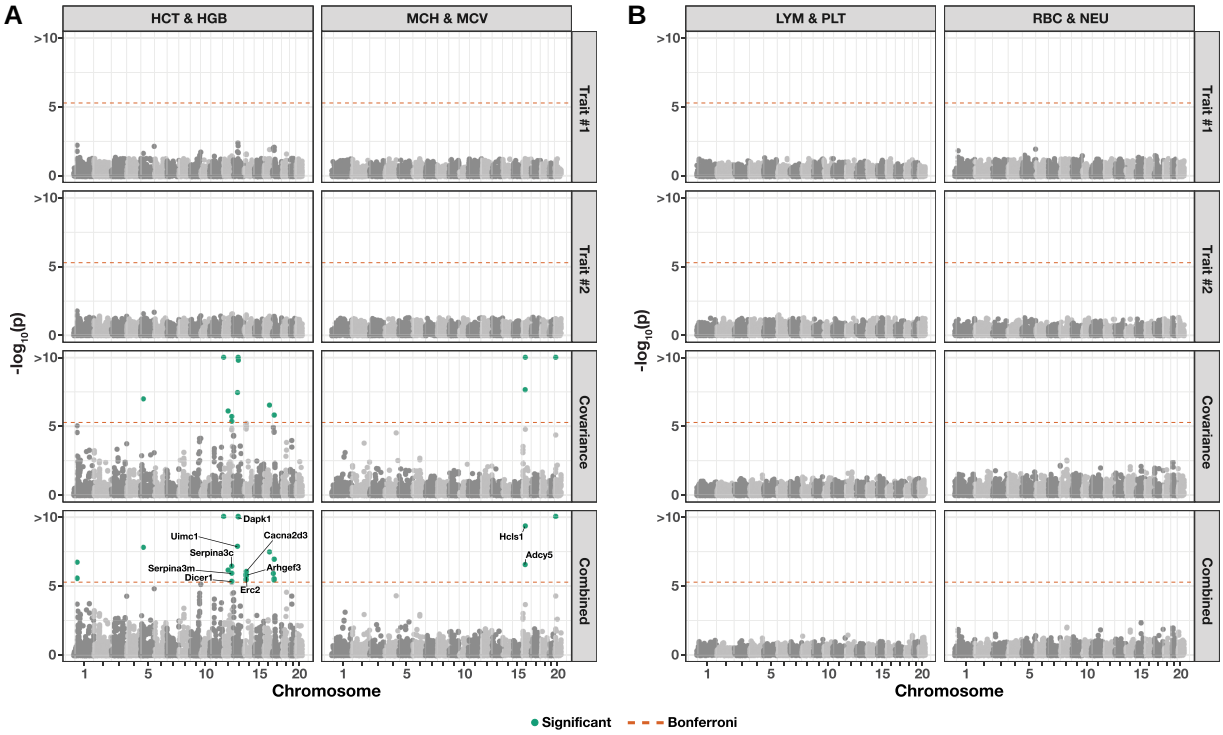


**Fig. 6.** Manhattan plots of a genome-wide interaction study for two pairs of hematology traits in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics (Valdar, Flint, *et al.* 2006; Valdar, Solberg, *et al.* 2006) using mvMAPIT with Fisher's method. Panel a) depicts results for trait pairs hematocrit (HCT) and hemoglobin (HGB) in the left column and mean corpuscular hemoglobin (MCH) and mean corpuscular volume (MCV) in the right column. Panel b) depicts results for trait pairs lymphocytes (LYM) and platelets (PLT) in the left column and red blood cell count (RBC) and neutrophils (NEU) in the right column. Here, we depict the *P*-values computed during each step of the mvMAPIT modeling pipeline. The dashed horizontal lines indicate a genome-wide Bonferroni corrected significance threshold ($P = 4.83 \times 10^{-6}$). The dots above the dashed horizontal lines are SNPs that have significant marginal epistatic effects after multiple test correction. Significant SNPs were mapped to the closest neighboring genes using the Mouse Genome Informatics database (http://www.informatics.jax.org) (Blake *et al.* 2020; Smith and Eppig 2009).

$H^2 \geq 0.6$ for susceptibility to common thrombosis, Souto *et al.* 2000).

As a final analysis, Supplementary Figs. S36 and 37 illustrate the possibility of using the mvMAPIT framework to jointly analyze any number of traits. These show both Manhattan plots and QQ plots corresponding to the application of mvMAPIT to a subset or all of the 15 hematology traits measured in the heterogeneous stock of mice data set. In this analysis, all univariate variance and all pairwise covariance test statistics are combined. Supplementary Figure S36 shows an inflation of smaller *P*-values that is reduced by excluding traits measuring abnormal lymphocytes and large immature cells (i.e. those traits that show strong signal in for the univariate analyses; see diagonal panels in Supplementary Figs. S27–S29). Results where these two phenotypes are excluded can be found in Supplementary Fig. S37 for direct comparison.

## Discussion

The marginal epistatic testing strategy offers an alternative to traditional epistatic mapping methods by seeking to identify variants that exhibit nonzero interaction effects with any other variant in the data (Crawford *et al.* 2017; Crawford and Zhou 2018; Turchin *et al.* 2020). This framework has been shown to drastically reduce the number of statistical tests needed to uncover evidence of significant nonadditive variation in complex traits and, as a result, alleviates much of the empirical power concerns and heavy computational burden associated with explicit search-based methods. Still, models testing for marginal epistasis can be underpowered when applied to traits with low heritability or to "polygenic" traits where the interactions between mutations have small effect sizes (Crawford *et al.* 2017). In this work, we present the "multivariate MArginal ePIstasis Test" (mvMAPIT), a multioutcome extension of the univariate marginal epistatic framework (Fig. 1). Theoretically, we formulate mvMAPIT as a multivariate linear mixed model (mvLMM) where its ability to jointly analyze any number of traits relies on a generalized "variance–covariance" component estimation algorithm (Zhou 2017). Through extensive simulations, we show that mvMAPIT preserves type I error rates and produces well-calibrated *P*-values under the null model when traits are generated only by additive effects (Fig. 2 and Supplementary Figs. S1–S8, and Table 1 and Supplementary Tables S1–S8). In these simulation studies, we also show that mvMAPIT improves upon the identification of epistatic variants over the univariate test when there is correlation between the effects of genetic interactions shared between multiple traits (Figs. 3, 4, and Supplementary Figs. S9–S24). By analyzing two real data sets, we demonstrated the ability of mvMAPIT to recover heavy-chain mutations known to be involved in epistatic interactions that affect binding against two influenza strains (Phillips *et al.* 2021) (Fig. 5, Supplementary Figs. S25 and S26, and Table S9) as well as to identify hematology trait relevant epistatic SNPs in heterogeneous stock of mice (Valdar, Flint, *et al.* 2006; Valdar, Solberg, *et al.* 2006) that have also been detected in

**Table 2.** Notable SNPs with marginal epistatic effects after applying the mvMAPIT framework to 15 hematology traits in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics (Valdar, Flint, *et al.* 2006; Valdar, Solberg, *et al.* 2006).

| SNP | Location | Trait #1 | Trait #2 | Trait #1 P-value | Trait #2 P-value | Cov. P-value | Comb. P-value | Gene | Genomic annotation | Reference |
|---|---|---|---|---|---|---|---|---|---|---|
| rs3699393 | 2:5887012 | MCV | PLT | 0.21 | 0.23 | $5.75 \times 10^{-7}$ | $4.9 \times 10^{-6}$ | *Upf2* | Anemia and abnormal bone marrow cell development | Weischenfeldt *et al.* (2008) |
| rs13478092 | 5:3601413 | LIC | PLT | 0.034 | 0.58 | $1.67 \times 10^{-10}$ | $1.26 \times 10^{-9}$ | *Pex1* | Abnormal venous thrombosis | Berendse *et al.* (2019) |
| rs3694887 | 5:102770070 | ALY | LIC | $1.26 \times 10^{-4}$ | 0.013 | $2.54 \times 10^{-6}$ | $1.55 \times 10^{-9}$ | *Aff1* | Abnormal B and T cell number and morphology | Isnard *et al.* (2000) |
| rs3694887 | 5:102770070 | LIC | PLT | 0.013 | 0.28 | $5.47 \times 10^{-27}$ | $4.49 \times 10^{-26}$ | *Aff1* | Abnormal B and T cell number and morphology | Isnard *et al.* (2000) |
| rs13478923 | 6:99475169 | ALY | LIC | $2.8 \times 10^{-4}$ | 0.12 | $1.79 \times 10^{-6}$ | $1.81 \times 10^{-8}$ | *Foxp1* | Abnormal B cell differentiation, physiology, count | Feng *et al.* (2010), Hu *et al.* (2006) |
| rs13478924 | 6:99571626 | ALY | LIC | $3.11 \times 10^{-4}$ | 0.12 | $2.70 \times 10^{-6}$ | $2.86 \times 10^{-8}$ | *Foxp1* | Abnormal B cell differentiation, physiology, count | Feng *et al.* (2010), Hu *et al.* (2006) |
| rs13478985 | 6:115245823 | MCV | WBC | 0.16 | 0.40 | $1.14 \times 10^{-81}$ | $1.34 \times 10^{-78}$ | *Atg7* | Decreased bone marrow cell count | Mortensen *et al.* (2011), Clarke *et al.* (2015) |
| rs3723163 | 11:103800737 | HCT | LYM | 0.072 | 0.30 | $3.99 \times 10^{-107}$ | $2.66 \times 10^{-104}$ | *Arf2* | Decreased fasting circulating glucose level | IMPC (2014) |
| rs3723163 | 11:103800737 | HGB | WBC | 0.069 | 0.25 | $1.85 \times 10^{-7}$ | $6.76 \times 10^{-7}$ | *Arf2* | Decreased fasting circulating glucose level | IMPC (2014) |
| rs3724260 | 12:100163212 | HGB | HCT | 0.030 | 0.062 | $1.44 \times 10^{-5}$ | $4.58 \times 10^{-6}$ | *Dicer1* | Anemia | Raaijmakers *et al.* (2010) |
| rs3692165 | 14:27756640 | HCT | HGB | 0.026 | 0.037 | $9.9 \times 10^{-6}$ | $1.8 \times 10^{-6}$ | *Cacna2d3* | Decreased circulating glucose level | IMPC (2014) |
| rs13482117 | 14:27614362 | HCT | HGB | 0.023 | 0.03 | $5.9 \times 10^{-6}$ | $9.0 \times 10^{-7}$ | *Cacna2d3* | Decreased circulating glucose level | IMPC (2014) |
| rs13482288 | 14:81840412 | ALY | BAS | 0.036 | 0.65 | $1.78 \times 10^{-8}$ | $1.1 \times 10^{-7}$ | *Tdrd3* | Abnormal B cell differentiation and physiology | Yang *et al.* (2014) |
| rs4173870 | 16:35764290 | MCH | MCV | 0.14 | 0.71 | $1.20 \times 10^{-11}$ | $4.89 \times 10^{-10}$ | *Hcls1* | Differentiation of erythrocytes | Castro-Ochoa *et al.* (2019) |
| rs4212102 | 16:84204704 | PLT | WBC | 0.17 | 0.35 | $1.16 \times 10^{-10}$ | $2.44 \times 10^{-9}$ | *App* | Increased susceptibility to induced thrombosis | Xu *et al.* (2017), Souto *et al.* (2000) |
| rs4212186 | 16:84273330 | PLT | WBC | 0.17 | 0.36 | $5.88 \times 10^{-11}$ | $1.31 \times 10^{-9}$ | *App* | Increased susceptibility to induced thrombosis | Xu *et al.* (2017), Souto *et al.* (2000) |
| rs3711994 | 19:45078018 | ALY | LYM | $3.71 \times 10^{-4}$ | 0.10 | $1.04 \times 10^{-12}$ | $2.80 \times 10^{-14}$ | *Btrc* | Abnormal lymphocyte morphology | Nakayama *et al.* (2003) |

In the first two columns, we list SNPs and their genetic location according to the mouse assembly NCBI build 34 (accessed from (Shifman *et al.* 2006)) in the format `Chromosome:Basepair`. Next, we give the results stemming from univariate analyses on traits #1 and #2, respectively, the covariance (cov) test, and the overall *P*-value derived by mvMAPIT using Fisher's method. The last columns detail the closest neighboring genes found using the Mouse Genome Informatics database (http://www.informatics.jax.org) (Smith and Eppig 2009; Blake *et al.* 2020), a short summary of the suggested annotated function for those genes, and the reference to the source of the annotation. See Supplementary Table S7 for the complete list of SNP and SNP-set level results.

previous publications and functional validation studies (Fig. 6 and Supplementary Figs. S27–S37, and Table 2 and Supplementary Table S10). Lastly, we have made mvMAPIT an open-source R software package with documentation to facilitate its use by the greater scientific community.

We want to highlight an important caveat for mapping epistasis in real data: in genome-wide association studies, statistically inferred interactions can sometimes be explained by same-locus additive effects (Wood *et al.* 2014). This means that conclusions made by mvMAPIT, as well as any other computational method aimed at the detection of epistasis, could be confounded by additive effects of untyped or uncontrolled variants in the same region. Notably, it is difficult to control for the additive fixed effects of all variants in the same locus (Crawford *et al.* 2017). This caveat needs to be considered to guard against overinterpreting the results of mvMAPIT. The results that we present in this work (e.g. in the study of hematology trait architecture within the heterogeneous stock of mice) should be understood as an illustration of how multivariate regression frameworks can be leveraged as powerful hypothesis generating tools that can help towards resolving the true contribution of genetic effects to phenotypic variation in complex traits.

The current implementation of the mvMAPIT framework offers many directions for future development and applications. First, like other marginal epistatic mapping methods, mvMAPIT is unable to directly identify detailed interaction pairs despite being able to identify SNPs that are involved in epistasis. As shown through our simulations and real data analyses, being able to identify SNPs involved in epistasis allows us to come up with an initial (likely) set of variants that are worth further exploration, and thus represents an important first step towards identifying and understanding detailed epistatic associations. In previous studies (Zhang *et al.* 2010; Lewinger *et al.* 2013; Crawford *et al.* 2017; Pecanka and Jonker 2021), two-step ad hoc procedures have been suggested where, in our case, we would first run mvMAPIT and then focus on significant SNPs from the first step to further test all of the pairwise interactions among them in order to identify specific epistatic interaction pairs. While this approach has been shown to be effective in univariate (single-trait) analyses, this two-step procedure is still ad hoc in nature and could miss important epistatic associations. Exploring robust ways unify these two steps in a joint fashion would be an interesting area for future research. Second, in its current implementation, mvMAPIT can be computationally expensive for data sets with large sample sizes (e.g. hundreds of thousands of individuals in a biobank scale study). In this study, we develop a "variance–component component" extension to the MQS algorithm to estimate parameters in the mvMAPIT model. Theoretically, MQS is based on the method of moments and produces estimates that are mathematically identical to the Haseman-Elston (HE) cross-product regression (Lee *et al.* 2011; Zhou 2017; Zhu and Zhou 2020). In practice, MQS is not only computationally more efficient than HE regression but also provides a simple, analytic estimation form that allows for exact *P*-value computation—thus alleviating the need for jackknife re-sampling procedures (Golan *et al.* 2014) that both are computationally expensive and rely on assumptions of independence across individuals in the data (Churchill and Doerge 2008). Exploring different ways to reliably fit large-scale mvLMMs with multiple random effects is a consideration for future work. For example, as an alternative, recent studies have proposed randomized multicomponent versions of HE regression for heritability estimation which scale up to data sets with millions of individuals and variants, respectively (Wu and Sankararaman 2018; Kerin and Marchini 2020a; Pazokitoroudi *et al.* 2020). It would be interesting to develop a well-calibrated hypothesis test within the randomized HE regression setting so that it may be implemented within the mvMAPIT software for association mapping.

In the future, we plan to expand the mvMAPIT framework to also identify individual variants contributing other sources of nonadditive genetic variation such as gene-by-environment (G×E) or gene-by-sex (G×Sex) interactions. We can do this by manipulating the marginal epistatic covariance matrix in equation (10) to encode how loci interact with one or more environmental instruments (Moore *et al.* 2019; Kerin and Marchini 2020a, 2020b; Zhu *et al.* 2022). Lastly, we have focused here on applying mvMAPIT to simple quantitative traits. However, there are many important traits with significant nonadditive genetic components in plants, animals, and humans that cannot be easily reduced to simple scalar values. Examples include longitudinal traits such as growth curves (Campbell *et al.* 2018), metabolic traits such as the relative concentrations of different families of metabolites (Chan *et al.* 2011), and morphological traits such as shape or color (Demmings *et al.* 2019). Indeed, each of these traits can be decomposed into vectors of interrelated components, but treating these components as independent phenotypes within existing univariate epistatic mapping tools would be inefficient because of their statistical dependence. As an alternative, the mvMAPIT framework can be used to make joint inferences about epistasis across any number of correlated phenotypic components—which, in the case of longitudinal studies for example (Couto Alves *et al.* 2019; Wu *et al.* 2019; Cousminer *et al.* 2021; Ko *et al.* 2022), could be used to interrogate how nonadditive variation of trait architecture changes or evolves over time.

## Data availability

Source code, tutorials, and tutorials for implementing the "multivariate MArginal ePIstasis Test" are publicly available as an R package which is available online at https://github.com/lcrawlab/mvMAPIT. We use the `CompQuadForm` R package (Duchesne and de Micheaux 2010) to compute *P*-values from the Davies method. The Davies method can sometimes yield a *P*-value equal exactly to 0 when the true *P*-value is extremely small (Duchesne and de Micheaux 2010). If this is of concern, one can compute the *P*-values for MAPIT using Kuonen's saddlepoint method (Kuonen 1999) or Satterthwaite's approximation equation (Satterthwaite 1946). In the current implementation of mvMAPIT, the saddlepoint approximation is performed if the Davies method returns with error. We wrote our own function to combine *P*-values using Fisher's method which is largely inspired by functions in the `metap` R package (Dewey 2022). We use the `harmonicmeanp` R package (Wilson 2019a, 2019c) to combine *P*-values using the harmonic mean. We also write our own function to perform the Cauchy combination test based on work developed in Liu and Xie (2020). Full package documentation can be found at https://lcrawlab.github.io/mvMAPIT/. Data to reproduce figures for the broadly neutralizing antibodies as well as the mice study can be found at https://doi.org/10.7910/DVN/WPFIGU (Stamp 2022).

Data about the binding affinity landscapes for neutralizing antibodies were downloaded directly from Phillips *et al.* (2021). Information about mice data set from the Wellcome Trust Centre for Human Genetics (Valdar, Flint, *et al.* 2006; Valdar, Solberg, *et al.* 2006) can be found at http://mtweb.cs.ucl.ac.uk/mus/www/mouse/index.shtml. The genotypes from this study were downloaded using the `BGLR-R` package (Perez and de los Campos 2014). Details about the mice phenotypes can be found

at http://mtweb.cs.ucl.ac.uk/mus/www/mouse/HS/index.shtml and hematological traits can be downloaded from the mvMAPIT package. In the real data analyses, SNPs were mapped to the closest neighboring genes using the Mouse Genome Informatics database (http://www.informatics.jax.org) (Blake *et al.* 2020).

## Acknowledgements

We thank members of the Weinreich and Crawford Labs for insightful comments on earlier versions of this manuscript as well as Sohini Ramachandran for helpful discussions.

## Funding

This research was conducted in part using computational resources and services at the Center for Computation and Visualization at Brown University. This research was supported in part by a David & Lucile Packard Fellowship for Science and Engineering awarded to L. Crawford. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of any of the funders.

## Conflicts of interest

The author(s) declare no conflict of interests.

## Author contributions

LC conceived the study. DW and LC supervised the project and provided resources. JS, AD, and LC developed the methods. JS developed the software and performed the analyses. All authors wrote and revised the manuscript.

## Literature cited

Bayat A, Hosking B, Jain Y, Hosking C, Kodikara M, Reti D, Twine NA, Bauer DC. Fast and accurate exhaustive higher-order epistasis search with BitEpi. Sci Rep. 2021;11(1):15923. doi:10.1038/s41598-021-94959-y

Berendse K, Boek M, Gijbels M, Van der Wel NN, Klouwer FC, van den Bergh-Weerman MA, Babaji Shinde A, Ofman R, Tien Poll-The B, Houten SM, *et al.* Liver disease predominates in a mouse model for mild human Zellweger spectrum disorder. Biochim Biophys Acta—Mol Basis Dis. 2019;1865(10):2774–2787. doi:10.1016/j.bbadis.2019.06.013

Blake JA, Baldarelli R, Kadin JA, Richardson JE, Smith CL, Bult CJ. Mouse genome database (MGD): knowledgebase for mouse–human comparative biology. Nucleic Acids Res. 2020;49:D981–D987. doi:10.1093/nar/gkaa1083

Bloom JS, Ehrenreich I M, Loo WT, Lite T-LV, Kruglyak L. Finding the sources of missing heritability in a yeast cross. Nature. 2013;494(7436):234–237. doi:10.1038/nature11867

Brem RB, Storey JD, Whittle J, Kruglyak L. Genetic interactions between polymorphisms that affect gene expression in yeast. Nature. 2005;436(7051):701–703. doi:10.1038/nature03865

Brown AA, Buil A, Viñuela A, Lappalainen T, Zheng H-F, Richards JB, Small KS, Spector TD, Dermitzakis ET, Durbin R. Genetic interactions affecting human gene expression identified by variance association mapping. eLife. 2014;3:e01381. doi:10.7554/eLife.01381

Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, Price AL, Neale BM. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47(3):291–295. doi:10.1038/ng.3211

Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand WH, Samani NJ, *et al.* Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661–678. doi:10.1038/nature05911

Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, *et al.* The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203–209. doi:10.1038/s41586-018-0579-z

Campbell M, Walia H, Morota G. Utilizing random regression models for genomic prediction of a longitudinal trait derived from high-throughput phenotyping. Plant Direct. 2018;2(9):e00080. doi:10.1002/pld3.80

Castro-Ochoa KF, Guerrero-Fonseca IM, Schnoor M. Hematopoietic cell-specific lyn substrate (HCLS1 or HS1): a versatile actin-binding protein in leukocytes. J Leukoc Biol. 2019;105(5):881–890. doi:10.1002/JLB.MR0618-212R

Chan EKF, Rowe HC, Corwin JA, Joseph B, Kliebenstein DJ. Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. PLoS Biol. 2011;9(8):e1001125. doi:10.1371/journal.pbio.1001125

Chang Y-C, Wu J-T, Hong M-Y, Tung Y-A, Hsieh P-H, Yee SW, Giacomini KM, Oyang Y-J, Chen C-Y, Weiner MW, *et al.* GenEpi: gene-based epistasis discovery using machine learning. BMC Bioinformatics. 2020;21(1):68. doi:10.1186/s12859-020-3368-2

Chari S, Dworkin I. The conditional nature of genetic interactions: the consequences of wild-type backgrounds on mutational interactions in a genome-wide modifier screen. PLoS Genet. 2013;9(8):e1003661. doi:10.1371/journal.pgen.1003661

Chen G, Zhang F, Xue W, Wu R, Xu H, Wang K, Zhu J. An association study revealed substantial effects of dominance, epistasis and substance dependence co-morbidity on alcohol dependence symptom count. Addict Biol. 2017;22(6):1475–1485. doi:10.1111/adb.12402

Churchill GA, Doerge RW. Naive application of permutation testing leads to inflated type I error rates. Genetics. 2008;178(1):609–610. doi:10.1534/genetics.107.074609

Clarke AJ, Ellinghaus U, Cortini A, Stranks A, Katharina Simon A, Botto M, Vyse TJ. Autophagy is activated in systemic lupus erythematosus and required for plasmablast development. Ann Rheum Dis. 2015;74(5):912–920. doi:10.1136/annrheumdis-2013-204343

Collins SR, Schuldiner M, Krogan NJ, Weissman JS. A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. Genome Biol. 2006;7(7):R63. doi:10.1186/gb-2006-7-7-r63

Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JLY, Toufighi K, Mostafavi S, *et al.* The genetic landscape of a cell. Science. 2010;327(5964):425–431. doi:10.1126/science.1180823

Cousminer DL, Wagley Y, Pippin JA, Elhakeem A, Way GP, Pahl MC, McCormack SE, Chesi A, Mitchell JA, Kindler JM, *et al.* Genome-wide association study implicates novel loci and reveals candidate effector genes for longitudinal pediatric bone accrual. Genome Biol. 2021;22(1):1. doi:10.1186/s13059-020-02207-9

Couto Alves A, De Silva NMG, Karhunen V, Sovio U, Das S, Taal HR, Warrington NM, Lewin AM, Kaakinen M, Cousminer DL, *et al.* GWAS on longitudinal growth traits reveals different genetic factors influencing infant, child, and adult BMI. Sci Adv. 2019;5(9):eaaw3095. doi:10.1126/sciadv.aaw3095

Crawford L, Wood KC, Zhou X, Mukherjee S. Bayesian approximate kernel regression with variable selection. J Am Stat Assoc. 2018;113(524):1710–1721. doi:10.1080/01621459.2017.1361830

Crawford L, Zeng P, Mukherjee S, Zhou X. Detecting epistasis with the marginal epistasis test in genetic mapping studies of quantitative traits. PLoS Genet. 2017;13(7):e1006869. doi:10.1371/journal.pgen.1006869

Crawford L, Zhou X. Genome-wide marginal epistatic association mapping in case-control studies. 2018. https://doi.org/10.1101/374983 Accessed: May 4, 2022.

Crow JF. On epistasis: why it is unimportant in polygenic directional selection. Philos Trans R Soc B Biol Sci. 2010;365(1544):1241–1244. doi:10.1098/rstb.2009.0275

Darnell G, Smith SP, Udwin D, Ramachandran S, Crawford L. Partitioning tagged non-additive genetic effects in summary statistics provides evidence of pervasive epistasis in complex traits. 2022. https://www.biorxiv.org/content/10.1101/2022.07.21.501001v3 Accessed: October 2, 2022.

Davies RB. Algorithm AS 155: the distribution of a linear combination of $\chi^2$ random variables. J R Stat Soc C (Appl Stat). 1980;29(3):323–333. doi:10.2307/2346911

Demmings EM, Williams BR, Lee C-R, Barba P, Yang S, Hwang C-F, Reisch BI, Chitwood DH, Londo JP. Quantitative trait locus analysis of leaf morphology indicates conserved shape loci in grapevine. Front Plant Sci. 2019;10:1373. doi:10.3389/fpls.2019.01373

Deutschbauer AM, Davis RW. Quantitative trait loci mapped to single-nucleotide resolution in yeast. Nat Genet. 2005;37(12):1333–1340. doi:10.1038/ng1674

Dewey M. metap: meta-analysis of significance values. R package version 1.8. 2022. https://cran.r-project.org/web/packages/metap/citation.html. Accessed: February 10, 2022.

Duchesne P, de Micheaux PL. Computing the distribution of quadratic forms: further comparisons between the Liu-Tang-Zhang approximation and exact methods. Comput Stat Data Anal. 2010;54:858–862. doi:10.1016/j.csda.2009.11.025

Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010;11(6):446–450. doi:10.1038/nrg2809

Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han B, Park YR, Raychaudhuri S, Pouget JG, Hübenthal M, *et al.* Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nat Genet. 2016;48(5):510–518. doi:10.1038/ng.3528

Fang G, Wang W, Paunic V, Heydari H, Costanzo M, Liu X, Liu X, VanderSluis B, Oately B, Steinbach M, *et al.* Discovering genetic interactions bridging pathways in genome-wide association studies. Nat Commun. 2019;10(1):4274. doi:10.1038/s41467-019-12131-7

Feng X, Ippolito GC, Tian L, Wiehagen K, Oh S, Sambandam A, Willen J, Bunte RM, Maika SD, Harriss JV, *et al.* Foxp1 is an essential transcriptional regulator for the generation of quiescent naive T cells during thymocyte development. Blood. 2010;115(3):510–518. doi:10.1182/blood-2009-07-232694

Fergus P, Montañez CC, Abdulaimma B, Lisboa P, Chalmers C, Pineles B. Utilizing deep learning and genome wide association studies for epistatic-driven preterm birth classification in African-American women. IEEE/ACM Trans Comput Biol Bioinformatics. 2020;17(2):668–678. doi:10.1109/TCBB.2018.2868667

Fisher RA. Statistical Methods for Research Workers. 5 ed. Edinburgh: Oliver and Boyd; 1925. http://www.haghish.com/resources/materials/Statistical˙Methods˙for˙Research˙Workers.pdf.

Flint J, Mackay TFC. Genetic architecture of quantitative traits in mice, flies, and humans. Genome Res. 2009;19(5):723–733. doi:10.1101/gr.086660.108

Forsberg SKG, Bloom JS, Sadhu MJ, Kruglyak L, Carlborg Ö. Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. Nat Genet. 2017;49(4):497–503. doi:10.1038/ng.3800

Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, Ma C, Fontanillas P, Moutsianas L, McCarthy DJ, *et al.* The genetic architecture of type 2 diabetes. Nature. 2016;536(7614):41–47. doi:10.1038/nature18642

Gaertner Byn E, Parmenter MD, Rockman MV, Kruglyak L, Phillips PC. More than the sum of its parts: a complex epistatic network underlies natural variation in thermal preference behavior in *Caenorhabditis elegans*. Genetics. 2012;192(4):1533–1542. doi:10.1534/genetics.112.142877

Gerke J, Lorenz K, Cohen B. Genetic interactions between transcription factors cause natural variation in yeast. Science. 2009;323(5913):498–501. doi:10.1126/science.1166426

Golan D, Lander ES, Rosset S. Measuring missing heritability: inferring the contribution of common variants. Proc Natl Acad Sci USA. 2014;111(49):E5272–E5281. doi:10.1073/pnas.1419064111

Guo Y, Cheng H, Yuan Z, Liang Z, Wang Y, Du D. Testing gene-gene interactions based on a neighborhood perspective in genome-wide association studies. Front Genet. 2021;12:801261. doi:10.3389/fgene.2021.801261

Guo Y, Zhong Z, Yang C, Hu J, Jiang Y, Liang Z, Gao H, Liu J. Epi-GTBN: an approach of epistasis mining based on genetic Tabu algorithm and Bayesian network. BMC Bioinformatics. 2019;20(1):444. doi:10.1186/s12859-019-3022-z

Gusareva ES, Carrasquillo MM, Bellenguez C, Cuyvers E, Colon S, Graff-Radford NR, Petersen RC, Dickson DW, John JMM, Bessonov K, *et al.* Genome-wide association interaction analysis for Alzheimer's disease. Neurobiol Aging. 2014;35(11):2436–2443. doi:10.1016/j.neurobiolaging.2014.05.014

Gusareva ES, Twizere J-C, Sleegers K, Dourlen P, Abisambra JF, Meier S, Cloyd R, Weiss B, Dermaut B, Bessonov K. Male-specific epistasis between WWC1 and TLN2 genes is associated with Alzheimer's disease. Neurobiol Aging. 2018;72:188.e3–188.e12. doi:10.1016/j.neurobiolaging.2018.08.001

Gusareva ES, Van Steen K. Practical aspects of genome-wide association interaction analysis. Hum Genet. 2014;133(11):1343–1358. doi:10.1007/s00439-014-1480-y

He X, Qian W, Wang Z, Li Y, Zhang J. Prevalent positive epistasis in *Escherichia coli* and *Saccharomyces cerevisiae* metabolic networks. Nat Genet. 2010;42(3):272–276. doi:10.1038/ng.524

Hemani G, Theocharidis A, Wei W, Haley C. EpiGPU: exhaustive pairwise epistasis scans parallelized on consumer level graphics cards. Bioinformatics. 2011;27(11):1462–1465. doi:10.1093/bioinformatics/btr172

Hill WG, Goddard ME, Visscher PM. Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 2008;4(2):e1000008. doi:10.1371/journal.pgen.1000008

Hivert V, Sidorenko J, Rohart F, Goddard ME, Yang J, Wray NR, Yengo L, Visscher PM. Estimation of non-additive genetic variance in human complex traits from a large sample of unrelated individuals. Am J Hum Genet. 2021;108(5):786–798. doi:10.1016/j.ajhg.2021.02.014

Horn T, Sandmann T, Fischer B, Axelsson E, Huber W, Boutros M. Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. Nat Methods. 2011;8(4):341–346. doi:10.1038/nmeth.1581

Hu H, Wang B, Borde M, Nardone J, Maika S, Allred L, Tucker PW, Rao A. Foxp1 is an essential transcriptional regulator of B cell development. Nat Immunol. 2006;7(8):819–826. doi:10.1038/ni1358

Huang W, Mackay TFC. The genetic architecture of quantitative traits cannot be inferred from variance component analysis. PLoS Genet. 2016;12(11):e1006421. doi:10.1371/journal.pgen.1006421

IMPC. Obtaining and loading phenotype annotations from the international mouse phenotyping consortium (IMPC) database; 2014. [accessed 2022 Aug 2]. http://www.informatics.jax.org/reference/J:211773

Isnard P, Coré N, Naquet P, Djabali M. Altered lymphoid development in mice deficient for the mAF4 proto-oncogene. Blood. 2000;96(2): 705–710. doi:10.1182/blood.V96.2.705

Jarvis JP, Cheverud JM. Mapping the epistatic network underlying murine reproductive fatpad variation. Genetics. 2011;187(2): 597–610. doi:10.1534/genetics.110.123505

Jiang Y, Reif JC. Modeling epistasis in genomic selection. Genetics. 2015;201(2):759–768. doi:10.1534/genetics.115.177907

Kerin M, Marchini J. A non-linear regression method for estimation of gene–environment heritability. Bioinformatics. 2020a;36(24): 5632–5639. doi:10.1093/bioinformatics/btaa1079

Kerin M, Marchini J. Inferring gene-by-environment interactions with a Bayesian whole-genome regression model. Am J Hum Genet. 2020b;107(4):698–713. doi:10.1016/j.ajhg.2020.08.009

Kirino Y, Bertsias G, Ishigatsubo Y, Mizuki N, Tugal-Tutkun I, Seyahi E, Ozyazgan Y, Sevgi Sacli F, Erer B, Inoko H. Genome-wide association analysis identifies new susceptibility loci for Behçet's disease and epistasis between HLA-b*51 and ERAP1. Nat Genet. 2013; 45(2):202–207. doi:10.1038/ng.2520

Ko S, German CA, Jensen A, Shen J, Wang A, Mehrotra DV, Sun YV, Sinsheimer JS, Zhou H, Zhou JJ. GWAS of longitudinal trajectories at biobank scale. Am J Hum Genet. 2022;109(3):433–445. doi:10.1016/j.ajhg.2022.01.018

Kroymann J, Mitchell-Olds T. Epistasis and balanced polymorphism influencing complex trait variation. Nature. 2005;435(7038): 95–98. doi:10.1038/nature03480

Kuonen D. Saddlepoint approximations for distributions of quadratic forms in normal variables. Biometrika. 1999;86(4):929–935. doi:10.1093/biomet/86.4.929

Leamy L, Gordon R, Pomp D. Sex-, diet-, and cancer-dependent epistatic effects on complex traits in mice. Front Genet. 2011;2:71. doi:10.3389/fgene.2011.00071

Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet. 2011;88(3):294–305. doi:10.1016/j.ajhg.2011.02.002

Lehner B, Crombie C, Tischler J, Fortunato A, Fraser AG. Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. Nat Genet. 2006;38(8):896–903. doi:10.1038/ng1844

Lewinger JP, Morrison JL, Thomas DC, Murcray CE, Conti DV, Li D, Gauderman WJ. Efficient two-step testing of gene-gene interactions in genome-wide association studies. Genet Epidemiol. 2013;37(5):440–451. doi:10.1002/gepi.21720

Lishout FV, Mahachie John JM, Gusareva ES, Urrea V, Cleynen I, Théâtre E, Charloteaux B, Calle ML, Wehenkel L, Steen KV. An efficient algorithm to perform multiple testing in epistasis screening. BMC Bioinformatics. 2013;14(1):138. doi:10.1186/1471-2105-14-138

Liu Y, Xie J. Cauchy combination test: a powerful test with analytic P-value calculation under arbitrary dependency structures. J Am Stat Assoc. 2020;115(529):393–402. doi:10.1080/01621459.2018.1554485

Long AD, Langley CH. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. Genome Res. 1999;9(8):720–731. doi:10.1101/gr.9.8.720

Loos RJF. 15 years of genome-wide association studies and no signs of slowing down. Nat Commun. 2020;11(1):5900. doi:10.1038/s41467-020-19653-5

Ma L, Clark AG, Keinan A. Gene-based testing of interactions in association studies of quantitative traits. PLoS Genet. 2013;9(2): e1003321. doi:10.1371/journal.pgen.1003321

Marchini J, Donnelly P, Cardon LR. Genome-wide strategies for detecting multiple loci that influence complex diseases. Nat Genet. 2005;37(4):413–417. doi:10.1038/ng1537

Monnahan PJ, Kelly JK. Epistasis is a major determinant of the additive genetic variance in *Mimulus guttatus*. PLoS Genet. 2015;11(5): e1005201. doi:10.1371/journal.pgen.1005201

Moore R, Casale FP, Bonder MJ, Horta D, Franke L, Barroso I, Stegle O. A linear mixed-model approach to study multivariate gene–environment interactions. Nat Genet. 2019;51(1):180–186. doi:10.1038/s41588-018-0271-0

Mortensen M, Soilleux EJ, Djordjevic G, Tripp R, Lutteropp M, Sadighi-Akha E, Stranks AJ, Glanville J, Knight S, Jacobsen S-EW, *et al.* The autophagy protein Atg7 is essential for hematopoietic stem cell maintenance. J Exp Med. 2011;208(3):455–467. doi:10.1084/jem.20101145

Muñoz PR, Resende MFR, Jr, Gezan SA, Resende MDV, de los Campos G, Kirst M, Huber D, Peter GF. Unraveling additive from nonadditive effects using genomic relationship matrices. Genetics. 2014; 198(4):1759–1768. doi:10.1534/genetics.114.171322

Nagai A, Hirata M, Kamatani Y, Muto K, Matsuda K, Kiyohara Y, Ninomiya T, Tamakoshi A, Yamagata Z, Mushiroda T, *et al.* Overview of the BioBank Japan Project: study design and profile. J Epidemiol. 2017;27(Suppl. 3):S2–S8. doi:10.1016/j.je.2016.12.005

Nakayama K, Hatakeyama S, Maruyama S-i, Kikuchi A, Onoé K, Good RA, Nakayama KI. Impaired degradation of inhibitory subunit of NF-κb (*Ikappa*B) and β-catenin as a result of targeted disruption of the β-TrCP1 gene. Proc Natl Acad Sci USA. 2003;100(15): 8752–8757. doi:10.1073/pnas.1133216100

Norris JM, Rich SS. Genetics of glucose homeostasis. Arterioscler Thromb Vasc Biol. 2012;32(9):2091–2096. doi:10.1161/ATVBAHA.112.255463

Onge RPS, Mani R, Oh J, Proctor M, Fung E, Davis RW, Nislow C, Roth FP, Giaever G. Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. Nat Genet. 2007; 39(2):199–206. doi:10.1038/ng1948

Patel RA, Musharoff SA, Spence JP, Pimentel H, Tcheandjieu C, Mostafavi H, Sinnott-Armstrong N, Clarke SL, Smith CJ, Durda PP, *et al.* Genetic interactions drive heterogeneity in causal variant effect sizes for gene expression and complex traits. Am J Hum Genet. 2022;109(7):1286–1297. doi:10.1016/j.ajhg.2022.05.014

Pazokitoroudi A, Chiu AM, Burch KS, Pasaniuc B, Sankararaman S. Quantifying the contribution of dominance deviation effects to complex trait variation in biobank-scale data. Am J Hum Genet. 2021;108(5):799–808. doi:10.1016/j.ajhg.2021.03.018

Pazokitoroudi A, Wu Y, Burch KS, Hou K, Zhou A, Pasaniuc B, Sankararaman S. Efficient variance components analysis across millions of genomes. Nat Commun. 2020;11(1):4020. doi:10.1038/s41467-020-17576-9

Pecanka J, Jonker MA. Two-stage testing for epistasis: screening and verification. In: Wong K-C, editor. Epistasis: Methods and Protocols. New York (NY): Springer US; 2021. p. 69–92; https://doi.org/10.1007/978-1-0716-0947-7˙6.

Perez P, de los Campos G. Genome-wide regression and prediction with the BGLR statistical package. Genetics. 2014;198(2):483–495.

Peripato AC, De Brito RA, Matioli SR, Pletscher LS, Vaughn TT, Cheverud JM. Epistasis affecting litter size in mice. J Evol Biol. 2004;17(3):593–602. doi:10.1111/j.1420-9101.2004.00702.x

Pettersson M, Besnier F, Siegel PB, Carlborg Ö. Replication and explorations of high-order epistasis using a large advanced intercross line pedigree. PLoS Genet. 2011;7(7):e1002180. doi:10.1371/journal.pgen.1002180

Phillips AM, Lawrence KR, Moulana A, Dupic T, Chang J, Johnson MS, Cvijovic I, Mora T, Walczak AM, Desai MM. Binding affinity

landscapes constrain the evolution of broadly neutralizing anti-influenza antibodies. eLife. 2021;10:e71393. doi:10.7554/eLife.71393

Prabhu S, Pe'er I. Ultrafast genome-wide scan for SNP–SNP interactions in common complex disease. Genome Res. 2012;22(11):2230–2240. doi:10.1101/gr.137885.112

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–575. doi:10.1086/519795

Raaijmakers MHGP, Mukherjee S, Guo S, Zhang S, Kobayashi T, Schoonmaker JA, Ebert BL, Al-Shahrour F, Hasserjian RP, Scadden EO, *et al.* Bone progenitor dysfunction induces myelodysplasia and secondary leukaemia. Nature. 2010;464(7290):852–857. doi:10.1038/nature08851

Ripke S, Neale BM, Corvin A, Walters JTR, Farh K-H, Holmans PA, Lee P, Bulik-Sullivan B, Collier DA, Huang H, *et al.* Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511(7510):421–427. doi:10.1038/nature13595

Runcie DE, Qu J, Cheng H, Crawford L. MegaLMM: mega-scale linear mixed models for genomic predictions with thousands of traits. Genome Biol. 2021;22(1):213. doi:10.1186/s13059-021-02416-w

Satterthwaite FE. An approximate distribution of estimates of variance components. Biom Bull. 1946;2(6):110–114. doi:10.2307/3002019

Schüpbach T, Xenarios I, Bergmann S, Kapur K. FastEpistasis: a high performance computing solution for quantitative trait epistasis. Bioinformatics. 2010;26(11):1468–1469. doi:10.1093/bioinformatics/btq147

Searle SR. Estimating multivariate variance and covariance components using quadratic and bilinear forms. Biom J. 1979;21(4):389–398. doi:10.1002/bimj.4710210407

Shao H, Burrage LC, Sinasac DS, Hill AE, Ernest SR, O'Brien W, Courtland H-W, Jepsen KJ, Kirby A, Kulbokas EJ, *et al.* Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. Proc Natl Acad Sci USA. 2008;105(50):19910–-19914. doi:10.1073/pnas.0810388105

Sheppard B, Rappoport N, Loh P-R, Sanders SJ, Zaitlen N, Dahl A. A model and test for coordinated polygenic epistasis in complex traits. Proc Natl Acad Sci USA. 2021;118(15):e1922305118. doi:10.1073/pnas.1922305118

Shifman S, Bell JT, Copley RR, Taylor MS, Williams RW, Mott R, Flint J. A high-resolution single nucleotide polymorphism genetic map of the mouse genome. PLoS Biol. 2006;4(12):e395. doi:10.1371/journal.pbio.0040395

Smith CL, Eppig JT. The mammalian phenotype ontology: enabling robust annotation and comparative analysis. Wiley Interdiscip Rev Syst Biol Med. 2009;1(3):390–399. doi:10.1002/wsbm.44

Souto JC, Almasy L, Borrell M, Blanco-Vaca F, Mateo J, Soria JM, Coll I, Felices R, Stone W, Fontcuberta J, *et al.* Genetic susceptibility to thrombosis and its relationship to physiological risk factors: the GAIT study. Am J Hum Genet. 2000;67(6):1452–1459. doi:10.1086/316903

Stamp J. Leveraging the genetic correlation between traits improves the detection of epistasis in genome-wide association studies. 2022. https://doi.org/10.7910/DVN/WPFIGU Accessed: January 31, 2023.

Szappanos B, Kovács K, Szamecz B, Honti F, Costanzo M, Baryshnikova A, Gelius-Dietrich G, Lercher MJ, Jelasity M, Myers CL, *et al.* An integrated approach to characterize genetic interaction networks in yeast metabolism. Nat Genet. 2011;43(7):656–662. doi:10.1038/ng.846

Tang W, Wu X, Jiang R, Li Y. Epistatic module detection for case-control studies: a Bayesian model with a Gibbs sampling strategy. PLoS Genet. 2009;5(5):e1000464. doi:10.1371/journal.pgen.1000464

Tao S, Feng J, Webster T, Jin G, Hsu F-C, Chen S-H, Kim S-T, Wang Z, Zhang Z, Zheng SL, *et al.* Genome-wide two-locus epistasis scans in prostate cancer using two European populations. Hum Genet. 2012;131(7):1225–1234. doi:10.1007/s00439-012-1148-4

Tong AHY, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M. Global mapping of the yeast genetic interaction network. Science. 2004;303(5659):808–813. doi:10.1126/science.1091317

Turchin MC, Darnell G, Crawford L, Ramachandran S. Pathway analysis within multiple human ancestries reveals novel signals for epistasis in complex traits. 2020. https://doi.org/10.1101/2020.09.24.312421 Accessed: October 14, 2022.

Valdar W, Flint J, Mott R. Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. Genetics. 2006;172(3):1783–1797. doi:10.1534/genetics.104.039313

Valdar W, Solberg LC, Gauguier D, Burnett S, Klenerman P, Cookson WO, Taylor MS, Rawlins JP, Mott R, Flint J. Genome-wide genetic association of complex traits in heterogeneous stock mice. Nat Genet. 2006;38(8):879–887. doi:10.1038/ng1840

van de Haar J, Canisius S, Yu MK, Voest E E, Wessels LFA, Ideker T. Identifying epistasis in cancer genomes: a delicate affair. Cell. 2019;177(6):1375–1383. doi:10.1016/j.cell.2019.05.005

Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet. 2012;90(1):7–24. doi:10.1016/j.ajhg.2011.11.029

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet. 2017;101(1):5–22. doi:10.1016/j.ajhg.2017.06.005

Wainschtein P, Jain D, Zheng Z, Cupples LA, Shadyab AH, McKnight B, Shoemaker BM, Mitchell BD, Psaty BM, Kooperberg C, *et al.* Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. Nat Genet. 2022;54(3):263–273. doi:10.1038/s41588-021-00997-7

Wan X, Yang C, Yang Q, Xue H, Fan X, Tang NLS, Yu W. BOOST: a fast approach to detecting gene-gene interactions in genome-wide case-control studies. Am J Hum Genet. 2010;87(3):325–340. doi:10.1016/j.ajhg.2010.07.021

Wanstrat A, Wakeland E. The genetics of complex autoimmune diseases: non-MHC susceptibility genes. Nat Immunol. 2001;2(9):802–809. doi:10.1038/ni0901-802

Wei W-H, Hemani G, Gyenesei A, Vitart V, Navarro P, Hayward C, Cabrera CP, Huffman JE, Knott SA, Hicks AA. Genome-wide analysis of epistasis in body mass index using multiple human populations. Eur J Hum Genet. 2012;20(8):857–862. doi:10.1038/ejhg.2012.17

Weinreich DM, Delaney NF, DePristo MA, Hartl DL. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science. 2006;312(5770):111–114. doi:10.1126/science.1123539

Weischenfeldt J, Damgaard I, Bryder D, Theilgaard-Mónch K, Thoren LA, Nielsen FC, Jacobsen SEW, Nerlov C, Torben Porse B. NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. Genes Dev. 2008;22(10):1381–1396. doi:10.1101/gad.468808

Wentzell AM, Rowe HC, Hansen BG, Ticconi C, Halkier BA, Kliebenstein DJ. Linking metabolic QTLs with network and cis-eQTLs controlling biosynthetic pathways. PLoS Genet. 2007;3(9):e162. doi:10.1371/journal.pgen.0030162

Wilson DJ. Harmonic mean *P*-values and model averaging by mean maximum likelihood [R package version 3.0]. 2019b. https://CRAN.R-project.org/package=harmonicmeanp Accessed: May 8, 2023.

Wilson DJ. The harmonic mean *P*-value for combining dependent tests. Proc Natl Acad Sci USA. 2019a;116(4):1195–1200. doi:10.1073/pnas.1814092116

Wood AR, Tuke MA, Nalls MA, Hernandez DG, Bandinelli S, Singleton AB, Melzer D, Ferrucci L, Frayling TM, Weedon MN. Another explanation for apparent epistasis. Nature. 2014;514(7520):E3–E5. doi:10.1038/nature13691

Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011;89(1):82–93. doi:10.1016/j.ajhg.2011.05.029

Wu Y, Sankararaman S. A scalable estimator of SNP heritability for biobank-scale data. Bioinformatics. 2018;34(13):i187–i194. doi:10.1093/bioinformatics/bty253

Wu W, Wang Z, Xu K, Zhang X, Amei A, Gelernter J, Zhao H, Justice AC, Wang Z. Retrospective association analysis of longitudinal binary traits identifies important loci and pathways in cocaine use. Genetics. 2019;213(4):1225–1236. doi:10.1534/genetics.119.302598

Xu F, Davis J, Hoos M, Van Nostrand WE. Mutation of the Kunitz-type proteinase inhibitor domain in the amyloid β-protein precursor abolishes its anti-thrombotic properties in vivo. Thromb Res. 2017;155:58–64. doi:10.1016/j.thromres.2017.05.003

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, *et al.* Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42(7):565–569. doi:10.1038/ng.608

Yang Y, McBride KM, Hensley S, Lu Y, Chedin F, Bedford MT. Arginine methylation facilitates the recruitment of TOP3b to chromatin to prevent R loop accumulation. Mol Cell. 2014;53(3):484–497. doi:10.1016/j.molcel.2014.01.011

Zhang X, Huang S, Zou F, Wang W. TEAM: efficient two-locus epistasis tests in human genome-wide association study. Bioinformatics. 2010;26(12):i217–i227. doi:10.1093/bioinformatics/btq186

Zhang Y, Liu JS. Bayesian inference of epistatic interactions in case-control studies. Nat Genet. 2007;39(9):1167–1173. doi:10.1038/ng2110

Zhang Y, Zhang J, Liu JS. Block-based Bayesian epistasis association mapping with application to WTCCC type 1 diabetes data. Ann Appl Stat. 2011;5(3):2052–2077. doi:10.1214/11-AOAS469

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

Zhou X, Carbonetto P, Stephens M. Polygenic modeling with Bayesian sparse linear mixed models. PLoS Genet. 2013;9(2):e1003264. doi:10.1371/journal.pgen.1003264

Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat Methods. 2014;11(4):407–409. doi:10.1038/nmeth.2848

Zhou J, Wong MS, Chen W-C, Krainer AR, Kinney JB, McCandlish DM. Higher-order epistasis and phenotypic prediction. Proc Natl Acad Sci USA. 2022;119(39):e2204233119. doi:10.1073/pnas.2204233119

Zhu S, Fang G. MatrixEpistasis: ultrafast, exhaustive epistasis scan for quantitative traits with covariate adjustment. Bioinformatics. 2018;34(14):2341–2348. doi:10.1093/bioinformatics/bty094

Zhu C, Ming MJ, Cole JM, Kirkpatrick M, Harpak A. Amplification is the primary mode of gene-by-sex interaction in complex human traits. 2022. https://doi.org/10.1101/2022.05.06.490973 Accessed: November 19, 2022.

Zhu H, Zhou X. Statistical methods for SNP heritability estimation and partition: a review. Comput Struct Biotechnol J. 2020;18:1557–1568. doi:10.1016/j.csbj.2020.06.011

Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: genetic interactions create phantom heritability. Proc Natl Acad Sci USA. 2012;109(4):1193–1198. doi:10.1073/pnas.1119675109

*Editor: D.-J. de Koning*