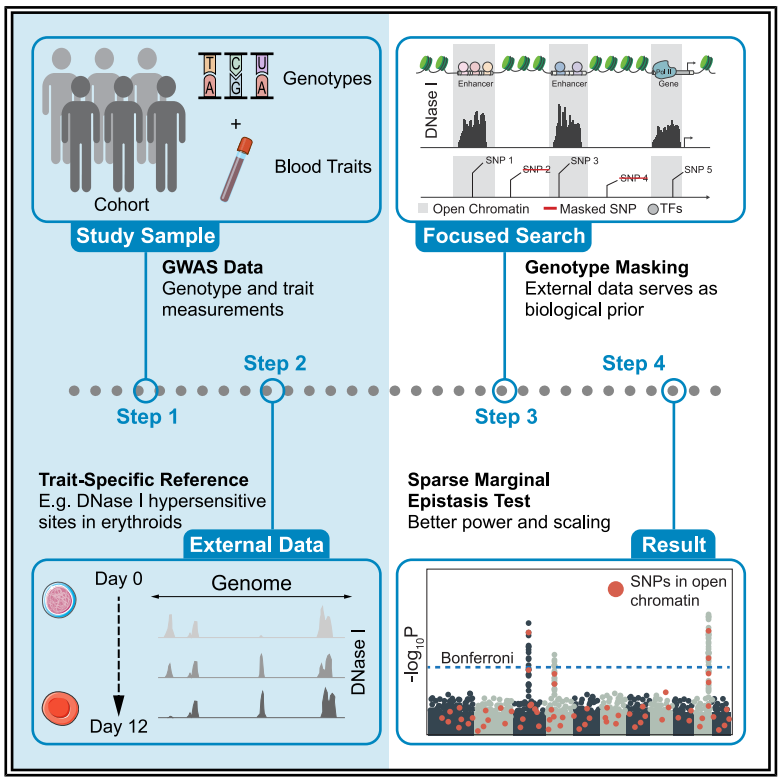
Sparse modeling of interactions enables fast detection of genome-wide epistasis in biobank-scale studies

Graphical abstract



Authors

Julian Stamp, Samuel Pattillo Smith, Daniel Weinreich, Lorin Crawford

Correspondence

julian_stamp@brown.edu (J.S.), lcrawford@microsoft.com (L.C.)

The sparse marginal epistasis test overcomes computational limitations of previous mapping approaches by focusing its epistatic search to regions of the genome that have some known functional relationship with a trait. This strategy both significantly improves its statistical power and allows it to scale genome wide for large human biobank studies.

Stamp et al., 2025, The American Journal of Human Genetics *112*, 2198–2212
September 4, 2025
https://doi.org/10.1016/j.ajhg.2025.07.004

CellPress

ARTICLE

Sparse modeling of interactions enables fast detection of genome-wide epistasis in biobank-scale studies

Julian Stamp,[1,*] Samuel Pattillo Smith,[2,3,6] Daniel Weinreich,[1,4] and Lorin Crawford[5,*]

Summary

The lack of computational methods capable of detecting epistasis in biobanks has led to uncertainty about the role of non-additive genetic effects on complex trait variation. The marginal epistasis framework is a powerful approach because it estimates the likelihood of a SNP being involved in any interaction, thereby reducing the multiple testing burden. Current implementations of this approach have failed to scale genome wide in large human studies. To address this, we present the sparse marginal epistasis (SME) test, which concentrates the scans for epistasis to regions of the genome that have known functional enrichment for a quantitative trait of interest. By leveraging the sparse nature of this modeling setup, we develop a statistical algorithm that allows SME to run 10–90 times faster than state-of-the-art epistatic mapping methods. In a study of complex traits measured in 349,411 individuals from the UK Biobank, we show that reducing searches of epistasis to variants in functionally enriched regions facilitates the identification of genetic interactions associated with regulatory genomic elements.

Introduction

Genome-wide association studies (GWASs) have identified thousands of genetic loci linked with various complex traits and common diseases, offering valuable insights into the genetic foundations of phenotypic variation.[1] As of late, there have been many efforts to estimate proportions of genetic variance beyond what is attributable to additive effects.[2–6] Epistasis, which refers to interactions between genetic loci, is thought to play a key role in constituting the genetic basis of evolution.[7,8] While many studies have shown epistasis to be pervasive in model organisms,[9–11] controversies remain with respect to its role in humans.[12] For example, some epistatic interactions identified in association mapping studies can be explained by additive effects of unobserved variants.[13] Though previous studies have shown that genetic variance is mainly additive,[9,14] these conclusions have recently been challenged.[5]

Numerous statistical methods have been developed to identify single-nucleotide polymorphisms (SNPs) that contribute to epistasis. Traditional approaches focus on explicitly detecting significant interactions through exhaustive or probabilistic searches utilizing frequentist tests, Bayesian inference, and machine learning techniques.[15–18] With advancements in sequencing technologies, many contemporary GWASs are conducted on biobank-scale datasets comprising hundreds of thousands of individuals genotyped at millions of markers and phenotyped for thousands of traits.[1,19,20] This is crucial, as the effect of epistatic interactions is hypothesized to be small for many traits,[12,14] and traditional search algorithms are known to be most powered when large training datasets are available.[14,15] However, despite efficient computational improvements, exploring large combinatorial domains continues to pose a challenge for epistatic mapping studies. With a lack of *a priori* knowledge about which epistatic loci to prioritize, exploring all possible combinations of genetic variants can result in low statistical power after correcting for multiple hypothesis tests (e.g., there are J choose 2 possible pairwise combinations for a study with J SNPs).

As an alternative to traditional exhaustive search methods, the marginal epistasis framework was developed to estimate the combined pairwise interaction effects between a focal SNP and all other variants in the dataset. The "marginal epistasis test" (MAPIT) evaluates each SNP individually and identifies candidates involved in epistasis without requiring the identification of their exact interacting partners.[2] Recently, the concept of marginal epistasis has been leveraged to estimate the contribution of non-additive heritability in complex traits using GWAS summary statistics.[5] It has also been extended to explore the importance of genetic interactions across multiple traits simultaneously.[3] Theoretically, MAPIT is formulated as a linear mixed model where the random effects and corresponding variance components are estimated using a method-of-moments (MoM) algorithm.[21,22] Although MAPIT mitigates the reduction of -power due to the multiple testing burden, its

[1]Center for Computational Molecular Biology, Brown University, Providence, RI, USA; [2]Department of Integrative Biology, University of Texas at Austin, Austin, TX, USA; [3]Department of Population Health, University of Texas at Austin, Austin, TX, USA; [4]Department of Ecology, Evolution, and Organismal Biology, Brown University, Providence, RI, USA; [5]Microsoft Research, Cambridge, MA, USA
[6]Present address: Genomics Ltd., Oxford, UK
*Correspondence: julian_stamp@brown.edu (J.S.), lcrawford@microsoft.com (L.C.)
https://doi.org/10.1016/j.ajhg.2025.07.004.

implementation on datasets with large sample sizes remains computationally intensive.[2] Specifically, the computational complexity scales linearly with the number of SNPs and (at best) quadratically with the number of individuals, making it suitable for moderately sized GWAS applications but infeasible for biobank-scale studies.[23,24] Efforts have been made to address this limitation, such as the "fast marginal epistasis test" (FAME),[4] which leverages a stochastic MoM framework and introduces both computationally efficient stochastic trace estimators[25] and innovative methods to expedite matrix multiplication.[26] However, despite these advancements, further work is necessary to scale the method to genome-wide applications.

This work introduces the sparse marginal epistasis (SME) test, which focuses on searching for epistasis in regions of the genome with known functional enrichment[27] related to a quantitative trait of interest. This method has two main advantages. First, it prioritizes candidate regions likely to involve epistatic gene action. Studies have indicated that variants in coding regions account for less than 10% of the phenotypic variance in many traits and diseases.[28] Consequently, the remaining heritability is attributed to regions expected to play a regulatory role[27–29] and that are active in trait-specific tissue.[30,31] Second, the sparse nature of this approach leads to more efficient estimators for model parameters[22,32] and allows SME to operate significantly faster than existing methods, such as MAPIT and FAME. Through detailed simulations, SME demonstrates effective type I error control and improved power compared to previous approaches. Furthermore, utilizing information from DNase I-hypersensitivity sites in *ex vivo* human erythroid differentiation[33] and GWAS summary statistics, we use SME to analyze complex traits in individuals from the UK Biobank[19] and identify genetic interactions associated with regulatory genomic elements.

Material and methods

The SME test

The SME test performs a genome-wide search for SNPs involved in genetic interactions while conditioning on information derived from functional genomic data (Figure 1A). Consider a GWAS with N individuals who have been genotyped for J SNPs encoded as $\{0, 1, 2\}$ copies of a reference allele at each locus. Also assume that we have access to an external reference $\mathcal{S}$ that encodes some additional biological information about the quantitative trait being studied. The marginal epistasis test aims to identify genetic variants that are involved in epistasis without exhaustively searching over all possible interactions.[2] By examining one SNP at a time (indexed by j), SME fits the following linear mixed model:

$$\mathbf{y} = \mu + \sum_l \mathbf{x}_l\beta_l + \sum_{l\neq j}(\mathbf{x}_j \circ \mathbf{x}_l)\alpha_l \cdot 1_{\mathcal{S}}(w_l) + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \tau^2\mathbf{I}) \quad \text{(Equation 1)}$$

where $\mathbf{y}$ is an N-dimensional quantitative trait vector measured for each individual in the study; μ is an intercept term; $\mathbf{X}$ is the $N \times J$ matrix of allele counts that have been column standardized across individuals with $\mathbf{x}_l$ representing an N-dimensional vector for the l-th SNP; β_l is the additive effect for the l-th SNP; $\mathbf{x}_j \circ \mathbf{x}_l$ is the Hadamard (elementwise) product of the two genotypic vectors with corresponding interaction effect size α_l; $\boldsymbol{\varepsilon}$ is a normally distributed error term with mean zero and scale variance term τ^2; and $\mathbf{I}$ denotes an $N \times N$ identity matrix. The key to this formulation is that the inclusion of the interaction between the j-th and l-th SNPs is based on an indicator function

$$1_{\mathcal{S}}(w_l) = \begin{cases} 1 & \text{if} \quad w_l \in \mathcal{S} \\ 0 & \text{if} \quad w_l \notin \mathcal{S} \end{cases}, \quad \text{(Equation 2)}$$

where w_l encodes information about the l-th SNP. For example, if testing for epistatic effects in red blood cell traits, we can incorporate information about regulatory regions during erythroid differentiation into the model (Figure 1B). In this case, $\mathcal{S}$ could be a set of genomic regions for which DNase sequencing (DNase-seq) implicates chromatin accessibility and w_l could encode the physical location of the l-th SNP on the genome. Here, $1_{\mathcal{S}}(w_l) = 1$ if the l-th SNP is located in one of these regions (i.e., $w_l \in \mathcal{S}$). This means that while all SNPs are tested for marginal epistasis, only their interactions with SNPs included in the mask resulting from Equation 2 are considered.

The benefits of this sparse approach are 2-fold. First, by limiting the search to regions of the genome that are most likely to be functionally associated with a phenotype, SME produces significantly more efficient estimators, which leads to an increase in power (Figure 1C). Second, by masking out sets of variants for each test on a focal SNP, SME leverages a fast MoM algorithm to substantially improve its scalability for genome-wide analyses (Figure 1D). Specifically, SME introduces an approximation to the efficient stochastic trace estimator,[4,23,25,34] which allows the algorithm to avoid repeating costly matrix computations when estimating model parameters across each SNP that is being tested (Figures S1–S3).

Variance component model formulation

For biobank-scale data, there are often more SNPs than individuals. To overcome an undetermined system in Equation 1, the marginal epistasis framework assumes that the effect sizes follow univariate normal distributions where $\beta_l \sim \mathcal{N}(0, \omega^2/J)$ and $\alpha_l \sim \mathcal{N}(0, \sigma^2/J^*)$, with $J^* = \sum_{l\neq j} 1_{\mathcal{S}}(w_l)$ representing the number of interactions considered in the model.[2,32,35–37] Assuming that the phenotype has been mean centered and scaled, these normal assumptions allow Equation 1 to be rewritten as

$$\mathbf{y} = \mathbf{m} + \mathbf{g}_j + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \tau^2\mathbf{I}), \quad \text{(Equation 3)}$$

where $\mathbf{m} = \sum_l \mathbf{x}_l\beta_l$ is the combined additive effects from all SNPs and $\mathbf{g}_j = \sum_{l\neq j}(\mathbf{x}_j \circ \mathbf{x}_l)\alpha_l \cdot 1_{\mathcal{S}}(w_l)$ represents the effects of a subset of pairwise interactions involving the j-th SNP.

Probabilistically, Equation 3 translates to SME assuming that $\mathbf{m} \sim \mathcal{N}(\mathbf{0}, \omega^2\mathbf{K})$, where the covariance matrix $\mathbf{K} = \mathbf{X}\mathbf{X}^\intercal/J$ accounts for the relatedness between individuals in the data and the corresponding component ω^2 models the phenotypic variance explained (PVE) by additive effects. The second term can be written as $\mathbf{g}_j \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{G}_j)$, where $\mathbf{G}_j = \mathbf{D}_j\mathbf{X}_{-j}\mathbf{W}_j\mathbf{X}_{-j}^\intercal\mathbf{D}_j/J^*$, with $\mathbf{X}_{-j}$ denoting the genotype matrix without the j-th SNP and $\mathbf{D}_j = \text{diag}(\mathbf{x}_j)$ representing an $N \times N$ diagonal matrix. Importantly, $\mathbf{W}_j = \text{diag}[1_{\mathcal{S}}(w_1), \ldots, 1_{\mathcal{S}}(w_{J-1})]$ has binary diagonal elements that only equate to 1 if the l-th SNP satisfies the criteria from the external data source $\mathcal{S}$. Altogether, these results show

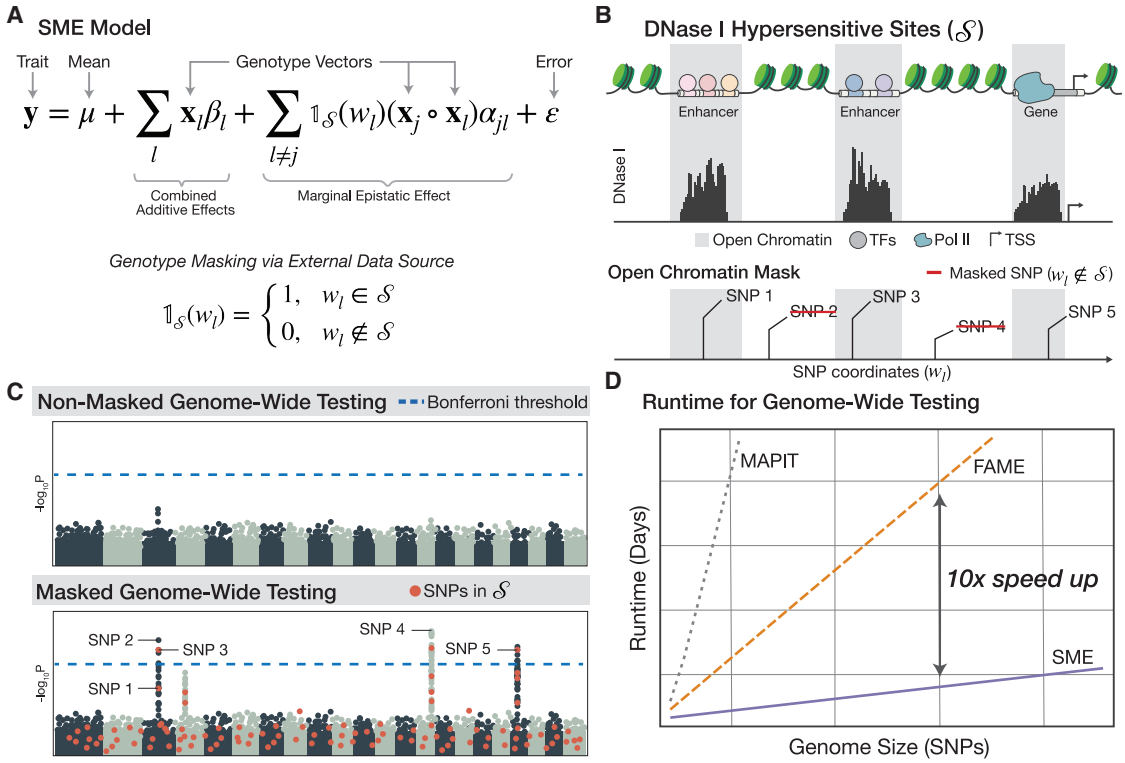


Figure 1. Schematic overview of the sparse marginal epistasis test

(A) Sparse marginal epistasis (SME) examines one SNP at a time and estimates marginal epistatic effects—the combined pairwise interaction effects between a j-th focal SNP and other variants on the genome (indexed by $l \neq j$). The key to SME is that it incorporates genomic data $\mathcal{S}$ through a binary indicator function $1_{\mathcal{S}}(w_l)$, where w_l provides information about the l-th background SNP. This creates a mask to only search for interactions in regions of the genome with known functional enrichment related to a trait of interest.

(B) As an example, let data on DNase I-hypersensitive sites (DHS) be used for $\mathcal{S}$. In this case, SME restricts the marginal epistasis test to assessing interactions between each focal SNP and variants with genomic coordinates that fall within open chromatin and regulatory regions. The DNase-seq signal is converted into a binary mask, excluding variants located in regions with closed chromatin (i.e., variants with coordinates $w_l \notin \mathcal{S}$).

(C) SME tests every SNP genome wide. Masking results in improved power to detect marginally epistatic variants versus the traditional non-masking approach.

(D) SME uses computationally efficient estimators to enable genome-wide testing on biobank-scale datasets. It achieves runtimes $10\times$ –$90\times$ faster than current state-of-the-art methods.

that the covariance matrix $\mathbf{G}_j$ represents all pairwise interactions involving the j-th SNP that have not been masked out according to the set of indicator functions $\{1_{\mathcal{S}}(w_l)\}_{l\neq j}$. The main takeaway from the variance component formulation of SME is that the term σ^2 measures SNP-specific contribution to the non-additive genetic variance.

Point estimates and hypothesis testing

The model in Equation 3 has three variance components that can be estimated using a computationally efficient MoM algorithm.[22] In expectation,

$$\mathbb{E}[\mathbf{y}^\intercal\mathbf{A}\mathbf{y}] = \sum_{k=1}^{3} \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}_{jk})\delta_k, \qquad \text{(Equation 4)}$$

with $\mathbf{A}$ being a symmetric and non-negative definite matrix used to create weighted second moments, $\mathrm{tr}(\cdot)$ denotes the trace of a matrix, and we use shorthand to represent $[\boldsymbol{\Sigma}_{j1};\boldsymbol{\Sigma}_{j2};\boldsymbol{\Sigma}_{j3}] = [\mathbf{K};\mathbf{G}_j;\mathbf{I}]$ 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 $\mathbf{y}^\intercal\mathbf{A}\mathbf{y}$. We also use the realized covariance matrices in place of the arbitrary $\mathbf{A}$. The point estimates for each variance component are then given as

$$\widehat{\delta}_{jk} = \mathbf{y}^\intercal\mathbf{H}_{jk}\mathbf{y}, \qquad \text{(Equation 5)}$$

where $\mathbf{H}_{jk} = \sum_{t=1}^{3}(\mathbf{S}_j^{-1})_{kt}\boldsymbol{\Sigma}_{jt}$ and $\mathbf{S}_j$ is a 3×3 matrix with elements $(\mathbf{S}_j)_{kt} = \mathrm{tr}(\boldsymbol{\Sigma}_{jk}\boldsymbol{\Sigma}_{jt})$.

SME tests for non-zero marginal epistasis using a one-sided Z score or normal test. This is equivalent to assessing the null hypothesis $H_0 : \sigma^2 = 0$ for each SNP in the data. We derive a test statistic with the estimate $\widehat{\sigma}_j^2$ using Equation 5 and compute an approximate standard error

$$\mathbb{V}\left[\widehat{\sigma}_j^2\right] \approx 2\mathbf{y}^\intercal\mathbf{H}_j^\intercal\mathbf{V}_j\mathbf{H}_j\mathbf{y}, \qquad \text{(Equation 6)}$$

where $\mathbf{V}_j = \widehat{\omega}_j^2\mathbf{K} + \widehat{\sigma}_j^2\mathbf{G}_j + \widehat{\tau}_j^2\mathbf{I}$. Note that the point estimates from Equations 4 and 5 are unbiased and can lead to negative values when the true variance component is zero.[22] The one-sided hypothesis test formalizes the constraint that only positive estimates of $\widehat{\sigma}_j^2$ can be indicative of marginal epistasis.

Scalable computation via stochastic MoM

The right-hand side of Equation 5 involves computing traces of matrix products. If each covariance matrix is held in memory, then this can be done efficiently (without matrix multiplication)

using the Frobenius inner product. However, holding large covariance matrices in memory itself prevents scalability of the method. Naively multiplying two $N \times N$ matrices requires N^3 field operations. This too can be impractical for biobank-scale data with large sample sizes. To enable genome-wide testing, SME makes use of a stochastic MoM approach through the implementation of Hutchinson's stochastic trace estimator and the Mailman algorithm. The stochastic trace implementation enables block-wise processing of genotype data. With this approach, SME computes the traces of all covariance matrix products without ever needing to explicitly estimate the covariance matrices themselves. By making the block size (i.e., the number of SNPs processed simultaneously) configurable, the computation can be performed using as little as 1 gigabyte (GB) of read access memory (RAM).

Hutchinson's stochastic trace estimator

Hutchinson's stochastic trace estimator approximates the trace of a matrix product via the following[4,23,25,34]:

$$\mathrm{tr}\left(\boldsymbol{\Sigma}_{jr}\boldsymbol{\Sigma}_{js}\right) \approx \frac{1}{B}\sum_{b=1}^{B}\mathbf{z}_b^{\intercal}\boldsymbol{\Sigma}_{jr}\boldsymbol{\Sigma}_{js}\mathbf{z}_b, \qquad \text{(Equation 7)}$$

where $\mathbf{z}_b \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is a normally distributed vector and B is the number of random draws used to approximate the trace. This operation only depends on a series of matrix-by-vector products and has time complexity $\mathcal{O}(BNJ)$. Essentially, we choose an order of operations such that computing the quadratic forms $\mathbf{z}_b^T\mathbf{X}\mathbf{X}^{\intercal}\mathbf{X}\mathbf{X}^{\intercal}\mathbf{z}_b = \|\mathbf{X}\mathbf{X}^{\intercal}\mathbf{z}_b\|^2$ is reduced by (1) applying $\mathbf{X}^{\intercal}$ to the vector $\mathbf{z}_b$ and then (2) applying $\mathbf{X}$ to the resulting vector $\mathbf{X}^{\intercal}\mathbf{z}_b$ for all B random vectors. Importantly, Equation 7 can be set up algorithmically such that the approximation is done block-wise over the individual-level genotype data. Using this approach, SME avoids having to compute any of the covariance matrices directly and alleviates the need to load the entire genotype matrix into memory all at once.

The Mailman algorithm

An additional computational speedup can be achieved by making use of the discrete encoding for each SNP. The Mailman algorithm allows for an $N \times J$ matrix to be multiplied by any real vector in $\mathcal{O}(NJ\,/\log_{\Omega}(\max\{N,J\}))$ time if it has elements defined over a finite alphabet size Ω.[4,23,26,34] A standardized genotype matrix can be written as $\mathbf{X} = (\mathbf{A} - \mathbf{U})\mathbf{Q}^{-1}$, where $\mathbf{A}$ is an $N\times J$ allele count matrix with elements $a_{ij} \in \{0,1,2\}$ over finite size $\Omega = 3$, $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_J]$ is a matrix where the j-th column contains the sample mean for the j-th SNP and the variance of each SNP as the diagonal entries. With this specification, we can write $\mathbf{X}^{\intercal}\mathbf{z}_b = \mathbf{Q}^{-1}(\mathbf{A}^{\intercal} - \mathbf{U}^{\intercal})\mathbf{z}_b$. The first term, $\mathbf{Q}^{-1}\mathbf{A}^{\intercal}\mathbf{z}_b$, can be solved in $\mathcal{O}(NJ\,/\log_3(\max\{N,J\}))$ time. The second term, $\mathbf{Q}^{-1}\mathbf{U}^{\intercal}\mathbf{z}_b$, corresponds to scaling the random N-dimensional vector $\mathbf{z}_b$, which can be computed in time $\mathcal{O}(N+J)$.[4,23,34]

Shared random vectors and parallelization

With the stochastic trace estimator and the Mailman algorithm, it is feasible to estimate the variance components for each focal SNP even when a study has a large number of individuals. Still, testing every focal SNP against all variants genome wide remains a challenge. In SME, we propose randomly selecting subsets of focal SNPs and having them share the same random vectors $\mathbf{z}_b$ when performing the stochastic trace estimation. This limits the number of computations that need to be performed while maintaining unbiasedness in the point estimates (Figure S4).

Since the error terms in Equation 3 are assumed to be independent, the only two intermediate products that need to be computed in Equation 7 are $\mathbf{K}\mathbf{z}_b$ and $\mathbf{G}_j\mathbf{z}_b$. With these terms, we can compute all combinations of traces of matrix products that are required to fit SME (e.g., $\mathrm{tr}(\mathbf{K}^2) \approx \|\mathbf{K}\mathbf{z}_b\|^2$ and $\mathrm{tr}(\mathbf{K}\mathbf{G}_j) \approx \mathbf{z}_b^{\intercal}\mathbf{K}\mathbf{G}_j\mathbf{z}_b$). For a subset of L focal SNPs and fixed $\mathbf{z}_b$, the term $\mathbf{K}\mathbf{z}_b$ is constant, and only $\mathbf{G}_j\mathbf{z}_b$ changes. This reduces the effective time needed to compute $\mathbf{K}\mathbf{z}_b$ per test by a factor of $1/L$. Figure S1 illustrates the idea of sharing random vectors.

As previously mentioned, reading biobank-scale genotype data into memory requires non-negligible overhead. The `R` implementation of SME reads in genotypes once for each subset of focal SNPs that share the same random vectors (web resources). The computation of $\mathbf{K}\mathbf{z}_b$ and each $\mathbf{G}_j\mathbf{z}_b$ can then be done in parallel using multithreading. While sharing random vectors helps accelerate this task, it also requires storing larger quantities of intermediate results in memory.

Masking further reduces the effective size of data

A direct benefit of masking is that it is equivalent to removing entire columns from the genotype matrix. This reduction contributes to a significantly faster runtime when computing $\mathbf{G}_j\mathbf{z}_b$ (e.g., Figures S2 and S3). Concretely, applying a mask to an $N \times J$ genotype matrix reduces the number of columns to $J^* \leq J$. If the mask is sparse enough such that $J^* \leq N$, the time complexity of the Mailman algorithm is also reduced to $\mathcal{O}(BNJ^*\,/\log_3(N))$.

Preprocessing the UK Biobank

Genotype data for 488,377 individuals in the UK Biobank were downloaded and converted using the `ukbgene` and `ukbconv` tools, respectively. Continuous traits were also downloaded using the `ukbgene` tool and were adjusted for age and sex. Individuals identified as having high heterozygosity, excessive relatedness, or aneuploidy were removed (1,550 individuals). After separating individuals into self-identified ancestral cohorts using data field `21000`, unrelated individuals were selected by randomly choosing one person from each related pair. This resulted in $N = 349{,}411$ White British individuals to be included in our analysis. We downloaded imputed SNP data from the UK Biobank for all remaining individuals and removed SNPs with an information score below 0.8. Information scores for each SNP are provided by the UK Biobank (web resources).

Quality control for the remaining genotyped and imputed 1,933,118 variants was then performed on each cohort separately using the following steps. All structural variants were first removed, leaving only SNPs in the genotype data. Next, all AT/CG SNPs were removed to avoid possible confounding due to sequencing errors. Then, SNPs with a minor-allele frequency less than 1% were removed using the `PLINK 2.0`[38] command `–maf 0.01`. We then removed all SNPs found to be out of Hardy-Weinberg equilibrium, using the `PLINK –hwe 0.000001` flag to remove all SNPs with a Fisher's exact test $p < 10^{-6}$. Finally, SNPs with any missingness were removed using the `PLINK 2.0 –geno 0.00` flag. This left a total of $J = 543{,}813$ SNPs for our study.

Assessing replicability and robustness of the study

To confirm that the marginal epistatic signal identified by SME in the UK Biobank is robust to sample composition, we randomly split the data into two distinct subsets of equal size. This resulted in two cohorts with $N = 174{,}705$ individuals and $J = 543{,}813$ SNPs. Additionally, we assessed the scale invariance of the signal identified by SME. Here, we reran the analysis for all significant marginal epistatic associations using quantile-normalized

versions of each trait. The quantile normalization was performed using the `qqnorm` function in `R`.

GWAS summary statistics

The summary statistics used to compare against marginal epistatic results for each trait in the UK Biobank were downloaded (web resources). These summary statistics were first filtered to match the same set of SNPs that passed our quality control. SNPs that were reported as being associated with a trait at genome-wide significance ($p < 5 \times 10^{-8}$) in the UK Biobank European cohort are highlighted in our analyses.

Generating masks from external data sources

Below is a description of the datasets that we used to generate the masks for SME when analyzing individuals and quantitative traits from the UK Biobank.

Masks using DNase I-hypersensitive sites

The chromatin accessibility-based masks were derived from DNase I-hypersensitive sites (DHSs) measured over 12 days of *ex vivo* erythroid differentiation.[33] The DHS intervals were reported using the hg38 human reference genome. To map correspondence to the UK Biobank data (which use the hg19 as reference), we performed a lift over using `CrossMap`.[39] We mapped each SNP in the UK Biobank to the genomic intervals in the DHS data using the `R` software package `GenomicRanges`.[40] The resulting mask comprised $J^* = 4{,}952$ SNPs.

Masks using GWAS summary statistics

Many complex traits have genetic signatures that are not tissue specific, and in many cases, biologically informative annotations from external functional studies may not be available. As a more general strategy, in the absence of trait-specific biological information, we induce sparsity within the SME framework using GWAS summary statistics. The motivation behind this approach is 2-fold. First, variant-level associations from genome-wide studies can serve as proxies for more targeted, biologically informed priors. For example, genomic regions identified by DNase-seq have been shown to be enriched for non-coding variants associated with common diseases and complex traits.[27] Second, GWAS summary statistics have been suggested to tag non-additive genetic effects contributing to trait architecture.[5] In practice, this results in masks of varying degrees of sparsity and mask sizes (using a genome-wide significance threshold $p < 5 \times 10^{-8}$). In our quality-controlled data from the UK Biobank, $J^* = 16{,}142$ SNPs are associated with body height, $J^* = 7{,}778$ SNPs are associated with mean corpuscular hemoglobin (MCH), $J^* = 3{,}536$ SNPs are associated with uric acid (or urate), and $J^* = 547$ SNPs are associated with vitamin D levels (VITDs). To match the sparsity observed in DNase I-hypersensitivity data, we select the top 5,000 SNPs ranked by strength of association (lowest p values) for height and MCH and include all significant SNPs for urate and VITDs.

Small note on linkage disequilibrium blocks

To control the type I error rate, variants in the same linkage disequilibrium (LD) block as the j-th focal SNP are also masked. The LD blocks used for this study were approximately independent and derived using European individuals.[41]

Simulation studies

To characterize the behavior of the SME test, we generate quantitative traits using chromosome 1 of the White British cohort from the UK Biobank.[2,3,5] These data consisted of $N = 349{,}411$ individuals and $J = 43{,}332$ SNPs. Here, we sample 10% of the SNPs in the data to be causal and simulate traits using the following linear model:

$$\mathbf{y} = \sum_{a \in \mathcal{A}} \mathbf{x}_a \beta_a + \sum_{g_1 \in \mathcal{G}_1} \sum_{g_2 \in \mathcal{G}_2} (\mathbf{x}_{g_1} \circ \mathbf{x}_{g_2}) \alpha_{g_1 g_2} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \lambda^2 \mathbf{I}), \quad \text{(Equation 8)}$$

where $\mathbf{y}$ is an N-dimensional phenotype vector; $\mathcal{A}$ represents the set of causal SNPs with additive effects; $\mathbf{x}_a$ is the genotype for the a-th causal SNP, which has been standardized to have mean zero and variance one across individuals; β_a is the additive effect sizes for the a-th SNP; both $\mathcal{G}_1$ and $\mathcal{G}_2$ are sets of epistatic SNPs that are non-overlapping subsets of $\mathcal{A}$; $\alpha_{g_1 g_2}$ is the interaction effect sizes between $\mathbf{x}_{g_1}$ and $\mathbf{x}_{g_2}$; and $\boldsymbol{\varepsilon}$ is a vector of normally distributed environmental noise. We sample the effect sizes from standard normal distributions and rescale them so that the additive and epistatic effects explain a desired proportion of the trait variance. Specifically, the additive and epistatic variance components make up the broad-sense heritability of the trait $H^2 = h^2_{\mathcal{A}} + h^2_{\mathcal{G}}$. Similarly, the environmental noise matrix is also rescaled, such that it explains the remaining $1 - H^2$ proportion of the trait variance.

In this simulation design, the epistatic causal SNPs interact between sets. All SNPs in $\mathcal{G}_1$ interact with all SNPs in $\mathcal{G}_2$ but do not interact with variants in their own group (and vice versa). Note that we use this setup because the ability to detect interacting variants in the marginal epistasis framework depends on the proportion of phenotypic variance that they marginally explain. The parameters that determine the PVE by a single SNP are the epistatic heritability $h^2_{\mathcal{G}}$ and the cardinality of the set to which the SNP belongs. For example, an SNP in $\mathcal{G}_1$ will explain, on average, $h^2_{\mathcal{G}}/|\mathcal{G}_1|$ of the total phenotypic variance.

To simulate masks, we select some proportion of the non-epistatic SNPs to zero out of the interaction covariance matrix. For example, when analyzing the j-th SNP, 95% masking corresponds to excluding 41,165 out of the 43,332 SNPs when computing $\mathbf{G}_j$. Similarly, only 433 SNPs are used when using a 99% masking strategy. To create masks that induce "uniform sparsity," we randomly sample from all SNPs in the dataset with uniform probability. In real data, masks are likely not sampled uniformly. An obvious potential source of complication for non-uniform masking can come from LD between SNPs. Therefore, to simulate masks that induce "localized sparsity," we randomly sample a seed SNP and define a genomic window around it. SNPs outside that window are then masked.

Unless otherwise specified, we used the following hyperparameters when applying SME and FAME to the simulated data: $B = 100$ random vectors for the stochastic trace estimator, and we block-wise processed 100 SNPs at a time and shared random vectors across $L = 25$ SNPs for the SME implementation. To ensure fair comparisons between methods, all were applied to the same SNPs, and synthetic traits within each simulation replicate. Note that causal SNP sets varied across replicates due to random sampling.

Software tools and data sources

Software for running the FAME is freely available at https://github.com/sriramlab/FAME. The version of FAME that we implemented for this work (GitHub commit hash prefix cfdd03f; date May 7, 2024) has been archived at https://doi.org/10.5281/zenodo.14607997. The original MAPIT was implemented using the `mvmapit` software package in `R` and is available both on CRAN

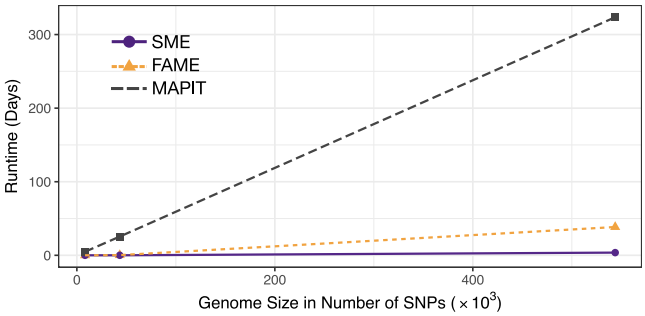


Figure 2. Computational time for running SME and other marginal epistatic approaches on biobank-scale data as a function of the genome size

The other methods compared include FAME[4] and MAPIT.[2] Here, we analyze genotype data from a fixed set of 349,411 individuals from the UK Biobank and vary the genome size. All results were computed on a single core of an Intel Xeon Platinum 8268 central processing unit (CPU). Total runtime was calculated based on the average runtime per SNP and parallel processing on a cluster with 960 CPUs available. Both SME and FAME were set to have the same hyperparameter configurations (e.g., the number of random vectors). SME also used a binary mask that contained 5,000 unmasked SNPs, and its stochastic trace approximation was applied such that sets of 250 focal SNPs shared the same random vectors. MAPIT could not be directly compared due to its excessive memory requirements for datasets of this size. Instead, the runtime for MAPIT was measured on smaller sample sizes (up to 20,000 individuals) and extrapolated to a sample size of 349,411 individuals. This extrapolation assumed quadratic scaling with the number of individuals and linear scaling with the number of SNPs.

(https://cran.r-project.org/package=mvMAPIT) and GitHub (https://github.com/lcrawlab/mvMAPIT). All software for SME, FAME, and MAPIT were fit using their default settings unless otherwise stated in the main text. Data from the UK Biobank Resource was made available under application number 14649 and can be accessed by direct application to the UK Biobank. GWAS summary statistics were downloaded from https://www.nealelab.is/uk-biobank, and corresponding LD maps were taken from https://bitbucket.org/nygcresearch/ldetect-data/. The DHS data are available at https://doi.org/10.5281/zenodo.5291736. We used `CrossMap` (https://crossmap.readthedocs.io/) and `GenomicRanges` (https://doi.org/10.18129/B9.bioc.GenomicRanges) to map SNPs from the UK Biobank to genomic intervals in the DHS data. The chain file for the `CrossMap` liftover tool can be found at http://hgdownload.soe.ucsc.edu/goldenPath/hg38/liftOver/.

Results

SME scales to biobank GWASs

To compare the expected central processing unit (CPU) computation for conducting a biobank-scale genome-wide analysis using SME, FAME,[4] and MAPIT,[2] we measured the average runtime per SNP for each method on an Intel Xeon Platinum 8268 CPU using a single core (i.e., no parallel processing). Here, we used genotype data from 349,411 individuals of self-identified European ancestry in the UK Biobank with 543,813 SNPs after quality control (material and methods). The memory requirements for MAPIT are prohibitively high, requiring resources on the order of terabytes for biobank-scale datasets with hundreds of thousands of observations. By using Hutchinson's stochastic trace estimator, both SME and FAME achieve configurable resource requirements and can effectively operate with only a few GB of memory at biobank-scale data.

We find that SME performs genome-wide testing 10× faster than FAME and 90× faster than MAPIT (Figure 2). While analyzing the complete dataset with a single core, SME requires only 3.7 days of runtime compared to FAME and MAPIT, which require 38.4 and 324 days, respectively. The greatest speedup in SME is achieved by approximating the stochastic trace when estimating model parameters (Figure S1). This allows for computations involving large genetic relatedness matrices to be reused across multiple tests for different focal SNPs. Even with the stochastic trace approximations, performing matrix calculations at biobank scale still takes minutes. However, the ability to share these computations across multiple variants significantly reduces the overall computational burden (Figures S2 and S3). The proposed approach enables SME to be effectively applied to the UK Biobank, facilitating GWASs of epistasis. For the real data application in this study, SME effectively computed hundreds of tests simultaneously with less than 85 GB of RAM for a dataset consisting of $N = 349{,}411$ individuals.

SME is a well-calibrated test and conserves type I error rates

We generate synthetic phenotypes using a linear model with real genotypes from chromosome 1 of White British individuals in the UK Biobank.[2,3,5] After quality control, we had a dataset of 349,411 individuals and 43,332 SNPs (material and methods). Under the null model, we simulate traits consisting of only additive effects. Here, we randomly sample 10% of the SNPs and scale their effect sizes such that they explain 40% of the total phenotypic variance.

We simulate external data sources (S) to be used when generating a mask for the marginal epistatic covariance matrix $\mathbf{G}_j$ in SME. Recall that these external data sources are intended to give alternative insight into the importance of SNPs and are used to induce sparsity in the modeled gene interactions by dropping interactions with "unimportant" variants. We consider two scenarios in our simulations (Figure S5). In the first, SNPs deemed important in the external data source are sparsely sampled with uniform probability from all variants. As a result, the modeled gene interactions are evenly distributed along the chromosome. We will refer to this scenario as inducing uniform sparsity in the SME model. In the second scenario, we randomly sample one central seed SNP and define variants in a block around it as important. We will refer to this scenario as one that induces localized sparsity. In these simulation experiments, we assess the calibration of SME using both types of external data sources as a function of the percentage of total variants that are masked (varying between 0%, 95%, and 99%) and the

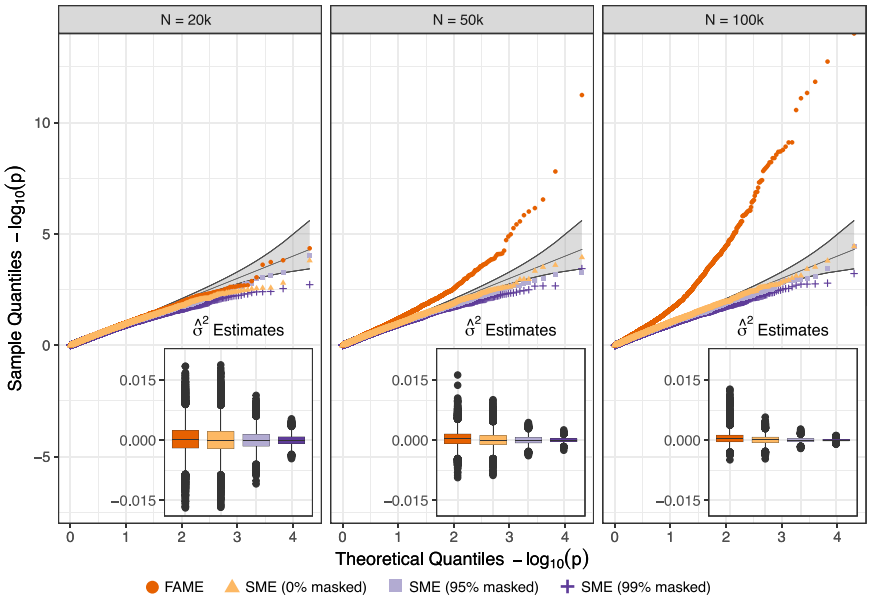


Figure 3. While using a mask that induces uniform sparsity, SME is well calibrated under the null hypothesis and does not identify epistasis when traits are generated by only additive effects

Synthetic traits were simulated with only additive effects using chromosome 1 from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled using sample sizes of 20,000, 50,000, and 100,000 individuals. We randomly selected 10% of all variants to be causal with additive effects, and we assume that they explain 40% of the phenotypic variance for each trait. Data were analyzed using both FAME (as a baseline) and SME under varying percentages of SNPs that are masked (0%, 95%, and 99%, respectively). The small insets in each plot show the distribution of the estimated marginal epistatic variance components across all experiments. For reference, under the null hypothesis $H_0 : \sigma^2 = 0$. Results are based on 100 simulated traits per scenario.

number of individuals being analyzed (varying between 20,000, 50,000, 100,000, and 300,000 randomly subsampled individuals).

Under the null model, we find that SME produces well-calibrated p values and unbiased variance component estimates (Figure 3; Table S1) using a uniformly sparse mask. Specifically, higher levels of sparsity lead to more accurate estimates of the marginal epistatic variance components. We also see the precision in its estimates improve as the sample size increases, which is expected since SME uses a normal test to compute p values for each SNP. Overall, this translates to SME preserving empirical type I error rates estimated at significance levels $\alpha = 0.05$, 0.01, and 0.001, respectively (Table 1). In contrast, FAME produces inflated test statistics as the number of samples in a dataset grows. Note that we do not include a comparison with MAPIT here due to its inability to scale to biobank settings (see Crawford et al.[2] for an assessment of its calibration on small-to-moderately sized data).

Notably, using an external data source with a localized sparse masking scheme introduces a slight negative bias in variance component estimates produced by SME, leading to fewer significant p values and more conservative inference (Figure S6). While type I error control remains conservative (Table S2), this also means that the test may have reduced power when traits are indeed simulated under the alternative with non-zero epistatic effects. We will explore this behavior further in the next section.

The masking strategy in SME leads to improved power in simulations

To assess the power of SME, we again generate synthetic continuous traits using real genotypes from chromosome 1 of White British individuals in the UK Biobank.[2,3,5] These data were subsampled using sample sizes of 50,000, 100,000, and 300,000 individuals. Here, we assume that 10% of all SNPs are causal and have additive effects that collectively explain 30% of the trait variance. Next, we fix the epistatic contribution to the trait variance to be 5%, making the total broad-sense heritability 35%. We select a set of epistatic variants from the causal SNPs and divide them into two equally sized groups. Each SNP in one group is simulated such that they only interact with SNPs in the other group. This simulation design gives control over the epistatic PVE by the individual variants. In this analysis, we select 10, 20, 50, and 100 of the causal SNPs to be epistatic, which corresponds to per-SNP epistatic PVE values equal to 1%, 0.5%, 0.2%, and 0.1% of the trait variance.

Once again, we analyze SME using two different external data source types that induce uniform and localized sparsity—masking out 0%, 95%, and 99% of the possible interactive partners when constructing the marginal epistatic covariance matrix $\mathbf{G}_j$ for each j-th focal SNP being tested (Figure S5A). We compare the empirical power of SME to FAME as a baseline by assessing the respective abilities of both models to identify causal epistatic SNPs at a genome-wide significance threshold $p < 5 \times 10^{-8}$.[42]

We find that using the uniformly sparse masking scheme significantly enhances the power of SME, with greater levels of sparsity leading to better method performance (Figure 4). When analyzing 300,000 individuals, the 99%-masked SME identifies at least 85.1% of the causal epistatic SNPs even when they contribute as little as 0.1% to the trait variance. This is compared to FAME and a non-masked SME, which only detect at most 1% of causal SNPs with very small PVE. When epistatic variants have larger effect sizes and individually account for 1% of the trait variance, the 99%-masked SME shows

Table 1. While using a mask that induces uniform sparsity, SME controls type I error rates when synthetic traits are generated under the null model

Method	Sample size	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
FAME	20,000	0.0531 (0.0214)	0.0128 (0.0124)	0.0022 (0.0044)
FAME	50,000	0.0720 (0.0237)	0.0200 (0.0127)	0.0053 (0.0078)
FAME	100,000	0.1208 (0.0318)	0.0526 (0.0223)	0.0241 (0.0133)
SME (0% masked)	20,000	0.0379 (0.0185)	0.0056 (0.0074)	0.0001 (0.0010)
SME (0% masked)	50,000	0.0459 (0.0207)	0.0084 (0.0090)	0.0004 (0.0020)
SME (0% masked)	100,000	0.0492 (0.0217)	0.0085 (0.0090)	0.0007 (0.0029)
SME (0% masked)	300,000	0.0537 (0.0209)	0.0090 (0.0090)	0.0012 (0.0036)
SME (95% masked)	20,000	0.0359 (0.0163)	0.0046 (0.0061)	0.0005 (0.0022)
SME (95% masked)	50,000	0.0416 (0.0168)	0.0060 (0.0075)	0.0002 (0.0014)
SME (95% masked)	100,000	0.0445 (0.0197)	0.0078 (0.0080)	0.0004 (0.0020)
SME (95% masked)	300,000	0.0474 (0.0213)	0.0073 (0.0081)	0.0003 (0.0017)
SME (99% masked)	20,000	0.0310 (0.0147)	0.0029 (0.0052)	0.0000 (0.0000)
SME (99% masked)	50,000	0.0355 (0.0186)	0.0037 (0.0065)	0.0001 (0.0010)
SME (99% masked)	100,000	0.0387 (0.0178)	0.0042 (0.0062)	0.0001 (0.0010)
SME (99% masked)	300,000	0.0404 (0.0189)	0.0059 (0.0070)	0.0002 (0.0014)

Synthetic traits were simulated with only additive effects using chromosome 1 from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled using sample sizes of 20,000, 50,000, 100,000, and 300,000 individuals. A total of 100 causal additive variants were randomly selected for each trait, and their effects were assumed to explain 40% of the phenotypic variance. Data were analyzed using both FAME (as a baseline) and SME under varying percentages of SNPs that are masked (0%, 95%, and 99%, respectively). Empirical size for the analyses used significance thresholds of $\alpha = 0.05$, 0.01, and 0.001. Values in the parentheses are the standard deviations of the estimates. Results are based on 100 simulations per scenario. Due to computational constraints, the data with 300,000 individuals were only analyzed with SME.

99.8% power even with relatively small sample sizes (e.g., 50,000 individuals). Again, this is compared to FAME and a non-masked SME, which each only have approximately 35% power in this scenario. To examine the sensitivity of the SME to the specification of an external data source, we conducted a simulation in which the model was provided with a weight matrix that incorrectly masked true interacting partners.[5] Here, we observed that the SME framework protects against the false discovery of non-additive genetic effects and underestimates the marginal epistatic variance component (σ^2) when causal SNPs involved in pairwise interactions were unobserved (Figure S7).

Similar to the null simulation study, we see that SME produces negatively biased variance component estimates when using an external data source that induces localized sparsity in the model. Indeed, overconcentrating the search for potential interacting pairs to a select number of correlated variants leads to reduced empirical power compared to the masking that results in uniform sparsity (Figure S8). For example, when analyzing 300,000 individuals, the localized 99%-masked SME has just 6.5% power to identify epistatic variants that explain 0.1% of the trait variance. To overcome this issue in practice, we propose a strategy in which we take an external data source with localized genomic information and randomly unmask "unimportant" variants with uniform probability along the genome (essentially making the localized sparsity look more uniform, as shown in Figure S5B). As a demonstration of this idea, we implement a version of SME where we include an additional 1% and 5% of initially disregarded interactions back into the construction of $\mathbf{G}_j$ for each $j = 1, \ldots, J$ tested focal SNPs in the dataset. We find that adding this "noise" back into the mask reduces the negative bias of the variance component estimates and recovers as much as 28% of the power that was lost with respect to the uniformly sparse models (Figure S9). For future users of the SME software, we want to note that there is likely an application-specific trade-off between adding SNPs to a mask to reduce potential bias and finding the degree of sparsity needed for an optimally powered test.

SME uses chromatin information to identify epistasis in hematology traits

We apply SME to four hematology traits—MCH, mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV), and hematocrit (HCT)—assayed in 349,411 White British individuals in the UK Biobank[19] and genotyped at 543,813 SNPs genome wide. As an external data source, we leverage DHS data measured over 12 days of *ex vivo* erythroid differentiation.[33] Of the

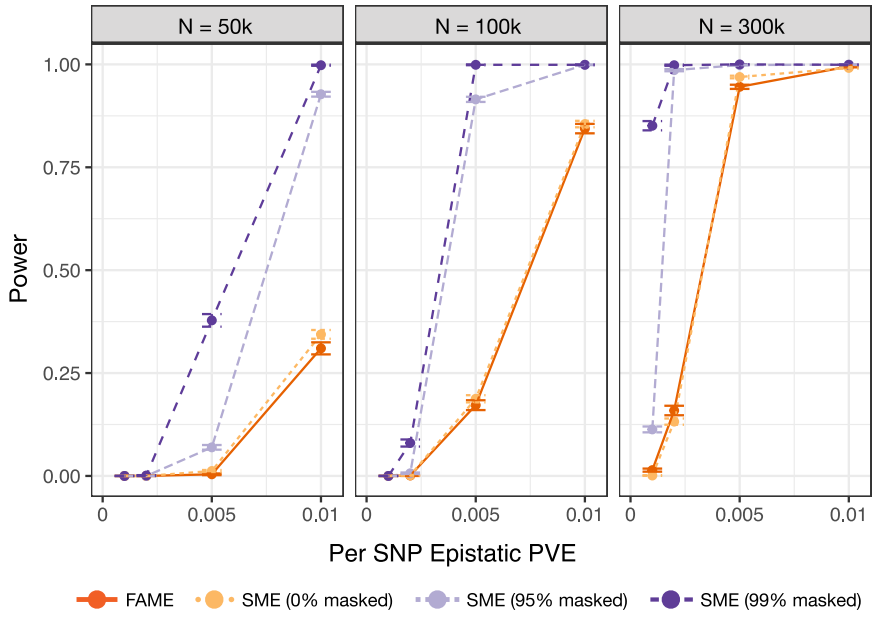


Figure 4. Uniform sparse modeling of interactions enhances the empirical power of SME

Synthetic traits were simulated with both additive and pairwise epistatic effects using chromosome 1 from individuals of self-identified European ancestry in the UK Biobank. Data were subsampled using sample sizes of 50,000, 100,000, and 300,000 individuals. We randomly selected 10% of all variants to have additive effects that collectively explained 30% of the trait variance. We then fixed the total epistatic variance to 5%. The per-SNP epistatic phenotypic variance explained (PVE) was adjusted by varying the number of interacting SNPs (chosen to be 10, 20, 50, or 100 SNPs). Data were analyzed using both FAME (as a baseline) and SME under varying percentages of variants that are excluded from consideration as potential interaction partners for each focal SNP (0%, 95%, and 99% masking, respectively). Empirical power was determined using the significance threshold $p < 5 \times 10^{-8}$. Results are based on 100 simulations per scenario, with error bars representing the standard deviation across replicates.

quality-controlled SNPs in our data, 4,932 of them are located in DHS regions enriched for transcriptional activity.[27] Since previous GWAS results have found genes associated with MCH, MCHC, and MCV to also be implicated in erythroid differentiation,[43] we expect that conditioning SME to test over regulatory mechanisms gathered during erythropoiesis will be helpful in identifying epistatic variants for these traits. On the other hand, HCT is a phenotype that measures the percentage of red blood cells in an individual. Since the regulation for this trait has little to do with DHS sites and more to do with oxygen available in the blood,[44] we would expect a mask derived from functional data on erythropoiesis to not be helpful in enabling SME to detect epistasis.

For each trait, we use Manhattan plots to visually display the variant-level mapping results across each of the four traits, where chromosomes are shown in alternating colors for clarity (Figures 5 and S10–S12). Corresponding genes that have SNPs with p values below the genome-wide significance threshold to correct for multiple testing ($p < 5 \times 10^{-8}$) are also highlighted. Importantly, many of the marginal epistatic variants identified by SME are supported by multiple published studies that have investigated non-additive gene action related to erythropoiesis and red blood cell traits (Table 2).

For example, when analyzing MCH, the strongest association identified by SME is the SNP rs4711092 ($p = 1.41 \times 10^{-11}$), which maps to the gene secretagogin (*SCGN*). SCGN regulates exocytosis by interacting with two soluble N-ethylmaleimide sensitive fustion attachment proteins (SNAP-25 and SNAP-23) and is critical for cell growth in some tissues.[45] For MCH, SME also identified five significantly associated SNPs (e.g., rs9366624 with $p = 1.8 \times 10^{-9}$) in the gene capping protein regulator and myosin 1 linker 1 (*CARMIL1*). CARMIL1 is known to interact with and regulate the capping protein (CP), which plays a role via protein-protein interactions in regulating erythropoiesis.[48] Specifically, CARMIL proteins regulate actin dynamics by regulating the activity of the CP.[55,56] Erythropoiesis leads to modifications in the expression of membrane and cytoskeletal proteins, whose interactions impact cell structure and function.[57,58] Both *SCGN* and *CARMIL1* have previously been associated with hemoglobin concentration.[43,46] A complete list of the results for all traits is listed in Tables S4–S7. As a baseline for comparison, we also applied SME without an external data source (i.e., 0% masked) and MAPIT to all significant SNPs. The point of this analysis was to explore whether these traditional methods would have also identified the same sets of epistatic variants. Due to computational constraints, MAPIT was only implemented on a random subset of 10,000 individuals. Importantly, neither baseline identified any genome-wide significant associations (Table 2).

Conditioning SME on GWAS variants reveals epistasis in other complex traits

Next, we apply SME using a different external data source to four traits assayed in the same 349,411 White British individuals in the UK Biobank[19] and genotyped at the 543,813 SNPs genome wide. These traits include body height, MCH, uric acid (which we refer to as urate), and VITDs. As an external data source, we leverage significant trait associations from GWAS summary statistics (material and methods). Here, we select the top 5,000 SNPs ranked by strength of association (i.e., lowest p values) for height and MCH and include all significant SNPs for urate (3,536 SNPs) and VITDs (547 SNPs). For each trait, we again use Manhattan plots to visually display the variant-level

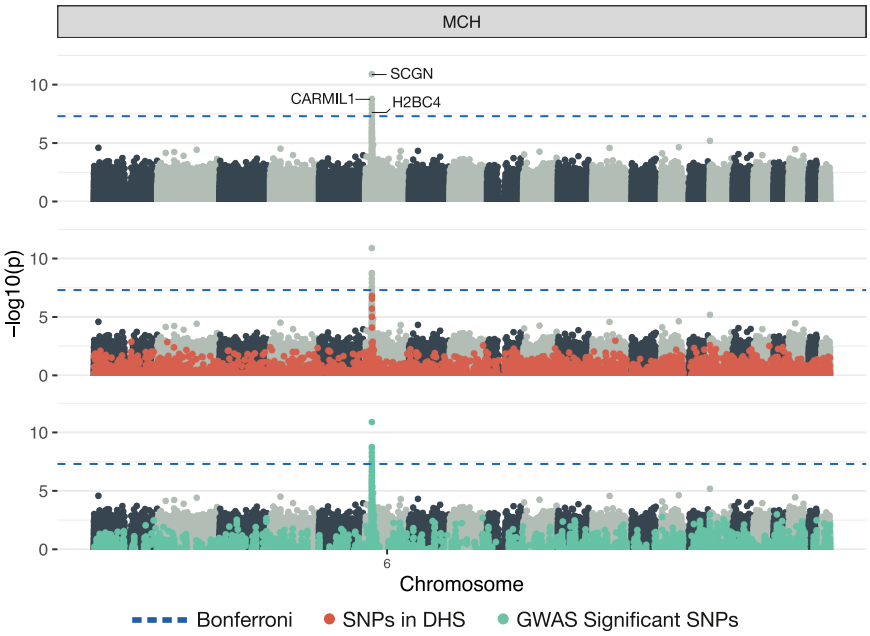


Figure 5. Manhattan plots of a genome-wide interaction analysis using SME to study mean corpuscular hemoglobin assayed in individuals in the UK Biobank

As a mask in this study, we leveraged DNase I-hypersensitive site (DHS) data measured over 12 days of *ex vivo* erythroid differentiation.[27,33] This means that while all SNPs are tested for marginal epistasis, only their interactions with SNPs in DHS regions are considered. Here, $-\log_{10}$-transformed p values from SME are plotted for each SNP against their genomic positions. Chromosomes are shown in alternating colors for clarity. The dashed blue line represents the genome-wide significance threshold ($p < 5 \times 10^{-8}$). Each image shows the same plot with different aspects of the result highlighted. The first simply shows the names of the closest neighboring genes to significant epistatic SNPs. The second image highlights the SNPs that fall in DHS regions, and the third image highlights SNPs that are also found to have a significant (additive) association with the trait according to a GWAS (material and methods).

mapping results, with chromosomes shown in alternating colors for clarity (Figures S13–S16). The nearest genes mapped to the lead SNPs of peaks are highlighted.

Importantly, even when conditioning on GWAS summary statistics, SME identifies significant marginal epistatic variants for both height and urate (see Table S3). For example, when analyzing height, SME finds significant epistasis for the SNP rs9467442 ($p = 5.49 \times 10^{-9}$), which maps to the gene cytidine monophospho-N-acetylneuraminic acid hydroxylase, pseudogene (*CMAHP*). Recently, *CMAHP* has been shown to be associated with body height for populations of European ancestry.[59] A complete list of the results for all traits is listed in Tables S8–S11. Again, as a baseline for comparison, we apply SME without an external data source (i.e., 0% masked) and MAPIT to all significant SNPs identified by SME with masking. Due to computational constraints, MAPIT was only applied to a random subset of 10,000 individuals. Neither baseline had enough power to identify any significant associations at the genome-wide threshold (Table S3).

Findings with SME are robust to sample composition and phenotypic scaling

As a final analysis, we assess the replicability and robustness of the marginal epistatic variants identified by SME. First, for all traits with significant marginal epistasis (MCH, MCV, height, and urate), we replicate the application of SME using the respective external data sources (DHS and GWAS) in two independent subsamples of 174,705 White British individuals from the UK Biobank[19] genotyped at 543,813 SNPs genome-wide. Across the split-half analyses, the overall results remained consistent, highlighting that the genomic loci selected by SME had stable marginal epistasis associations (Figures S17–S20).

Next, for all significant marginal epistasis associations, we apply two additional variations of SME and FAME (Table S12). First, we assess the scale invariance of the signal by quantile normalizing the traits. Second, we increase the stringency for which variants are excluded in the mask—here, in addition to removing variants within the LD block surrounding a focal SNP, we also exclude all variants on the same chromosome. After quantile normalization, most of the marginal epistasis signal remained significant; however, SME lost all power after excluding the entire chromosome where the focal SNP is located. As part of future work, it will be important to further distinguish statistically whether the observed marginal epistatic effects estimated by SME in these traits arise from *cis*-chromosome interactions or same-locus additive effects.[60] Lastly, despite its inflated observed type I error, FAME does not find significant marginal epistasis at any of these loci.

Discussion

The marginal epistasis framework is an alternative to detect gene interactions. It derives its power by modeling the combined effect between a focal SNP and all other variants, thus alleviating the need to test every possible interaction separately. Still, current methods seeking to identify marginal epistasis struggle to scale to biobank-scale data and can be underpowered when non-additive genetic effects only explain a small portion of the overall trait variance.[2,3] SME overcomes these limitations by inducing sparsity, essentially limiting the combined interaction for

Table 2. SME identifies marginal epistasis in hematology traits from individuals in the UK Biobank

Trait	ID	Coordinates	*p* value	PVE	*p* value (SME 0% masked)	*p* value (MAPIT)	Gene	Reference
MCH	rs4711092	chr6:25684405	1.41×10^{-11}	0.007	4.78×10^{-4}	0.47	*SCGN*	Qin et al.,[45] Timoteo et al.[46] Bauer et al.[47]
MCH	rs9366624	chr6:25439492	1.80×10^{-9}	0.011	5.75×10^{-7}	9.37×10^{-3}	*CARMIL1*	Ding et al.,[43] Ray et al.,[48] Timoteo et al.,[46] Edwards et al.,[49] Yang et al.[56]
MCH	rs9461167	chr6:25418571	2.34×10^{-9}	0.007	1.40×10^{-5}	3.05×10^{-3}	*CARMIL1*	Ding et al.,[43] Ray et al.,[48] Timoteo et al.,[46] Edwards et al.,[49] Yang et al.[50]
MCH	rs9379764	chr6:25414023	5.53×10^{-9}	0.012	4.58×10^{-6}	0.21	*CARMIL1*	Ding et al.,[43] Ray et al.,[48] Timoteo et al.,[46] Edwards et al.,[49] Yang et al.,[50] Vuckovic et al.,[51] Zhang et al.[52]
MCH	rs441460	chr6:25548288	1.20×10^{-8}	0.008	3.10×10^{-6}	0.36	*CARMIL1*	Ding et al.,[43] Ray et al.,[48] Timoteo et al.,[46] Edwards et al.,[49] Yang et al.[50]
MCH	rs198834	chr6:26114372	2.77×10^{-8}	0.008	2.17×10^{-6}	1.83×10^{-3}	*H2BC4*	Vuckovic et al.,[53] Zhang et al.[54]
MCH	rs13203202	chr6:25582771	3.17×10^{-8}	0.012	1.90×10^{-4}	0.13	*CARMIL1*	Ding et al.,[43] Ray et al.,[48] Timoteo et al.,[46] Edwards et al.,[49] Yang et al.[50]
MCV	rs9276	chr6:33053577	9.09×10^{-10}	0.002	7.27×10^{-1}	0.18	*HLA-DPB1*	–
MCV	rs9366624	chr6:25439492	1.86×10^{-8}	0.008	2.63×10^{-7}	0.16	*CARMIL1*	Ding et al.,[43] Ray et al.,[48] Timoteo et al.,[46] Edwards et al.,[49] Yang et al.[50]

Here, we analyze 349,411 White British individuals in the UK Biobank genotyped at 543,813 SNPs genome-wide. Traits in this analysis included mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV), and hematocrit (HCT). As a mask, we leveraged DNase I-hypersensitive site (DHS) data measured over 12 days of *ex vivo* erythroid differentiation.[27,33] Listed are only results corresponding to SNPs that have marginal epistatic *p* values below a genome-wide significance threshold to correct for multiple testing ($p < 5 \times 10^{-8}$). In the second and third columns, we list SNPs and their genomic location in the format `chromosome:base pair`. Next, we give the *p* value and marginal epistatic phenotypic variance explained (PVE) for each SNP as estimated by SME. The next two columns report the resulting *p* values when using SME without an external data source (i.e., 0% masked) and MAPIT. The last columns detail the closest neighboring gene as well as a reference that has previously suggested some level of association or enrichment between each gene and the traits of interest. Due to computational resource constraints, MAPIT was only applied to a random subset of 10,000 individuals.

a focal SNP to just regions of the genome that have some known functional relationship with the quantitative trait of interest. This approach not only results in more efficient estimators but also offers a mechanism that allows the method to perform genome-wide analyses on modern datasets with runtimes that are magnitudes faster than previous approaches. Through extensive simulations, we show that SME controls type I error rates and produces calibrated *p* values. We also show that SME has the power to detect SNPs involved in epistasis even when they explain very small fractions of the trait variance. By analyzing hematology traits from participants in the UK Biobank, we illustrate that SME, informed by DNase-seq data, identifies statistical epistasis in variants for which previous research has also found interaction pathways. Split-half analyses on distinct subsets of the UK Biobank also show that the non-additive signal in these hematological traits is robust. Lastly, to showcase SME in the absence of biologically informative priors, we illustrate that SME identifies significant marginal epistasis in height and urate with sparsity induced by GWAS summary statistics. We make SME available as an open-source `R` software package to enable the broader community to easily use it in their research.

The current implementation of SME offers many directions for future development and applications. For example, the key to SME is that it relies on external data sources to induce sparsity in the model. This reliance on biologically informative priors to induce sparsity could serve as a limitation, as appropriate external data may be hard to identify in practice. While simulations show that misspecified or localized sparsity does not jeopardize the ability to control false positive rates, SME currently does not provide instructions on how to best format the external data for a particular analysis. In simulations, we show that some choices can induce a structure that leads to negative bias in the model estimates. We also show that adding random "noise" to the data can reduce this bias. As part of future work, we will explore how to automatically balance this trade-off within the software.

An important consideration when mapping epistasis in real data is that statistically inferred interactions in GWASs may arise from same-locus additive effects.[60] Consequently, SME—like any other computational method for epistasis detection—may be confounded by additive effects from untyped or unmodeled variants in the same genomic region. For example, it was found in Hemani et al.[13] that an initial set of signals pointing toward evidence of genetic interactions were better explained using linear models of unobserved variants in the same haplotype.[5] The analysis of real traits from the UK Biobank presented in this work primarily serves to illustrate potential use cases and demonstrate the scalability of SME. In future work, we hope that SME will contribute to characterizing the role of epistasis in human traits. Such analyses should encompass a broader range of traits across multiple cohorts and incorporate various sparsity-inducing external data sources.

While SME uses an efficient model-fitting algorithm, its current implementation has a non-negligible input/output overhead from repeatedly needing to read in (often large) genotype data into memory. Future development that optimizes this file read bottleneck has the potential to further improve the scalability of the method. Lastly, the method currently only models quantitative traits. Future work could extend the advantages of sparse modeling of marginal epistasis to case-control traits.

Data and code availability

Source code and tutorials for implementing the SME are publicly available as an `R` package, which is available online on CRAN (https://cran.r-project.org/package=smer) and GitHub (https://github.com/lcrawlab/sme). The full list of summary statistics from the genome-wide interaction analysis using SME to study hematology traits in UK Biobank is publicly available at https://doi.org/10.5281/zenodo.14607997.

Acknowledgments

We thank members of the Weinreich and Crawford labs for insightful comments on earlier versions of this manuscript, as well as Bogdan Pasaniuc (University of Pennsylvania) and Roberta DeVito (Brown University) for helpful discussions. We also thank Ashok Ragavendran (Brown University) and the Computational Biology Core (CBC) for advice on software development. Lastly, we are incredibly grateful to Boyang Fu (UCLA) and Sriram Sankararaman (UCLA) for both technical and conceptual discussions about the implementation of FAME. This research was supported by a David & Lucile Packard Fellowship for Science and Engineering awarded to L.C. This research was conducted in part using computational resources and services at the Center for Computation and Visualization at Brown University. This research was also conducted using the UK Biobank Resource under application number 14649. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any of the funders.

Author contributions

J.S. and L.C. conceived the study. J.S. and L.C. developed the methods. S.P.S. preprocessed and provided the data. J.S. developed the software and performed the analyses. D.W. and L.C. supervised the project and provided resources. All authors wrote and revised the manuscript.

Declaration of interests

L.C. is an employee of Microsoft Research and holds equity in Microsoft. S.P.S. is an employee of and holds equity in Genomics Ltd.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2025.07.004.

Web resources

Chain file for genome liftover, http://hgdownload.soe.ucsc.edu/goldenPath/hg38/liftOver/
DHS data, https://doi.org/10.5281/zenodo.5291736
FAME, https://github.com/sriramlab/FAME
FAME (GitHub commit hash prefix cfdd03f), https://doi.org/10.5281/zenodo.14607997
LD maps, https://bitbucket.org/nygcresearch/ldetect-data/
MAPIT, https://github.com/lcrawlab/mvMAPIT
SME, https://github.com/lcrawlab/sme
UK Biobank, http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=1967
UKB GWAS summary statistics, https://www.nealelab.is/uk-biobank

Received: January 24, 2025
Accepted: July 7, 2025
Published: July 29, 2025

References

1. Abdellaoui, A., Yengo, L., Verweij, K.J.H., and Visscher, P.M. (2023). 15 years of GWAS discovery: Realizing the promise. Am. J. Hum. Genet. *110*, 179–194. https://www.sciencedirect.com/science/article/pii/S0002929722005456.
2. Crawford, L., Zeng, P., Mukherjee, S., and Zhou, X. (2017). Detecting epistasis with the marginal epistasis test in genetic mapping studies of quantitative traits. PLoS Genet. *13*, e1006869. https://dx.plos.org/10.1371/journal.pgen.1006869.
3. Stamp, J., DenAdel, A., Weinreich, D., and Crawford, L. (2023). Leveraging the Genetic Correlation between Traits Improves the Detection of Epistasis in Genome-wide Association Studies. G3 (Bethesda) *13*, jkad118. https://doi.org/10.1093/g3journal/jkad118.
4. Fu, B., Pazokitoroudi, A., Xue, A., Anand, A., Anand, P., Zaitlen, N., and Sankararaman, S. (2024). A biobank-scale test of marginal epistasis reveals genome-wide signals of polygenic epistasis. Preprint at bioRxiv. https://doi.org/10.1101/2023.09.10.557084v1.
5. Pattillo Smith, S., Darnell, G., Udwin, D., Stamp, J., Harpak, A., Ramachandran, S., and Crawford, L. (2024). Discovering non-additive heritability using additive GWAS summary statistics. eLife *13*, e90459. https://doi.org/10.7554/eLife.90459.
6. Balvert, M., Cooper-Knock, J., Stamp, J., Byrne, R.P., Mourragui, S., van Gils, J., Benonisdottir, S., Schlüter, J., Kenna, K., Abeln, S., et al. (2024). Considerations in the search for epistasis. Genome Biol. *25*, 296. https://doi.org/10.1186/s13059-024-03427-z.
7. Weinreich, D.M., Delaney, N.F., DePristo, M.A., and Hartl, D. L. (2006). Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. Science *312*, 111–114. https://www.science.org/doi/10.1126/science.1123539.
8. Fröhlich, C., Bunzel, H.A., Buda, K., Mulholland, A.J., van der Kamp, M.W., Johnsen, P.J., Leiros, H.-K.S., and Tokuriki, N. (2024). Epistasis arises from shifting the rate-limiting step during enzyme evolution of aβ-lactamase. Nat. Catal. *7*, 499–509. https://www.nature.com/articles/s41929-024-01117-4.
9. Mackay, T.F.C., and Anholt, R.R.H. (2024). Pleiotropy, epistasis and the genetic architecture of quantitative traits. Nat. Rev. Genet. *25*, 639–657. https://www.nature.com/articles/s41576-024-00711-3.

10. Weatherly, S.M., Collin, G.B., Charette, J.R., Stone, L., Damkham, N., Hyde, L.F., Peterson, J.G., Hicks, W., Carter, G.W., Naggert, J.K., et al. (2022). Identification of Arhgef12 and Prkci as genetic modifiers of retinal dysplasia in the Crb1rd8 mouse model. PLoS Genet. *18*, e1009798. https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1009798.
11. Zwarts, L., Magwire, M.M., Carbone, M.A., Versteven, M., Herteleer, L., Anholt, R.R.H., Callaerts, P., and Mackay, T.F.C. (2011). Complex genetic architecture of Drosophila aggressive behavior. Proc. Natl. Acad. Sci. USA *108*, 17070–17075. https://www.pnas.org/doi/full/10.1073/pnas.1113877108.
12. Polderman, T.J.C., Benyamin, B., de Leeuw, C.A., Sullivan, P.F., van Bochoven, A., Visscher, P.M., and Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat. Genet. *47*, 702–709. https://www.nature.com/articles/ng.3285.
13. Hemani, G., Powell, J.E., Wang, H., Shakhbazov, K., Westra, H.J., Esko, T., Henders, A.K., McRae, A.F., Martin, N.G., Metspalu, A., et al. (2021). Phantom epistasis between unlinked loci. Nature *596*, E1–E3. https://www.nature.com/articles/s41586-021-03765-z.
14. Hivert, V., Sidorenko, J., Rohart, F., Goddard, M.E., Yang, J., Wray, N.R., Yengo, L., and Visscher, P.M. (2021). Estimation of non-additive genetic variance in human complex traits from a large sample of unrelated individuals. Am. J. Hum. Genet. *108*, 786–798. https://www.sciencedirect.com/science/article/pii/S0002929721000562.
15. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet. *81*, 559–575. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1950838/.
16. Schüpbach, T., Xenarios, I., Bergmann, S., and Kapur, K. (2010). FastEpistasis: a high performance computing solution for quantitative trait epistasis. Bioinformatics *26*, 1468–1469. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2872003/.
17. Prabhu, S., and Pe'er, I. (2012). Ultrafast genome-wide scan for SNP–SNP interactions in common complex disease. Genome Res. *22*, 2230–2240. https://genome.cshlp.org/content/22/11/2230.
18. Wan, X., Yang, C., Yang, Q., Xue, H., Fan, X., Tang, N.L.S., and Yu, W. (2010). BOOST: A Fast Approach to Detecting Gene-Gene Interactions in Genome-wide Case-Control Studies. Am. J. Hum. Genet. *87*, 325–340. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2933337/.
19. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209. https://www.nature.com/articles/s41586-018-0579-z.
20. Nagai, A., Hirata, M., Kamatani, Y., Muto, K., Matsuda, K., Kiyohara, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z., Mushiroda, T., et al. (2017). Overview of the BioBank Japan Project: Study design and profile. J. Epidemiol. *27*, S2–S8. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5350590/.
21. Haseman, J.K., and Elston, R.C. (1972). The investigation of linkage between a quantitative trait and a marker locus. Behav. Genet. *2*, 3–19. https://doi.org/10.1007/BF01066731.
22. Zhou, X. (2017). A unified framework for variance component estimation with summary statistics in genome-wide association studies. Ann. Appl. Stat. *11*, 2027–2051. https://projecteuclid.org/euclid.aoas/1514430276.
23. Wu, Y., and Sankararaman, S. (2018). A scalable estimator of SNP heritability for biobank-scale data. Bioinformatics *34*, i187–i194. https://doi.org/10.1093/bioinformatics/bty253.
24. Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J.A., Ziyatdinov, A., Benner, C., O'Dushlaine, C., Barber, M., Boutkov, B., et al. (2021). Computationally efficient whole-genome regression for quantitative and binary traits. Nat. Genet. *53*, 1097–1103. https://www.nature.com/articles/s41588-021-00870-7.
25. Hutchinson, M.F. (1989). A Stochastic Estimator of the Trace of the Influence Matrix for Laplacian Smoothing Splines. Commun. Stat. Simulat. Comput. *18*, 1059–1076. http://www.tandfonline.com/doi/abs/10.1080/03610918908812806.
26. Liberty, E., and Zucker, S.W. (2009). The Mailman algorithm: A note on matrix–vector multiplication. Inf. Process. Lett. *109*, 179–182. https://www.sciencedirect.com/science/article/pii/S0020019008002949.
27. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science *337*, 1190–1195. https://www.science.org/doi/10.1126/science.1222794.
28. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. *47*, 1228–1235. https://www.nature.com/articles/ng.3404.
29. Cano-Gamez, E., and Trynka, G. (2020). From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. Front. Genet. *11*, 424. https://www.frontiersin.org/articles/10.3389/fgene.2020.00424/full.
30. Finucane, H.K., Reshef, Y.A., Anttila, V., Slowikowski, K., Gusev, A., Byrnes, A., Gazal, S., Loh, P.R., Lareau, C., Shoresh, N., et al. (2018). Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat. Genet. *50*, 621–629. https://www.nature.com/articles/s41588-018-0081-4.
31. Boix, C.A., James, B.T., Park, Y.P., Meuleman, W., and Kellis, M. (2021). Regulatory genomic circuitry of human disease loci by integrative epigenomics. Nature *590*, 300–307. https://www.nature.com/articles/s41586-020-03145-z.
32. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. *42*, 565–569. https://www.nature.com/articles/ng.608.
33. Georgolopoulos, G., Psatha, N., Iwata, M., Nishida, A., Som, T., Yiangou, M., Stamatoyannopoulos, J.A., and Vierstra, J. (2021). Discrete regulatory modules instruct hematopoietic

lineage commitment and differentiation. Nat. Commun. *12*, 6790. https://www.nature.com/articles/s41467-021-27159-x.

34. Wu, Y., Burch, K.S., Ganna, A., Pajukanta, P., Pasaniuc, B., and Sankararaman, S. (2022). Fast estimation of genetic correlation for biobank-scale data. Am. J. Hum. Genet. *109*, 24–32. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8764132/.
35. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295. https://www.nature.com/articles/ng.3211.
36. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test. Am. J. Hum. Genet. *89*, 82–93. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3135811/.
37. Zhou, X., Carbonetto, P., and Stephens, M. (2013). Polygenic Modeling with Bayesian Sparse Linear Mixed Models. PLoS Genet. *9*, e1003264. https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1003264.
38. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 7.
39. Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P., and Wang, L. (2014). CrossMap: a versatile tool for coordinate conversion between genome assemblies. Bioinformatics *30*, 1006–1007. https://doi.org/10.1093/bioinformatics/btt730.
40. Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for Computing and Annotating Genomic Ranges. PLoS Comput. Biol. *9*, e1003118. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003118.
41. Berisa, T., and Pickrell, J.K. (2016). Approximately independent linkage disequilibrium blocks in human populations. Bioinformatics *32*, 283–285. https://doi.org/10.1093/bioinformatics/btv546.
42. Fadista, J., Manning, A.K., Florez, J.C., and Groop, L. (2016). The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. Eur. J. Hum. Genet. *24*, 1202–1205. https://www.nature.com/articles/ejhg2015269.
43. Ding, K., Shameer, K., Jouni, H., Masys, D.R., Jarvik, G.P., Kho, A.N., Ritchie, M.D., McCarty, C.A., Chute, C.G., Manolio, T.A., and Kullo, I.J. (2012). Genetic Loci Implicated in Erythroid Differentiation and Cell Cycle Regulation Are Associated With Red Blood Cell Traits. Mayo Clin. Proc. *87*, 461–474. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3538470/.
44. Bhoopalan, S.V., Huang, L.J.-s., and Weiss, M.J. (2020). Erythropoietin regulation of red blood cell production: from bench to bedside and back. F1000Res. *9*. F1000 Faculty Rev-1153. https://pmc.ncbi.nlm.nih.gov/articles/PMC7503180/.
45. Qin, J., Liu, Q., Liu, Z., Pan, Y.Z., Sifuentes-Dominguez, L., Stepien, K.P., Wang, Y., Tu, Y., Tan, S., Wang, Y., et al. (2020). Structural and mechanistic insights into secretagogin-mediated exocytosis. Proc. Natl. Acad. Sci. USA *117*, 6559–6570. https://www.pnas.org/doi/10.1073/pnas.1919698117.
46. Timoteo, V.J., Chiang, K.-M., Yang, H.-C., and Pan, W.-H. (2023). Common and ethnic-specific genetic determinants of hemoglobin concentration between Taiwanese Han Chinese and European Whites: findings from comparative two-stage genome-wide association studies. J. Nutr. Biochem. *111*, 109126. https://www.sciencedirect.com/science/article/pii/S0955286322001942.
47. Bauer, M.C., O'Connell, D.J., Maj, M., Wagner, L., Cahill, D. J., and Linse, S. (2011). Identification of a high-affinity network of secretagogin-binding proteins involved in vesicle secretion. Mol. Biosyst. *7*, 2196–2204. https://pubs.rsc.org/en/content/articlelanding/2011/mb/c0mb00349b.
48. Ray, S., Chee, L., Zhou, Y., Schaefer, M.A., Naldrett, M.J., Alvarez, S., Woods, N.T., and Hewitt, K.J. (2022). Functional requirements for a Samd14-capping protein complex in stress erythropoiesis. eLife *11*, e76497. https://doi.org/10.7554/eLife.76497.
49. Edwards, M., Liang, Y., Kim, T., and Cooper, J.A. (2013). Physiological role of the interaction between CARMIL1 and capping protein. Mol. Biol. Cell *24*, 3047–3055. https://www.molbiolcell.org/doi/10.1091/mbc.e13-05-0270.
50. Yang, C., Pring, M., Wear, M.A., Huang, M., Cooper, J.A., Svitkina, T.M., and Zigmond, S.H. (2005). Mammalian CARMIL Inhibits Actin Filament Capping by Capping Protein. Dev. Cell *9*, 209–221. https://www.sciencedirect.com/science/article/pii/S1534580705002510.
51. Plotnikov, D., Huang, Y., Khawaja, A.P., Foster, P.J., Zhu, Z., Guggenheim, J.A., and He, M. (2022). High Blood Pressure and Intraocular Pressure: A Mendelian Randomization Study. Investig. Ophthalmol. Vis. Sci. *63*, 29.
52. Galesloot, T.E., Verweij, N., Traglia, M., Barbieri, C., van Dijk, F., Geurts-Moespot, A.J., Girelli, D., Kiemeney, L.A.L.M., Sweep, F.C.G.J., Swertz, M.A., et al. (2016). Meta-GWAS and Meta-Analysis of Exome Array Studies Do Not Reveal Genetic Determinants of Serum Hepcidin. PLoS One *11*, e0166628. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0166628.
53. Vuckovic, D., Bao, E.L., Akbari, P., Lareau, C.A., Mousas, A., Jiang, T., Chen, M.H., Raffield, L.M., Tardaguila, M., Huffman, J.E., et al. (2020). The Polygenic and Monogenic Basis of Blood Traits and Diseases. Cell *182*, 1214–1231.e11. https://www.sciencedirect.com/science/article/pii/S0092867420309995.
54. Zhang, W., Duan, S., Kistner, E.O., Bleibel, W.K., Huang, R.S., Clark, T.A., Chen, T.X., Schweitzer, A.C., Blume, J.E., Cox, N. J., and Dolan, M.E. (2008). Evaluation of Genetic Variation Contributing to Differences in Gene Expression between Populations. Am. J. Hum. Genet. *82*, 631–640. https://www.cell.com/ajhg/abstract/S0002-9297(08)00136-5.
55. Jung, G., Pan, M., Alexander, C.J., Jin, T., and Hammer, J.A. (2022). Dual regulation of the actin cytoskeleton by CARMIL-GAP. J. Cell Sci. *135*, jcs258704. https://doi.org/10.1242/jcs.258704.
56. Stark, B.C., Lanier, M.H., and Cooper, J.A. (2017). CARMIL family proteins as multidomain regulators of actin-based motility. Mol. Biol. Cell *28*, 1713–1723. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5491179/.

57. Nigra, A.D., Casale, C.H., and Santander, V.S. (2020). Human erythrocytes: cytoskeleton and its origin. Cell. Mol. Life Sci. *77*, 1681–1694. https://doi.org/10.1007/s00018-019-03346-4.
58. Gokhin, D.S., and Fowler, V.M. (2016). Feisty filaments: actin dynamics in the red blood cell membrane skeleton. Curr. Opin. Hematol. *23*, 206–214. https://journals.lww.com/co-hematology/fulltext/2016/05000/feisty_filaments__actin_dynamics_in_the_red_blood.5.aspx.
59. Yengo, L., Vedantam, S., Marouli, E., Sidorenko, J., Bartell, E., Sakaue, S., Graff, M., Eliasen, A.U., Jiang, Y., Raghavan, S., et al. (2022). A saturated map of common genetic variants associated with human height. Nature *610*, 704–712.
60. Wood, A.R., Tuke, M.A., Nalls, M.A., Hernandez, D.G., Bandinelli, S., Singleton, A.B., Melzer, D., Ferrucci, L., Frayling, T. M., and Weedon, M.N. (2014). Another explanation for apparent epistasis. Nature *514*, E3–E5. https://www.nature.com/articles/nature13691.

The American Journal of Human Genetics, Volume *112*

Supplemental information

Sparse modeling of interactions enables fast detection of genome-wide epistasis in biobank-scale studies

Julian Stamp, Samuel Pattillo Smith, Daniel Weinreich, and Lorin Crawford

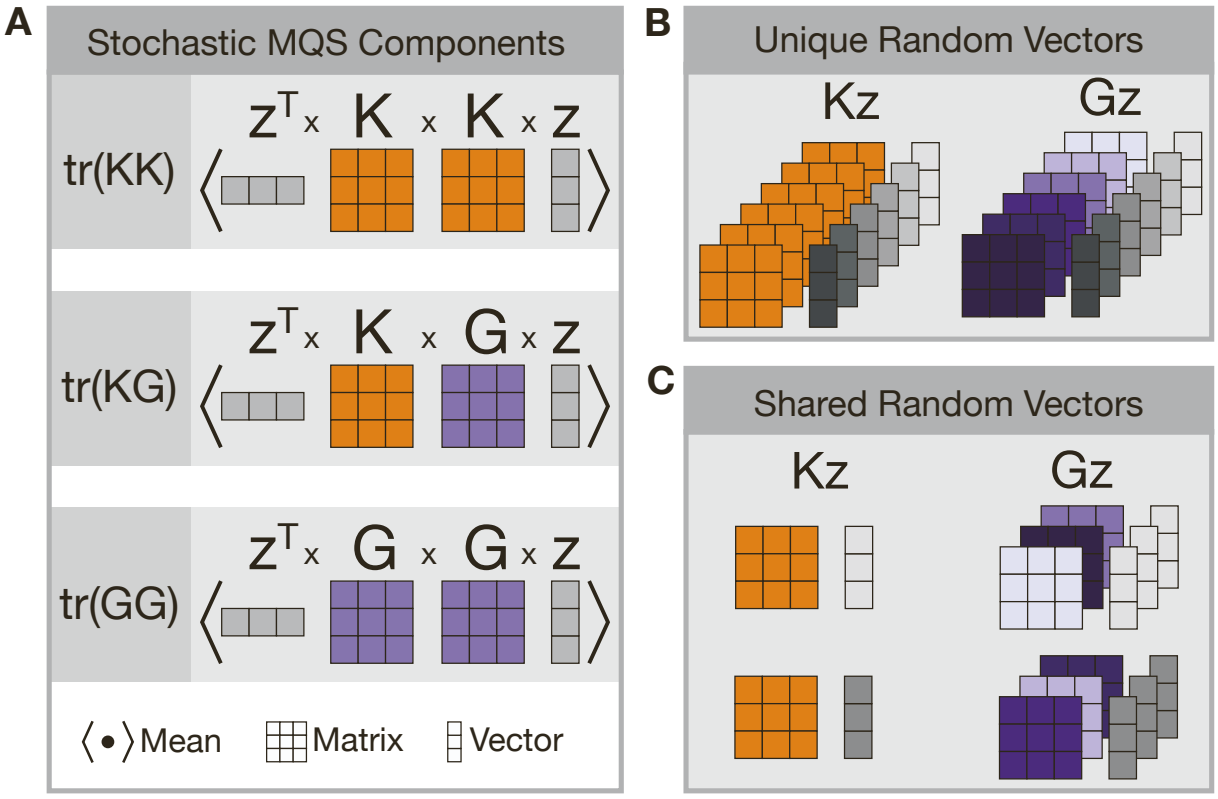


Figure S1. Schematic overview illustrating the approximation to the stochastic trace estimator used in SME. (A) Computing the exact trace of a product for two covariance matrices is computationally infeasible for large studies with many individuals. SME overcomes this limitation by computing point estimates for variance components using a method-of-moments (MoM) algorithm that features traces of matrix products with random vectors $\mathbf{z}$. In the stochastic trace estimates, we can identify reusable matrix-by-vector products. We see that the matrix-by-vector products of the form $\mathbf{Az}$ with $\mathbf{A} \in \{\mathbf{K}, \mathbf{G}_j\}$ and combinations thereof appear in multiple traces. **(B)** The genetic relatedness matrix $\mathbf{K}$ is the same for all focal SNPs. Using unique random vectors in this computation for every focal SNP, we compute the stochastic approximation repeatedly. Computing the matrix-by-vector products $\mathbf{Kz}$ constitutes about half of the computation time of the point estimates when no mask is applied. It constitutes most of the computation time when a mask induces sparsity. **(C)** By sharing random vectors $\mathbf{z}$ between a set of focal SNPs, computing $\mathbf{Kz}$ can be done once for the entire set. With this, we greatly reduce the computational burden of computing $\mathbf{Kz}$.

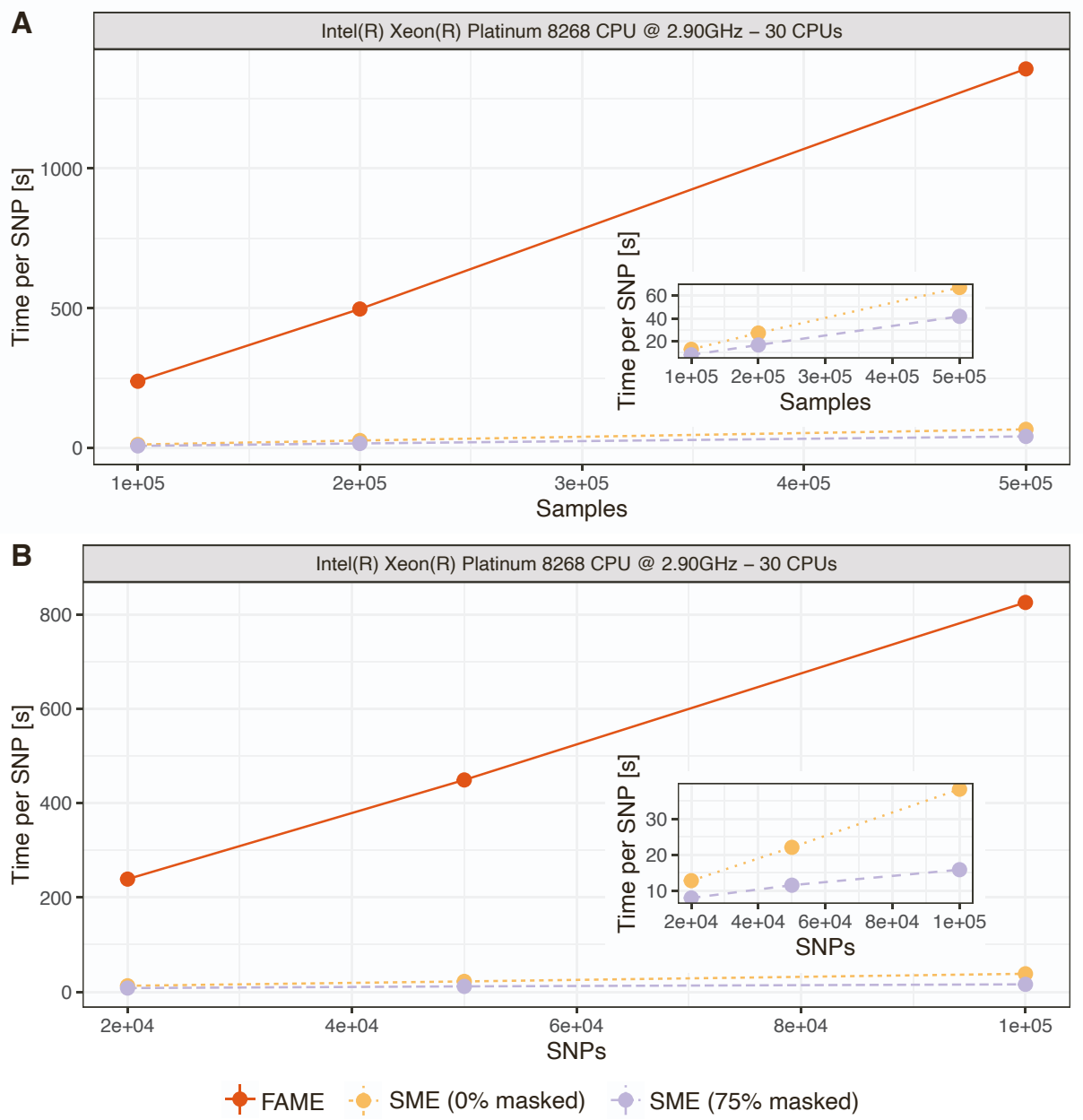


Figure S2. Computational time needed to run a single test with SME and FAME[1] as a function of sample size and total number of SNPs. The solid orange line represents the runtime using FAME. The dotted yellow line shows the runtime of SME without a mask. The dashed purple line represents the runtime of SME with 75% of variants masked. In all cases, the number of random vectors is fixed at 100. The number of variants sharing random vectors in SME is fixed at 90. Computations were performed on an Intel(R) Xeon(R) Platinum 8268 CPU with 30 cores. **(A)** Computational time per test (in seconds) as a function of sample size, with the genome size fixed at 20,000 SNPs. **(B)** Computational time per test (in seconds) as a function of the genome size in number of SNPs, with the sample size fixed at 100,000 individuals.

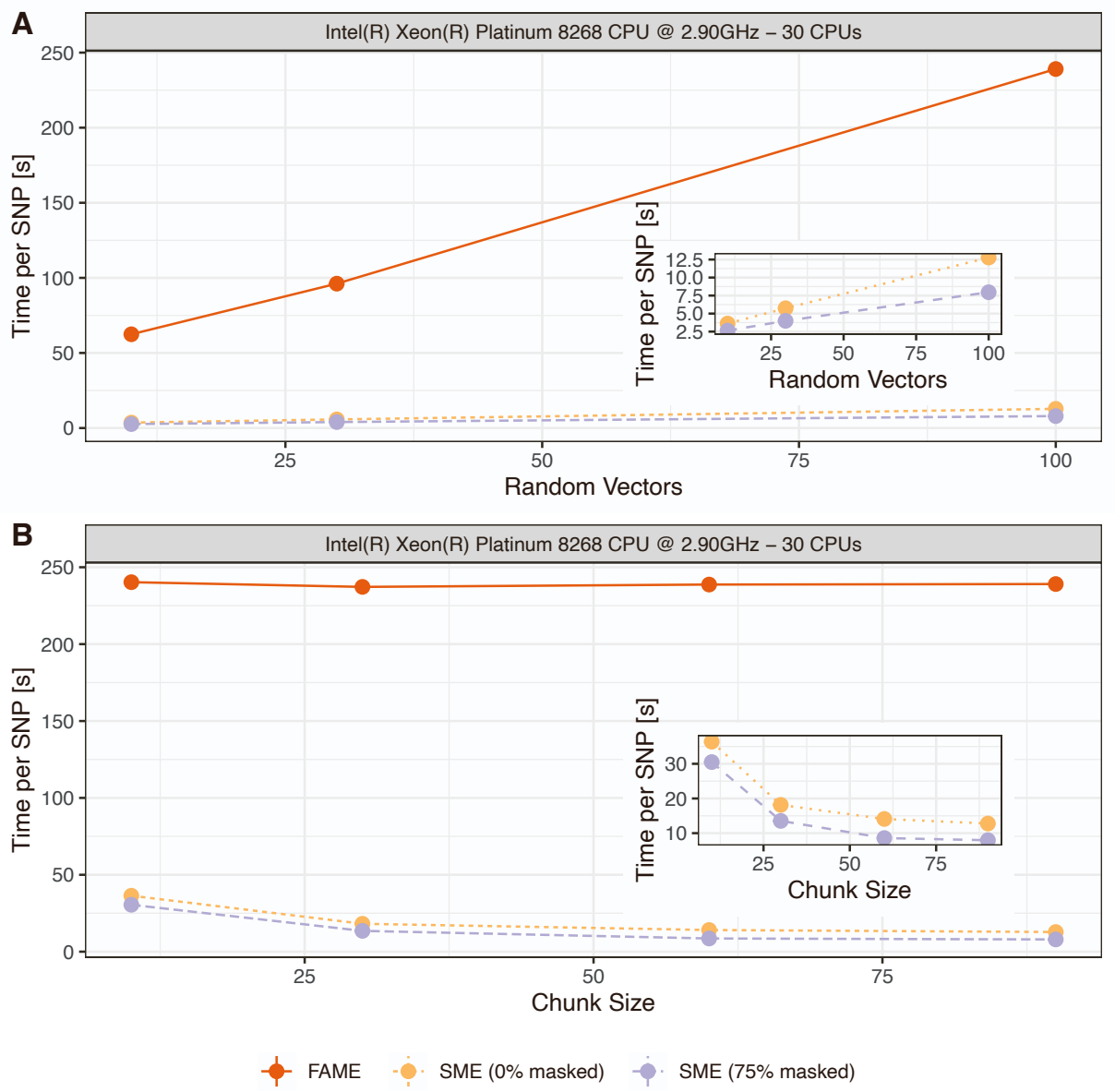


Figure S3. Computational time needed to run a single test with SME and FAME[1] as a function of the number of random vectors and number of SNPs sharing random vectors. The solid orange line represents the runtime using the command-line tool FAME. The dotted yellow line shows the runtime using the `R` function of SME without applying a mask. The dashed purple line represents the runtime using SME with 75% of the variants masked. The analyzed data had a genome size of 20,000 SNPs and a sample size of 100,000 individuals. All computations were performed on an Intel(R) Xeon(R) Platinum 8268 CPU with 30 cores. **(A)** Computational time per test as a function of the number of random vectors used in the stochastic trace estimator. The number of focal variants sharing random vectors (chunk size) in SME was fixed at 90. **(B)** Computational time per test as a function of the number of focal variants sharing random vectors (chunk size). The number of random vectors was fixed at 100.

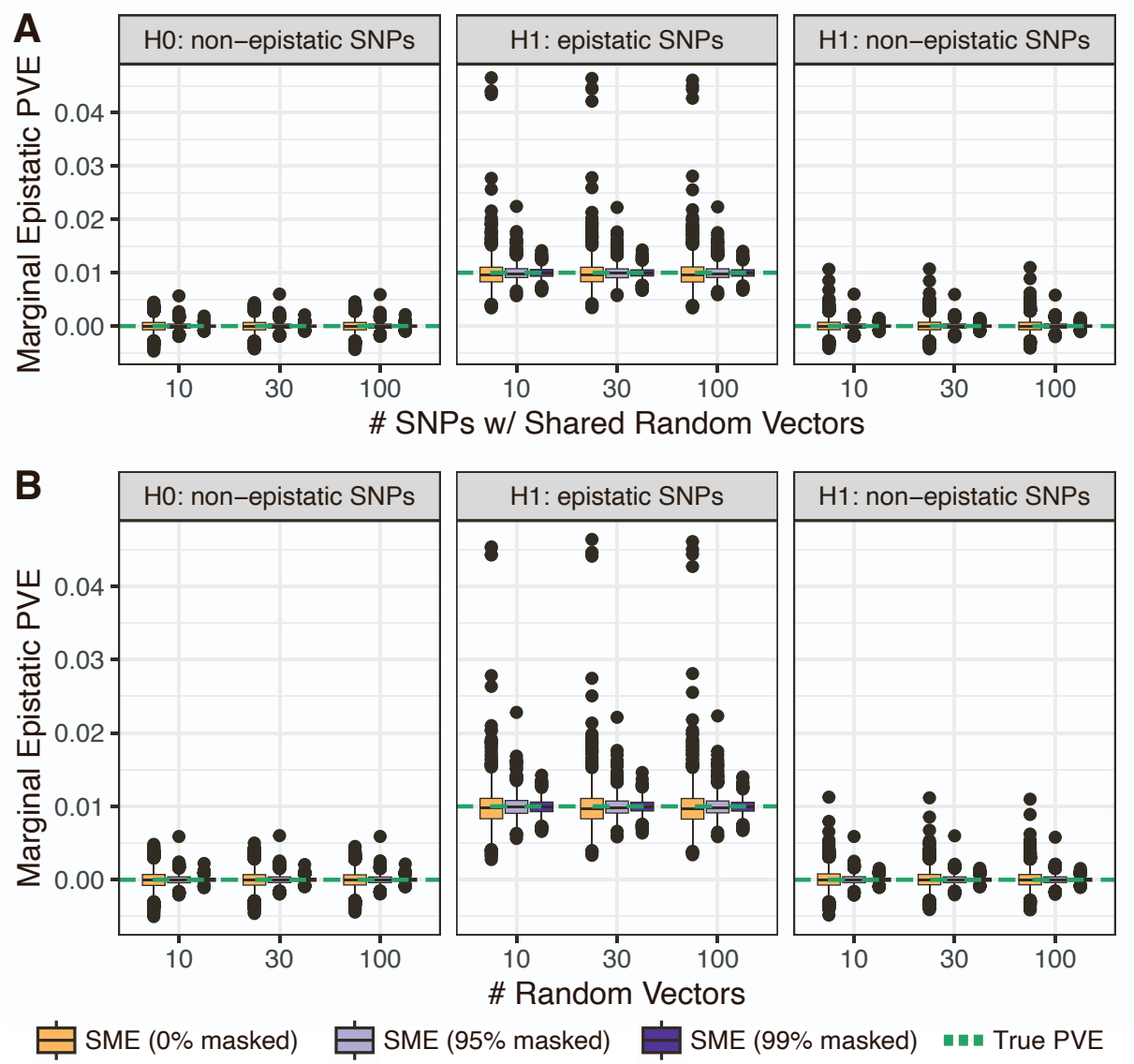


Figure S4. The SME stochastic method-of-moments algorithm produces estimates that are robust to the choice of number of random vectors and the number of SNPs sharing random vectors. Synthetic traits were simulated under the null hypothesis (H_0) of no epistasis and the alternaitve hypothesis (H_1) using chromosome 1 from individuals of self-identified European ancestry in the UK Biobank 100k individuals. We randomly selected 10% of all variants to have additive effects that collectively explained 30% of the trait variance. For simulations under H_1, we then fixed the total epistatic variance to 5%. The per SNP epistatic phenotypic variance explained (PVE) was adjusted by randomly choosing 10 epistatic SNPs. **(A)** Marginal epistatic variance component estimates as a function of the number of SNPs sharing random vectors in the stochastic trace estimates (also see Fig. S1). The number of random vectors was fixed at 100. **(B)** Marginal epistatic variance component estimates as a function of number of random vectors in the stochastic trace estimates. The number of SNPs sharing random vectors was fixed at 100. In both panels, the orange box plots show the results of using SME with no masking applied. The light and dark purple show the results for implementing SME with 95% and 99% masking, respectively. The dashed green line indicates the true marginal epistatic PVE used in the simulations.

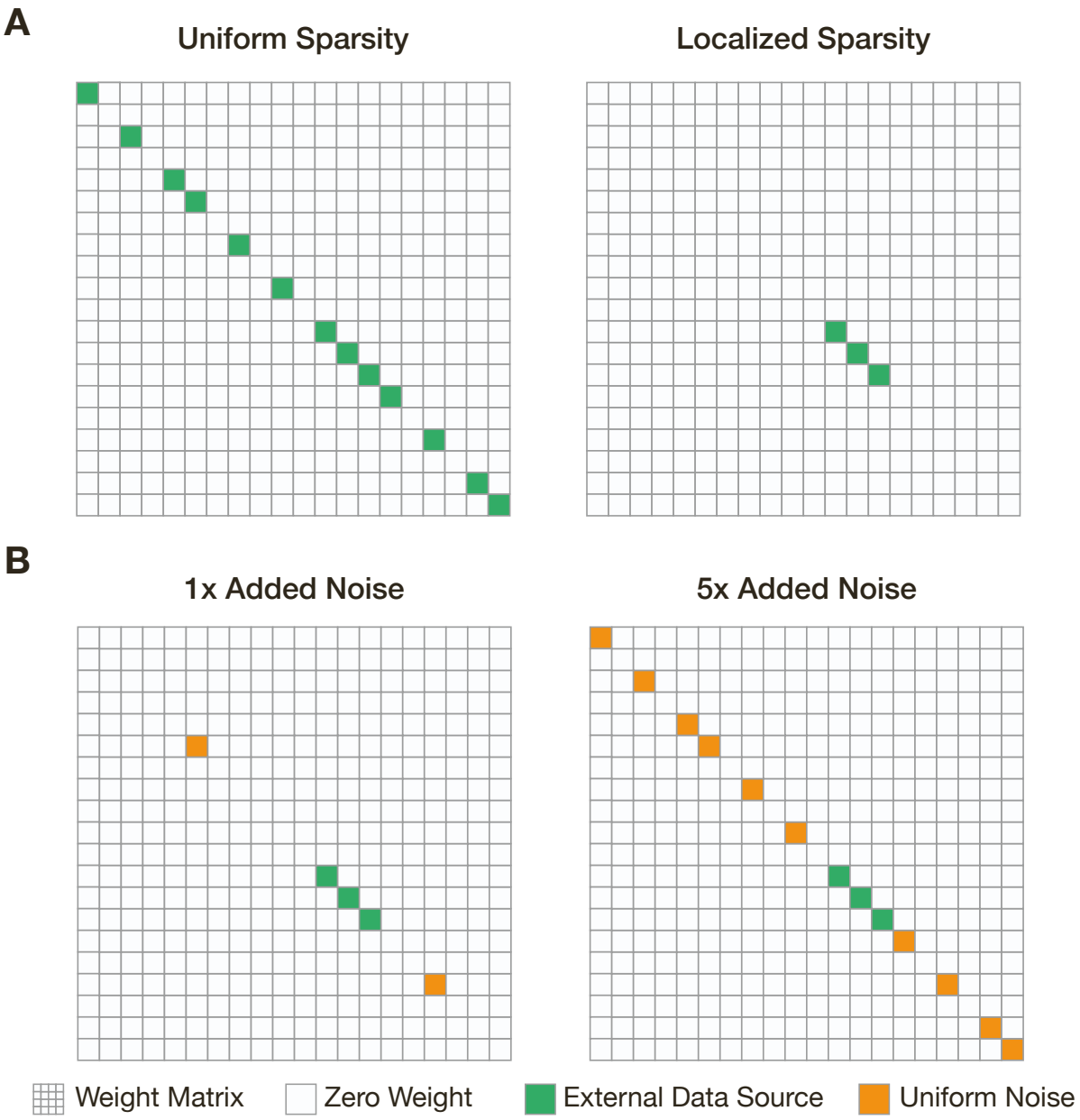


Figure S5. Schematic overview illustrating weight matrices with uniform and localized sparsity. **(A)** The simulations consider two scenarios that result in different distributions of sparsely modeled gene-interactions. The weight matrices are binary everywhere except for places on the diagonal where variants satisfy some criteria from an external data source (e.g., it is located in a position on the genome that overlaps with open chromatin regions). The first column illustrates *uniform sparsity* where non-zero entries are evenly distributed along the diagonal. As a result the modeled gene-interactions are evenly distributed along the genome. In the second scenario, the external data source induces *localized sparsity* where the modeled gene-interactions are concentrated in a small genomic window illustrated as a block structure in the weight matrix. **(B)** SME produces negatively biased variance component estimates when using an external data source that induces localized sparsity in the model. To overcome this issue in practice, we propose a strategy in which we take an external data source with localized genomic information and randomly unmask "unimportant" variants with uniform probability along the genome (making the localized sparsity look more uniform). The cartoon illustrates the distribution of non-zero weights in the weight matrix after including an additional $1\times$ and $5\times$ of initially disregarded interactions.

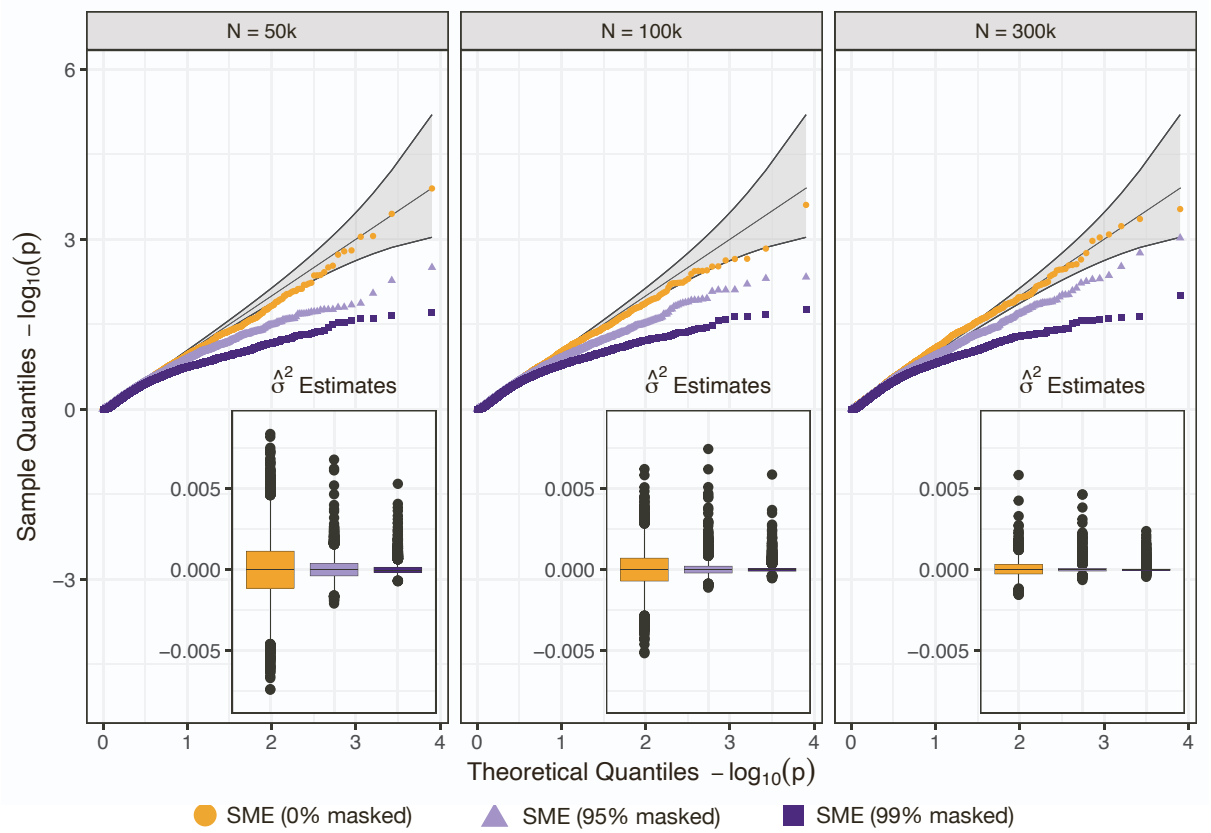


Figure S6. While using a mask that induces localized sparsity, SME is negatively biased under the null hypothesis. Synthetic traits were simulated with only additive effects using chromosome 1 from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled using sample sizes of 50k, 100k, and 300k individuals. We randomly selected 10% of all variants to be causal with additive effects and we assume that they explain 40% of the phenotypic variance for each trait. To simulate localized sparsity in the mask, we randomly sample a seed SNP and define a block around it. All SNPs not in that block are masked. Data were analyzed with SME under varying percentages of SNPs that are masked (0%, 95%, and 99%, respectively). The small insets in each plot show the distribution of the estimated marginal epistatic variance components across all experiments. For reference, under the null hypothesis $H_0 : \sigma^2 = 0$. Results are based on 100 simulations per scenario.

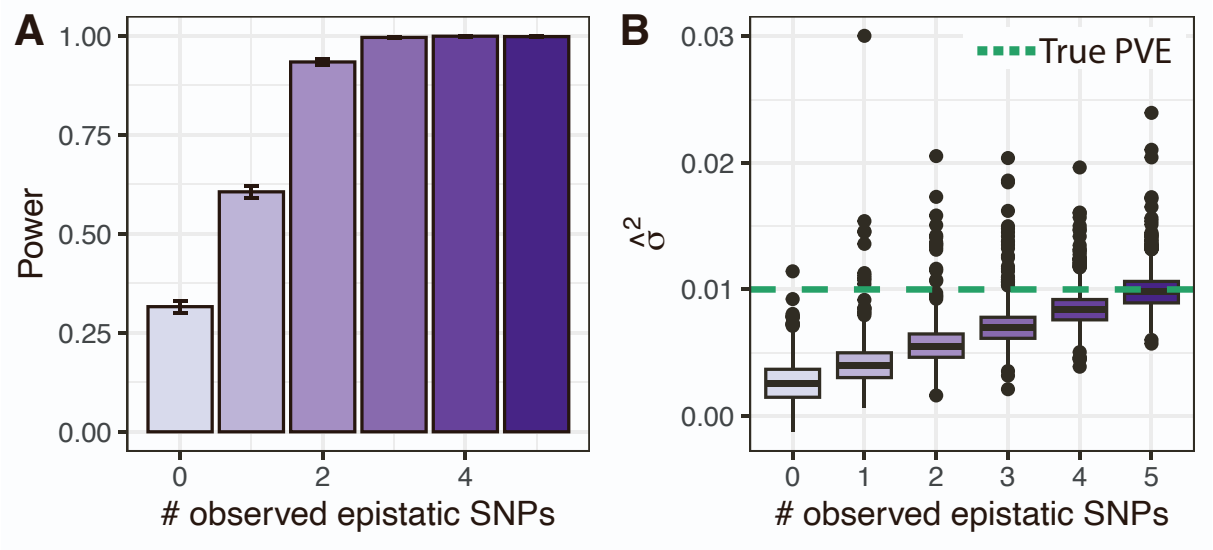


Figure S7. Consequence of having a misspecified mask on power and variance component estimates in SME. Synthetic traits were simulated using chromosome 1 from individuals of self-identified European ancestry in the UK Biobank. Data were subsampled to 100k individuals. We randomly selected 10% of all variants to have additive effects that collectively explained 30% of the trait variance. We then fixed the total epistatic variance to 5%. Each epistatic SNP was simulated to have 5 interaction partners. SME was given a weight matrix in which 95% of all SNPs were masked. In this simulation, we investigate a scenario where the weight matrix incorrectly masked 0 to 5 of true interacting partners for each causal SNP. **(A)** Empirical power computed as the function of masked true interactive partners. Here, significant SNPs were evaluated at a genome-wide threshold $P < 5 \times 10^{-8}$. Error bars represent the standard deviation across 100 replicate experiments. **(B)** Marginal epistatic variance component estimates ($\hat{\sigma}^2$) as a function of masked true interactive partners. The dashed green line represents the true simulated per SNP epistatic phenotypic variance explained (PVE).

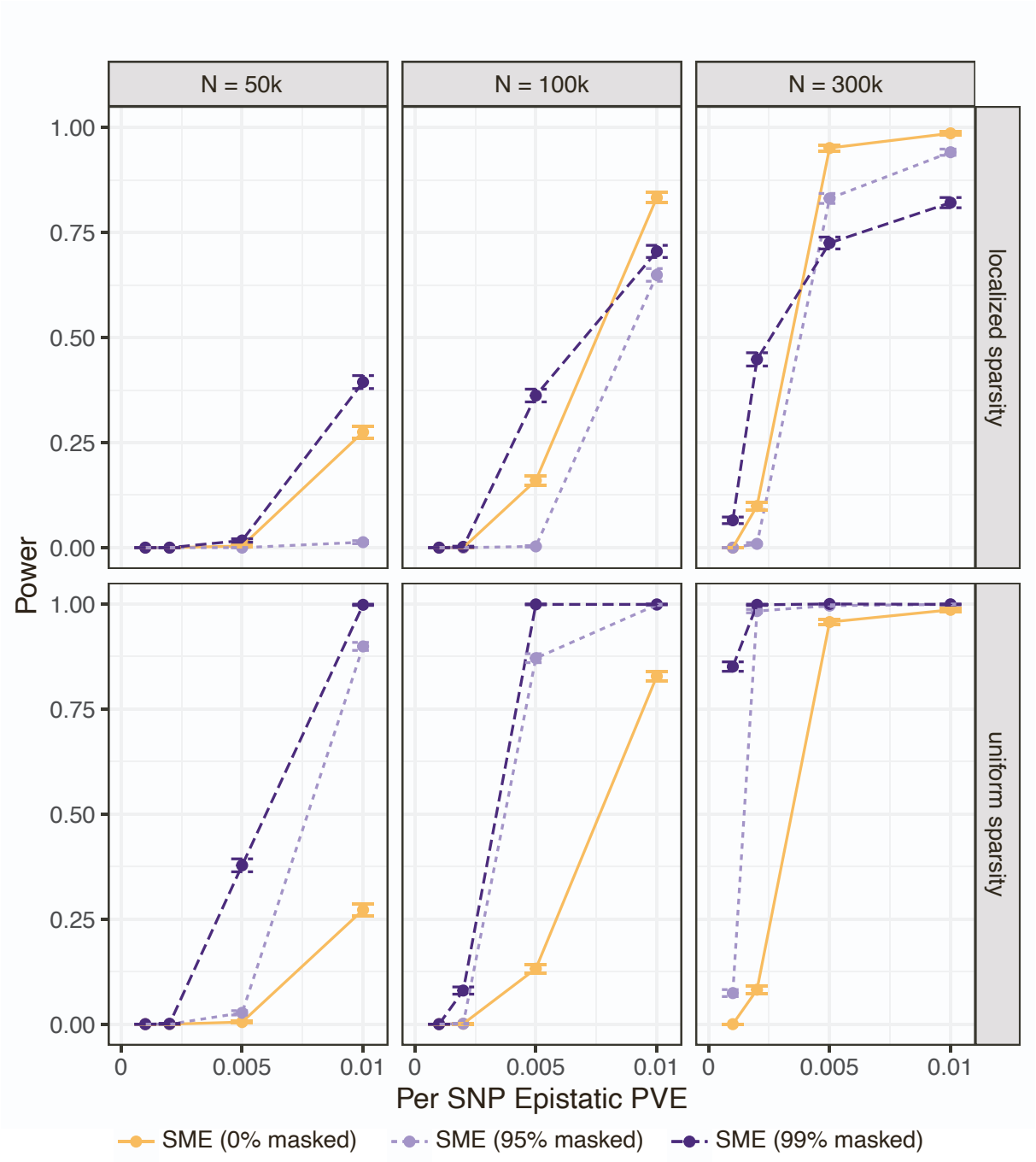


Figure S8. SME exhibits less power when using a mask that induces localized sparsity. Synthetic traits were simulated with both additive and pairwise epistatic effects using chromosome 1 from individuals of self-identified European ancestry in the UK Biobank. Data were subsampled using sample sizes of 50k, 100k, and 300k individuals. We randomly selected 10% of all variants to have additive effects that collectively explained 30% of the trait variance. We then fixed the total epistatic variance to 5%. The per SNP epistatic phenotypic variance explained (PVE) was adjusted by varying the number of interacting SNPs (chosen to be 10, 20, 50, or 100 SNPs). Data were analyzed using SME under varying percentages of variants that are excluded from consideration as potential interaction partners for each focal SNP (0%, 95%, and 99% masking, respectively). Empirical power was determined using the significance threshold $P < 5 \times 10^{-8}$. Results for uniform sparsity are given on the bottom as reference. We conducted 100 simulations per scenario, with error bars representing the standard deviation across replicates.

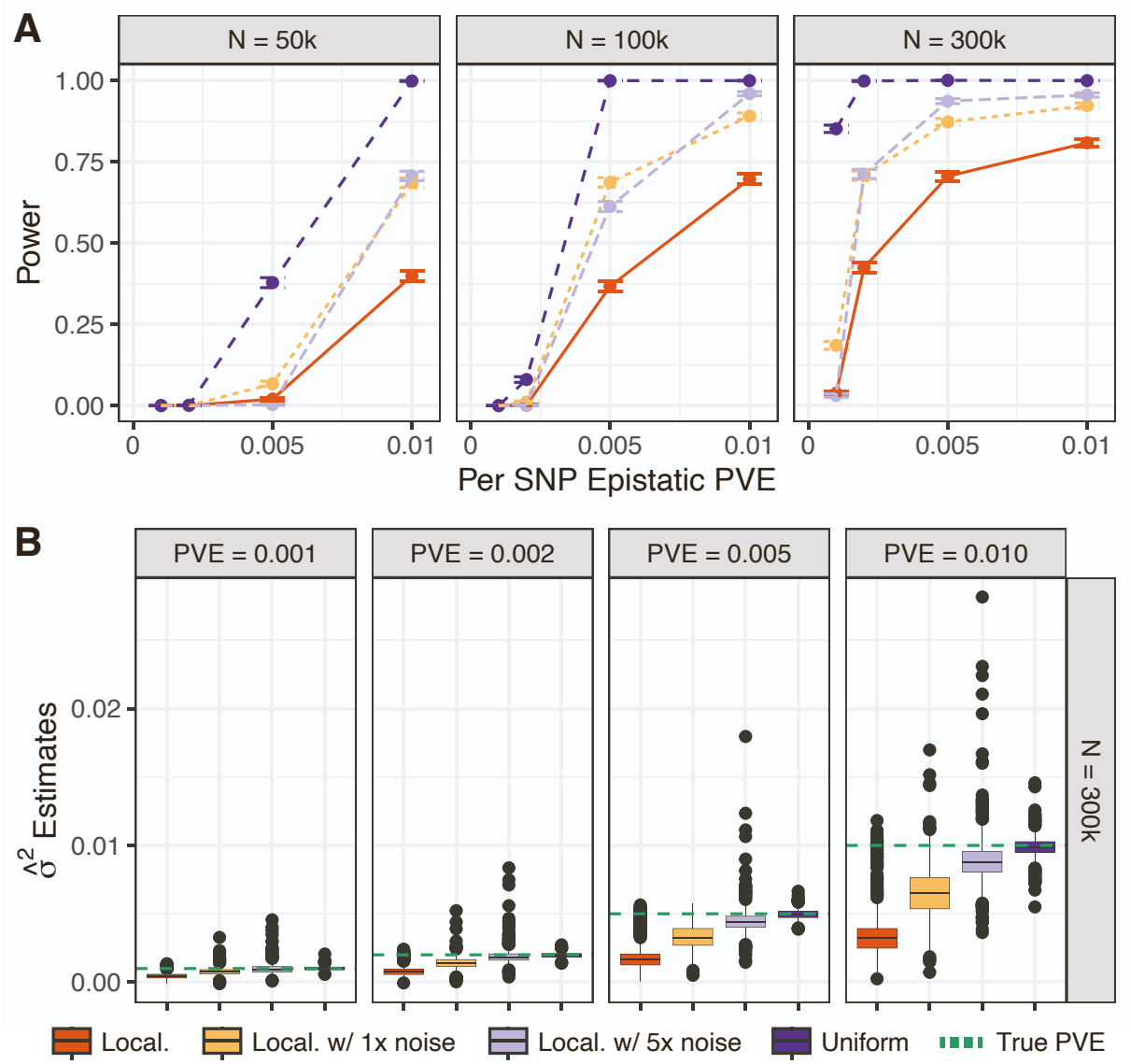


Figure S9. Adding random noise to external data sources that induce localized sparsity recoveres empirical power and reduces the negative biase in variance component estimates. Synthetic traits were simulated with both additive and pairwise epistatic effects using chromosome 1 from individuals of self-identified European ancestry in the UK Biobank. Data were subsampled using sample sizes of 50k, 100k, and 300k individuals. We randomly selected 10% of all variants to have additive effects that collectively explained 30% of the trait variance. We then fixed the total epistatic variance to 5%. The per SNP epistatic phenotypic variance explained (PVE) was adjusted by varying the number of interacting SNPs (chosen to be 10, 20, 50, or 100 SNPs). We simulated external data sources that induced both *uniform sparsity* (dark purple) and *localized sparsity* (dark orange) resulting in 99% of variants being masked. To counteract the negative bias of the localized sparsity, we randomly unmask 1% and 5% additional SNPs with uniform probability along the entire chromosome—making the localized sparsity look more uniform (also see Fig. S5). **(A)** Empirical power at a genome-wide significance threshold $P < 5 \times 10^{-8}$, shown as a function of the per SNP epistatic phenotypic variance explained (PVE) by causal variants across three different sample sizes. Error bars represent the standard deviation across 100 replicate experiments. **(B)** Marginal epistatic variance component estimates ($\hat{\sigma}^2$) on simulated traits with 300k individuals. The dashed green line represents the true per SNP epistatic PVE.

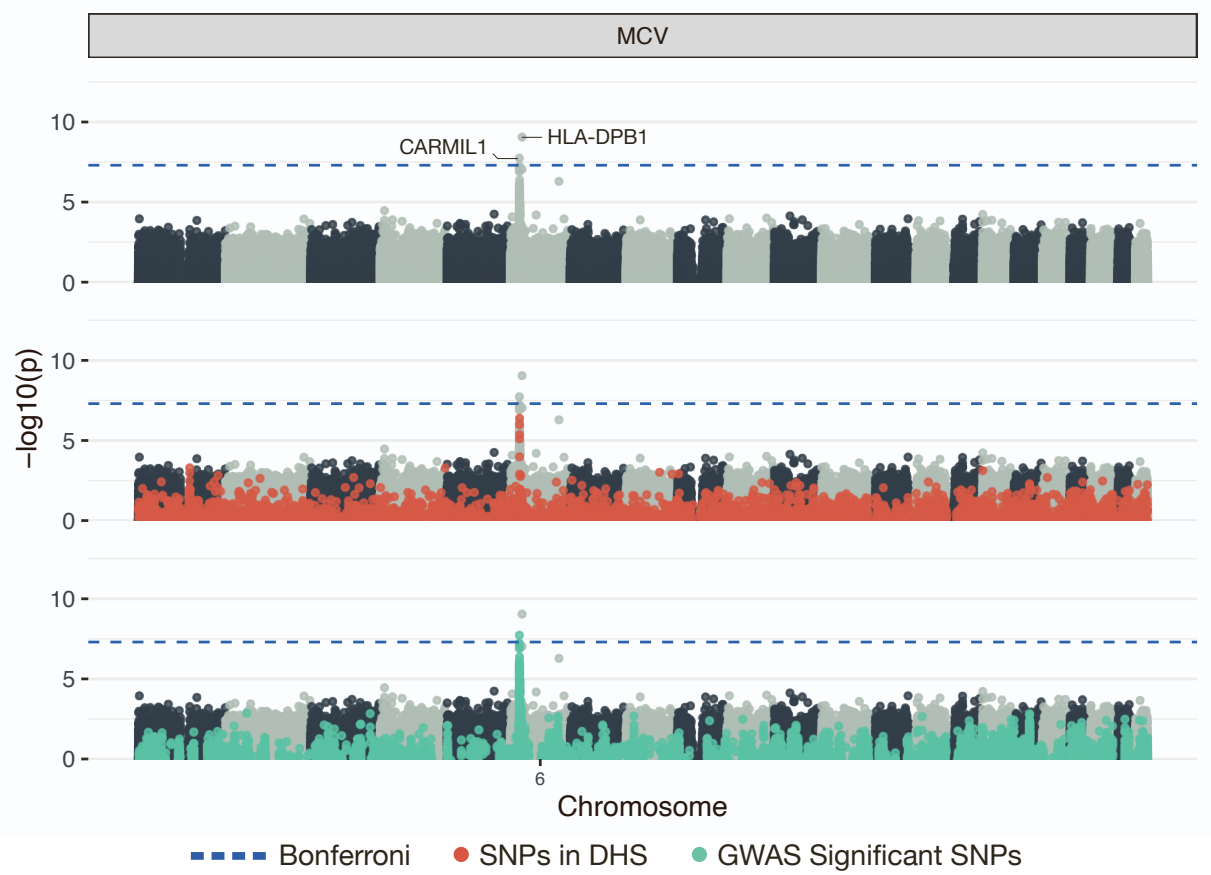


Figure S10. Manhattan plots of a genome-wide interaction analysis using SME to study mean corpuscular volume (MCV) assayed in individuals in the UK Biobank. As a mask in this study, we leveraged DNase I-hypersensitive sites (DHS) data measured over 12 days of *ex vivo* erythroid differentiation[2,3]. This means that, while all SNPs are tested for marginal epistasis, only their interactions with SNPs in DHS regions are considered. Here, $-\log_{10}$ transformed P-values from SME are plotted for each SNP against their genomic positions. Chromosomes are shown in alternating colors for clarity. The dashed blue line represents the genome-wide significance threshold ($P < 5 \times 10^{-8}$). Each panel shows the same plot with different aspects of the result highlighted. The first simply shows the names of the closest neighboring genes to significant epistatic SNPs. The second panel highlights the SNPs that fall in DHS regions, and the third panel highlights SNPs that were also found to have a significant (additive) association with the trait according to a GWAS.

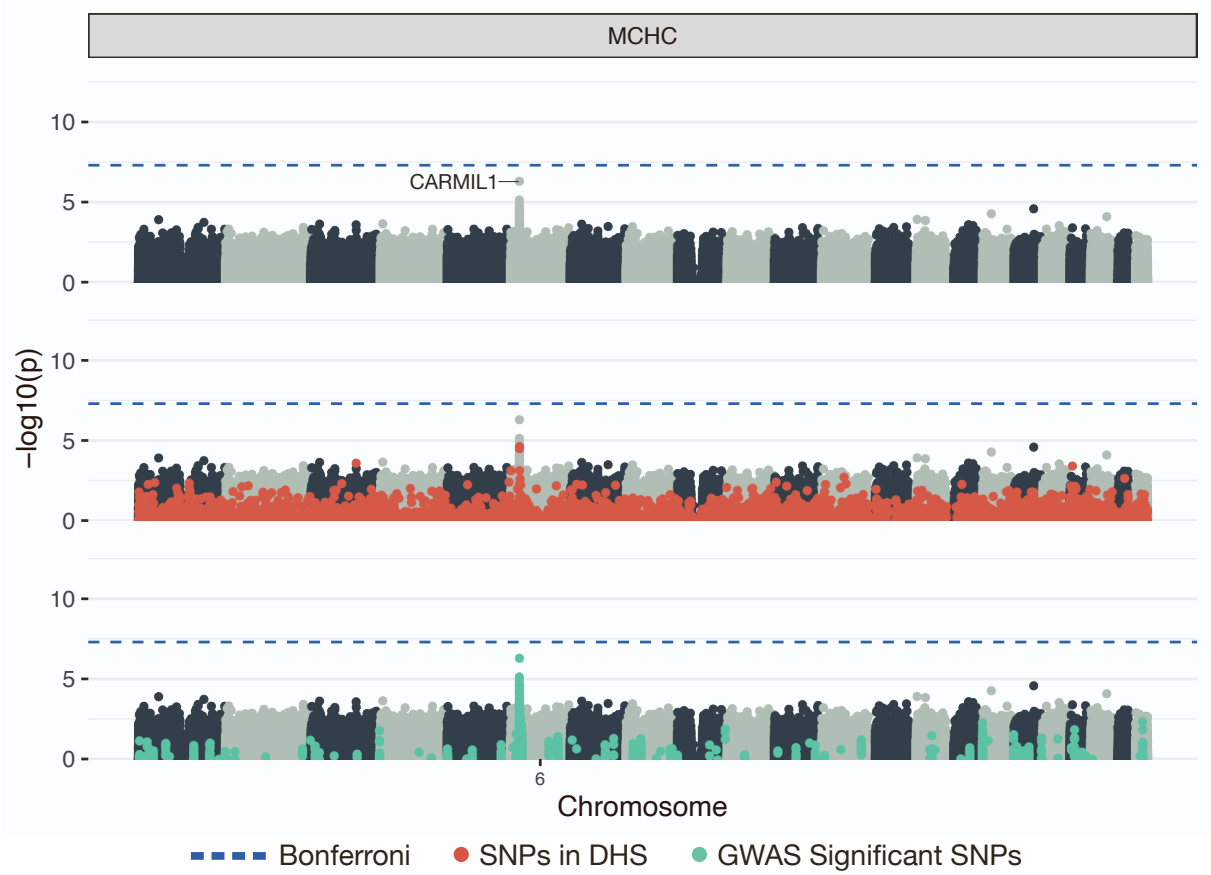


Figure S11. Manhattan plots of a genome-wide interaction analysis using SME to study mean corpuscular hemoglobin concentration (MCHC) assayed in individuals in the UK Biobank. As a mask in this study, we leveraged DNase I-hypersensitive sites (DHS) data measured over 12 days of *ex vivo* erythroid differentiation[2,3]. This means that, while all SNPs are tested for marginal epistasis, only their interactions with SNPs in DHS regions are considered. Here, $-\log_{10}$ transformed P-values from SME are plotted for each SNP against their genomic positions. Chromosomes are shown in alternating colors for clarity. The dashed blue line represents the genome-wide significance threshold ($P < 5 \times 10^{-8}$). Each panel shows the same plot with different aspects of the result highlighted. The first simply shows the names of the closest neighboring genes to significant epistatic SNPs. The second panel highlights the SNPs that fall in DHS regions, and the third panel highlights SNPs that were also found to have a significant (additive) association with the trait according to a GWAS.

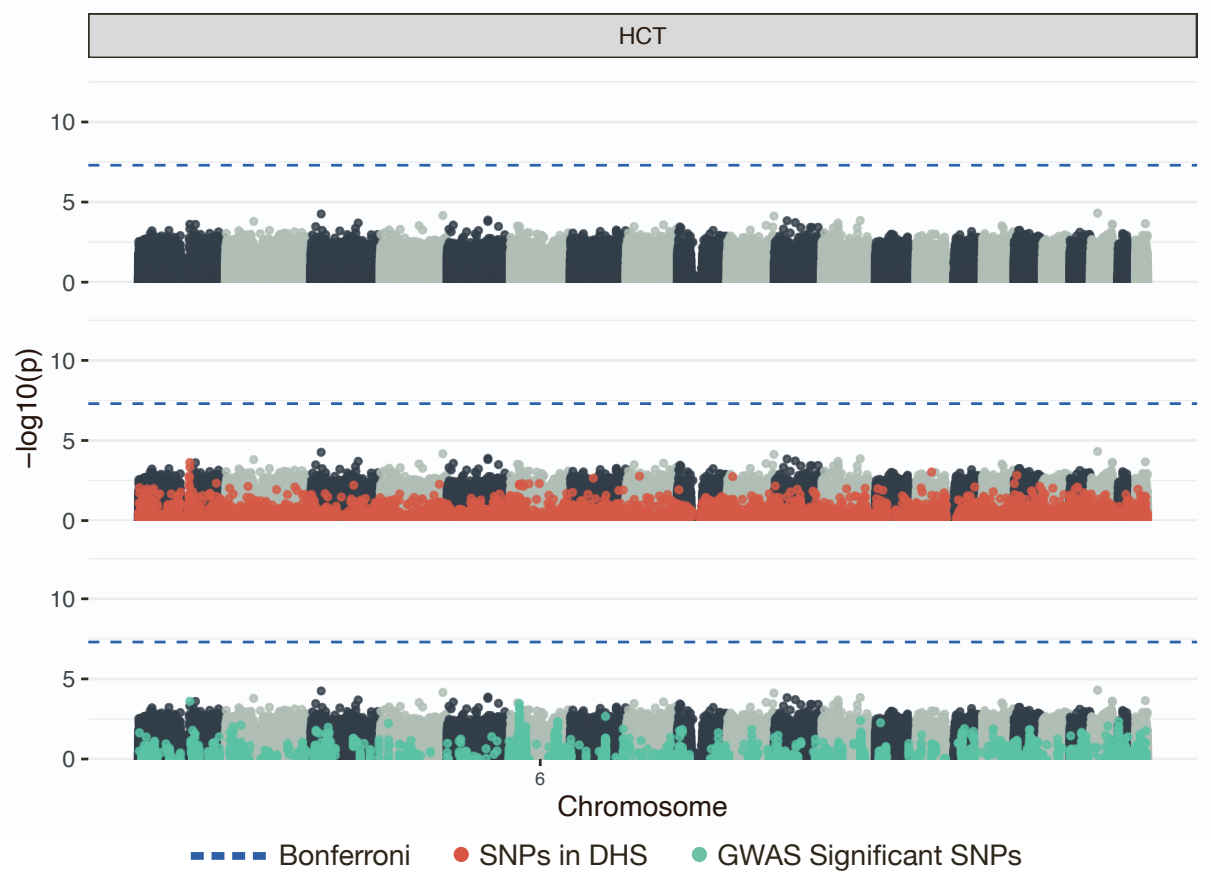


Figure S12. Manhattan plots of a genome-wide interaction analysis using SME to study hematocrit (HCT) assayed in individuals in the UK Biobank. As a mask in this study, we leveraged DNase I-hypersensitive sites (DHS) data measured over 12 days of *ex vivo* erythroid differentiation[2,3]. This means that, while all SNPs are tested for marginal epistasis, only their interactions with SNPs in DHS regions are considered. Here, $-\log_{10}$ transformed P-values from SME are plotted for each SNP against their genomic positions. Chromosomes are shown in alternating colors for clarity. The dashed blue line represents the genome-wide significance threshold ($P < 5 \times 10^{-8}$). Each panel shows the same plot with different aspects of the result highlighted. The first simply shows the names of the closest neighboring genes to significant epistatic SNPs. The second panel highlights the SNPs that fall in DHS regions, and the third panel highlights SNPs that were also found to have a significant (additive) association with the trait according to a GWAS.

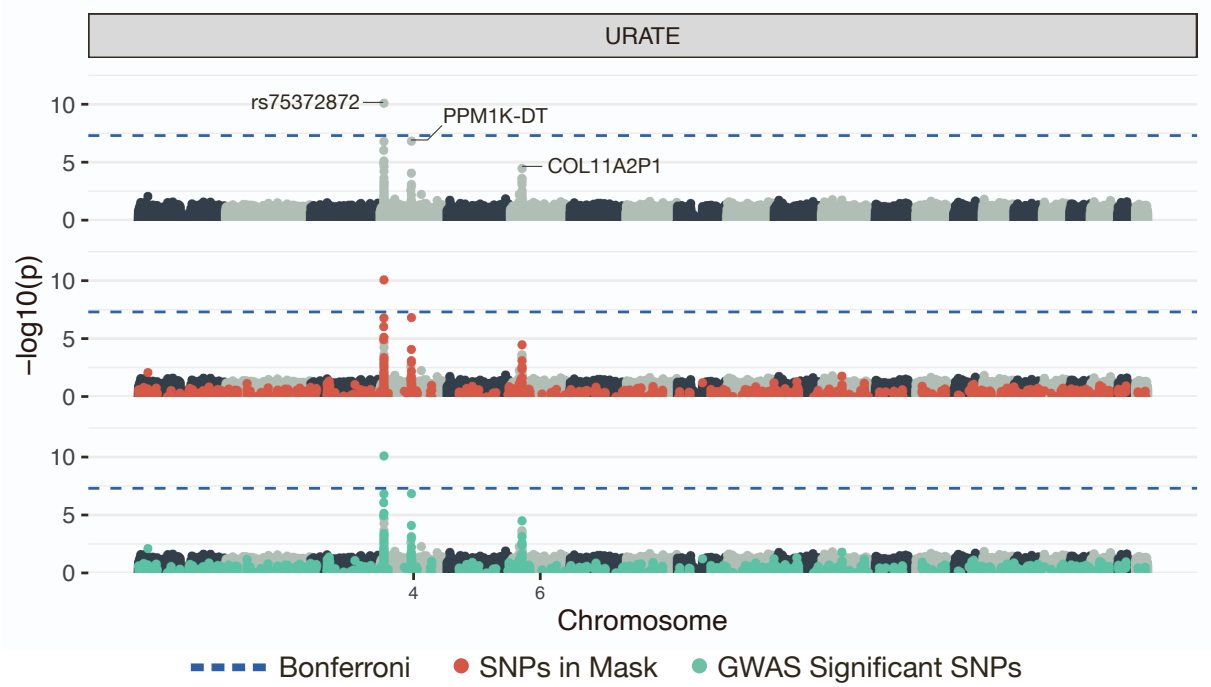


Figure S13. Manhattan plots of a genome-wide interaction analysis using SME to study uric acid (urate) assayed in individuals in the UK Biobank. As a mask in this study, we leveraged 3,536 significant GWAS variants. This means that, while all SNPs are tested for marginal epistasis, only their interactions with SNPs identified by GWAS are considered. Here, $-\log_{10}$ transformed P-values from SME are plotted for each SNP against their genomic positions. Chromosomes are shown in alternating colors for clarity. The dashed blue line represents the genome-wide significance threshold ($P < 5 \times 10^{-8}$). Each panel shows the same plot with different aspects of the result highlighted. The first simply shows the names of the closest neighboring genes to significant epistatic SNPs. The second panel highlights the SNPs that are modeled to interact with the focal SNP, and the third panel highlights SNPs that were also found to have a significant (additive) association with the trait according to a GWAS.

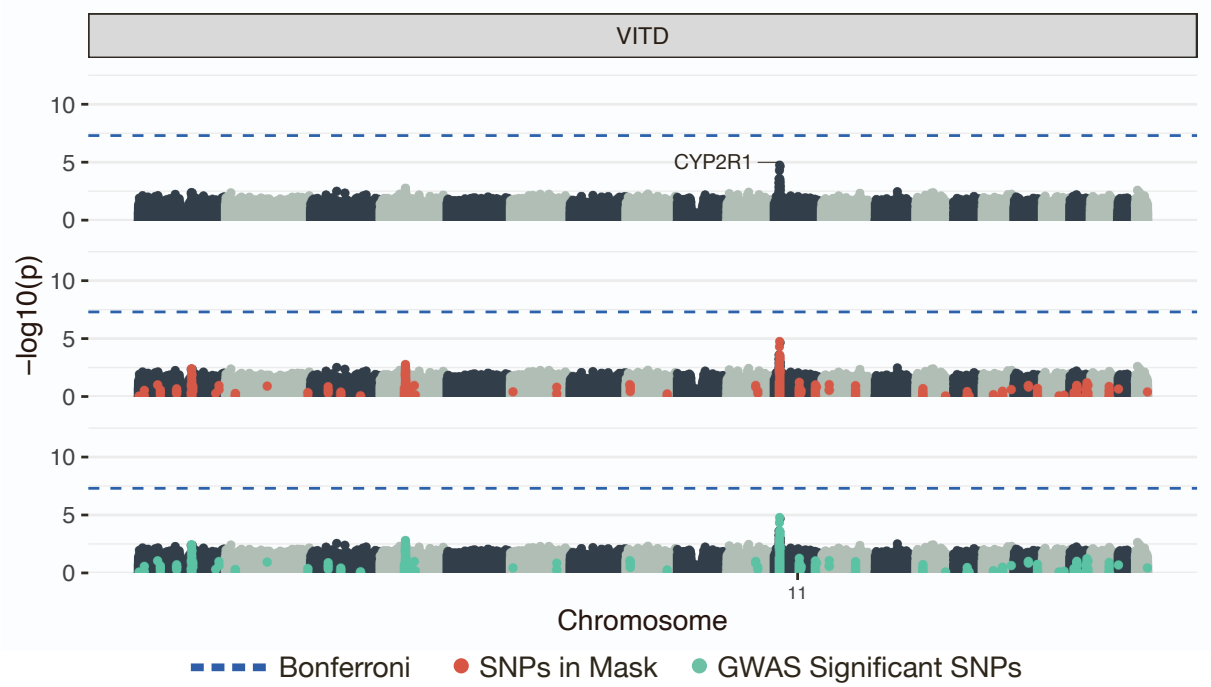


Figure S14. Manhattan plots of a genome-wide interaction analysis using SME to study vitamin D levels (VITD) assayed in individuals in the UK Biobank. As a mask in this study, we leveraged 547 significant GWAS variants. This means that, while all SNPs are tested for marginal epistasis, only their interactions with SNPs identified by GWAS are considered. Here, $-\log_{10}$ transformed P-values from SME are plotted for each SNP against their genomic positions. Chromosomes are shown in alternating colors for clarity. The dashed blue line represents the genome-wide significance threshold ($P < 5 \times 10^{-8}$). Each panel shows the same plot with different aspects of the result highlighted. The first simply shows the names of the closest neighboring genes to significant epistatic SNPs. The second panel highlights the SNPs that are modeled to interact with the focal SNP, and the third panel highlights SNPs that were also found to have a significant (additive) association with the trait according to a GWAS.

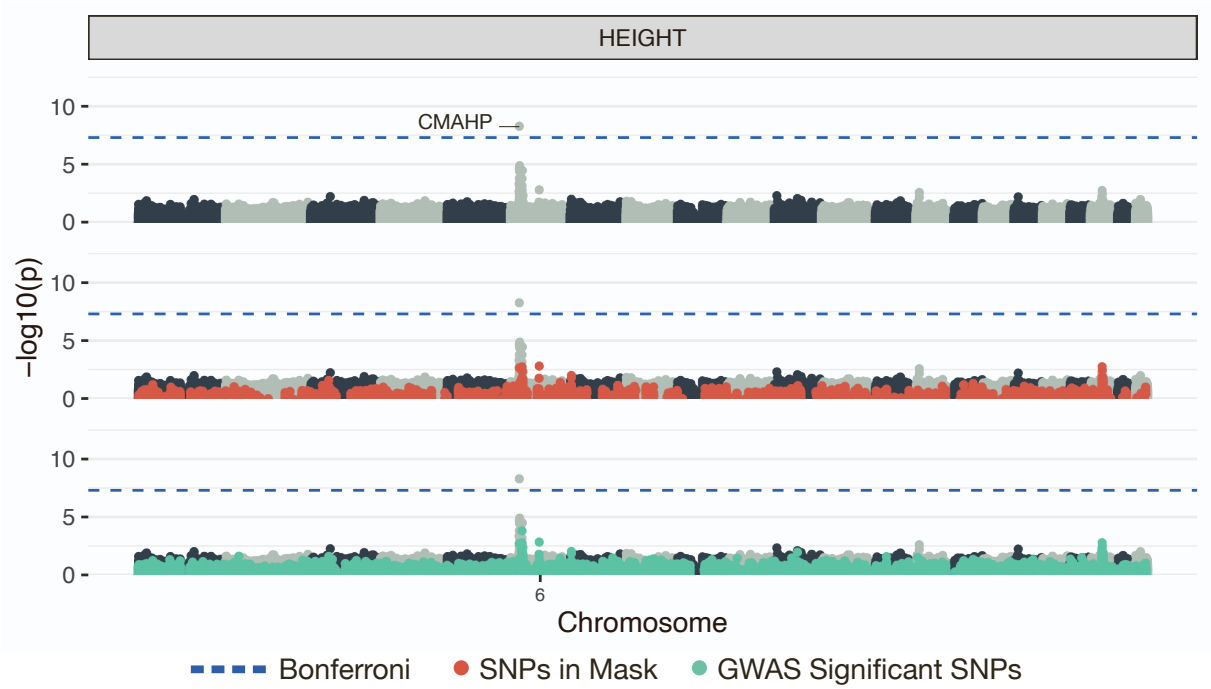


Figure S15. Manhattan plots of a genome-wide interaction analysis using SME to study body height assayed in individuals in the UK Biobank. As a mask in this study, we leveraged the top 5,000 significant GWAS variants ranked by strength of association (i.e., lowest P-values). This means that, while all SNPs are tested for marginal epistasis, only their interactions with the top SNPs identified by GWAS are considered. Here, $-\log_{10}$ transformed P-values from SME are plotted for each SNP against their genomic positions. Chromosomes are shown in alternating colors for clarity. The dashed blue line represents the genome-wide significance threshold ($P < 5 \times 10^{-8}$). Each panel shows the same plot with different aspects of the result highlighted. The first simply shows the names of the closest neighboring genes to significant epistatic SNPs. The second panel highlights the SNPs that are modeled to interact with the focal SNP, and the third panel highlights SNPs that were also found to have a significant (additive) association with the trait according to a GWAS.

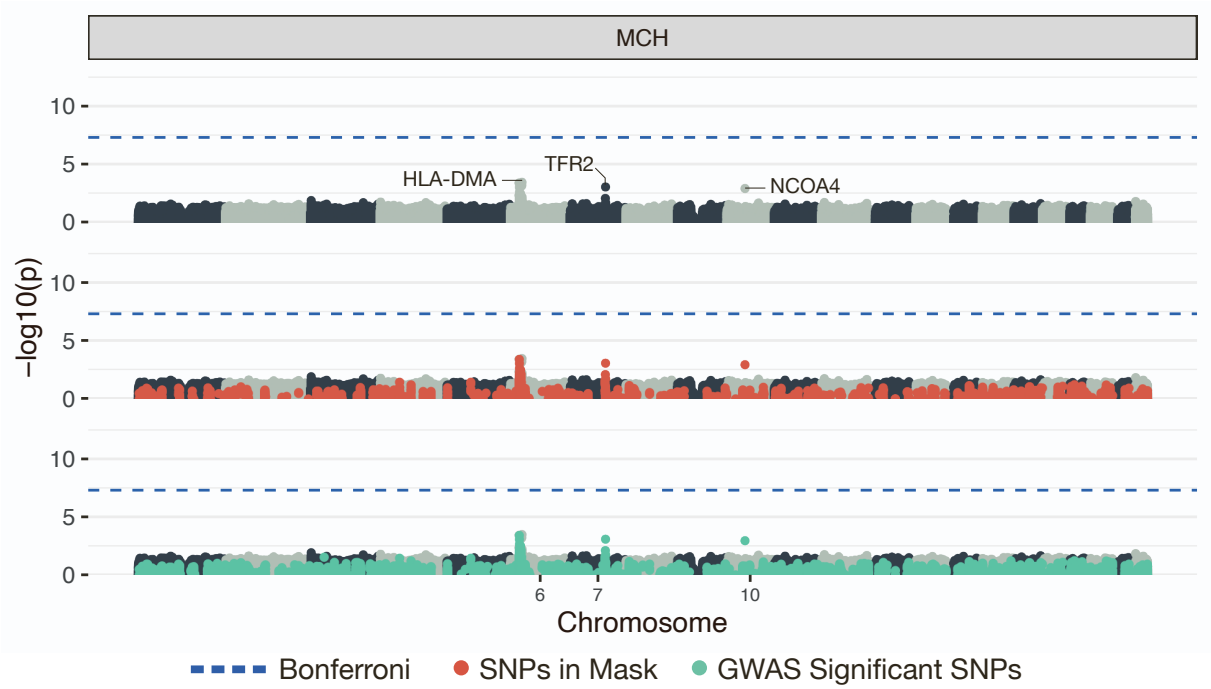


Figure S16. Manhattan plots of a genome-wide interaction analysis using SME to study mean corpuscular hemoglobin (MCH) assayed in individuals in the UK Biobank. As a mask in this study, we leveraged the top 5,000 significant GWAS variants ranked by strength of association (i.e., lowest P-values). This means that, while all SNPs are tested for marginal epistasis, only their interactions with the top SNPs identified by GWAS are considered. Here, $-\log_{10}$ transformed P-values from SME are plotted for each SNP against their genomic positions. Chromosomes are shown in alternating colors for clarity. The dashed blue line represents the genome-wide significance threshold ($P < 5 \times 10^{-8}$). Each panel shows the same plot with different aspects of the result highlighted. The first simply shows the names of the closest neighboring genes to significant epistatic SNPs. The second panel highlights the SNPs that are modeled to interact with the focal SNP, and the third panel highlights SNPs that were also found to have a significant (additive) association with the trait according to a GWAS.

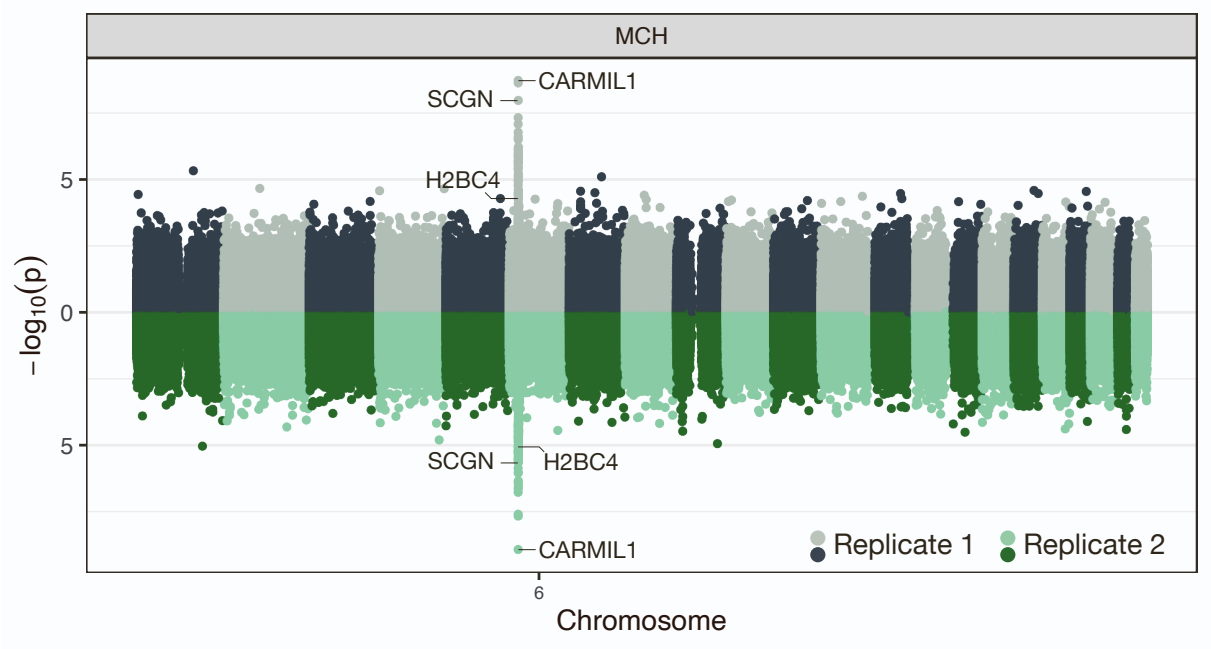


Figure S17. Manhattan plots of a genome-wide interaction split-half analysis using SME to study mean corpuscular hemoglobin (MCH) assayed in two distinct subsets of 174,705 individuals each taken from the UK Biobank. As a mask in this study, we leveraged DNase I-hypersensitive sites (DHS) data measured over 12 days of *ex vivo* erythroid differentiation[2,3]. This means that, while all SNPs are tested for marginal epistasis, only their interactions with SNPs in DHS regions are considered. Here, $-\log_{10}$ transformed P-values from SME are plotted for each SNP against their genomic positions. Chromosomes are shown in alternating colors for clarity. Variant-level results for the first replicate is shown in gray tones with a positive y-axis, while results for the second cohort is mirrored along the negative half of the y-axis and colored in green.

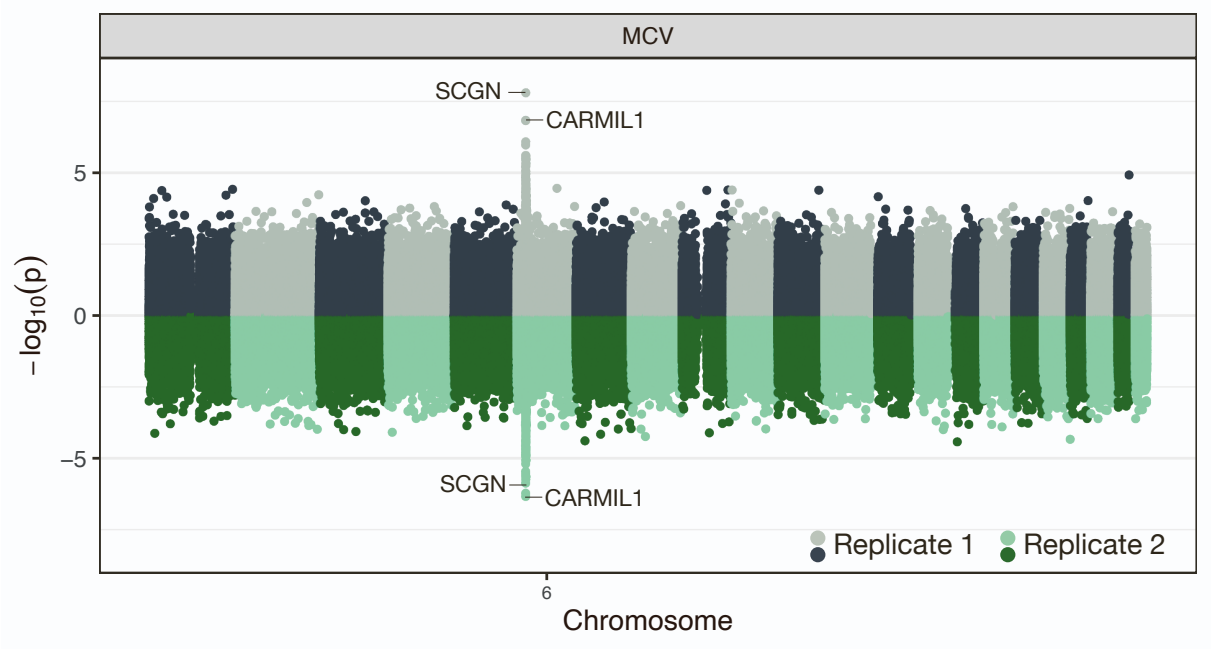


Figure S18. Manhattan plots of a genome-wide interaction split-half analysis using SME to study mean corpuscular volume (MCV) assayed in two distinct subsets of 174,705 individuals each taken from the UK Biobank. As a mask in this study, we leveraged DNase I-hypersensitive sites (DHS) data measured over 12 days of *ex vivo* erythroid differentiation[2,3]. This means that, while all SNPs are tested for marginal epistasis, only their interactions with SNPs in DHS regions are considered. Here, $-\log_{10}$ transformed P-values from SME are plotted for each SNP against their genomic positions. Chromosomes are shown in alternating colors for clarity. Variant-level results for the first replicate is shown in gray tones with a positive y-axis, while results for the second cohort is mirrored along the negative half of the y-axis and colored in green.

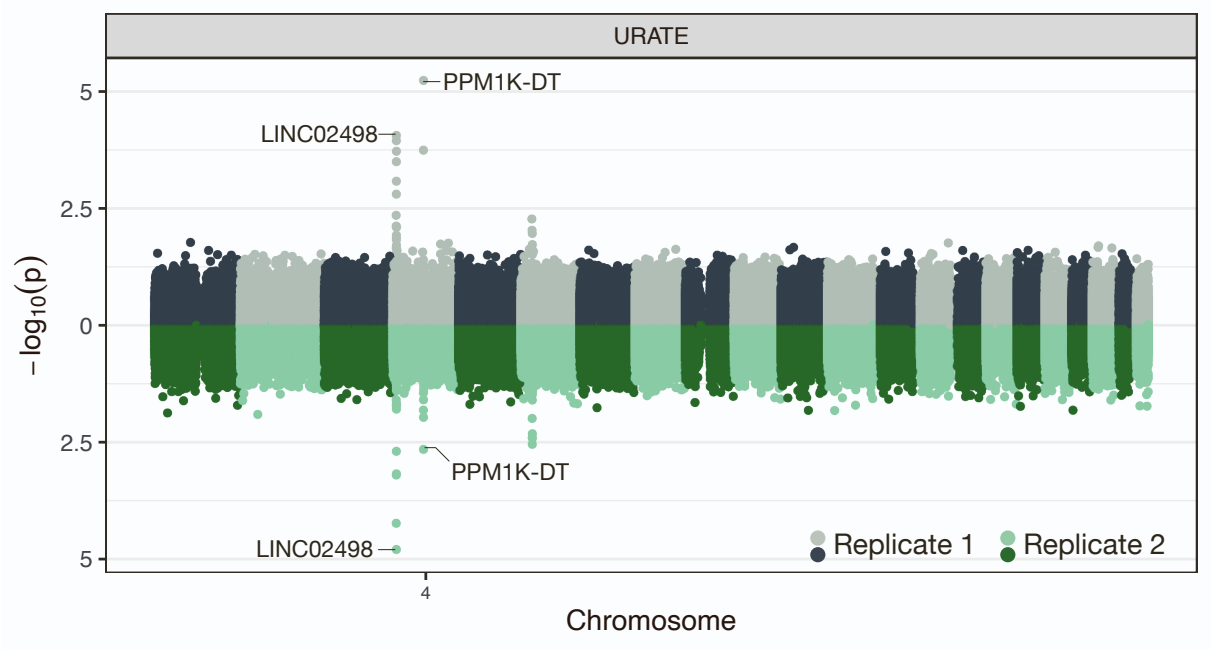


Figure S19. Manhattan plots of a genome-wide interaction split-half analysis using SME to study uric acid (urate) assayed in two distinct subsets of 174,705 individuals each taken from the UK Biobank. As a mask in this study, we leveraged GWAS summary statistics reported in the UK Biobank. This means that, while all SNPs are tested for marginal epistasis, only their interactions with SNPs associated with the trait are considered. Here, $-\log_{10}$ transformed P-values from SME are plotted for each SNP against their genomic positions. Chromosomes are shown in alternating colors for clarity. Variant-level results for the first replicate is shown in gray tones with a positive y-axis, while results for the second cohort is mirrored along the negative half of the y-axis and colored in green.

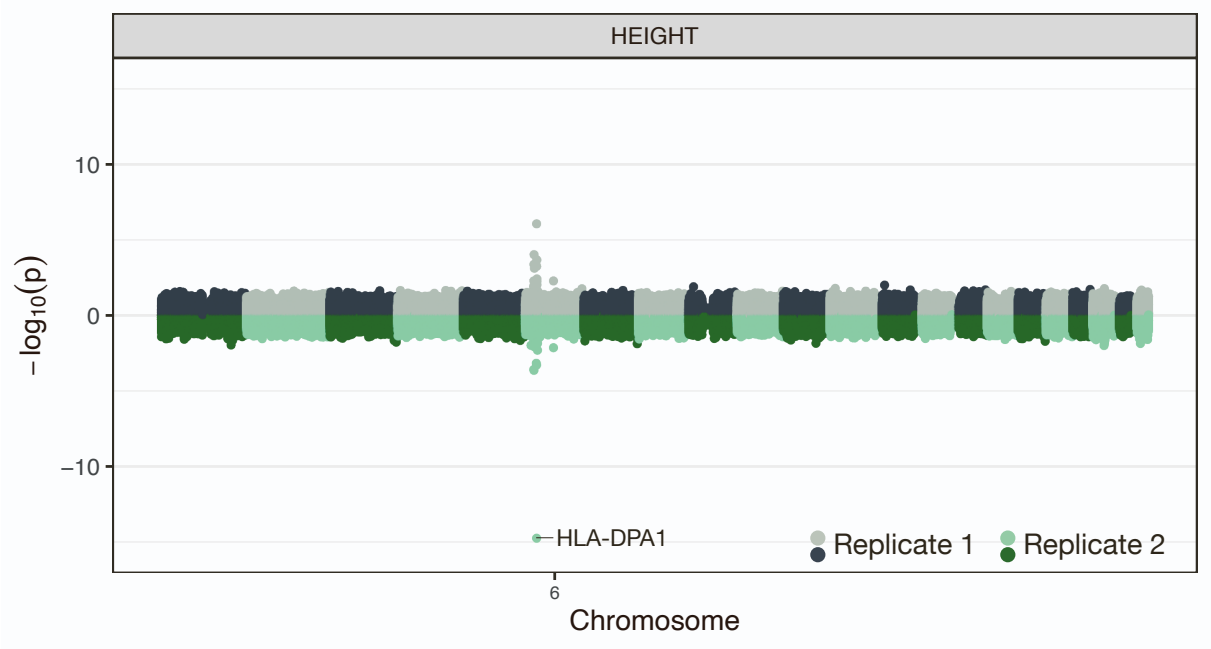


Figure S20. Manhattan plots of a genome-wide interaction split-half analysis using SME to study body height assayed in two distinct subsets of 174,705 individuals each taken from the UK Biobank. As a mask in this study, we leveraged GWAS summary statistics reported in the UK Biobank. This means that, while all SNPs are tested for marginal epistasis, only their interactions with SNPs associated with the trait are considered. Here, $-\log_{10}$ transformed P-values from SME are plotted for each SNP against their genomic positions. Chromosomes are shown in alternating colors for clarity. Variant-level results for the first replicate is shown in gray tones with a positive y-axis, while results for the second cohort is mirrored along the negative half of the y-axis and colored in green.

Table S1. While using a mask that induces uniform sparsity, SME estimates zero marginal epistasis under the null hypothesis when traits are generated by only additive effects. Synthetic traits were simulated with only additive effects using chromosome 1 from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled using sample sizes of 20k, 50k, and 100k individuals. We randomly selected 10% of all variants to be causal with additive effects and we assume that they explain 40% of the phenotypic variance for each trait. Data were analyzed using both FAME (as a baseline) and SME under varying percentages of SNPs that are masked (0%, 95%, and 99%, respectively). For reference, under the null hypothesis $H_0 : \sigma^2 = 0$. Results are based on 100 simulated traits per scenario and values in the parentheses are the standard deviations of the variance component estimates across replicates. The results in this table correspond to the data in the insets of Fig. 3 in the main text.

Method	Sample Size	$\hat{\sigma}^2$ (Standard Deviation)
FAME	20k	0.000245 (0.003912)
	50k	0.000349 (0.002053)
	100k	0.000519 (0.001491)
SME (0% masked)	20k	0.000046 (0.003872)
	50k	0.000038 (0.001934)
	100k	0.000065 (0.001120)
	300k	0.000039 (0.000504)
SME (95% masked)	20k	0.000024 (0.002355)
	50k	0.000021 (0.001064)
	100k	0.000022 (0.000571)
	300k	0.000010 (0.000211)
SME (99% masked)	20k	-0.000010 (0.001318)
	50k	0.000011 (0.000562)
	100k	0.000009 (0.000285)
	300k	0.000003 (0.000101)

Table S2. While using a mask that induces localized sparsity, SME produces deflated type I error rates when synthetic traits are generated under the null model. Synthetic traits were simulated with only additive effects using chromosome 1 from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled using sample sizes of 20k, 50k, 100k, and 300k individuals. A total of 100 causal additive variants were randomly selected for each trait and their effects were assumed to explain 40% of the phenotypic variance. Data were analyzed using SME under varying percentages of SNPs that are masked (0%, 95%, and 99%, respectively). Empirical size for the analyses used significance thresholds of $\alpha = 0.05$, 0.01, and 0.001. Values in the parentheses are the standard deviations of the estimates. Results are based on 100 simulations per scenario.

Method	Sample Size	α=0.05	α=0.01	α=0.001
SME (0% masked)	20k	0.0414 (0.0181)	0.0052 (0.0081)	0.0000 (0.0000)
	50k	0.0427 (0.0199)	0.0073 (0.0095)	0.0005 (0.0022)
	100k	0.0474 (0.0234)	0.0073 (0.0081)	0.0006 (0.0024)
	300k	0.0614 (0.0279)	0.0082 (0.0099)	0.0011 (0.0032)
SME (95% masked)	20k	0.0204 (0.0148)	0.0009 (0.0029)	0.0000 (0.0000)
	50k	0.0232 (0.0178)	0.0007 (0.0026)	0.0000 (0.0000)
	100k	0.0278 (0.0164)	0.0021 (0.0043)	0.0000 (0.0000)
	300k	0.0359 (0.0201)	0.0030 (0.0055)	0.0005 (0.0021)
SME (99% masked)	20k	0.0044 (0.0066)	0.0000 (0.0000)	0.0000 (0.0000)
	50k	0.0039 (0.0069)	0.0000 (0.0000)	0.0000 (0.0000)
	100k	0.0048 (0.0069)	0.0000 (0.0000)	0.0000 (0.0000)
	300k	0.0077 (0.0080)	0.0002 (0.0015)	0.0000 (0.0000)

Table S3. SME identifies marginal epistasis in complex traits from individuals in the UK Biobank. Here, we analyze 349,411 white British individuals in the UK Biobank genotyped at 543,813 SNPs genome-wide. Traits in this analysis included: body height, mean corpuscular hemoglobin (MCH), uric acid (urate), and vitamin D levels (VITD). As a mask, we leveraged significant trait associations from GWAS summary statistics. Listed are only results corresponding to SNPs that have marginal epistatic P-values below a genome-wide significance threshold to correct for multiple testing ($P < 5 \times 10^{-8}$). In the second and third columns, we list SNPs and their genomic location in the format `Chromosome:Basepair`. Next, we give the P-value and marginal epistatic phenotypic variance explained (PVE) for each SNP as estimated by SME. The next two columns report the P-values for SME without external data source (i.e., 0% masked) and MAPIT. The last columns detail the closest neighboring gene and a reference that have previously suggested some level of association or enrichment between each gene and the traits of interest. Due to computational resource constraints, MAPIT was applied to a random subset of 10k individuals.

Trait	ID	Coordinates	P-Value	PVE	P-Value (0% Masked)	P-Value (MAPIT)	Gene	Ref.
Urate	rs75372872	4:10835207	8.81×10^{-11}	0.00031	2.72×10^{-3}	0.22	–	–
Height	rs9467442	6:25224925	5.49×10^{-9}	0.001	1	0.07	*CMAHP*	[4]

Table S4. Complete list of marginal epistasis results from running SME on mean corpuscular hemoglobin (MCH) assayed in individuals in the UK Biobank. Here, we analyze 349,411 white British individuals in the UK Biobank genotyped at 543,813 SNPs genome-wide. As a mask, we leveraged DNase I-hypersensitive sites (DHS) data measured over 12 days of *ex vivo* erythroid differentiation[2,3].In the first three columns, we list the identifier of the SNPs, their chromosome, and their genomic location (basepair position). Next, we give the P-value and marginal epistatic phenotypic variance explained (PVE) for each SNP as estimated by SME. In the last two columns, we give the abbreviation for the trait that is analyzed and the type of the external data source used to induce sparsity. These results are saved as comma-separated value (CSV) files and can be downloaded directly from https://doi.org/10.5281/zenodo.14607997.

Table S5. Complete list of marginal epistasis results from running SME on mean corpuscular hemoglobin concentration (MCHC) assayed in individuals in the UK Biobank. Here, we analyze 349,411 white British individuals in the UK Biobank genotyped at 543,813 SNPs genome-wide. As a mask, we leveraged DNase I-hypersensitive sites (DHS) data measured over 12 days of *ex vivo* erythroid differentiation[2,3]. In the first three columns, we list the identifier of the SNPs, their chromosome, and their genomic location (basepair position). Next, we give the P-value and marginal epistatic phenotypic variance explained (PVE) for each SNP as estimated by SME. In the last two columns, we give the abbreviation for the trait that is analyzed and the type of the external data source used to induce sparsity. These results are saved as comma-separated value (CSV) files and can be downloaded directly from https://doi.org/10.5281/zenodo.14607997.

Table S6. Complete list of marginal epistasis results from running SME on mean corpuscular volume (MCV) assayed in individuals in the UK Biobank. Here, we analyze 349,411 white British individuals in the UK Biobank genotyped at 543,813 SNPs genome-wide. As a mask, we leveraged DNase I-hypersensitive sites (DHS) data measured over 12 days of *ex vivo* erythroid differentiation[2,3]. In the first three columns, we list the identifier of the SNPs, their chromosome, and their genomic location (basepair position). Next, we give the P-value and marginal epistatic phenotypic variance explained (PVE) for each SNP as estimated by SME. In the last two columns, we give the abbreviation for the trait that is analyzed and the type of the external data source used to induce sparsity. These results are saved as comma-separated value (CSV) files and can be downloaded directly from https://doi.org/10.5281/zenodo.14607997.

Table S7. Complete list of marginal epistasis results from running SME on hematocrit (HCT) assayed in individuals in the UK Biobank. Here, we analyze 349,411 white British individuals in the UK Biobank genotyped at 543,813 SNPs genome-wide. As a mask, we leveraged DNase I-hypersensitive sites (DHS) data measured over 12 days of *ex vivo* erythroid differentiation[2,3]. In the first three columns, we list the identifier of the SNPs, their chromosome, and their genomic location (basepair position). Next, we give the P-value and marginal epistatic phenotypic variance explained (PVE) for each SNP as estimated by SME. In the last two columns, we give the abbreviation for the trait that is analyzed and the type of the external data source used to induce sparsity. These results are saved as comma-separated value (CSV) files and can be downloaded directly from https://doi.org/10.5281/zenodo.14607997.

Table S8. Complete list of marginal epistasis results from running SME on body height assayed in individuals in the UK Biobank. Here, we analyze 349,411 white British individuals in the UK Biobank genotyped at 543,813 SNPs genome-wide. As a mask, we leveraged the top 5,000 significant GWAS variants ranked by strength of association (i.e., lowest P-values). In the first three columns, we list the identifier of the SNPs, their chromosome, and their genomic location (basepair position). Next, we give the P-value and marginal epistatic phenotypic variance explained (PVE) for each SNP as estimated by SME. In the last two columns, we give the abbreviation for the trait that is analyzed and the type of the external data source used to induce sparsity. These results are saved as comma-separated value (CSV) files and can be downloaded directly from https://doi.org/10.5281/zenodo.14607997.

Table S9. Complete list of marginal epistasis results from running SME on mean corpuscular hemoglobin (MCH) assayed in individuals in the UK Biobank. Here, we analyze 349,411 white British individuals in the UK Biobank genotyped at 543,813 SNPs genome-wide. As a mask, we leveraged the top 5,000 significant GWAS variants ranked by strength of association (i.e., lowest P-values). In the first three columns, we list the identifier of the SNPs, their chromosome, and their genomic location (basepair position). Next, we give the P-value and marginal epistatic phenotypic variance explained (PVE) for each SNP as estimated by SME. In the last two columns, we give the abbreviation for the trait that is analyzed and the type of the external data source used to induce sparsity. These results are saved as comma-separated value (CSV) files and can be downloaded directly from https://doi.org/10.5281/zenodo.14607997.

Table S10. Complete list of marginal epistasis results from running SME on uric acid (or urate) assayed in individuals in the UK Biobank. Here, we analyze 349,411 white British individuals in the UK Biobank genotyped at 543,813 SNPs genome-wide. As a mask, we leveraged 3,536 significant trait associations from GWAS summary statistics. In the first three columns, we list the identifier of the SNPs, their chromosome, and their genomic location (basepair position). Next, we give the P-value and marginal epistatic phenotypic variance explained (PVE) for each SNP as estimated by SME. In the last two columns, we give the abbreviation for the trait that is analyzed and the type of the external data source used to induce sparsity. These results are saved as comma-separated value (CSV) files and can be downloaded directly from https://doi.org/10.5281/zenodo.14607997.

Table S11. Complete list of marginal epistasis results from running SME on vitamin D levels (VITD) assayed in individuals in the UK Biobank. Here, we analyze 349,411 white British individuals in the UK Biobank genotyped at 543,813 SNPs genome-wide. As a mask, we leveraged 547 significant trait associations from GWAS summary statistics. In the first three columns, we list the identifier of the SNPs, their chromosome, and their genomic location (basepair position). Next, we give the P-value and marginal epistatic phenotypic variance explained (PVE) for each SNP as estimated by SME. In the last two columns, we give the abbreviation for the trait that is analyzed and the type of the external data source used to induce sparsity. These results are saved as comma-separated value (CSV) files and can be downloaded directly from https://doi.org/10.5281/zenodo.14607997.

Table S12. Assessing the robustness of SME to phenotypic scaling and masking composition for traits analyzed in the UK Biobank. We analyzed 349,411 unrelated white British individuals in the UK Biobank, genotyped at 543,813 SNPs genome-wide. For mean corpuscular hemoglobin (MCH) and mean corpuscular volume (MCV), we used DNase I-hypersensitive site (DHS) data measured over 12 days of *ex vivo* erythroid differentiation as the masking annotation[2,3]. For uric acid (urate) and body height, we used significant trait associations from GWAS summary statistics to construct the mask. The table reports only SNPs for which SME identified marginal epistatic P-values below the genome-wide significance threshold ($P < 5 \times 10^{-8}$). The second and third columns list the SNP and corresponding SME P-value leveraging external data to induce sparsity. Columns four to six show SME P-values based on quantile-normalized traits, exclusion of interactions with SNPs on the focal chromosome, and no external data (0% masked). The final two columns present P-values from FAME and MAPIT. Due to computational resource constraints, MAPIT was applied to a random subset of 10k individuals. NA: test resulted in a missing variance estimate for the marginal epistatic variance component.

Trait	ID	P-Value	P-Value (QQ-Norm)	P-Value (Excl. Chr.)	P-Value (0% Masked)	P-Value (FAME)	P-Value (MAPIT)
MCH	rs4711092	1.41×10^{-11}	8.20×10^{-12}	0.78	4.78×10^{-4}	5.37×10^{-6}	0.47
MCH	rs9366624	1.80×10^{-9}	7.37×10^{-9}	1.21×10^{-3}	5.75×10^{-7}	7.16×10^{-7}	9.37×10^{-3}
MCH	rs9461167	2.34×10^{-9}	3.37×10^{-9}	0.73	1.40×10^{-5}	5.73×10^{-7}	3.05×10^{-3}
MCH	rs9379764	5.53×10^{-9}	1.02×10^{-8}	0.62	4.58×10^{-6}	4.53×10^{-6}	0.21
MCH	rs441460	1.20×10^{-8}	9.18×10^{-9}	0.52	3.07×10^{-6}	5.67×10^{-6}	0.36
MCH	rs198834	2.77×10^{-8}	1.36×10^{-7}	0.17	2.17×10^{-6}	1.49×10^{-7}	1.83×10^{-3}
MCH	rs13203202	3.17×10^{-8}	6.31×10^{-8}	3.92×10^{-2}	1.90×10^{-4}	9.02×10^{-5}	0.13
MCV	rs9276	9.09×10^{-10}	1	2.94×10^{-2}	0.73	NA	0.18
MCV	rs9366624	1.86×10^{-8}	1.27×10^{-8}	4.06×10^{-3}	2.63×10^{-7}	8.64×10^{-7}	0.16
Urate	rs75372872	8.81×10^{-11}	3.82×10^{-11}	0.93	2.72×10^{-3}	1.55×10^{-3}	0.22
Height	rs9467442	5.49×10^{-9}	NA	0.65	1	1	0.07

References

1. Fu, B., Pazokitoroudi, A., Xue, A., Anand, A., Anand, P., Zaitlen, N. & Sankararaman, S. *A biobank-scale test of marginal epistasis reveals genome-wide signals of polygenic epistasis* en. 2023. https://www.biorxiv.org/content/10.1101/2023.09.10.557084v1 (2024).
2. Georgolopoulos, G., Psatha, N., Iwata, M., Nishida, A., Som, T., Yiangou, M., Stamatoyannopoulos, J. A. & Vierstra, J. (2021). Discrete regulatory modules instruct hematopoietic lineage commitment and differentiation. *Nature Communications* **12,** 6790. https://www.nature.com/articles/s41467-021-27159-x.
3. Maurano, M. T. *et al.* (2012). Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **337,** 1190–1195. https://www.science.org/doi/10.1126/science.1222794.
4. Yengo, L. *et al.* (2022). A saturated map of common genetic variants associated with human height. *Nature* **610,** 704–712.