

Statistical and Computational Advances for Detecting Nonlinear Contributions to Complex Traits

By

Julian Stamp

Dissertation
submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy of Computer Science & Computational Biology

Center for Computational Molecular Biology
Brown University
Providence, Rhode Island
February 2026

© Copyright 2026 Julian Stamp

This dissertation by Julian Stamp is accepted in its present form by the Center for Computational Molecular Biology as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

Date _____
Lorin Crawford, Co-Advisor

Date _____
Daniel Weinreich, Co-Advisor

Recommended to the Graduate Council

Date _____
Roberta De Vito, Reader

Date _____
Bogdan Pasaniuc, Reader

Approved by the Graduate Council

Date _____
Janet A. Blume, Interim Dean of the Graduate School

Curriculum Vitae

Education

2021–2025

Ph.D., Computer Science & Computational Biology

M.Sc., Computational Biology

Brown University (Providence, RI, USA)

2016–2019

M.Sc., Physics

Ludwig Maximilian University (Munich, BY, Germany)

2012–2016

B.Sc., Physics

University of Konstanz (Konstanz, BW, Germany)

Publications

Stamp, J., Pattillo Smith, S., Weinreich, D., Crawford, L. (2025). Sparse modeling of interactions enables fast detection of genome-wide epistasis in biobank-scale studies. *AJHG*

Balvert, M., Cooper-Knock, J., **Stamp, J.** et al. Considerations in the search for epistasis. *Genome Biol* 25, 296 (2024).

Li, K., Chaguza, C., **Stamp, J.** et al. (2024). Genome-wide association study between SARS-CoV-2 single nucleotide polymorphisms and virus copies during infections. *PLOS Computational Biology*, 20(9), e1012469.

Kim, I. E., Jr., Oduor, C., **Stamp, J.**, Luftig, M. A., Moormann, A. M., Crawford, L., Bailey, J. (2024). Incorporation of Epstein-Barr viral variation implicates significance of LMP1 in survival prediction and prognostic subgrouping in Burkitt lymphoma. *Int. Journal of Cancer*

Pattillo Smith S., Darnell G., Udwin, D., **Stamp, J.**, Harpak, A., Ramachandran S., Crawford L. (2024). Discovering nonadditive heritability using additive GWAS summary statistics. *eLife*

Stamp, J., DenAdel A., Weinreich D., Crawford L. (2023). Leveraging the Genetic Correlation between Traits Improves the Detection of Epistasis in Genome-wide Association Studies. *G3 Genes/Genomes/Genetics*

Edeleva, E., Salditt, A., **Stamp, J.**, Schwintek, P., Boekhoven, J., Braun, D. (2019). Continuous nonenzymatic cross-replication of DNA strands with in-situ activated DNA oligonucleotides. *Chemical Science*

Software

Stamp, J., Crawford L. (2025). smer: Sparse Marginal Epistasis Test. <https://github.com/lcrawlab/sme> (R Package)

Stamp, J., Crawford L. (2025). cphgxe: Cox Proportional Hazards Gene-by-Environment Interaction Test. <https://github.com/lcrawlab/cphgxe> (R Package)

Stamp, J., Crawford L. (2023). mvMAPIT: Multivariate Genome-Wide Marginal Epistasis Test. <https://github.com/lcrawlab/mvMAPIT> (R Package)

Professional Development

2019 - 2020

Software Consultant

TNG Technology Consulting GmbH (Munich, BY, Germany)

Conference Talks

Stamp, J., Pattillo Smith, S., Weinreich, D., Crawford, L. (2025) Sparse modeling of interactions enables fast detection of genome-wide epistasis in biobank-scale studies., *Intelligent Systems for Molecular Biology* (Liverpool, UK)

Stamp, J., DenAdel, A., Weinreich, D., Crawford, L. (2023) Leveraging the Genetic Correlation between Traits Improves the Detection of Epistasis in Genome-wide Association Studies., *Intelligent Systems for Molecular Biology* (Lyon, France)

Stamp, J., Crawford, L. (2023) Partitioning the nonadditive variation of complex traits. *Lorentz Center, Leiden University* (Leiden, Netherlands)

Fellowships

2021-2022, Vartan Gregorian Fellowship

Brown University (Providence, RI, USA)

2018, Erasmus Fellowship

Ludwig Maximilian University (Munich, BY, Germany)

2016, PROMOS Fellowship

University of Konstanz (Konstanz, BW, Germany)

Acknowledgments

The completion of this thesis that marks the end of my dissertation research would not have been possible without the support, guidance and encouragement of my advisors, collaborators and friends. I began my graduate studies under unusual circumstances. I was admitted to the PhD program only weeks before the COVID-19 pandemic brought international travel to a halt. From the very beginning, I learned that the faculty and staff at the Center for Computational Molecular Biology (CCMB) at Brown University have created an exceptional environment for graduate students. No effort was spared to make sure that I had a smooth transition into the program.

It was my great privilege and pleasure to have been afforded the opportunity to conduct my research at the CCMB under the guidance of my advisors, Lorin Crawford and Daniel Weinreich. Both Lorin and Dan have been a constant source of support throughout my time at Brown. Their mentorship and guidance have laid the foundation for my research and career. In particular, I am immensely grateful to Lorin for his unwavering belief in my abilities and his encouragement even when I doubted myself. Lorin's academic rigor, expertise in statistics, and writing skills have inspired me to strive for excellence in my

own research and writing. Dan's outstanding teaching skills and profound expertise in theoretical biology inspired me to engage more deeply with the biology I aimed to model.

I want to thank my committee members, Lorin Crawford, Daniel Weinreich, Roberta De Vito, and Bogdan Pasaniuc for their insightful feedback and suggestions that guided the development of this dissertation. Beyond their support in my academic journey, I am deeply grateful for their wisdom in helping me think about career opportunities and for encouraging an approach to research that values both academic growth and personal well-being.

Looking back on my time at the Brown University, I am filled with gratitude for the many people who were part of this community. I want to thank my friends and colleagues who walked this path alongside me. Thank you Chibu, Ria, and Dave for the many hours we spent discussing research, life, and everything in between. Thank you to the current and former members of the Crawford lab, Cecile, Alex, Shevaughn, Emily, Wei-Shing, Alan, Whitney, Dana, Ashley for all the immediate feedback on my research and the breadth of research interests and depth of knowledge you brought to our lab meetings that let me peer into many corners of the field of computational biology. Thank you to my cohort, Vivek, Cole, Hyeyeon, Mayra, for the camaraderie and shared dedication throughout the program. Working together to meet expectations, take classes, and prepare for the qualifying exam made the entire journey rewarding and memorable.

Thinking about pivotal moments that led me to where I am today, I am profoundly grateful for my dear friend Andrew, who sparked my interest in research and introduced me to my first research mentor, Hoi-Chung Leung. Andrew and Hoi-Chung provided me with perspectives and opportunities that I was unaware of. Without them, I may not have pursued a career in research.

I am particularly grateful to my partner, Bianca. Her presence is a source of peace and stability that helped me find the stamina and focus to bring this dissertation to

completion. She has remained patient and supportive when my research consumed all of my time and energy and reminded me to take breaks and enjoy life with her. I am also very grateful to Helmut and Keiko, who have been incredibly generous and kind, inviting me into their home and family, enabling me to pursue research at Columbia University and sparing no effort to support me through transitions in my career.

Lastly, I want to express my deepest gratitude to the bedrock that has supported me throughout my life: my family. My parents endowed me with the courage and resilience to pursue my ambitions and have always been there to offer advice and encouragement. My brother has always been a source of inspiration and motivation that I look up to, encouraging me to think critically and creatively about the world around me while extending kindness and generosity to others. I dedicate this thesis to my parents Dagmar and Wolfgang and my brother Fabian.

The text of Chapter 1 of this dissertation is an adaptation of the material as it appears in Stamp et al. (2023). *Leveraging the Genetic Correlation between Traits Improves the Detection of Epistasis in Genome-wide Association Studies*. G3 Genes|Genomes|Genetics under the terms of the Creative Commons Attribution (CC BY) licence. <https://creativecommons.org/licenses/by/4.0/>

The text of Chapter 3 of this dissertation is an adaptation of the material as it appears in Stamp et al. (2025). *Sparse modeling of interactions enables fast detection of genome-wide epistasis in biobank-scale studies*. The American Journal of Human Genetics under the terms of the Creative Commons Attribution (CC BY NC ND) licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Contents

Acknowledgments	vii
List of figures	xvi
List of tables	xxiv
Introduction	1
1 Leveraging the genetic correlation between traits improves the detection of epistasis in genome-wide association studies	7
1.1 Introduction	7
1.2 Materials and methods	11
1.2.1 The marginal epistasis test for single traits	11
1.2.2 Derivation of the multivariate marginal epistasis test	16
1.2.3 Hypothesis testing in the mvMAPIT framework	17
1.2.4 Simulation study design	21
1.2.5 Preprocessing of the heterogeneous stock of mice data set	24

1.3	Results	26
1.3.1	mvMAPIT produces calibrated P -values and conservative type I error rates	26
1.3.2	Improved detection of epistatic variants using mvMAPIT in simulations	27
1.3.3	Synergistic epistasis in binding affinity landscapes for neutralizing antibodies	34
1.3.4	Joint modeling of hematology traits yields epistatic signal in stock of mice	36
1.4	Discussion	41
1.5	URLs	44
1.5.1	Data and software availability	44
2	Frailty model approach for detecting gene-by-environment interactions in time-to-event association analysis	46
2.1	Introduction	46
2.2	Methods	48
2.2.1	The Cox proportional hazards gene-by-environment interaction test	48
2.2.2	The likelihood function for censored time-to-event data	49
2.2.3	Auxiliary Poisson GLMM for time-to-event data	52
2.2.4	Fitting the null model	53
2.2.5	Hypothesis testing using the efficient score statistic	55
2.2.6	Simulating time-to-event data under the frailty model	56
2.2.7	Stabilizing measures in the iterative null fit and computational improvements	58
2.2.8	Preprocessing the UK Biobank	58
2.3	Results	59

2.3.1	CphGxE produces asymptotically calibrated P -values at high censoring rates	60
2.3.2	CphGxE estimates of heritable frailty variance and fixed effects are asymptotically unbiased	62
2.3.3	CphGxE is well powered and correctly prioritizes gene-by-environment interaction variants	63
2.3.4	The iterative null fit in CphGxE is sensitive to censoring rates and initial parameter values	65
2.3.5	Caching of the null fit improves the runtime of the $G \times E$ test . . .	67
2.4	Discussion	68
2.5	Data and software availability	69
2.6	Web resources	69

3 Sparse modeling of interactions enables fast detection of genome-wide epistasis in biobank-scale studies 70

3.1	Introduction	71
3.2	Material and methods	73
3.2.1	The sparse marginal epistasis test	73
3.2.2	Variance component model formulation	76
3.2.3	Point estimates and hypothesis testing	77
3.2.4	Scalable computation via stochastic MoM	78
3.2.5	Preprocessing the UK Biobank	80
3.2.6	GWAS summary statistics	81
3.2.7	Generating masks from external data sources	82
3.2.8	Simulation studies	83
3.2.9	Software tools and data sources	85
3.3	Results	85
3.3.1	SME scales to biobank genome-wide association studies	85

3.3.2	SME is a well-calibrated test and conserves type I error rates . . .	87
3.3.3	The masking strategy in SME leads to improved power in simulations	91
3.3.4	SME uses chromatin information to identify epistasis in hematology traits	94
3.3.5	Conditioning SME on GWAS variants reveals epistasis in other complex traits	96
3.3.6	Findings with SME are robust to sample composition and pheno- typic scaling	98
3.4	Discussion	99
3.5	Data and software availability	101
3.6	Web resources	101
3.7	Acknowledgments	102
	Conclusion	103
A	Supplementary information for Chapter 1	109
A.1	Supplementary figures	110
A.2	Supplementary tables	147
B	Supplementary information for Chapter 2	157
B.1	The survival and hazard functions	158
B.2	The Breslow estimator	158
B.3	Approximation of Gaussian integral using Laplace’s method	159
B.4	Derivation of the penalized log-likelihood	160
B.5	The permutation test ensures well-calibrated test statistics for small sample sizes	161
B.6	Supplementary figures	163
B.7	Supplementary tables	172
C	Supplementary information for Chapter 3	175

C.1	Supplementary figures	176
C.2	Supplementary tables	196
	Bibliography	204

List of figures

1.1	Schematic of the types of shared interactions modeled by the multivariate marginal epistasis test.	22
1.2	The mvMAPIT framework using Fisher's method produces well-calibrated P -values when traits are generated by only additive effects.	28
1.3	Empirical power of mvMAPIT with Fisher's method to detect epistatic variants across complex traits with moderate broad-sense heritability. . .	30
1.4	Receiver operating characteristic (ROC) curves comparing the ability of mvMAPIT with Fisher's method to the univariate MAPIT model in detecting epistatic variants across complex traits.	32
1.5	Applying mvMAPIT with Fisher's method to broadly neutralizing antibodies recovers heavy-chain mutations known to be involved in epistatic interactions that affect binding against two influenza strains.	35
1.6	Manhattan plots of a genome-wide interaction study for two pairs of hematology traits in the heterogeneous stock of mice data set using mvMAPIT with Fisher's method.	37

2.1	QQ-plots for CphGxE under the null model.	60
2.2	Empirical power as function of per SNP $G \times E$ frailty variance explained.	64
2.3	ROC curves for CphGxE under the alternative model as function of frailty variance explained.	66
3.1	Schematic overview of the sparse marginal epistasis (SME) test.	74
3.2	Computational time for running SME and other marginal epistatic approaches on biobank-scale data as a function of the genome size.	86
3.3	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.	89
3.4	Uniform sparse modeling of interactions enhances the empirical power of SME.	93
3.5	Manhattan plots of a genome-wide interaction analysis using SME to study mean corpuscular hemoglobin (MCH) assayed in individuals in the UK Biobank.	95
A.1	The mvMAPIT framework using Fisher’s method produces well-calibrated P -values when traits are generated by only additive effects (sample size $N = 1,000$ individuals).	110
A.2	The mvMAPIT framework using Fisher’s method produces well-calibrated P -values when traits are generated by only additive effects (sample size $N = 1,750$ individuals).	111
A.3	The mvMAPIT framework using the harmonic mean produces well-calibrated P -values when traits are generated by only additive effects (sample size $N = 1,000$ individuals).	112
A.4	The mvMAPIT framework using the harmonic mean produces well-calibrated P -values when traits are generated by only additive effects (sample size $N = 1,750$ individuals).	113

A.5	The mvMAPIT framework using the harmonic mean produces well-calibrated P -values when traits are generated by only additive effects (sample size $N = 2,500$ individuals).	114
A.6	The mvMAPIT framework using the Cauchy combined test produces well-calibrated P -values when traits are generated by only additive effects (sample size $N = 1,000$ individuals).	115
A.7	The mvMAPIT framework using the Cauchy combined test produces well-calibrated P -values when traits are generated by only additive effects (sample size $N = 1,750$ individuals).	116
A.8	The mvMAPIT framework using the Cauchy combined test produces well-calibrated P -values when traits are generated by only additive effects (sample size $N = 2,500$ individuals).	117
A.9	Empirical power of mvMAPIT with Fisher's method to detect group #1 (10) and group #2 (20) epistatic variants across complex traits with moderate broad-sense heritability.	118
A.10	Empirical power of mvMAPIT with Fisher's method to detect group #1 (10) and group #2 (10) epistatic variants across complex traits with different levels of broad-sense heritability.	119
A.11	Empirical power of mvMAPIT with Fisher's method to detect group #1 (10) and group #2 (20) epistatic variants across complex traits with different levels of broad-sense heritability.	120
A.12	Empirical power of mvMAPIT with the harmonic mean combination approach to detect group #1 (10) and group #2 (10) epistatic variants across complex traits with moderate broad-sense heritability.	121
A.13	Empirical power of mvMAPIT with the harmonic mean combination approach to detect group #1 (10) and group #2 (20) epistatic variants across complex traits with moderate broad-sense heritability.	122

A.14 Empirical power of mvMAPIT with the harmonic mean combination approach to detect group #1 (10) and group #2 (10) epistatic variants across complex traits with different levels of broad-sense heritability.	123
A.15 Empirical power of mvMAPIT with the harmonic mean combination approach to detect group #1 (10) and group #2 (20) epistatic variants across complex traits with different levels of broad-sense heritability.	124
A.16 Empirical power of mvMAPIT with the Cauchy combination test approach to detect group #1 (10) and group #2 (10) epistatic variants across complex traits with moderate broad-sense heritability.	125
A.17 Empirical power of mvMAPIT with the Cauchy combination test approach to detect group #1 (10) and group #2 (20) epistatic variants across complex traits with moderate broad-sense heritability.	126
A.18 Empirical power of mvMAPIT with the Cauchy combination test approach to detect group #1 (10) and group #2 (10) epistatic variants across complex traits with different levels of broad-sense heritability.	127
A.19 Empirical power of mvMAPIT with the Cauchy combination test approach to detect group #1 (10) and group #2 (20) epistatic variants across complex traits with different levels of broad-sense heritability.	128
A.20 Receiver operating characteristic (ROC) curves comparing the ability of mvMAPIT using the harmonic mean to the univariate MAPIT model in detecting group #1 (10) and group #2 (10) epistatic variants across complex traits.	129
A.21 Receiver operating characteristic (ROC) curves comparing the ability of mvMAPIT using the Cauchy combination test to the univariate MAPIT model in detecting group #1 (10) and group #2 (10) epistatic variants across complex traits.	130

A.22 Receiver operating characteristic (ROC) curves comparing the ability of mvMAPIT with Fisher’s method to the univariate MAPIT model in detecting group #1 (10) and group #2 (20) epistatic variants across complex traits.	131
A.23 Receiver operating characteristic (ROC) curves comparing the ability of mvMAPIT using the harmonic mean to the univariate MAPIT model in detecting group #1 (10) and group #2 (20) epistatic variants across complex traits.	132
A.24 Receiver operating characteristic (ROC) curves comparing the ability of mvMAPIT using the Cauchy combination test to the univariate MAPIT model in detecting group #1 (10) and group #2 (20) epistatic variants across complex traits.	133
A.25 Applying mvMAPIT with the harmonic mean to broadly neutralizing antibodies recovers heavy-chain mutations known to be involved in epistatic interactions that affect binding against two influenza strains.	134
A.26 Applying mvMAPIT with the Cauchy combination test to broadly neutralizing antibodies recovers heavy-chain mutations known to be involved in epistatic interactions that affect binding against two influenza strains. . .	135
A.27 Manhattan plot of genome-wide interaction study for all trait pairs in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics using mvMAPIT with Fisher’s method.	136
A.28 Manhattan plot of genome-wide interaction study for all trait pairs in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics using mvMAPIT with the harmonic mean.	137
A.29 Manhattan plot of genome-wide interaction study for all trait pairs in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics using mvMAPIT with the Cauchy combination test. . .	138

A.30	Manhattan plots of a genome-wide interaction study for two pairs of hematology traits in the heterogeneous stock of mice data set using mvMAPIT with the harmonic mean.	139
A.31	Manhattan plots of a genome-wide interaction study for two pairs of hematology traits in the heterogeneous stock of mice data set using mvMAPIT with the Cauchy combination test.	140
A.32	Manhattan plots of a genome-wide interaction study for two pairs of hematology traits in the heterogeneous stock of mice data set using mvMAPIT with Fisher’s method.	141
A.33	Manhattan plots of a genome-wide interaction study for two pairs of hematology traits in the heterogeneous stock of mice data set using mvMAPIT with the harmonic mean.	142
A.34	Manhattan plots of a genome-wide interaction study for two pairs of hematology traits in the heterogeneous stock of mice data set using mvMAPIT with the Cauchy combination test.	143
A.35	Empirical correlations for all trait pairs in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics.	144
A.36	Manhattan and QQ-plots of a genome-wide interaction study combining all hematology traits using mvMAPIT.	145
A.37	Manhattan and QQ-plots of a genome-wide interaction study combining many hematology traits using mvMAPIT.	146
B.1	Frailty variance component estimates due to genetic relatedness are asymptotically unbiased.	163
B.2	Frailty variance component estimates due to environmental exposures are unbiased.	164
B.3	Frailty covariate fixed effect estimates are unbiased.	165
B.4	Empirical power of CphGxE as function of the censoring rate.	166

B.5	Empirical power of CphGxE as function of the censoring rate.	167
B.6	ROC curves for CphGxE under the alternative model.	168
B.7	Null fit convergence is sensitive to initial parameter values and censoring rates.	169
B.8	Runtime of the null fit in CphGxE as function of sample size.	170
B.9	Runtime of the G×E test in CphGxE as function of sample size.	171
C.2	Computational time needed to run a single test with SME and FAME as a function of sample size and total number of SNPs.	176
C.3	Computational time needed to run a single test with SME and FAME as a function of the number of random vectors and number of SNPs sharing random vectors.	177
C.4	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.	178
C.5	Schematic overview illustrating weight matrices with uniform and localized sparsity.	179
C.6	While using a mask that induces localized sparsity, SME is negatively biased under the null hypothesis.	180
C.7	Consequence of having a misspecified mask on power and variance component estimates in SME.	181
C.8	SME exhibits less power when using a mask that induces localized sparsity.	182
C.9	Adding random noise to external data sources that induce localized sparsity recovers empirical power and reduces the negative bias in variance component estimates.	183
C.10	Manhattan plots of a genome-wide interaction analysis using SME to study mean corpuscular volume (MCV) assayed in individuals in the UK Biobank.	184

C.11	Manhattan plots of a genome-wide interaction analysis using SME to study mean corpuscular hemoglobin concentration (MCHC) assayed in individuals in the UK Biobank.	185
C.12	Manhattan plots of a genome-wide interaction analysis using SME to study hematocrit (HCT) assayed in individuals in the UK Biobank.	186
C.13	Manhattan plots of a genome-wide interaction analysis using SME to study uric acid (urate) assayed in individuals in the UK Biobank.	187
C.14	Manhattan plots of a genome-wide interaction analysis using SME to study vitamin D levels (VITD) assayed in individuals in the UK Biobank. . . .	188
C.15	Manhattan plots of a genome-wide interaction analysis using SME to study body height assayed in individuals in the UK Biobank.	189
C.16	Manhattan plots of a genome-wide interaction analysis using SME to study mean corpuscular hemoglobin (MCH) assayed in individuals in the UK Biobank.	190
C.17	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.	191
C.18	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.	192
C.19	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.	193
C.20	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.	194
C.1	Schematic overview illustrating the approximation to the stochastic trace estimator used in SME.	195

List of tables

1.1	The mvMAPIT framework using Fisher’s method preserves type I error rates under the null model when traits are generated by only additive effects (sample size $N = 2,500$ individuals).	29
1.2	Notable SNPs with marginal epistatic effects after applying the mvMAPIT framework to 15 hematology traits in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics.	40
2.1	Empirical type I error rates for CphGxE with 90% censoring.	62
3.1	While using a mask that induces uniform sparsity, SME controls type I error rates when synthetic traits are generated under the null model. . .	90
3.2	SME identifies marginal epistasis in hematology traits from individuals in the UK Biobank.	97
A.1	The mvMAPIT framework using Fisher’s method preserves type I error rates under the null model when traits are generated by only additive effects (sample size $N = 1,000$ individuals).	147

A.2	The mvMAPIT framework using Fisher’s method preserves type I error rates under the null model when traits are generated by only additive effects (sample size $N = 1,750$ individuals).	148
A.3	The mvMAPIT framework using the harmonic mean preserves type I error rates under the null model when traits are generated by only additive effects (sample size $N = 1,000$ individuals).	149
A.4	The mvMAPIT framework using the harmonic mean preserves type I error rates under the null model when traits are generated by only additive effects (sample size $N = 1,750$ individuals).	150
A.5	The mvMAPIT framework using the harmonic mean preserves type I error rates under the null model when traits are generated by only additive effects (sample size $N = 2,500$ individuals).	151
A.6	The mvMAPIT framework using the Cauchy combination test preserves type I error rates under the null model when traits are generated by only additive effects (sample size $N = 1,000$ individuals).	152
A.7	The mvMAPIT framework using the Cauchy combination test preserves type I error rates under the null model when traits are generated by only additive effects (sample size $N = 1,750$ individuals).	153
A.8	The mvMAPIT framework using the Cauchy combination test preserves type I error rates under the null model when traits are generated by only additive effects (sample size $N = 2,500$ individuals).	154
A.9	Complete summary of the marginal epistatic results after applying the mvMAPIT framework to protein sequence data from a nearly combinatorially complete library of two broadly neutralizing anti-influenza antibodies.	155
A.10	Complete summary of the marginal epistatic results after applying the mvMAPIT framework to 15 hematology traits in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics. .	155
A.11	Notable SNPs with marginal epistatic effects (Part 2).	156

B.1	Empirical type I error rates for CphGxE with 50% censoring.	172
B.2	Empirical type I error rates for CphGxE with 10% censoring.	173
B.3	Area under the ROC curve for 20k samples.	174
C.1	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.	196
C.2	While using a mask that induces localized sparsity, SME produces deflated type I error rates when synthetic traits are generated under the null model.	197
C.3	SME identifies marginal epistasis in complex traits from individuals in the UK Biobank.	198
C.4	Complete list of marginal epistasis results from running SME on mean corpuscular hemoglobin (MCH) assayed in individuals in the UK Biobank.	199
C.5	Complete list of marginal epistasis results from running SME on mean corpuscular hemoglobin concentration (MCHC) assayed in individuals in the UK Biobank.	199
C.6	Complete list of marginal epistasis results from running SME on mean corpuscular volume (MCV) assayed in individuals in the UK Biobank. . .	199
C.7	Complete list of marginal epistasis results from running SME on hematocrit (HCT) assayed in individuals in the UK Biobank.	200
C.8	Complete list of marginal epistasis results from running SME on body height assayed in individuals in the UK Biobank.	200
C.9	Complete list of marginal epistasis results from running SME on mean corpuscular hemoglobin (MCH) assayed in individuals in the UK Biobank.	200
C.10	Complete list of marginal epistasis results from running SME on uric acid (or urate) assayed in individuals in the UK Biobank.	201
C.11	Complete list of marginal epistasis results from running SME on vitamin D levels (VITD) assayed in individuals in the UK Biobank.	201

C.12 Assessing the robustness of SME to phenotypic scaling and masking composition for traits analyzed in the UK Biobank.	202
C.13 SME identifies marginal epistasis in hematology traits from individuals in the UK Biobank. (Part 2)	203

Introduction

In recent decades, the field of statistical genetics has developed rapidly (Visscher et al. 2012; Visscher et al. 2017; Abdellaoui et al. 2023), driven by factors such as increases in high performance computing resources (Robinson et al. 2021; Castrignanò et al. 2020), advances in sequencing technologies (Jonghe et al. 2024; Warburton et al. 2023), and global efforts of curating large biobanks (Bycroft et al. 2018; Nagai et al. 2017; Kurki et al. 2023; All of Us Research Program Investigators et al. 2019). With the growing scale and availability of data, new directions of research emerged — ranging from estimating phenotypic heritability and discovering single-variant associations (Yang et al. 2010) to studying polygenic architectures (Bulik-Sullivan et al. 2015; Boyle et al. 2017), multi-ancestry analysis (Ding et al. 2023; Johnson et al. 2023; Hu et al. 2025; Zhao et al. 2025), and integrating functional genomics data (Cano-Gamez et al. 2020).

The plethora of available genetic data and statistical methods have promising applications to medicine. By estimating the contribution of genetic variants to phenotypic variation, statistical genetics provides principled methodology to identify associations between genes and phenotypes, guiding the development of risk assessment and therapeutic

hypotheses (McDonagh et al. 2024). Drug targets supported by genetic evidence are substantially more likely to lead to successful therapeutics (Minikel et al. 2020; Minikel et al. 2024). Beyond target identification, statistical genetics enables quantitative risk prediction through polygenic models that aggregate information across the genome. These strategies offer new opportunities for disease prevention, early detection, and individualized treatment (Lennon et al. 2024; Kachuri et al. 2024; Ding et al. 2023). Together with gene-editing technologies, insights from statistical genetics hold a transformative potential for heritable disease (Visscher et al. 2025).

To identify associations between genetic variants and traits, researchers have developed many analytical approaches including genome-wide association studies (GWAS) for common-variants (Visscher et al. 2017), rare variant burden tests (Wu et al. 2011; Lee et al. 2014), as well as more complex models such as gene-by-environment ($G \times E$) interaction (Kraft et al. 2007; Zhong et al. 2023; Kerin et al. 2020b), and gene-by-gene ($G \times G$) interaction models (Crawford et al. 2017). Often, methods developed for the identification of interactions perform explicit tests using various search strategies such as exhaustive search (Purcell et al. 2007; Schüpbach et al. 2010; Ma et al. 2013), probabilistic search (Prabhu et al. 2012), or prioritization based on a predefined set of biological annotations (Lewinger et al. 2013; Guo et al. 2021) and a number of statistical paradigms such as frequentist tests (Purcell et al. 2007; Wan et al. 2010; Lishout et al. 2013), Bayesian inference (Zhang et al. 2007; Zhang et al. 2011; Guo et al. 2019; Tang et al. 2009), and machine learning (Chang et al. 2020; Fergus et al. 2020) have been tried. However, even with computationally efficient methods (Hemani et al. 2011; Prabhu et al. 2012; Wan et al. 2010; Bayat et al. 2021; Zhu et al. 2018), searching the large combinatorial space of possible interactions is computationally challenging and tanks statistical power due to the multiple testing burden (Crawford et al. 2017).

Despite the rapid growth in data and numerous methodological advances, our understanding of the genetic architecture of complex traits and diseases remains incomplete

(Lee et al. 2025). Many complex traits are highly polygenic, often driven by thousands of variants of small effect (Zhang et al. 2020). Moreover, pleiotropy, functional epistasis, and gene-by-environment interactions can obscure links between variants and traits (Carlborg et al. 2006). In addition, the structure in the genetic data of natural populations, arising from evolutionary and reproductive processes, can further complicate the identification of causal variants and pathways that contribute to disease risk and progression (Lappalainen et al. 2024; Yang et al. 2025). The predictive reliability of polygenic risk scores deteriorates for predictions across ancestries (Ding et al. 2023; Moreno-Grau et al. 2024). Likely reasons for this observation include different linkage disequilibrium patterns across ancestries, differences in environmental exposures, but also differences in the genetic architecture of traits as may be caused by gene-by-gene or gene-by-environment interactions (Hou et al. 2023; Patel et al. 2022).

Although it appears that most of the heritable trait variation is additive (Hivert et al. 2021; Mackay et al. 2024; Mäki-Tanila et al. 2014), molecular biology reveals that gene interactions are ubiquitous (Emani et al. 2024; Kamimoto et al. 2023). Therefore, developing better statistical approaches capable of detecting gene interactions could contribute to overcoming the disconnect between statistical associations and biological causality. The marginal effect sizes of most associated common variants are small, and higher order contributions such as gene-by-gene or gene-by-environment interactions are expected to be even smaller (Mäki-Tanila et al. 2014; Hivert et al. 2021), while variants with large effects are usually rare (Fiziev et al. 2023). In addition, many of the associations identified by GWAS lie in non-coding regions (Maurano et al. 2012), making it difficult to interpret their biological role. Improving the detection and characterization of true causal variants consequently requires ever-larger sample sizes, richer phenotype definitions, and multi-omic integration, as well as approaches capable of integrating data at scale (Chatelain et al. 2025; Baranger et al. 2023). While GWAS associations may point to a locus, establishing causal links remains challenging and conclusive evidence often requires

systematic perturbation experiments (Morris et al. 2023). A comprehensive understanding of the mode of action and context of a genetic variant, e.g. the relevant tissue or cell type, is crucial for the translation of genetic associations into meaningful biological insight (McDonagh et al. 2024).

Existing statistical methods are evolving rapidly, but are often limited in their ability to detect nonlinear relationships or leverage apriori knowledge from functional genomics in their inference. Statistical models going beyond additive effects face prohibitive computational and statistical challenges due to multiple testing burden for exhaustive search approaches or compute intensive matrix operations (Crawford et al. 2017). Further development is needed to help better understand nonlinear contributions to the genetic trait architectures such as gene-by-gene interactions, gene-by-environment interactions, as gene action and regulation are rarely additive (Hivert et al. 2021; Kamimoto et al. 2023; Emani et al. 2024).

This dissertation sets out to address some of the methodological gaps in statistical genetics. Specifically, I will develop statistical methods to study nonlinear contributions to phenotypic variation that account for pleiotropy, gene-by-gene interactions, gene-by-environment interactions, as well as offer approaches to integrate multi-omic functional annotations, and extend the target phenotypes that can be studied for nonlinear trait architectures.

In Chapter 1, I develop the “multivariate MArginal ePIstasis Test” (mvMAPIT) (Stamp et al. 2023) — a multi-outcome generalization of the “MArginal ePIstasis Test” (MAPIT) (Crawford et al. 2017), a gene-by-gene interaction detection method which seeks to detect *marginal epistasis* or the combined pairwise interaction effects between a given variant and all other variants. By searching for marginal epistatic effects, one can identify genetic variants that are involved in epistasis without the need to identify the exact partners with which the variants interact — thus, potentially alleviating much of the statistical and computational burden associated with conventional explicit search-based methods.

The developed method builds upon this strategy by taking advantage of correlation structure between traits to improve the identification of variants involved in epistasis. I formulate mvMAPIT as a multivariate linear mixed model and develop a multi-trait variance component estimation algorithm for efficient parameter inference and P -value computation. Together with reasonable model approximations, our proposed approach is scalable to moderately sized GWA studies. With simulations, I illustrate the benefits of mvMAPIT over univariate (or single-trait) epistatic mapping strategies. I also apply the mvMAPIT framework to protein sequence data from two broadly neutralizing anti-influenza antibodies (Phillips et al. 2021) and approximately 2,000 heterogeneous stock of mice from the Wellcome Trust Centre for Human Genetics (Valdar et al. 2006a; Valdar et al. 2006b).

In Chapter 2, I develop a statistical method to detect gene-by-environment interactions in time-to-event outcomes, the Cox proportional hazards gene-by-environment interaction test (CphGxE). While much of genetic association testing has centered around quantitative or binary phenotypes, time-to-event outcomes pose additional challenges due to censoring and the need for proper modeling of hazard functions. I introduce a framework for studying time-to-event traits that partitions the heritable variance of frailty into polygenic background, environmental exposure, and gene-by-environment interactions. Our approach leverages the correspondence of the Cox proportional hazards model and a modified Poisson generalized linear mixed model in order to deal with random effects. The naive implementation of this approach scales to moderately sized biobank-scale data, although the iterative fitting strategy employed in the presented implementation can be scaled to larger biobank-scale data. I offer several avenues to improve scalability such as caching of the null fit and the use of an efficient variance component score test. Through extensive simulations, I demonstrate how this framework offers a powerful approach to detect gene-by-environment interactions with an effect on disease progression and survival.

In Chapter 3, I develop 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 (Stamp et al. 2025). The lack of computational methods capable of detecting epistasis in biobanks has led to uncertainty about the role of nonadditive genetic effects on complex trait variation (Hill et al. 2008; Hivert et al. 2021; Pattillo Smith et al. 2024). While the marginal epistasis framework is a powerful approach, current implementations of this approach have failed to scale genome-wide in large human studies. By leveraging the biological priors to induce sparsity, I develop a statistical algorithm that allows SME to run 10 to 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, I show that reducing searches of epistasis to variants in functionally enriched regions facilitates the identification of genetic interactions associated with regulatory genomic elements.

CHAPTER 1

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

The text in this chapter is an adaptation of the material as it appears in Stamp et al. (2023). *Leveraging the Genetic Correlation between Traits Improves the Detection of Epistasis in Genome-wide Association Studies*. G3 Genes|Genomes|Genetics. This article is licensed under the Creative Commons Attribution (CC BY <https://creativecommons.org/licenses/by/4.0/>) licence. The coauthors listed in this publication are Julian Stamp, Alan DenAdel, Daniel Weinreich, and Lorin Crawford. Lorin Crawford and Daniel Weinreich supervised the research which forms the basis for the dissertation.

1.1 Introduction

Genome-wide association (GWA) studies have contributed substantially in the discovery of genetic markers associated with the architecture of disease phenotypes (Ripke et al.

2014; Ellinghaus et al. 2016; Fuchsberger et al. 2016; Visscher et al. 2012; Visscher et al. 2017; Loos 2020). Epistasis, commonly defined as the interaction between genetic loci, has long been thought to play a key role in defining the genetic architecture underlying many complex traits and common diseases (Gusareva et al. 2018; Kirino et al. 2013; Tao et al. 2012; Wei et al. 2012; Chen et al. 2017). Indeed, previous studies have detected pervasive epistasis in many model organisms (Pattillo Smith et al. 2024; Peripato et al. 2004; Tong et al. 2004; Brem et al. 2005; Deutschbauer et al. 2005; Kroymann et al. 2005; Collins et al. 2006; Lehner et al. 2006; Onge et al. 2007; Wentzell et al. 2007; Shao et al. 2008; Flint et al. 2009; Gerke et al. 2009; Costanzo et al. 2010; He et al. 2010; Horn et al. 2011; Jarvis et al. 2011; Leamy et al. 2011; Pettersson et al. 2011; Szappanos et al. 2011; Gaertner et al. 2012; Bloom et al. 2013; Chari et al. 2013; Monnahan et al. 2015). Substantial contributions of epistasis to phenotypic variance have been revealed for many complex traits (Huang et al. 2016; Crawford et al. 2018b) and have been suggested to constitute an important component of evolution (Weinreich et al. 2006). Furthermore, modeling epistasis in addition to additive and dominant effects has been argued to increase phenotypic prediction accuracy in model organisms (Forsberg et al. 2017; Zhou et al. 2022) and facilitate genomic selection in breeding programs (Runcie et al. 2021; Muñoz et al. 2014; Jiang et al. 2015). Despite a longstanding and currently ongoing debate about the contribution of nonadditive effects on the architecture of human complex traits (Long et al. 1999; Marchini et al. 2005; Hill et al. 2008; Shao et al. 2008; Crow 2010; Zuk et al. 2012; Hivert et al. 2021; Pazokitoroudi et al. 2021; Wainschtein et al. 2022), recent genetic mapping studies have also identified evidence of epistatic interactions that significantly contribute to quantitative traits and diseases (Brown et al. 2014; Fang et al. 2019; Haar et al. 2019; Sheppard et al. 2021), and some have recently argued that gene-by-gene interactions can drive heterogeneity of causal variant effect sizes across diverse human populations (Patel et al. 2022). Importantly, epistasis is often proposed as a key contributor to missing heritability — the proportion of heritability not explained by the top associated variants in GWA studies (Eichler et al. 2010; Gusareva et al. 2014b;

Gusareva et al. 2014a; Gusareva et al. 2018; Wanstrat et al. 2001).

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

As a departure from the explicit search strategy, the second category of epistatic mapping methods attempts to address the previously mentioned challenges by detecting

marginal epistasis. Specifically, instead of directly identifying individual pairwise or higher-order interactions, these approaches focus on identifying variants that have a non-zero interaction effect with any other variant in the data set. For example, MAPIT (Crawford et al. 2017) assesses each variant (in turn) and identifies candidate markers that are involved in epistasis without the need to identify the exact partners with which the variants interact — thus, alleviating much of the statistical power concerns and heavy computational burdens associated with explicit search-based methods. As a framework, the marginal epistatic strategy has been implemented in both linear mixed models and machine learning and has been used for case-control studies (Crawford et al. 2018a), pathway enrichment applications (Turchin et al. 2020), heritability estimation (Pattillo Smith et al. 2024), and even extended to explore different sources of nonadditive genetic variation (e.g., gene-by-environment interactions) (Moore et al. 2019; Kerin et al. 2020b). However, despite its wide adoption, this approach can still be underpowered for traits with low heritability or “polygenic” traits which are generated by many mutations of small effect (Crawford et al. 2017).

To date, both the explicit search and marginal epistasis detection methodologies have only focused on analyzing one phenotype at a time. However, many previous genetic association studies have extensively shown that jointly modeling multiple phenotypes can often dramatically increase power for variant detection (Zhou et al. 2014). In this work, we present the “multivariate MArginal ePIstasis Test” (mvMAPIT) — a multi-outcome generalization of the MAPIT model which aims leverage relations between traits to improve the identification of variants involved in epistasis. We formulate mvMAPIT as a multivariate linear mixed model (mvLMM) and extend a previously developed variance component estimation algorithm for efficient parameter inference and P -value computation in the multi-trait setting (Zhou 2017). Together with reasonable model approximations, our proposed approach is scalable to moderately sized GWA studies. With detailed simulations, we illustrate the benefits of mvMAPIT in terms of providing

effective type I error control and compare its power to the univariate (or single-trait) mapping strategy used in the original MAPIT model. Here, part of our main contribution is the demonstration that the power in our proposed multivariate approach is driven by the correlations between the effects of pairwise interactions on multiple traits. To close, we also apply the mvMAPIT framework to protein sequence data from a nearly combinatorially complete library of two broadly neutralizing anti-influenza antibodies (Phillips et al. 2021) and to 15 quantitative hematology traits assayed in a heterogeneous stock of mice from Wellcome Trust Centre for Human Genetics (Valdar et al. 2006a; Valdar et al. 2006b).

1.2 Materials and methods

1.2.1 The marginal epistasis test for single traits

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

$$\mathbf{y} = \mu + \mathbf{x}_j\beta_j + \sum_{l \neq j} \mathbf{x}_l\beta_l + \sum_{l \neq j} (\mathbf{x}_j \circ \mathbf{x}_l)\alpha_l + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \tau^2 \mathbf{I}) \quad (1.1)$$

where $\mathbf{y}$ is an N -dimensional vector of phenotypic states for a quantitative trait of

interest measured in the N individuals; μ is an intercept term; $\mathbf{X}$ denotes an $N \times J$ matrix of genotypes with $\mathbf{x}_j$ and $\mathbf{x}_l$ representing N -dimensional vectors for the j -th and l -th SNPs; β_j and β_l are the respective additive effects; $\mathbf{x}_j \circ \mathbf{x}_l$ denotes the Hadamard (element-wise) product between two genotypic vectors with corresponding interaction effect size α_l ; ε 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. For convenience, we will assume that both the genotype matrix (column-wise) and trait of interest have been mean-centered and standardized. It is also worth noting that, while we limit the above to the task of identifying second order (i.e., pairwise) interactions between genetic variants, extensions of MAPIT to higher-order epistatic and gene-by-environmental effects have been shown to be straightforward to implement (Moore et al. 2019; Kerin et al. 2020a; Kerin et al. 2020b; Zhu et al. 2022).

Variance component model formulation. Since many modern GWA applications present scenarios that would make Eq. 1.1 an underdetermined linear system (e.g., in modern biobank-scale studies where genotyped markers $J > N$ individuals), the MAPIT framework follows other standard approaches (Bulik-Sullivan et al. 2015; Yang et al. 2010; Wu et al. 2011; Zhou et al. 2013; Crawford et al. 2017) to ensure model identifiability by assuming that the additive and interaction effect sizes follow univariate normal distributions where $\beta_l \sim \mathcal{N}(0, \omega^2/(J-1))$ and $\alpha_l \sim \mathcal{N}(0, \sigma^2/(J-1))$ for $l \neq j$, respectively. This key normal assumption on the regression coefficients allows for Eq. 1.1 to be equivalently represented as the following variance component model

$$\mathbf{y} = \mu + \mathbf{x}_j\beta_j + \mathbf{m}_j + \mathbf{z}_j + \varepsilon, \quad \varepsilon \sim \mathcal{N}(\mathbf{0}, \tau^2\mathbf{I}) \quad (1.2)$$

where, in addition to previous notation, $\mathbf{m}_j = \sum_{l \neq j} \mathbf{x}_l\beta_l$ is the combined additive effects from all variants other than the j -th; and $\mathbf{z}_j = \sum_{l \neq j} (\mathbf{x}_j \circ \mathbf{x}_l)\alpha_l$ denote the summation of all pairwise interaction effects between the j -th variant and all other variants. Under the

variance component formulation in Eq. 1.2, the two random effects can also be expressed probabilistically as $\mathbf{m}_j \sim \mathcal{N}(\mathbf{0}, \omega^2 \mathbf{K}_j)$ where $\mathbf{K}_j = \mathbf{X}_{-j} \mathbf{X}_{-j}^\top / (J - 1)$ is an additive genetic relatedness matrix that is computed using all genotypes other than the j -th SNP, and $\mathbf{z}_j \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{G}_j)$ where $\mathbf{G}_j = \mathbf{D}_j \mathbf{K}_j \mathbf{D}_j$ is a nonadditive relatedness matrix computed based on all pairwise interaction terms involving the j -th SNP. Here, we let $\mathbf{D}_j = \text{diag}(\mathbf{x}_j)$ denote an $N \times N$ diagonal matrix with the j -th genotype as its only nonzero elements. It is also important to note that both $\mathbf{K}_j$ and $\mathbf{G}_j$ change with every new j -th marker that is tested.

Univariate point estimates. The key takeaway from the variance component model formulation in Eq. 1.2 is that σ^2 represents a measure on the marginal epistatic effect for each variant in the data. Therefore, in order to identify variants that have significant non-zero interaction effects, we must assess the null hypothesis $H_0 : \sigma^2 = 0$ for each variant in the data set. The original MAPIT framework uses a computationally efficient method of moments algorithm called MQS (Zhou 2017) to estimate model parameters and to carry out calibrated statistical tests. Briefly, MQS produces point estimates that are mathematically identical to the Haseman-Elston (HE) cross-product regression (Zhou 2017; Zhu et al. 2020; Lee et al. 2011). To implement this algorithm, we first specify a two-dimensional matrix $\mathbf{b}_j = [\mathbf{1}, \mathbf{x}_j]$ with $\mathbf{1}$ being an N -dimensional vector of ones. Next, we then multiply both sides of Eq. 1.2 by a variant-specific projection $\mathbf{P}_j = \mathbf{I} - \mathbf{b}_j (\mathbf{b}_j^\top \mathbf{b}_j)^{-1} \mathbf{b}_j^\top$ which maps the model onto a column space that is orthogonal to both the intercept and the genotypic vector $\mathbf{x}_j$. This process simplifies the model specification of MAPIT in Eq. 1.2 to the following

$$\mathbf{y}_j^* = \mathbf{m}_j^* + \mathbf{z}_j^* + \boldsymbol{\varepsilon}_j^*, \quad \mathbf{m}_j^* \sim \mathcal{N}(\mathbf{0}, \omega^2 \mathbf{K}_j^*), \quad \mathbf{z}_j^* \sim \mathcal{N}(\mathbf{0}, \sigma^2, \mathbf{G}_j^*), \quad \boldsymbol{\varepsilon}_j^* \sim \mathcal{N}(\mathbf{0}, \tau^2 \mathbf{P}_j) \quad (1.3)$$

where we denote $\mathbf{y}_j^* = \mathbf{P}_j \mathbf{y}$; $\mathbf{m}_j^* = \mathbf{P}_j \mathbf{m}_j$; $\mathbf{K}_j^* = \mathbf{P}_j \mathbf{K}_j \mathbf{P}_j$; $\mathbf{z}_j^* = \mathbf{P}_j^* \mathbf{z}_j$; $\mathbf{G}_j^* = \mathbf{P}_j \mathbf{G}_j \mathbf{P}_j$; and $\boldsymbol{\varepsilon}_j^* = \mathbf{P}_j \boldsymbol{\varepsilon}$, respectively. The estimators for the variance components in Eq. 1.3 are

naturally based on the second moment matching equations where, in expectation, we have

$$\mathbb{E}[\mathbf{y}_j^{*\top} \mathbf{H} \mathbf{y}_j] = \sum_{k=1}^3 \text{tr}(\mathbf{H} \boldsymbol{\Sigma}_{jk}) \delta_k \quad (1.4)$$

with $\mathbf{H}$ being a symmetric and non-negative definite matrix used to create weighted second moments, $\text{tr}(\bullet)$ denotes the trace of a matrix, and we use shorthand to represent $[\boldsymbol{\Sigma}_{j1}; \boldsymbol{\Sigma}_{j2}; \boldsymbol{\Sigma}_{j3}] = [\mathbf{K}_j^*; \mathbf{G}_j^*; \mathbf{P}_j]$ and $\boldsymbol{\delta} = (\omega^2, \sigma^2, \tau^2)$, respectively. In practice, we replace the left hand side of Eq. 1.4 with the realized value $\mathbf{y}_j^{*\top} \mathbf{H} \mathbf{y}_j$. Note that many choices of $\mathbf{H}$ will yield unbiased estimates for $(\omega^2, \sigma^2, \tau^2)$, but different choices of $\mathbf{H}$ can affect statistical efficiency of the estimates. The set of moment matching equations in MQS is generated by using the covariance matrices corresponding to the variance components in place of the arbitrary $\mathbf{H}$. This system of equations can then be rewritten as the following matrix multiplication

$$\boldsymbol{\delta} = \mathbf{S}^{-1} \mathbf{q}, \quad q_k = \mathbf{y}_j^{*\top} \boldsymbol{\Sigma}_{jk} \mathbf{y}_j, \quad \mathbf{S}_{rs} = \text{tr}(\boldsymbol{\Sigma}_{jr} \boldsymbol{\Sigma}_{js}) \quad (1.5)$$

where $\mathbf{q}$ is a 3-dimensional vector and $\mathbf{S}$ is a 3×3 dimensional matrix with $k, r, s \in \{1, 2, 3\}$ being indices to represent the different variance components. If we subset just to compute an estimate for the marginal epistatic variance component (i.e., for the second index), then Eq. 1.5 reduces to the following formula

$$\hat{\sigma}_j^2 = \mathbf{y}_j^{*\top} \mathbf{H}_j \mathbf{y}_j \quad (1.6)$$

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

Univariate hypothesis testing. In general, there are two ways to compute P -values in the MAPIT framework (Crawford et al. 2017). The first option uses a two-sided z-score or normal test. This particular test only requires the variance component estimate $\hat{\sigma}_j^2$

from Eq. 1.6 and its corresponding standard error, which is approximated in the MQS approach as

$$\mathbb{V}[\hat{\sigma}_j^2] \approx 2\mathbf{y}_j^{*\top} \mathbf{H}_j^\top \mathbf{V}_j \mathbf{H}_j \mathbf{y}_j^* \quad (1.7)$$

where $\mathbf{V}_j = \hat{\omega}_j^2 \mathbf{K}_j^* + \hat{\sigma}_j^2 \mathbf{G}_j^* + \hat{\tau}_j^2 \mathbf{P}_j$. The second option for deriving P -values in the MAPIT framework uses an exact test which is based on the fact that the MQS variance component estimate follows a mixture of chi-square distributions under the null hypothesis. This is derived from both the normality assumption on $\mathbf{y}^*$ and the quadratic form of the statistic in Eq. 1.6. Namely, $\hat{\sigma}_j^2 \sim \sum_{i=1}^N \lambda_i \chi_{1,i}^2$ where χ_1^2 are chi-square random variables with one degree of freedom and $(\lambda_1, \dots, \lambda_N)$ are the eigenvalues of the matrix

$$\left(\hat{\omega}_0^2 \mathbf{K}_j^* + \hat{\tau}_0^2 \mathbf{P}_j \right)^{1/2} \mathbf{H}_j \left(\hat{\omega}_0^2 \mathbf{K}_j^* + \hat{\tau}_0^2 \mathbf{P}_j \right)^{1/2} \quad (1.8)$$

with $(\hat{\omega}_0^2, \hat{\tau}_0^2)$ being the MQS estimates of (ω^2, τ^2) under the null hypothesis $H_0 : \sigma^2 = 0$. Several approaches have been proposed to obtain P -values under a mixture of chi-square distributions, including the Davies method (Davies 1980) (see Data and Software Availability). In practice, while the Davies method is an exact test and is expected to produce calibrated P -values, it can become computationally intensive since it scales cubically in the number of individuals N . On the other hand, while the normal test only scales quadratically in N because of the standard error approximation in Eq. 1.7, it has been shown to lead to mis-calibrated P -values for data sets with small sample sizes. As a result, MAPIT uses a hybrid procedure which implements the normal test by default, and then applies the Davies method when the P -value from the normal test is below the nominal threshold of 0.05 (Crawford et al. 2017).

1.2.2 Derivation of the multivariate marginal epistasis test

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

$$\mathbf{Y} = \mathbf{U} + \beta_j \mathbf{x}_j^\top + \sum_{l \neq j} \beta_l \mathbf{x}_l^\top + \sum_{l \neq j} \alpha_l (\mathbf{x}_j \circ \mathbf{x}_l)^\top + \mathbf{E} \quad \mathbf{E} \sim \mathcal{MN}(\mathbf{0}, \mathbf{V}_\varepsilon, \mathbf{I}) \quad (1.9)$$

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

Similar to the univariate setting, we need to make additional probabilistic assumptions to ensure model identifiability when Eq. 1.9 is an underdetermined linear system. To that end, let $\mathbf{B} = [\beta_l]_{l \neq j}$ and $\mathbf{A} = [\alpha_l]_{l \neq j}$ denote the collection of coefficients not involving the j -th variant of interest. Here, we will assume that these $D \times (J - 1)$ effect size matrices also follow matrix-variate normal distributions where $\mathbf{B} \sim \mathcal{MN}(\mathbf{0}, \mathbf{V}_\beta, \mathbf{I})$ and $\mathbf{A} \sim \mathcal{MN}(\mathbf{0}, \mathbf{V}_\alpha, \mathbf{I})$, respectively. Note that this formulation is largely similar to the

univariate case except with the additional property that the phenotypes being studied share some genetic covariance through $\mathbf{V}_\beta$ and $\mathbf{V}_\alpha$. This assumption, coupled with the affine transformation property of matrix normal distributions, allows for us to equivalently represent the mvMAPIT model in Eq. 1.9 as the following multivariate variance component model

$$\mathbf{Y} = \mathbf{U} + \beta_j \mathbf{x}_j^\top + \mathbf{M}_j + \mathbf{Z}_j + \mathbf{E} \quad \mathbf{E} \sim \mathcal{MN}(\mathbf{0}, \mathbf{V}_\varepsilon, \mathbf{I}) \quad (1.10)$$

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

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

1.2.3 Hypothesis testing in the mvMAPIT framework

The goal of the mvMAPIT framework still comes down to assessing the null hypothesis that tests for nonzero marginal epistatic effects. However, parameter estimation in

mvLMMs can present substantial computational challenges. For example, one common way in the literature to rewrite the model specified in Eq. 1.10 is to vectorize (or stack) the columns of each matrix in the regression such that $\mathbf{y} = \text{vec}(\mathbf{Y})$, $\boldsymbol{\mu} = \text{vec}(\mathbf{U})$, $\mathbf{m}_j = \text{vec}(\mathbf{M}_j)$, $\mathbf{z}_j = \text{vec}(\mathbf{Z}_j)$, and $\boldsymbol{\varepsilon} = \text{vec}(\mathbf{E})$. Under this reformulation, we could simply follow the procedures in Eqs. 1.3-1.8 to find significant variance components; but since $\mathbb{V}[\mathbf{m}_j] = \mathbf{K}_j \otimes \mathbf{V}_\beta$ and $\mathbb{V}[\mathbf{z}_j] = \mathbf{G}_j \otimes \mathbf{V}_\alpha$ are each $ND \times ND$ dimensions (via the Kronecker product $\otimes$), the per-iterative computation time for performing hypothesis testing on each j -th SNP would now increase both with the number of individuals (N) and with the number of phenotypes (D). This could make model fitting infeasible for large biobanks even with only two traits. As an alternative, we present a combinatorial approach which first fits univariate MAPIT models and then combines the resulting P -values with those stemming from a “covariance statistic” which looks for shared marginal epistatic effects between all pairwise combinations of the D traits. Importantly, our combinatorial approach does not make assumptions about the covariance structure between traits, which would need to be known (or assumed) in the Kronecker formulation.

To implement the multivariate marginal epistasis test, we follow a similar strategy used in the univariate MAPIT model and right multiply Eq. 1.10 by a variant-specific projection $\mathbf{P}_j = \mathbf{I} - \mathbf{b}_j(\mathbf{b}_j^\top \mathbf{b}_j)^{-1} \mathbf{b}_j^\top$ which maps the model onto a column space that is orthogonal to the population-level intercepts and the genotypic vector $\mathbf{x}_j$. This results in a simplified mvLMM of the following form

$$\mathbf{Y}_j^* = \mathbf{M}_j^* + \mathbf{Z}_j^* + \mathbf{E}_j^*, \quad \mathbf{E}_j^* \sim \mathcal{MN}(\mathbf{0}, \mathbf{V}_\varepsilon, \mathbf{P}_j). \quad (1.11)$$

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

d -th rows of the adjusted phenotypic matrix $\mathbf{Y}_j^*$, respectively. The general MQS estimates for the marginal epistatic effect is a generalization of Eq. 1.6 which is given in the following quadratic form

$$\hat{\sigma}_{j,(cd)}^2 = \mathbf{y}_c^{*\top} \mathbf{H}_j \mathbf{y}_d^*, \quad (1.12)$$

where $\mathbf{H}_j$ is as previously defined in the univariate MAPIT case and the trait-specific indices span between the $c, d \in 1, \dots, D$ phenotypes. Here, when $c = d$, the above is exactly equal to Eq. 1.6 and the variance component point estimate is computed using only one trait row in $\mathbf{Y}_j^*$. On the other hand, when $c \neq d$, Eq. 1.12 takes on a bilinear form where $\mathbb{E}[\mathbf{y}_c^{*\top} \mathbf{H}_j \mathbf{y}_d^*] = \text{tr}(\mathbf{H}_j \mathbf{V}_{j,(cd)})$ with $\mathbf{V}_{j,(cd)} = \mathbb{V}[\mathbf{y}_c^*, \mathbf{y}_d^*]$ being the covariance between any two traits of interest. The corresponding standard error of the bilinear covariance component can then be estimated via the following approximation (Searle 1979)

$$\mathbb{V}[\hat{\sigma}_{j,(cd)}^2] \approx \mathbf{y}_c^{*\top} \mathbf{H}_j^\top \mathbf{V}_{j,(cd)} \mathbf{H}_j \mathbf{y}_d^* + \mathbf{y}_c^{*\top} \mathbf{H}_j^\top \mathbf{V}_{j,(dd)} \mathbf{H}_j \mathbf{y}_c^*. \quad (1.13)$$

Once again, notice that when $c = d$, the term $\mathbf{V}_{j,(cd)} = \mathbf{V}_{j,(dd)}$ and the above approximation in Eq. 1.13 is equal to Eq. 1.7.

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

1. In the first step, the model fits univariate models for all D traits of interests (i.e., using Eqs. 1.3-1.8 from the MAPIT model or equivalently Eqs. 1.12 and 1.13 with $c = d$). Here, we use the proposed hybrid testing approach where we first implement a normal test by default, and then apply the exact Davies method when the P -value from the normal test is below the nominal significance threshold of 0.05 (Crawford et al. 2017).
2. In the second step, we derive P -values for the covariance components (i.e., using

Eqs. 1.12 and 1.13 when $c \neq d$) with a normal test. As we will show below, the P -values derived for the covariance components with the asymptotic normal approximation tend to be slightly deflated under the null hypothesis. While this leads to generally conservative behavior with respect to type I error control, the downside is that the test may result in reduced power under the alternative, especially after multiple correction for data sets with small sample sizes or for traits that have low genetic correlation. In these cases, deriving an exact test to obtain more calibrated P -values could be done; however, we do not explore this line of work here.

3. In the third and final step, mvMAPIT combines the P -values from the first two steps into an overall marginal epistatic P -value. Assume that we have $T = 3$ sets of P -values (two sets corresponding to marginal effects and one covariance set). The mvMAPIT software carries out the P -value combining procedure in three different ways. The first assumes that each of the $t = 1, \dots, T$ tests are (effectively) independent and implements Fisher’s method (Fisher 1925) which combines P -values into a single chi-square test statistic using the formula $\chi_{2T}^2 \sim -2 \sum_{t=1}^T \log(p_t)$ where p_t denotes the P -value from the t -th test. In Fisher’s method, the χ^2 test statistic will be large when P -values tend to be small (i.e., when the null hypothesis is not true for every test). The second approach assumes an unknown dependency structure between each of the T tests and computes a harmonic mean (Wilson 2019) P -value where $\hat{p} = \sum_t w_t / \sum_t w_t / p_t$. Here, $\sum_t w_t = 1$ are weights which we uniformly set to be $w_t = 1/T$ for all P -values. The last approach implements the Cauchy combination test which produces analytic P -values for any arbitrary dependency structure (Liu et al. 2020). Under this method, the Cauchy combination test statistic is defined as $CCT = \sum_t w_t \tan\{(0.5 - p_t)\pi\}$ where $\tan(\bullet)$ is the tangent function and again we set $w_t = 1/T$ to be uniform for all P -values. The test statistic follows a standard Cauchy distribution $CCT \sim C(0, 1)$.

In practice, epistatic effects are assumed to make small contributions to the overall

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

Note on settings where mvMAPIT is designed to be most powered. The formulation of the general estimates in Eqs. 1.12 and 1.13 highlight an important takeaway in that the mvMAPIT covariance statistic models epistatic pairs that together affect the architecture of multiple traits. It is not meant to identify individual SNPs that are involved in epistasis for multiple traits while being a member of different interacting pairs. To clarify this, consider two simple scenarios in Figure 1.1 where we have two phenotypes ($\mathbf{y}_1$ and $\mathbf{y}_2$) that are generated by a combination of four SNPs ($\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$). In the first scenario, we say that (in expectation) $\mathbb{E}[\mathbf{y}_1] = \mathbf{x}_1\beta_1 + (\mathbf{x}_2 \circ \mathbf{x}_3)\alpha_1$ and $\mathbb{E}[\mathbf{y}_2] = (\mathbf{x}_2 \circ \mathbf{x}_3)\alpha_2$ (Figure 1.1A), while, in the second scenario, $\mathbb{E}[\mathbf{y}_1] = \mathbf{x}_1\beta_1 + (\mathbf{x}_2 \circ \mathbf{x}_3)\alpha_1$ and $\mathbb{E}[\mathbf{y}_2] = (\mathbf{x}_3 \circ \mathbf{x}_4)\alpha_2$ (Figure 1.1B). The key to power in the mvMAPIT framework is that, in the first scenario, the interaction between $\mathbf{x}_2$ and $\mathbf{x}_3$ appears in both traits with nonzero correlation between the effect sizes α_1 and α_2 . This is in contrast to the second scenario where there is a common variant involved in epistasis but it is a member of two different sets of interactions that affect each trait. The mvMAPIT covariance statistic captures the situation illustrated in the first scenario (Figure 1.1A) but not in the second (Figure 1.1B).

1.2.4 Simulation study design

To test the utility of the mvMAPIT framework, we modified a frequently used simulation scheme (Crawford et al. 2017; Pattillo Smith et al. 2024) to generate collections of

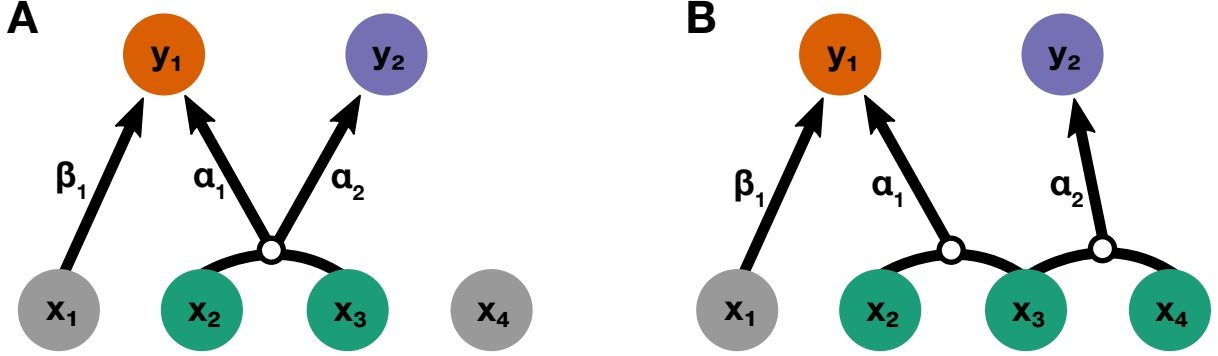


Figure 1.1: **Schematic of the types of shared interactions modeled by the multi-variate marginal epistasis test.** Consider two simple, proof-of-concept simulation scenarios where two traits (y_1, y_2) are generated by a combination of four SNPs (x_1, x_2, x_3, x_4). Panel (A) shows the first scenario where (in expectation) $\mathbb{E}[y_1] = x_1\beta_1 + (x_2 \circ x_3)\alpha_1$ and $\mathbb{E}[y_2] = (x_2 \circ x_3)\alpha_2$. Panel (B) shows the second scenario where $\mathbb{E}[y_1] = x_1\beta_1 + (x_2 \circ x_3)\alpha_1$ and $\mathbb{E}[y_2] = (x_3 \circ x_4)\alpha_2$. In both panels, variant x_1 only has an additive effect β_1 on trait y_1 . The mvMAPIT approach models correlations between the effects of a given interaction on multiple traits. Therefore, mvMAPIT is designed to identify SNPs involved in the first scenario where the interaction between variants x_2 and x_3 is shared between traits with nonzero correlated effect sizes α_1 and α_2 . This is in contrast to the second case, where variant x_3 is important to both traits but through distinct interactions with variants x_2 and x_4 , respectively.

synthetic quantitative traits under multiple genetic architectures using real genotypes from chromosome 22 of the control samples in the Wellcome Trust Case Control Consortium (WTCCC) 1 study. After preprocessing, considering this particular group of individuals and SNPs resulted in a data set consisting of $N = 2,938$ individuals and $J = 5,747$ markers. In these simulations, we randomly choose 1,000 causal SNPs to directly affect $D = 2$ phenotypes. We generate these synthetic traits via the following general multivariate linear model:

$$\mathbf{Y} = \sum_{g \in \mathcal{G}} \beta_g \mathbf{x}_g^\top + \mathbf{A}\mathbf{W}^\top + \mathbf{E}, \quad \mathbf{E} \sim \mathcal{MN}(\mathbf{0}, \mathbf{I}, \mathbf{I}) \quad (1.14)$$

where $\mathbf{Y}$ is an $D \times N$ matrix containing all the phenotypes; $\mathcal{G}$ represents the set of 1,000 causal SNPs; $\mathbf{x}_g$ is the genotype for the g -th causal SNP encoded as 0, 1, or 2 copies of a reference allele; β_g is a D -dimensional vector and represent the additive effect sizes for the g -th SNP in the D traits; $\mathbf{W}$ is an $N \times M$ matrix which holds pairwise interactions (i.e.,

Hadamard products) between some subset of causal SNPs; $\mathbf{A} = [\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_M]$ is a $D \times M$ matrix of interaction effect sizes with $\boldsymbol{\alpha}_m$ being D -dimensional epistatic coefficients for the m -th interaction in the d -th trait; and $\mathbf{E}$ is an $D \times N$ matrix of normally distributed environmental noise.

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

$$\boldsymbol{\beta}_g \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_\beta), \quad \boldsymbol{\alpha}_m \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_\alpha) \quad (1.15)$$

where $\mathbf{V}_\beta$ and $\mathbf{V}_\alpha$ are $D \times D$ covariance matrices for additive effects and pairwise interactions between the phenotypes. Once these coefficients are sampled, we rescale them so that they explain a fixed proportion of the broad-sense heritability H^2 . Similarly, the environmental noise matrix is rescaled such that it explains $1 - H^2$. When generating synthetic traits, we assume that the additive effects make up $\rho\%$ of the broad-sense heritability while the pairwise interactions make up the remaining $(1 - \rho)\%$. Alternatively, we say that the proportion of the heritability explained by additivity is ρH^2 , while the proportion of phenotypic variance explained by pairwise interactions is $(1 - \rho)H^2$. Setting $\rho = 1$ represents the null model where the variation of a trait is driven by solely additive effects. Here, we use the same simulation strategy used in previous studies (Crawford et al. 2017; Pattillo Smith et al. 2024) where we divide the causal variants into three groups where:

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

Here, the epistatic causal SNPs interact between sets. All SNPs in $\mathcal{G}_1$ interact with all

SNPs in the $\mathcal{G}_2$, but do not interact with variants in their own group (with the same rule applied to the second group). Again, each interaction between SNPs in the two groups has an epistatic effect size α_d for the d -th trait (see Fig. 1.1A). The correlation of epistatic effects between traits is determined via the correlation coefficient $v_{\alpha,12}$ (i.e., the off-diagonal elements of $\mathbf{V}_\alpha$). With this set up, one can think of the SNPs assigned to $\mathcal{G}_1$ as being the “hub nodes” in an interaction network. Note that we use this setup because it has been shown that the ability to detect two interacting variants depends on the proportion of phenotypic variance that they marginally explain. For example, in our case, this means that power is expected to depend on $\mathbb{V}[\mathbf{W}\boldsymbol{\alpha}]/|\mathcal{G}_1|$ and $\mathbb{V}[\mathbf{W}\boldsymbol{\alpha}]/|\mathcal{G}_2|$ for groups 1 and 2, respectively, where $|\mathcal{G}|$ denotes the cardinality of the set. Given different parameters for the generative model in Eq. 1.14, we simulate data mirroring a wide range of genetic architectures by varying the following parameters:

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

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

1.2.5 Preprocessing of the heterogeneous stock of mice data set

As part of the analyses, this work makes use of GWA data from the Wellcome Trust Centre for Human Genetics (Valdar et al. 2006a; Valdar et al. 2006b) (<http://mtweb.cs.ucl.ac.uk/mus/www/mouse/index.shtml>). The genotypes from this study

were downloaded directly using the BGLR-R package (Pérez et al. 2014). This study contains $N = 1,814$ heterogeneous stock of mice from 85 families (all descending from eight inbred progenitor strains) (Valdar et al. 2006a; Valdar et al. 2006b), and 131 quantitative traits that are classified into 6 broad categories including behavior, diabetes, asthma, immunology, haematology, and biochemistry. Phenotypic measurements for these mice can be found freely available online to download (details can be found at <http://mtweb.cs.ucl.ac.uk/mus/www/mouse/HS/index.shtml> and <https://github.com/lcrawlab/mvMAPIT>). In the analyses below, we focused on 15 hematological phenotypes including: atypical lymphocytes (ALY; Haem.ALYabs), basophils (BAS; Haem.BASabs), hematocrit (HCT; Haem.HCT), hemoglobin (HGB; Haem.HGB), large immature cells (LIC; Haem.LICabs), lymphocytes (LYM; Haem.LYMabs), mean corpuscular hemoglobin (MCH; Haem.MCH), mean corpuscular volume (MCV; Haem.MCV), monocytes (MON; Haem.MONabs), mean platelet volume (MPV; Haem.MPV), neutrophils (NEU; Haem.NEUabs), plateletcrit (PCT; Haem.PCT), platelets (PLT; Haem.PLT), red blood cell count (RBC; Haem.RBC), red cell distribution width (RDW; Haem.RDW), and white blood cell count (WBC; Haem.WBC). All phenotypes were previously corrected for sex, age, body weight, season, year, and cage effects (Valdar et al. 2006a; Valdar et al. 2006b). For individuals with missing genotypes, we imputed values by the mean genotype of that SNP in their corresponding family. Only polymorphic SNPs with minor allele frequency above 5% were kept for the analyses. This left a total of $J = 10,227$ autosomal SNPs that were available for all mice.

1.3 Results

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

In this section, we make use of a previously described simulation scheme (Crawford et al. 2017; Pattillo Smith et al. 2024) in order to investigate whether mvMAPIT and its combinatorial inference approach preserves the desired type I error rate and produces well-calibrated P -values under the null hypothesis. Here, we generate synthetic phenotypes using real genotypes from the 22nd chromosome of the control samples in the Wellcome Trust Case Control Consortium (WTCCC) 1 study (Burton et al. 2007). Altogether, these data consist of $N = 2,938$ individuals and $J = 5,747$ SNPs. Since the goal of mvMAPIT is to search for variants involved in epistatic interactions, we consider the null model to be satisfied when the phenotypic variation of the synthetic traits are solely driven by additive effects. Here, we first subsample the genotypes for $N = 1,000, 1,750$, and $2,500$ observations. Next, we randomly select 1,000 causal SNPs and simulate continuous phenotypes by using the linear model $\mathbf{Y} = \mathbf{B}\mathbf{X}^\top + \mathbf{E}$. The additive effect sizes for each causal SNP are drawn as $\beta \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_\beta)$ across traits, and then we scale all terms to ensure a narrow-sense heritability of 60%. In these simulations, we vary the correlation of the additive genetic effects such that we have traits with independent additive effects ($v_{\beta,12} = 0$), traits with highly correlated additive effects ($v_{\beta,12} = 0.8$), and traits with perfectly correlated additive effects ($v_{\beta,12} = 1$). We assess the calibration of the P -values that are produced by mvMAPIT during each of the three key steps in its combinatorial hypothesis testing procedure. That is, we evaluate (1) the P -values resulting from the univariate test on each trait, (2) the P -values derived from the covariance test, and (3) the final overall P -value that is computed by combining the first two sets of P -values via Fisher’s method, the harmonic mean, or the Cauchy combination procedure. Note that we expect the P -values from the first univariate test to be well-calibrated since it

is equivalent to the MAPIT model. Figures 1.2 and A.1-A.2 show the quantile-quantile (QQ) plots based on P -values combined using Fisher’s method, while Figures A.3-A.8 depict results while using the harmonic mean and the Cauchy combination test. Similarly, Tables 1.1 and A.1-A.8 show the empirical type I error rates estimated for mvMAPIT at significance levels $P = 0.05, 0.01,$ and $0.001,$ respectively.

Overall, mvMAPIT conservatively controls type 1 error rate, both in the presence of nonzero correlation between additive effects on the two traits and even with small sample sizes in the data. This result holds regardless of how P -values are combined in the model. The QQ-plots of the P -values for all three components in mvMAPIT follow the expected uniform distribution for the univariate and combined analysis. Notably, because of the approximations used to compute the standard error of the test statistic in Eq. 1.13, the multivariate extension of the MQS-based testing procedure in mvMAPIT can result in conservative P -values for the covariance components under the null.

1.3.2 Improved detection of epistatic variants using mvMAPIT in simulations

We test the power of mvMAPIT across different genetic trait architectures via an extensive simulation study (Materials and Methods). Once again, we generate synthetic phenotypes using real genotypes from the 22nd chromosome of the control samples in the WTCCC 1 study (Burton et al. 2007). As a reminder, these data consist of $N = 2,938$ individuals and $J = 5,747$ SNPs. In these simulations, we randomly choose 1,000 causal variants to directly affect the genetic architecture of $D = 2$ phenotypes. All causal SNPs are assumed to have a non-zero additive effect on both traits. Next, we randomly select a set of epistatic variants from the 1,000 causal SNPs and divide them into two interacting groups (again see Materials and Methods). We will denote these groups #1 and #2 as $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively, with $|\mathcal{G}|$ denoting the cardinality of the group. One may interpret the epistatic SNPs in $\mathcal{G}_1$ as being the “hub nodes” in an interaction network where each of

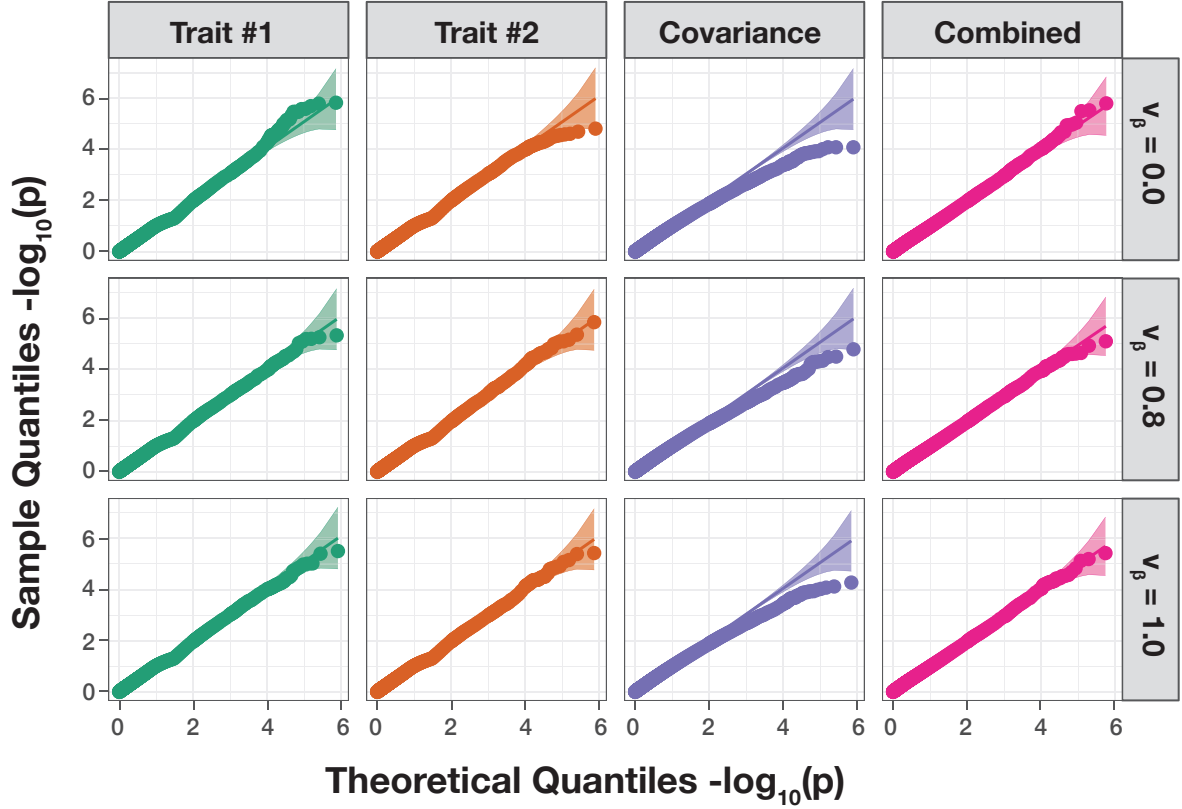


Figure 1.2: **The mvMAPIT framework using Fisher’s method produces well-calibrated P-values when traits are generated by only additive effects (sample size $N = 2,500$ individuals).** In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row of quantile-quantile (QQ) plots corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have independent traits ($v_\beta = 0$), highly correlated traits ($v_\beta = 0.8$), and perfectly correlated traits ($v_\beta = 1$). The first two columns show P -values resulting from the univariate MAPIT test on “trait #1” and “trait #2”, respectively. The third column depicts the “covariance” P -values which corresponds to assessing the pairwise interactions affecting both traits is. Lastly, the fourth column shows the final “combined” P -value which combines the P -values from the first three columns using Fisher’s method. The 95% confidence interval for the null hypothesis of no marginal epistatic effects is shown in grey. Each plot combines results from 100 simulated replicates.

these variants interact with all of the SNPs assigned to $\mathcal{G}_2$. We generate synthetic traits by using the multivariate linear model $\mathbf{Y} = \mathbf{B}\mathbf{X}^\top + \mathbf{A}\mathbf{W}^\top + \mathbf{E}$ where, in addition to previous notation, $\mathbf{W}$ is matrix of interactions between the SNPs assigned to the groups $\mathcal{G}_1$ and $\mathcal{G}_2$. The additive and interaction coefficients for causal SNP effects across traits are drawn

Table 1.1: **The mvMAPIT framework using Fisher’s method preserves type I error rates under the null model when traits are generated by only additive effects (sample size $N = 2,500$ individuals).** In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent additive effects ($v_\beta = 0$), traits with highly correlated additive effects ($v_\beta = 0.8$), and traits with perfectly correlated additive effects ($v_\beta = 1$). We assess the calibration of the P -values that are produced by mvMAPIT during each of the three key steps in its combinatorial hypothesis testing procedure (see Materials and Methods). We show type I error rates resulting from P -values taken from the “univariate” test on each trait independently, the “covariance” P -values which correspond to assessing the pairwise interactions affecting both traits, and the final “combined” P -value. Results are summarized over 100 simulated replicates. Values in the parentheses are the standard deviations across replicates.

Step	Add. Effect Corr.	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
Univariate	$v_\beta = 0.0$	0.030 (1×10^{-2})	0.009 (2×10^{-3})	0.0010 (9×10^{-4})
	$v_\beta = 0.8$	0.030 (1×10^{-2})	0.009 (2×10^{-3})	0.0010 (7×10^{-4})
	$v_\beta = 1.0$	0.030 (1×10^{-2})	0.009 (3×10^{-3})	0.0009 (7×10^{-4})
Covariance	$v_\beta = 0.0$	0.040 (1×10^{-2})	0.006 (2×10^{-3})	0.0003 (3×10^{-4})
	$v_\beta = 0.8$	0.040 (1×10^{-2})	0.007 (2×10^{-3})	0.0004 (5×10^{-4})
	$v_\beta = 1.0$	0.040 (1×10^{-2})	0.006 (2×10^{-3})	0.0003 (4×10^{-4})
Combined	$v_\beta = 0.0$	0.040 (1×10^{-2})	0.006 (2×10^{-3})	0.0003 (3×10^{-4})
	$v_\beta = 0.8$	0.040 (1×10^{-2})	0.007 (2×10^{-3})	0.0004 (5×10^{-4})
	$v_\beta = 1.0$	0.040 (1×10^{-2})	0.006 (2×10^{-3})	0.0003 (4×10^{-4})

as $\beta \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_\beta)$ and $\alpha \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_\alpha)$, respectively. As a final step, we scale all terms to ensure that all genetic effects explain a fixed proportion of the total phenotypic variation.

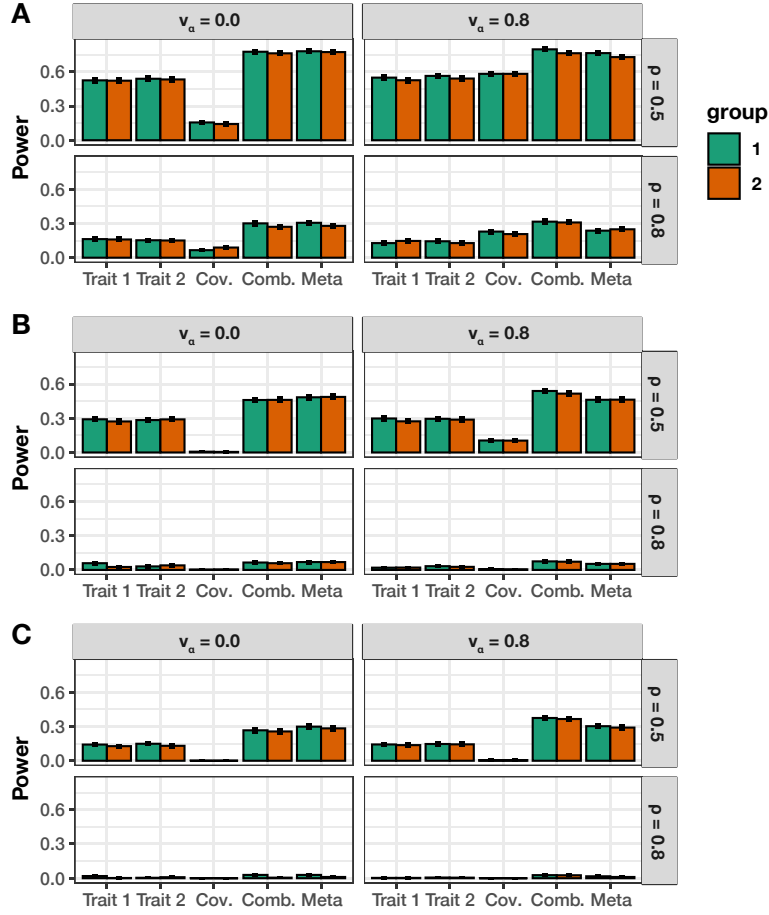


Figure 1.3: **Empirical power of mvMAPIT with Fisher's method to detect group #1 (10) and group #2 (10) epistatic variants across complex traits with moderate broad-sense heritability.** In these simulations, both quantitative traits are simulated to have broad-sense heritability $H^2 = 0.6$ with architectures made up of both additive and epistatic effects. The parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to a setting where the epistatic effects for interactive pairs have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent epistatic effects ($v_\alpha = 0$) and traits with highly correlated epistatic effects ($v_\alpha = 0.8$). This plot shows the empirical power of mvMAPIT at significance levels **(A)** $P = 5 \times 10^{-2}$, **(B)** $P = 5 \times 10^{-4}$, and **(C)** $P = 1 \times 10^{-5}$, respectively. Group #1 and #2 causal markers are colored in green and orange, respectively. For comparison, the “trait #1” and “trait #2” bars correspond to the univariate MAPIT model, the “cov” bars corresponds to power contributed by the covariance test, and “comb” details power from the overall association identified by mvMAPIT in the combination approach. Lastly, the “meta” bars shows power from the association identified by combining only the P -values of the two univariate cases “trait #1” and “trait #2”. Results are based on 100 simulations per parameter combination and the horizontal bars represent standard errors.

We assume a wide-range of simulation scenarios by varying the following parameters:

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

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

The main point of these simulations is to highlight the potential power gained from taking a multivariate approach to epistatic detection. To that end, in each of the simulation scenarios, we assess (i) the power of running the univariate MAPIT model on each trait individually, (ii) the marginal epistatic effects detected by the covariance test, (iii) the power from the overall association identified by mvMAPIT, and (iv) a baseline meta-analytic approach which combines only the P -values from the univariate MAPIT models on each trait. We will hereafter refer to that last model as “meta-MAPIT” for simplicity. Figures 1.3 and A.9-A.11 show the empirical power of the univariate MAPIT model, the covariance test, mvMAPIT, and meta-MAPIT while using Fisher’s method at various multiple hypothesis testing correction thresholds. Figures A.12-A.19 depict the same information but with mvMAPIT and meta-MAPIT using the harmonic mean and Cauchy combination test to combine P -values. We also compare each method’s ability to rank true positives over false positives via receiver operating characteristic (ROC) and precision-recall curves (Figures 1.4 and A.20-A.24). There are several key takeaways from these simulation results. Overall, the ability of the univariate MAPIT framework to detect group #1 and #2 causal variants depends on the proportion of nonadditive phenotypic variation that they explain. This has been shown in previous demonstrations

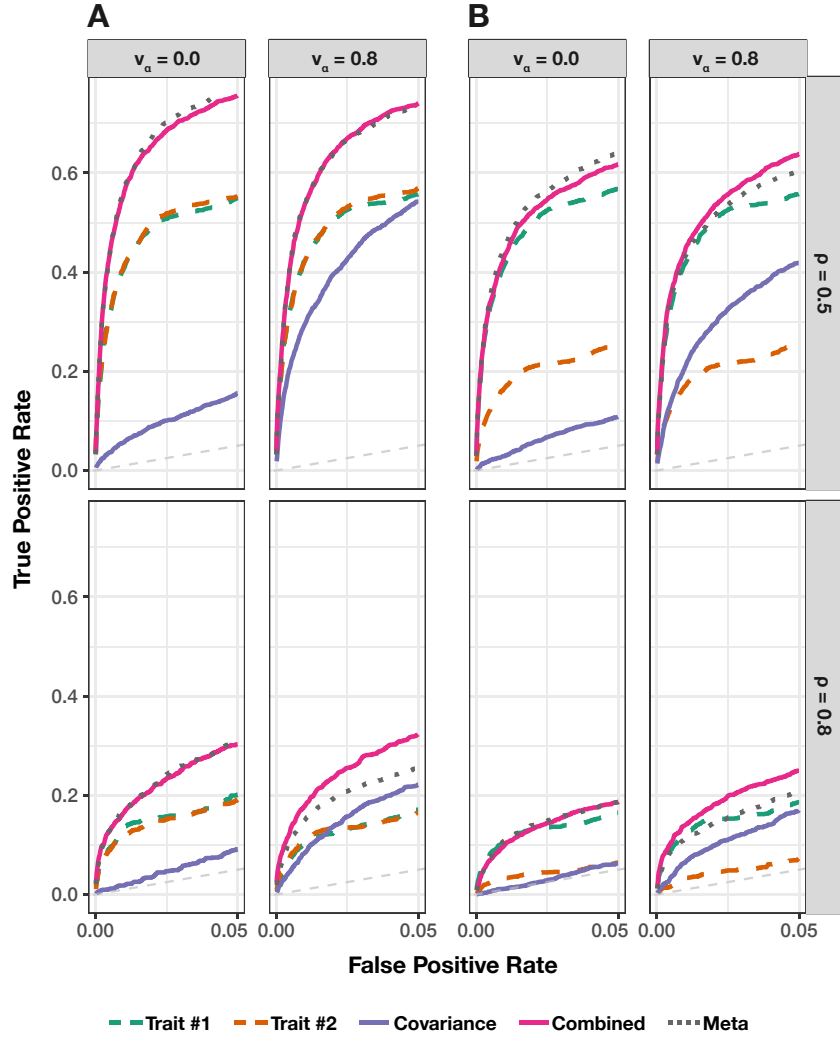


Figure 1.4: **Receiver operating characteristic (ROC) curves comparing the ability of mvMAPIT with Fisher’s method to the univariate MAPIT model in detecting group #1 (10) and group #2 (10) epistatic variants across complex traits.** In panel (A) both traits have broad-sense heritability $H^2 = 0.6$; while in panel (B) one of traits has broad-sense heritability $H^2 = 0.6$ and the other has heritability $H^2 = 0.3$. Across the rows, the parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to settings where the epistatic effects across traits are independent ($v_\alpha = 0$) or highly correlated ($v_\alpha = 0.8$). For comparison, the “trait #1” and “trait #2” dashed lines correspond to the univariate MAPIT model. The “covariance” solid purple line corresponds to power contributed by the covariance test. The “combined” pink line shows power from the overall association identified by mvMAPIT in the multivariate approach. Lastly, the “meta” gray dotted line shows power from the association identified by combining only the P -values of the two univariate cases “trait #1” and “trait #2”. Note that the upper limit of the x-axis (i.e., false positive rate) has been truncated at 0.05. All results are based on 100 simulated replicates.

of the method (Crawford et al. 2017). For example, when there are $|\mathcal{G}_2| = 10$ causal SNPs in group #2, each variant in the set is expected to explain $(1 - \rho)H^2/10\%$ of the genetic variance. As we increase that number of causal SNPs in group #2 to $|\mathcal{G}_2| = 20$, this proportion of variance explained by SNPs in group #2 will decrease which will make it more difficult to prioritize markers involved in interactions. Importantly, it is worth noting that the single-phenotypic test in MAPIT depends on the total interaction effects, rather than individual pairwise effects or the number of interacting pairs. An example of this can be seen by comparing Figure 1.3A to Figure A.9A where the ability to group #1 variants is independent of the number of variants in group #2.

Intuitively, the joint modeling approach of mvMAPIT provides a viable strategy for identifying SNPs contributing to nonadditive variation that would have otherwise gone undetected by univariate methods. The scenario where mvMAPIT shows significant gains over the univariate MAPIT modeling approach is when there is nonzero correlation between the effects of the epistatic interactions shared between traits (e.g., when $v_{\alpha,12} = 0.8$). The sensitivity of the covariance hypothesis test depends on the strength of this correlation which can help increase power when combining over P -values in the final step of mvMAPIT. This becomes increasingly relevant in the low heritability cases. When there is no correlation of interaction effects shared between pairs of traits, combining only the univariate P -values in the meta-MAPIT model performs equally well as mvMAPIT which includes contributions from the covariance statistic. Figures 1.4 and A.20-A.24 demonstrate that the sensitivity of the covariance statistic is comparable to the univariate statistic for highly correlated epistatic effects ($v_{\alpha} = 0.8$) despite genetic variance being predominantly explained by additivity ($\rho = 0.8$). In this case, the multivariate approach performs better than simply doing a meta-analysis on only the univariate P -values. In Figures 1.4, A.10, A.11, A.14, A.15, and A.18-A.24, we simulated synthetic traits such that one has a moderate broad-sense heritability $H^2 = 0.6$ and the other has heritability $H^2 = 0.3$. In these scenarios, detecting variants involved in interactions increased for the

trait with low heritability. In particular, the covariance component analysis is shown to play an important role in this improved detection (e.g., see Figure 1.4B).

1.3.3 Synergistic epistasis in binding affinity landscapes for neutralizing antibodies

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

Once again, we report results after running mvMAPIT with Fisher’s method, the harmonic mean, and the Cauchy combination approach (Table A.9). Figures 1.5A, A.25A, and A.26A show Manhattan plots for P -values corresponding to the trait-specific marginal epistatic tests (i.e., the univariate MAPIT model), the covariance test, and the mvMAPIT approach. Here, green colored dots are positions that have significant marginal epistatic effects beyond a Bonferroni corrected threshold for multiple testing ($P = 0.05/11 = 4.55 \times 10^{-3}$ for CR6261 and $P = 0.05/16 = 3.13 \times 10^{-3}$ for CR9114, respectively). Interestingly, while the univariate MAPIT approach was able to identify

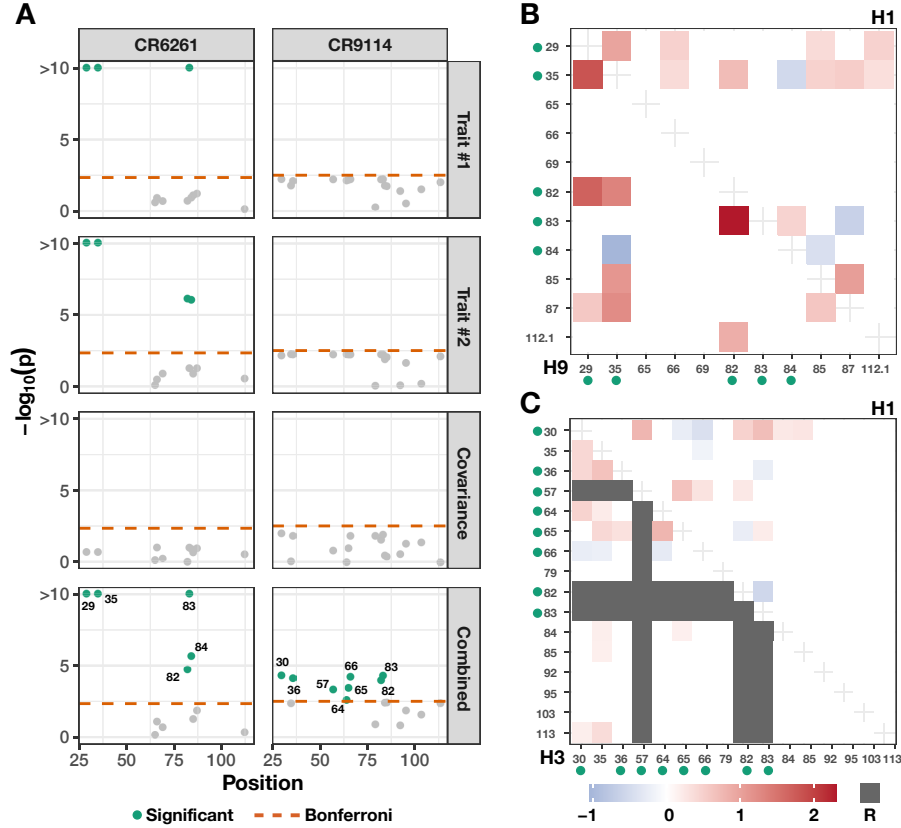


Figure 1.5: **Applying mvMAPIT with Fisher's method to broadly neutralizing antibodies recovers heavy-chain mutations known to be involved in epistatic interactions that affect binding against two influenza strains.** These results are based on protein sequence data from Phillips et al. (2021) who generated a nearly combinatorially complete library for two broadly neutralizing anti-influenza antibodies (bnAbs), CR6261 and CR9114. For each antibody, we assess binding affinity to different influenza strains. For CR6261, traits #1 and #2 are binding measurements to the antigens H_1 and H_9 ; while, for CR9114, we assess the same measurement for H_1 and H_3 . Panel (A) shows Manhattan plots for the different sets of P -values computed during the mvMAPIT analysis. The red horizontal lines indicate a chain-wide Bonferroni corrected significance threshold ($P = 4.55 \times 10^{-3}$ for CR6261 and $P = 3.13 \times 10^{-3}$ for CR9114, respectively). The green colored dots are positions that have significant marginal epistatic effects after multiple correction. Panels (B) and (C) reproduce exhaustive search results originally reported by Phillips et al. (2021). The green dots next to the mutation labels on the axes are the residues that are significant in the multivariate MAPIT analysis and correspond to panel (A). The shaded regions in panel (B) are the regression coefficients for pairwise interactions between positions when assessing binding of CR6261 with H_1 (upper right triangle) and H_9 (lower left triangle). Similarly, panel (C) shows the same information when assessing binding of CR9114 with H_1 (upper right triangle) and H_3 (lower left triangle). Required mutations (indicated by R) are plotted in gray and left out of the analysis (Phillips et al. 2021).

significant marginal epistatic effects for CR6261, it lacked the power to identify significant positions driving nonadditive variation in binding affinity for CR9114. Overall, the combined trait approach in mvMAPIT revealed marginal epistatic effects for positions 29, 35, 82, 83, and 84 in CR6261, and positions 30, 36, 57, 64, 65, 66, 82, and 83 for CR9114. Most notably, these same positions were also identified as contributing to pairwise epistasis by Phillips et al. (2021). In the original study, the authors first ran an exhaustive-search to statistically detect significant interactions and then conducted downstream analyses to find that these positions are likely responsible for the antibodies binding to the influenza surface protein hemagglutinin. The regression coefficients from the exhaustive search, as reported by Phillips et al. (2021), are illustrated in panels B and C of Figures 1.5, A.25, and A.26. Panel B illustrates interaction coefficients when assessing binding of CR6261 with H_1 (upper right triangle) and H_9 (lower left triangle). Panel C shows the same information when assessing binding of CR9114 with H_1 (upper right triangle) and H_9 (lower left triangle). Our results show that mvMAPIT identifies all required mutations in these systems as well as most positions involved in at least one epistatic pair.

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

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

levels of broad-sense heritability with varying contributions from nonadditive genetic effects.

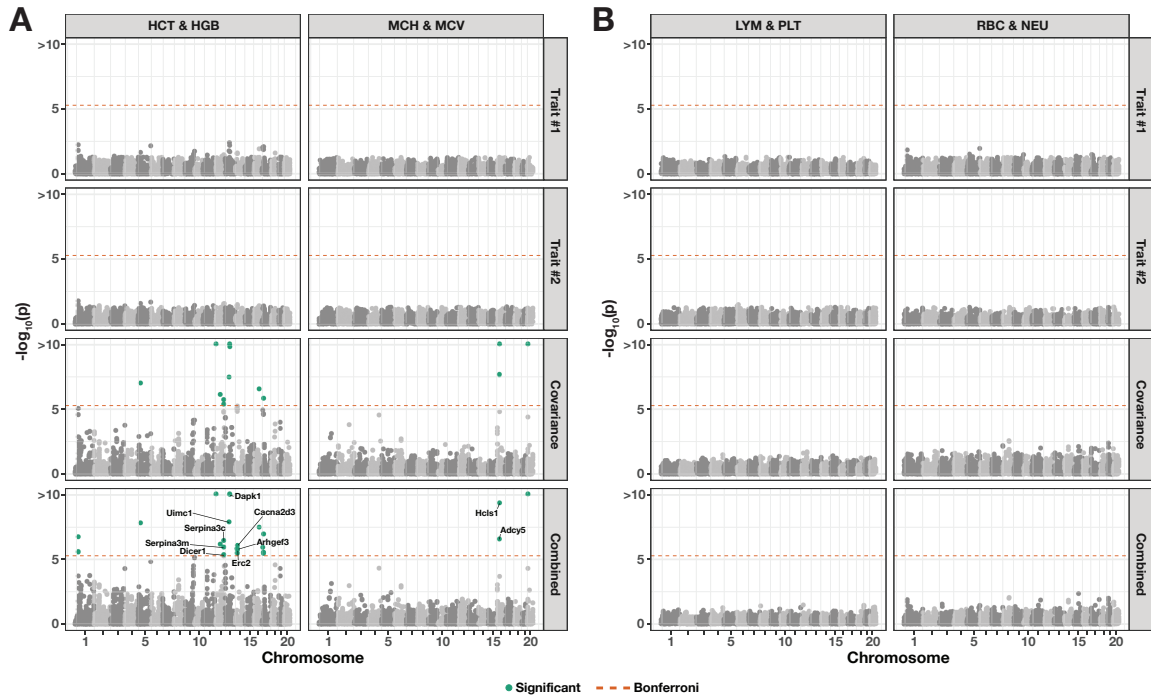


Figure 1.6: Manhattan plots of a genome-wide interaction study for two pairs of hematology traits in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics (Valdar et al. 2006a; Valdar et al. 2006b) using mvMAPIT with Fisher’s method. Panel (A) depicts results for trait pairs hematocrit (HCT) and hemoglobin (HGB) in the left column and mean corpuscular hemoglobin (MCH) and mean corpuscular volume (MCV) in the right column. Panel (B) depicts results for trait pairs lymphocytes (LYM) and platelets (PLT) in the left column and red blood cell count (RBC) and neutrophils (NEU) in the right column. Here, we depict the P -values computed during each step of the mvMAPIT modeling pipeline. The red horizontal lines indicate a genome-wide Bonferroni corrected significance threshold ($P = 4.83 \times 10^{-6}$). The green colored dots are SNPs that have significant marginal epistatic effects after multiple test correction. Significant SNPs were mapped to the closest neighboring genes using the Mouse Genome Informatics database (<http://www.informatics.jax.org>) (Blake et al. 2020; Smith et al. 2009).

For each pairwise trait analysis, we provide a summary table which lists the combined P -values after running mvMAPIT with Fisher’s method and the harmonic mean (Table A.10). We also include results corresponding to the univariate MAPIT model and the covariance test for comparison. Overall, the single-trait marginal epistatic test only identifies significant variants for the large immature cells (LIC) after Bonferroni correction

($P = 4.83 \times 10^{-6}$). A complete picture of this can be seen in Figures [A.27-A.29](#) which depict Manhattan plots of our genome-wide interaction study for all combinations of trait pairs. Here, we can see that most of the signal in the combined P -values from mvMAPIT likely stems from the covariance component portion of the model. This hypothesis holds true for the joint pairwise analysis of (i) hematocrit (HCT) and hemoglobin (HGB) and (ii) mean corpuscular hemoglobin (MCH) and mean corpuscular volume (MCV) (e.g., see the third and fourth rows of Figures [1.6A](#), [A.30A](#), and [A.31A](#)). One explanation for observing more signal in the covariance components over the univariate test could be derived from the traits having low heritability but high correlation between epistatic interaction effects. Recall that our simulation studies showed that the sensitivity of the covariance statistic increased for these cases. As a direct comparison, Figures [1.6B](#), [A.30B](#), [A.31B](#), and [A.32-A.34](#) give examples of analyses that do not identify significant marginal epistatic effects even after using the mvMAPIT approach. This further supports the claim that the signal provided by the covariance statistic is not likely due to inflation.

Notably, the nonadditive signal identified by the covariance test is not totally dependent on the empirical correlation between traits (see Figure [A.35](#)). Instead, as previously shown in our simulation study, the power of mvMAPIT over the univariate approach occurs when there is correlation between the effects of epistatic interactions shared between two traits. Importantly, many of the candidate SNPs selected by the mvMAPIT framework have been previously discovered by past publications as having some functional nonlinear relationship with the traits of interest. For example, the multivariate analysis with traits MCH and MCV show a significant SNP rs4173870 ($P = 4.89 \times 10^{-10}$) in the gene hematopoietic cell-specific Lyn substrate 1 (*Hcls1*) on chromosome 16 which has been shown to play a role in differentiation of erythrocytes (Castro-Ochoa et al. [2019](#)). Similarly, the joint analysis of HGB and HCT shows hits in multiple coding regions. One example here are the SNPs rs3692165 ($P = 1.82 \times 10^{-6}$) and rs13482117 ($P = 8.94 \times 10^{-7}$) in the gene calcium voltage-gated channel auxiliary subunit alpha2delta 3 (*Cacna2d3*) on chromosome

14, which has been associated with decreased circulating glucose levels (Database 2014), and SNP rs3724260 ($P = 4.58 \times 10^{-6}$) in the gene *Dicer1* on chromosome 12 which has been annotated for anemia both in humans and mice (Raaijmakers et al. 2010).

Tabs. 1.2 and A.11 list a select subset of SNPs in coding regions of genes that have been associated with phenotypes related to the hematopoietic system, immune system, or homeostasis and metabolism. Each of these are significant (after correction for multiple hypothesis testing) in the mvMAPIT analysis of related hematology traits. Some of these phenotypes have been reported as having large broad-sense heritability, which improves the ability of mvMAPIT to detect the signal. For example, the genes *Arf2* and *Cacna2d3* are associated with phenotypes related to glucose homeostasis, which has been reported to have a large heritable component (estimated $H^2 = 0.3$ for insulin sensitivity (Norris et al. 2012)). Similarly, the genes *App* and *Pex1* are associated with thrombosis where (an estimated) more than half of phenotypic variation has been attributed to genetic effects (estimated $H^2 \geq 0.6$ for susceptibility to common thrombosis (Souto et al. 2000)).

As a final analysis, Figures A.36 and A.37 illustrate the possibility of using the mvMAPIT framework to jointly analyze any number of traits. These show both Manhattan plots and QQ-plots corresponding to the application of mvMAPIT to a subset or all of the 15 hematology traits measured in the heterogeneous stock of mice data set. In this analysis, all univariate variance and all pairwise covariance test statistics are combined. Figure A.36 shows an inflation of smaller P -values that is reduced by excluding traits measuring abnormal lymphocytes and large immature cells (i.e., those traits that show strong signal in for the univariate analyses; see diagonal panels in Figures A.27-A.29). Results where these two phenotypes are excluded can be found in Figure A.37 for direct comparison.

Table 1.2: Notable SNPs with marginal epistatic effects after applying the mvMAPIT framework to 15 hematology traits in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics (Valdar et al. 2006a; Valdar et al. 2006b; *Wellcome Trust Centre for Human Genetics - Mouse Resources 2022*). In the first two columns, we list SNPs and their genetic location according to the mouse assembly NCBI build 34 (accessed from Shifman et al. (2006)) in the format Chromosome:Basepair. Next, we give the results stemming from univariate analyses on traits #1 and #2, respectively, the covariance (cov) test, and the overall P -value derived by mvMAPIT using Fisher’s method. The last columns detail the closest neighboring genes found using the Mouse Genome Informatics database (<http://www.informatics.jax.org>) (Blake et al. 2020; Smith et al. 2009), a short summary of the suggested annotated function for those genes, and the reference to the source of the annotation. See Table A.10 for the complete list of SNP and SNP-set level results. SNP location, genomic annotation and references can be found in Tab. A.11.

SNP	Trait #1	Trait #2	Trait #1 P-value	Trait #2 P-value	Cov. P-value	Comb. P-value	Gene
rs3699393	MCV	PLT	0.21	0.23	5.75×10^{-7}	4.9×10^{-6}	<i>Upf2</i>
rs13478092	LIC	PLT	0.034	0.58	1.67×10^{-10}	1.26×10^{-9}	<i>Pex1</i>
rs3694887	ALY	LIC	1.26×10^{-4}	0.013	2.54×10^{-6}	1.55×10^{-9}	<i>Aff1</i>
rs3694887	LIC	PLT	0.013	0.28	5.47×10^{-27}	4.49×10^{-26}	<i>Aff1</i>
rs13478923	ALY	LIC	2.8×10^{-4}	0.12	1.79×10^{-6}	1.81×10^{-8}	<i>Foxp1</i>
rs13478924	ALY	LIC	3.11×10^{-4}	0.12	2.70×10^{-6}	2.86×10^{-8}	<i>Foxp1</i>
rs13478985	MCV	WBC	0.16	0.40	1.14×10^{-81}	1.34×10^{-78}	<i>Atg7</i>
rs3723163	HCT	LYM	0.072	0.30	3.99×10^{-107}	2.66×10^{-104}	<i>Arf2</i>
rs3723163	HGB	WBC	0.069	0.25	1.85×10^{-7}	6.76×10^{-7}	<i>Arf2</i>
rs3724260	HGB	HCT	0.030	0.062	1.44×10^{-5}	4.58×10^{-6}	<i>Dicer1</i>
rs3692165	HCT	HGB	0.026	0.037	9.9×10^{-6}	1.8×10^{-6}	<i>Cacna2d3</i>
rs13482117	HCT	HGB	0.023	0.03	5.9×10^{-6}	9.0×10^{-7}	<i>Cacna2d3</i>
rs13482288	ALY	BAS	0.036	0.65	1.78×10^{-8}	1.1×10^{-7}	<i>Tdld3</i>
rs4173870	MCH	MCV	0.14	0.71	1.20×10^{-11}	4.89×10^{-10}	<i>Hcls1</i>
rs4212102	PLT	WBC	0.17	0.35	1.16×10^{-10}	2.44×10^{-9}	<i>App</i>
rs4212186	PLT	WBC	0.17	0.36	5.88×10^{-11}	1.31×10^{-9}	<i>App</i>
rs3711994	ALY	LYM	3.71×10^{-4}	0.10	1.04×10^{-12}	2.80×10^{-14}	<i>Btrc</i>

1.4 Discussion

The marginal epistatic testing strategy offers an alternative to traditional epistatic mapping methods by seeking to identify variants that exhibit non-zero interaction effects with any other variant in the data (Crawford et al. 2017; Crawford et al. 2018a; Turchin et al. 2020). This framework has been shown to drastically reduce the number of statistical tests needed to uncover evidence of significant nonadditive variation in complex traits and, as a result, alleviates much of the empirical power concerns and heavy computational burden associated with explicit search-based methods. Still, models testing for marginal epistasis can be underpowered when applied to traits with low heritability or to “polygenic” traits where the interactions between mutations have small effect sizes (Crawford et al. 2017). In this work, we present the “multivariate MArginal ePIstasis Test” (mvMAPIT), a multi-outcome extension of the univariate marginal epistatic framework (Figure 1.1). Theoretically, we formulate mvMAPIT as a multivariate linear mixed model (mvLMM) where its ability to jointly analyze any number of traits relies on a generalized “variance-covariance” component estimation algorithm (Zhou 2017). Through extensive simulations, we show that mvMAPIT preserves type I error rates and produces well-calibrated P -values under the null model when traits are generated only by additive effects (Figures 1.2 and A.1-A.8, and Tables 1.1 and A.1-A.8). In these simulation studies, we also show that mvMAPIT improves upon the identification of epistatic variants over the univariate test when there is correlation between the effects of genetic interactions shared between multiple traits (Figures 1.3, 1.4, and A.9-A.24). By analyzing two real data sets, we demonstrated the ability of mvMAPIT to recover heavy-chain mutations known to be involved in epistatic interactions that affect binding against two influenza strains (Phillips et al. 2021) (Figures 1.5, A.25, and A.26, and Table A.9) as well as to identify hematology trait relevant epistatic SNPs in heterogeneous stock of mice (Valdar et al. 2006a; Valdar et al. 2006b) that have also been detected in previous publications and functional validation studies (Figures 1.6 and A.27-A.37, and Tables 1.2, A.10 and A.10). Lastly, we have made

mvMAPIT an open-source R software package with documentation to facilitate its use by the greater scientific community.

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

The current implementation of the mvMAPIT framework offers many directions for future development and applications. First, like other marginal epistatic mapping methods, mvMAPIT is unable to directly identify detailed interaction pairs despite being able to identify SNPs that are involved in epistasis. As shown through our simulations and real data analyses, being able to identify SNPs involved in epistasis allows us to come up with an initial (likely) set of variants that are worth further exploration, and thus represents an important first step towards identifying and understanding detailed epistatic associations. In previous studies (Crawford et al. 2017; Lewinger et al. 2013; Zhang et al. 2010; Pecanka et al. 2021), two-step *ad hoc* procedures have been suggested where, in our case, we would first run mvMAPIT and then focus on significant SNPs from the first step to further test all of the pairwise interactions among them in order to identify specific epistatic interaction pairs. While this approach has been shown to be effective in univariate (single-trait) analyses, this two-step procedure is still *ad hoc* in nature and

could miss important epistatic associations. Exploring robust ways unify these two steps in a joint fashion would be an interesting area for future research. Second, in its current implementation, mvMAPIT can be computationally expensive for data sets with large sample sizes (e.g., hundreds of thousands of individuals in a biobank-scale study). In this study, we develop a “variance-covariance component” extension to the MQS algorithm to estimate parameters in the mvMAPIT model. Theoretically, MQS is based on the method of moments and produces estimates that are mathematically identical to the Haseman-Elston (HE) cross-product regression (Zhou 2017; Zhu et al. 2020; Lee et al. 2011). In practice, MQS is not only computationally more efficient than HE regression, but also provides a simple, analytic estimation form that allows for exact P -value computation — thus alleviating the need for jackknife re-sampling procedures (Golan et al. 2014) that both are computationally expensive and rely on assumptions of independence across individuals in the data (Churchill et al. 2008). Exploring different ways to reliably fit large-scale mvLMMs with multiple random effects is a consideration for future work. For example, as an alternative, recent studies have proposed randomized multi-component versions of HE regression for heritability estimation which scale up to data sets with millions of individuals and variants, respectively (Pazokitoroudi et al. 2020; Wu et al. 2018; Kerin et al. 2020a). It would be interesting to develop a well-calibrated hypothesis test within the randomized HE regression setting so that it may be implemented within the mvMAPIT software for association mapping.

In the future, we plan to expand the mvMAPIT framework to also identify individual variants contributing other sources of nonadditive genetic variation such as gene-by-environment ($G \times E$) or gene-by-sex ($G \times \text{Sex}$) interactions. We can do this by manipulating the marginal epistatic covariance matrix in Eq. 1.10 to encode how loci interact with one or more environmental instruments (Moore et al. 2019; Kerin et al. 2020a; Kerin et al. 2020b; Zhu et al. 2022). Lastly, we have focused here on applying mvMAPIT to simple quantitative traits. However, there are many important traits with significant

nonadditive genetic components in plants, animals, and humans that cannot be easily reduced to simple scalar values. Examples include longitudinal traits such as growth curves (Campbell et al. 2018), metabolic traits such as the relative concentrations of different families of metabolites (Chan et al. 2011), and morphological traits such as shape or color (Demmings et al. 2019). Indeed, each of these traits can be decomposed into vectors of interrelated components, but treating these components as independent phenotypes within existing univariate epistatic mapping tools would be inefficient because of their statistical dependence. As an alternative, the mvMAPIT framework can be used to make joint inferences about epistasis across any number of correlated phenotypic components—which, in the case of longitudinal studies for example (Ko et al. 2022; Wu et al. 2019; Couto Alves et al. 2019; Cousminer et al. 2021), could be used to interrogate how nonadditive variation of trait architecture changes or evolves over time.

1.5 URLs

Multivariate marginal epistasis test (mvMAPIT) software, <https://github.com/lcrawlab/mvMAPIT>; univariate marginal epistasis test (MAPIT) software, <https://github.com/lcrawlab/mvMAPIT>; Wellcome Trust Centre for Human Genetics, <http://mtweb.cs.ucl.ac.uk/mus/www/mouse/index.shtml>; Mouse Genome Informatics database, <http://www.informatics.jax.org>.

1.5.1 Data and software availability

Source code, tutorials, and tutorials for implementing the “multivariate MArginal ePIstasis Test” are publicly available as an R package which is available online at <https://github.com/lcrawlab/mvMAPIT>. We use the `CompQuadForm` R package (Duchesne et al. 2010) to compute P -values from the Davies method. The Davies method can sometimes yield a P -value equal exactly to 0 when the true P -value is extremely small (Duchesne et al. 2010). If this is of concern, one can compute the P -values for MAPIT using

Kuonen’s saddlepoint method (Kuonen 1999) or Satterthwaite’s approximation equation (Satterthwaite 1946). In the current implementation of mvMAPIT, the saddlepoint approximation is performed if the Davies method returns with error. We wrote our own function to combine P -values using Fisher’s method which is largely inspired by functions in the `metap` R package (Dewey 2022). We use the `harmonicmeanp` R package (Wilson 2019) to combine P -values using the harmonic mean. We also write our own function to perform the Cauchy combination test based on work developed in Liu et al. (2020). Full package documentation can be found at <https://lcrawl.github.io/mvMAPIT/>. Data to reproduce figures for the broadly neutralizing antibodies as well as the mice study can be found at <https://doi.org/10.7910/DVN/WPFIGU> (Stamp et al. 2022).

Data about the binding affinity landscapes for neutralizing antibodies were downloaded directly from Phillips et al. (2021). Information about mice data set from the Wellcome Trust Centre for Human Genetics (Valdar et al. 2006a; Valdar et al. 2006b) can be found at <http://mtweb.cs.ucl.ac.uk/mus/www/mouse/index.shtml>. The genotypes from this study were downloaded using the `BGLR-R` package (Pérez et al. 2014). Details about the mice phenotypes can be found <http://mtweb.cs.ucl.ac.uk/mus/www/mouse/HS/index.shtml> and hematological traits can be downloaded from the mvMAPIT package. In the real data analyses, SNPs were mapped to the closest neighboring genes using the Mouse Genome Informatics database (<http://www.informatics.jax.org>) (Blake et al. 2020).

Frailty model approach for detecting gene-by-environment interactions in time-to-event association analysis

2.1 Introduction

Genome-wide association studies (GWAS) have shaped our understanding of the genetic architecture of complex human traits and helped identify variants associated with disease risk (Visscher et al. [2017](#)). The majority of these studies employ models that assume genetic effects are constant across individuals within populations. While this modelling approach has proven computationally efficient, it ignores the possibility of the environment modulating genetic effects. Substantial evidence suggests that lifestyle factors such as physical activity and environmental exposures influence the genetic risk of common disease, e.g. obesity (Kim et al. [2024](#)), cancer (Cho et al. [2025](#)), and asthma (Koppelman et al. [2025](#)). A growing number of large population cohorts now combine genetic profiles with

detailed measurements of complex traits and environmental exposures (Bycroft et al. 2018; Bick et al. 2024), enabling systematic investigation of gene-environment ($G \times E$) interactions and how the effects of genetic variants on phenotypes depend on environmental context.

Regression-based $G \times E$ analyses evaluate one variant-environment pair at a time (Kraft et al. 2007). In contrast, variance-component models treat $G \times E$ effects as random terms, enabling joint inference across multiple environmental variables and alleviating the multiple-testing burden (Zhong et al. 2023; Zhou et al. 2025; Moore et al. 2019). Most of these approaches have been developed primarily for quantitative or binary traits, however, recent studies demonstrate that modeling time-to-event outcomes for disease onset can uncover associations that remain undetected in logistic mixed-effects frameworks (Dey et al. 2022).

In many diseases, patient survival time is a crucial clinical outcome of disease progression and evaluating potential therapeutic targets. The Cox proportional hazards model (Cox 1972) has been modified for genome-wide association studies (GWAS) to handle time-to-event outcomes (Dey et al. 2022). A major challenge in time-to-event studies is censoring, e.g. due to patient loss to follow-up, which creates incomplete survival records. Since many traits of interest develop over periods of time longer than the duration of studies, it becomes essential to utilize all available data, including information from patients with incomplete follow-up or no event observed.

In this study, we develop a method to investigate the role of $G \times E$ interactions on time-to-event outcomes such as time to disease onset or death by integrating genomic and clinical patient data with environmental data. We introduce the Cox proportional hazards gene-by-environment interaction test, a method that utilizes a multivariate Gaussian frailty model reformulated as a Poisson generalized linear mixed model (GLMM) (Whitehead 1980; Dey et al. 2022). Our approach employs a variance component score test to enhance computational efficiency for analysis (Therneau et al. 2000).

2.2 Methods

2.2.1 The Cox proportional hazards gene-by-environment interaction test

CphGxE performs association testing for gene-by-environment interactions in censored time-to-event data. Assume that we are analyzing a GWAS data set with N samples where $\mathbf{X}$ is a $N \times J$ matrix of patient genotypes with J denoting the number of SNPs (encoded as 0, 1, or 2 copies of a reference allele at each locus j), $\mathbf{E}$ is $N \times L$ matrix of environmental exposure variables, with L denoting the number of variables, and $\mathbf{Z}$ a $N \times M$ matrix containing the patient covariate data for M covariates. The time-to-event phenotype $(\boldsymbol{\delta}, \mathbf{t})$ consists of $\boldsymbol{\delta}$, an N -dimensional indicator vector and $\mathbf{t}$, an N -dimensional vector of observed times. The genetic information for the i -th sample to test for interactions involving the k -th genetic variant is incorporated into individual frailties η_{ik}

$$\eta_{ik} = \mathbf{z}_i^T \boldsymbol{\beta} + m_i + e_i + g_{ik} \quad (2.1)$$

where $\boldsymbol{\beta}$ are the fixed effects of the covariates $\mathbf{z}_i$. Both $\boldsymbol{\beta}$ and $\mathbf{z}_i$ are M -dimensional vectors. The frailty contributions m_i , e_i , and g_{ik} are the polygenic background, environmental exposure, and gene-by-environment interaction, respectively. The subscript i denotes the i -th sample and k denotes the model setup to test for interactions of the k -th genetic variant with the environmental exposure.

For the i -th subject, we can write a conditional hazard function as the baseline hazard $\lambda_0(t)$ at time t which is modulated by the individual frailties η_{ik}

$$\lambda_i(t|\eta_{ik}) = \lambda_0(t) \exp \{ \eta_{ik} \} \quad (2.2)$$

to test for interaction of the k -th genetic variant and the environment. The frailty model incorporates a linear mixed model (LMM) that controls for population structure and

confounding variables. Instead of using an intercept term in the LMM, the population mean is modeled by the baseline hazard. For CphGxE we test the null hypothesis of no interaction between the k -th genetic variant and the environmental exposure $H_0 : \sigma^2 = 0$.

Association testing in CphGxE proceeds in two steps. First, the null model is fit by iteratively estimating the variance components of the polygenic background and environmental exposure, as well as the fixed effect sizes of the covariates, under a frailty model without the interaction term. Second, CphGxE tests for gene-by-environment interaction by performing a hypothesis test for interaction between the k -th genetic variant and the environmental exposure.

To be able to deal with random effects, we reformulate the model likelihood as a Poisson generalized linear mixed model. Under the null hypothesis (no gene-by-environment interactions), we fit the model using an iterative average information (AI) residual maximum likelihood (REML) approach (Gilmour et al. 1995). For computational efficiency, rather than maximizing the full likelihood, we compute an efficient score statistic which follows a χ_1^2 distribution. For small sample sizes, the asymptotic assumptions do not hold, and the test statistic produces inflated type I error rates. We therefore recommend to perform a permutation test for calibrated P -values at small sample sizes (see Appendix B.5).

2.2.2 The likelihood function for censored time-to-event data

For each subject i , we observe the time-to-event (δ_i, t_i) where δ_i is an indicator variable (1 if the event occurred during follow-up, 0 if censored) and t_i represents either the time to event or censoring time. The individual observed times could for example be the survival of patients, the time to onset of a disease, or the time until symptoms resolve. Censoring occurs when the event has not been observed at the time of data collection, as well as when patients are lost to follow-up during the investigation.

More specifically, for the i -th subjects we let τ_i and ζ_i denote the time-to-event and

the censoring time respectively. The observed time is $t_i = \min(\tau_i, \zeta_i)$ and $\delta_i = \mathbb{1}(\tau_i \leq \zeta_i)$ with $\mathbb{1}(\cdot)$ the indicator function.

In the following, we assume the matrices $\mathbf{X}$, $\mathbf{E}$, and $\mathbf{Z}$ are already mean centered and scaled across all samples. The genetic information to test for interactions involving the k -th genetic variant is incorporated into frailties $\boldsymbol{\eta}_k$

$$\boldsymbol{\eta}_k = \mathbf{Z}\boldsymbol{\beta} + \mathbf{m} + \mathbf{e} + \mathbf{g}_k \quad (2.3)$$

where $\boldsymbol{\beta}$ are the fixed effects of the M covariates and $\mathbf{m}$, $\mathbf{e}$, and $\mathbf{g}_k$ are N -dimensional vectors of genetic frailties corresponding to the polygenic background, environmental exposure, and gene-by-environment interaction, respectively. These frailties are modeled as multivariate normal random effects $\mathbf{m} \sim \mathcal{N}(0, h^2\mathbf{K})$, $\mathbf{e} \sim \mathcal{N}(0, \tau^2\mathbf{H})$, and $\mathbf{g}_k \sim \mathcal{N}(0, \sigma^2\mathbf{G}_k)$. The covariance matrix $\mathbf{K} = \mathbf{X}\mathbf{X}^\top/J$ accounts for the relatedness between individuals in the data and the corresponding component h^2 models the narrow-sense heritability while $\mathbf{H} = \mathbf{E}\mathbf{E}^\top/L$ models the frailty variance explained by the similarity of environmental exposure between individuals. The covariance matrix $\mathbf{G}_k$ represents all pairwise gene-by-environment interactions involving the k -th genetic variant and is constructed from the exposure similarity matrix as $\mathbf{G}_k = \mathbf{D}_k\mathbf{H}\mathbf{D}_k$ where $\mathbf{D}_k = \text{diag}(\mathbf{x}_k)$ is an $N \times N$ diagonal matrix constructed from the genotype $\mathbf{x}_k$ to be tested.

For the derivations of the likelihood and the estimators, we define the random effect of the frailties as $\mathbf{r}_k = \mathbf{m} + \mathbf{e} + \mathbf{g}_k$ a mixture of polygenic background, environmental exposure, and gene-by-environment interactions that is multivariate normal distributed as $\mathbf{r}_k \sim \mathcal{N}(0, \mathbf{V}_k(\boldsymbol{\theta}))$ with $\boldsymbol{\theta} = (h^2, \tau^2, \sigma^2)$ and the covariance matrix $\mathbf{V}_k(\boldsymbol{\theta}) = h^2\mathbf{K} + \tau^2\mathbf{H} + \sigma^2\mathbf{G}_k$. Accordingly, the conditional hazard function depends on the random effect as given in Eq. 2.4.

$$\lambda(t|\mathbf{z}_i, \boldsymbol{\beta}, r_{ik}) = \lambda_0(t) \exp \left\{ \underbrace{\mathbf{z}_i^\top \boldsymbol{\beta} + r_{ik}}_{\eta_{ik}} \right\} \quad (2.4)$$

The frailty incorporates a linear mixed model (LMM) that controls for population structure, as well as confounding variables. Instead of using an intercept term in the LMM, the population mean is modeled by the baseline hazard. For CphGxE we test the null hypothesis of no interaction between the k -th genetic variant and the environmental exposure $H_0 : \sigma^2 = 0$.

To find the likelihood we not only need the hazard function but also the survival function $S(t|\mathbf{z}_i, \boldsymbol{\beta}, r_{ik})$. This function represents the probability that subject i survives beyond time t given their frailties η_{ik} and can be written as

$$S(t|\mathbf{z}_i, \boldsymbol{\beta}, r_{ik}) = \exp(-\Lambda_0(t) \cdot \exp(\mathbf{z}_i^T \boldsymbol{\beta} + r_{ik})) \quad (2.5)$$

with the cumulative baseline hazard function $\Lambda_0(t)$. The cumulative baseline hazard function $\Lambda_0(t)$ represents the accumulated risk of experiencing an event up to time t in the absence of any genetic effects or covariates. It is defined as the integral of the baseline hazard function $\lambda_0(t)$ from time 0 to time t :

$$\Lambda_0(t) = \int_0^t \lambda_0(u) du \quad (2.6)$$

This function characterizes the baseline risk profile of the population. In the context of our frailty model, it serves as the population decay rate which is modified by individual frailties and can be estimated from the data using Breslow's estimator (see Appendix B.2).

The probability distribution for observing a time to a given event is expressed as the product of the hazard and the survival likelihood (see Appendix B.1). The full likelihood of the observed data then is given by Eq. 2.7.

$$\begin{aligned}
L(\lambda_0, \boldsymbol{\beta}, \boldsymbol{\theta}) &= \int_{\mathbb{R}^N} \prod_{i=1}^N \left[\lambda(t_i; \mathbf{z}_i, \boldsymbol{\beta}, r_{ik})^{\delta_i} S(t_i; \mathbf{z}_i, \boldsymbol{\beta}, r_{ik}) \right] \\
&\quad \times (2\pi)^{-\frac{N}{2}} \frac{1}{\sqrt{|\mathbf{V}_k(\boldsymbol{\theta})|}} \exp\left(-\frac{1}{2} \mathbf{r}_k^T (\mathbf{V}_k(\boldsymbol{\theta}))^{-1} \mathbf{r}_k\right) d\mathbf{r}_k
\end{aligned} \tag{2.7}$$

In this likelihood, we marginalize over the frailty random effects $\mathbf{r}_k$. To arrive at a tractable expression, we can follow an approximate inference approach (Breslow et al. 1993; McCulloch et al. 2005) for generalized linear mixed models (GLMMs) to find the penalized log-likelihood in Eq. 2.8 (see Appendix B.3 and B.4).

$$\begin{aligned}
l_p(\lambda_0, \boldsymbol{\beta}, \boldsymbol{\theta}) &= \sum_{i=1}^n [\delta_i (\log \lambda_0(t_i) + \mathbf{z}_i^T \boldsymbol{\beta} + \hat{r}_{ik}) - \Lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta} + \hat{r}_{ik})] \\
&\quad - \frac{1}{2} \hat{\mathbf{r}}_k^T (\mathbf{V}_k(\boldsymbol{\theta}))^{-1} \hat{\mathbf{r}}_k
\end{aligned} \tag{2.8}$$

This likelihood can be maximized using the AI-REML algorithm to find the maximum likelihood estimators for the baseline hazard λ_0 and the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$.

2.2.3 Auxiliary Poisson GLMM for time-to-event data

The Cox Proportional Hazards model, while computationally efficient, lacks the ability to incorporate random effects. To overcome this limitation, we can employ an auxiliary Poisson GLMM for survival data analysis, which yields equivalent statistical inferences to the Cox model (Whitehead 1980). This approach involves partitioning the observed time intervals into small segments and modeling the event counts within each interval as a Poisson process.

In order to derive efficient estimators for the parameters in the penalized log-likelihood (Eq. 2.8), we make use of a working response vector derived as first order Taylor series expansion of the link function around the mean of the response variable (McCulloch et al.

2005). As such, the working response Y_i is given by

$$\begin{aligned} Y_i &= \eta_i + g'(\mu_i)(\delta_i - \mu_i) \\ &= \mathbf{z}_i^T \boldsymbol{\beta} + r_{ik} + \frac{\delta_i - \Lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta} + r_{ik})}{\Lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta} + r_{ik})} \end{aligned} \quad (2.9)$$

for a logarithmic link function $g(x) = \log(x)$ and $\mu_i = \Lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta} + r_{ik})$ (Breslow et al. 1993; McCulloch et al. 2005). The working response is useful for the iterative fitting algorithm as it linearizes the problem and allows for the use of iterative weighted least squares (IWLS) to find the maximum likelihood estimators for the baseline hazard λ_0 and the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ (Breslow et al. 1993; McCulloch et al. 2005; Lee et al. 1996; Stroup 2012; Jiang 2007).

2.2.4 Fitting the null model

To construct an efficient score test for the interaction variance component, we first fit the null model ($H_0 : \sigma^2 = 0$) to obtain the MLEs of the variance components $\boldsymbol{\theta}_0 = (h^2, \tau^2, 0)$, and the fixed-effect coefficients $\boldsymbol{\beta}$, using an iterative procedure (Chen et al. 2016). Under the Cox proportional hazards model, we use the working vector $\mathbf{Y} = \boldsymbol{\eta} + \mathbf{W}^{-1}(\boldsymbol{\delta} - \boldsymbol{\mu})$ where $\mathbf{W} = \text{diag}(\boldsymbol{\mu})$ and $\mu_i = \Lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta} + r_{ik})$.

The maximum likelihood estimators for the fixed and random effects can be found by computing the derivatives of the penalized log-likelihood (Eq. 2.8) and setting them to zero. They can be written as the solution to the following system of equations (Chen et al. 2016; Dey et al. 2022).

$$\begin{bmatrix} \mathbf{Z}^T \mathbf{W} \mathbf{Z} & \mathbf{Z}^T \mathbf{W} \\ \mathbf{W} \mathbf{Z} & \mathbf{W} + (\mathbf{V}(\boldsymbol{\theta}_0))^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{r}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{Z}^T \mathbf{W} \mathbf{Y} \\ \mathbf{W} \mathbf{Y} \end{bmatrix} \quad (2.10)$$

Let $\Sigma = \mathbf{W}^{-1} + \mathbf{V}(\boldsymbol{\theta}_0) = \mathbf{W}^{-1} + h^2 \mathbf{K} + \tau^2 \mathbf{H}$ to write the maximum likelihood estimators as given in Eq. 2.11.

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}^T \Sigma^{-1} \mathbf{Z})^{-1} \mathbf{Z}^T \Sigma^{-1} \mathbf{Y} \quad (2.11)$$

$$\hat{\mathbf{r}}_0 = \mathbf{V}(\boldsymbol{\theta}_0) \Sigma^{-1} (\mathbf{Y} - \mathbf{Z} \hat{\boldsymbol{\beta}}) \quad (2.12)$$

With $\mathbf{P} = \Sigma^{-1} - \Sigma^{-1} \mathbf{Z} (\mathbf{Z}^T \Sigma^{-1} \mathbf{Z})^{-1} \mathbf{Z}^T \Sigma^{-1}$ we can write the residual maximum likelihood (REML) as Eq. 2.13 (Chen et al. 2016; Dey et al. 2022; Breslow et al. 1993).

$$l_R(\lambda_0, \boldsymbol{\beta}, \boldsymbol{\theta}_0) = -\frac{1}{2} \log |\Sigma| - \frac{1}{2} \log |\mathbf{Z}^T \Sigma^{-1} \mathbf{Z}| - \frac{1}{2} \mathbf{Y}^T \mathbf{P} \mathbf{Y} \quad (2.13)$$

Score and (average) information for variance components The AI-REML updates require the score (first derivative) and an information matrix for the variance components. For component θ_i with associated covariance matrix $\mathbf{A}_i \in \{\mathbf{K}, \mathbf{H}\}$, the REML score is

$$U_{\theta_i} = \frac{1}{2} \left\{ \mathbf{Y}^T \mathbf{P} \mathbf{A}_i \mathbf{P} \mathbf{Y} - \text{tr}(\mathbf{P} \mathbf{A}_i) \right\}. \quad (2.14)$$

The observed and expected information entries are

$$\mathcal{I}_{ij}^{\text{obs}} = -\frac{\partial^2 l_R}{\partial \theta_i \partial \theta_j} = \frac{1}{2} \mathbf{Y}^T \mathbf{P} \mathbf{A}_i \mathbf{P} \mathbf{A}_j \mathbf{P} \mathbf{Y} - \frac{1}{2} \text{tr}(\mathbf{P} \mathbf{A}_i \mathbf{P} \mathbf{A}_j), \quad (2.15)$$

$$\mathcal{I}_{ij}^{\text{exp}} = \mathbb{E}[\mathcal{I}_{ij}^{\text{obs}}] = \frac{1}{2} \text{tr}(\mathbf{P} \mathbf{A}_i \mathbf{P} \mathbf{A}_j). \quad (2.16)$$

The *average information* (AI) matrix is the arithmetic mean of (2.15) and (2.16) and simplifies to the computationally efficient form

$$\text{AI}_{ij} = \frac{1}{2} \mathbf{Y}^\top \mathbf{P} \mathbf{A}_i \mathbf{P} \mathbf{A}_j \mathbf{P} \mathbf{Y}, \quad (2.17)$$

which avoids some trace evaluations inside the iteration (Gilmour et al. 1995). The Newton-type update is then

$$\boldsymbol{\theta}_0^{(t+1)} = \boldsymbol{\theta}_0^{(t)} + \mathbf{AI}(\boldsymbol{\theta}_0^{(t)})^{-1} \mathbf{U}(\boldsymbol{\theta}_0^{(t)}). \quad (2.18)$$

Iterative fitting algorithm

These two estimates are then iteratively updated in a Newton–Raphson scheme to obtain the maximum-likelihood estimates of the variance components.

Algorithm 1 AI–REML Iterative Fitting for the Null Frailty Model

- 1: Initialize under the null ($H_0 : \sigma^2 = 0$) and compute initial estimates $\hat{\boldsymbol{\beta}}^{(0)}$, $\hat{\mathbf{r}}_0^{(0)}$, and $\mathbf{Y}^{(0)}$.
 - 2: Compute the cumulative baseline hazard $\hat{\Lambda}_0(t)$ using Breslow’s estimator.
 - 3: **while** parameters not converged **do**
 - 4: Update variance components: $\boldsymbol{\theta}_0^{(t+1)} = \boldsymbol{\theta}_0^{(t)} + \mathbf{AI}(\boldsymbol{\theta}_0^{(t)})^{-1} \mathbf{U}(\boldsymbol{\theta}_0^{(t)})$.
 - 5: Update $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{r}}_0$ using $\mathbf{Y}$ and the updated $\boldsymbol{\theta}^{(t+1)}$.
 - 6: Recompute $\hat{\Lambda}_0(t)$ and the working response $\mathbf{Y}$.
 - 7: **end while**
-

2.2.5 Hypothesis testing using the efficient score statistic

For hypothesis testing, we obtain parameter estimates from the null fit, where $H_0 : \sigma^2 = 0$. Let $\hat{\boldsymbol{\theta}}_0 = (\hat{h}^2, \hat{\tau}^2)$ be the MLEs of the variance components from the AI–REML null fit.

To test H_0 we use the efficient score statistic for the variance component σ^2 . Evaluated at $\hat{\boldsymbol{\theta}}_0$, the efficient score statistic S is given by Eq. 2.19.

$$S = \mathbf{U}(\hat{\boldsymbol{\theta}}_0)^\top \mathbf{AI}(\hat{\boldsymbol{\theta}}_0)^{-1} \mathbf{U}(\hat{\boldsymbol{\theta}}_0) \quad (2.19)$$

where $\mathbf{U}$ is computed as the 3×1 REML score from Eq. 2.14 and $\mathbf{AI}$ is the 3×3 average-information matrix – computed using all three variance components, including the interaction covariance matrix of the test variant. Under H_0 , $S \sim \chi_1^2$ (Therneau et al. 2000). Because S is evaluated at $\hat{\boldsymbol{\theta}}_0$, it avoids iterative maximization under the alternative, necessary for reasonable scalability.

2.2.6 Simulating time-to-event data under the frailty model

To assess the statistical properties of CphGxE, we performed simulation studies focused on type I error control and power. We generated genetic frailties using both genotype data consisting of synthetic independent biallelic variants as well as genotypes from the UK Biobank. Only common variants were used, with all minor allele frequencies (MAF) of at least 5%.

Survival outcomes were generated under a Cox proportional hazards frailty model with an exponential baseline hazard (Dey et al. 2022). For each subject i , the true event time T_i was sampled as

$$T_i = \frac{\log(u_i)}{\lambda \exp(\eta_i)}, \quad (2.20)$$

where $u_i \sim \text{Uniform}(0, 1)$ is a random variable representing a uniform draw between 0 and 1, and η_i is the linear predictor incorporating both covariates as well as additive variant effects, additive exposure effects, and gene-by-environment interaction effects. Censoring times $C_i \sim \text{Exponential}(\kappa)$ were independently drawn from an exponential distribution. The observed time is given by $t_i = \min(T_i, C_i)$, and the event indicator

by $\delta_i = I(T_i \leq C_i)$. In order to simulate a desired censoring rate, for a fixed κ , λ was iteratively adjusted until the desired censoring rate was achieved.

The individual frailty was generated similar to a quantitative trait described in previous marginal epistasis methods (Crawford et al. 2017; Pattillo Smith et al. 2024; Stamp et al. 2023; Stamp et al. 2025). We sample 10% of the SNPs in the data to be causal and simulate traits using the following linear model

$$\boldsymbol{\eta} = \mathbf{Z}\boldsymbol{\gamma} + \sum_{a \in \mathcal{A}} \mathbf{x}_a \beta_a + \sum_{b \in \mathcal{B}} \mathbf{e}_b \alpha_b + \sum_{g \in \mathcal{G}} \sum_{h \in \mathcal{E}} (\mathbf{x}_g \circ \mathbf{e}_h) \xi_{gh} \quad (2.21)$$

Here, $\boldsymbol{\eta}$ denotes the N -dimensional vector of individual-specific frailty values. The matrix $\mathbf{Z}$ contains covariates (such as age, sex, or principal components), and $\boldsymbol{\gamma}$ is the corresponding vector of covariate effects. The set $\mathcal{A}$ indexes the causal SNPs with additive effects, where $\mathbf{x}_a$ is the genotype vector for SNP a (coded as 0, 1, or 2), and β_a is its additive effect size. The set $\mathcal{B}$ indexes the causal environmental exposure variables, where $\mathbf{e}_b$ is the exposure vector, and α_b is its environmental exposure effect size. The set $\mathcal{G}$ indexes the causal SNPs with gene-by-environment interaction effects, where $\mathbf{x}_g$ is the genotype vector for SNP g (coded as 0, 1, or 2), and ξ_{gh} is its gene-by-environment interaction effect size. Similarly, $\mathcal{E}$ indexes the causal environmental exposure variables.

All effect sizes (γ , β_a , α_b , and ξ_{gh}) are sampled from standard normal distributions and then rescaled so that the additive and gene-by-environment interaction components together account for a specified proportion of the total variance of frailties, $H^2 = h_{\mathcal{A}}^2 + h_{\mathcal{G}\mathcal{E}}^2$. The remainder $1 - H^2$ of the variance of frailties is explained by the covariates and exposure main effects and noise is introduced by sampling the time variables at random.

In this framework, interactions occur between all SNPs in $\mathcal{G}$ and all exposure variables in $\mathcal{E}$. This design allows us to control the marginal variance explained by each SNP in the context of the variance component gene-by-environment test. Specifically, the proportion

of interaction variance attributable to a single SNP depends on the $G \times E$ heritability $h_{G \times E}^2$ and the cardinality of the interaction SNP set; for instance, each SNP in $\mathcal{G}$ contributes on average $h_{G \times E}^2/|\mathcal{G}|$ to the total phenotypic variance.

2.2.7 Stabilizing measures in the iterative null fit and computational improvements

In the implemented algorithm, several measures were introduced to improve numerical stability throughout the computation. First, the genotype and exposure covariance matrices are explicitly symmetrised to prevent minor numerical asymmetries that can propagate and destabilize subsequent linear algebra operations. Second, constant genotype or exposure columns need to be removed to prevent singular covariance matrices. Finally, the REML fitting procedure was further stabilised by trace-normalising covariance matrices and adding a ridge regularisation term, and damped AI-REML updates (preventing negative variance component estimates as well as excessive updates), effectively preventing divergence of variance component estimates during optimisation. Additionally, the implemented method caches reusable results from the null fit procedure, in order to improve the runtime of the $G \times E$ test.

2.2.8 Preprocessing the UK Biobank

Genotype data for 488,377 individuals from the UK Biobank were downloaded and converted using the `ukbgene` and `ukbconv` tools, respectively. Continuous traits were obtained using `ukbgene` and adjusted for age and sex. Individuals exhibiting high heterozygosity, excessive relatedness, or aneuploidy were excluded (1,550 individuals). Participants were then divided into self-identified ancestral cohorts using data field `21000`, and unrelated individuals were selected by randomly retaining one individual from each related pair. This yielded $N = 349,411$ white British individuals for analysis. Imputed SNP data were downloaded for all remaining individuals, and variants with an information

score below 0.8 were excluded. Information scores for each SNP were provided by the UK Biobank (Web resources).

Quality control for the remaining 1,933,118 genotyped and imputed variants was performed separately within each cohort using the following steps. First, all structural variants were removed, retaining only single nucleotide polymorphisms (SNPs). Second, AT/CG SNPs were excluded to prevent potential confounding from sequencing errors. Third, variants with minor allele frequency below 1% were removed using the PLINK 2.0 (Chang et al. 2015) command `--maf 0.01`. Next, SNPs deviating from Hardy–Weinberg equilibrium were filtered out using the PLINK flag `--hwe 0.000001`, excluding those with Fisher’s exact test P -values $< 10^{-6}$. Finally, variants with any missing genotypes were excluded using the PLINK 2.0 command `--geno 0.00`. After these steps, a total of $J = 543,813$ SNPs remained.

To construct the genetic relatedness matrix for the simulation studies, we performed linkage disequilibrium (LD) pruning to obtain an approximately independent set of variants spanning the genome. LD pruning was conducted using PLINK 2.0 with the command `--indep-pairwise 50 5 0.01`, which applies a pairwise r^2 threshold of 0.01 to retain only markers in low LD. After pruning, approximately 15,000 variants remained.

2.3 Results

In this work, we introduce a method for gene-by-environment interaction analysis of censored time-to-event phenotypes. Our approach is based on the Cox Proportional Hazards model, incorporating covariates, population structure, environmental exposures, and marginal gene-by-environment interaction effects as frailties.

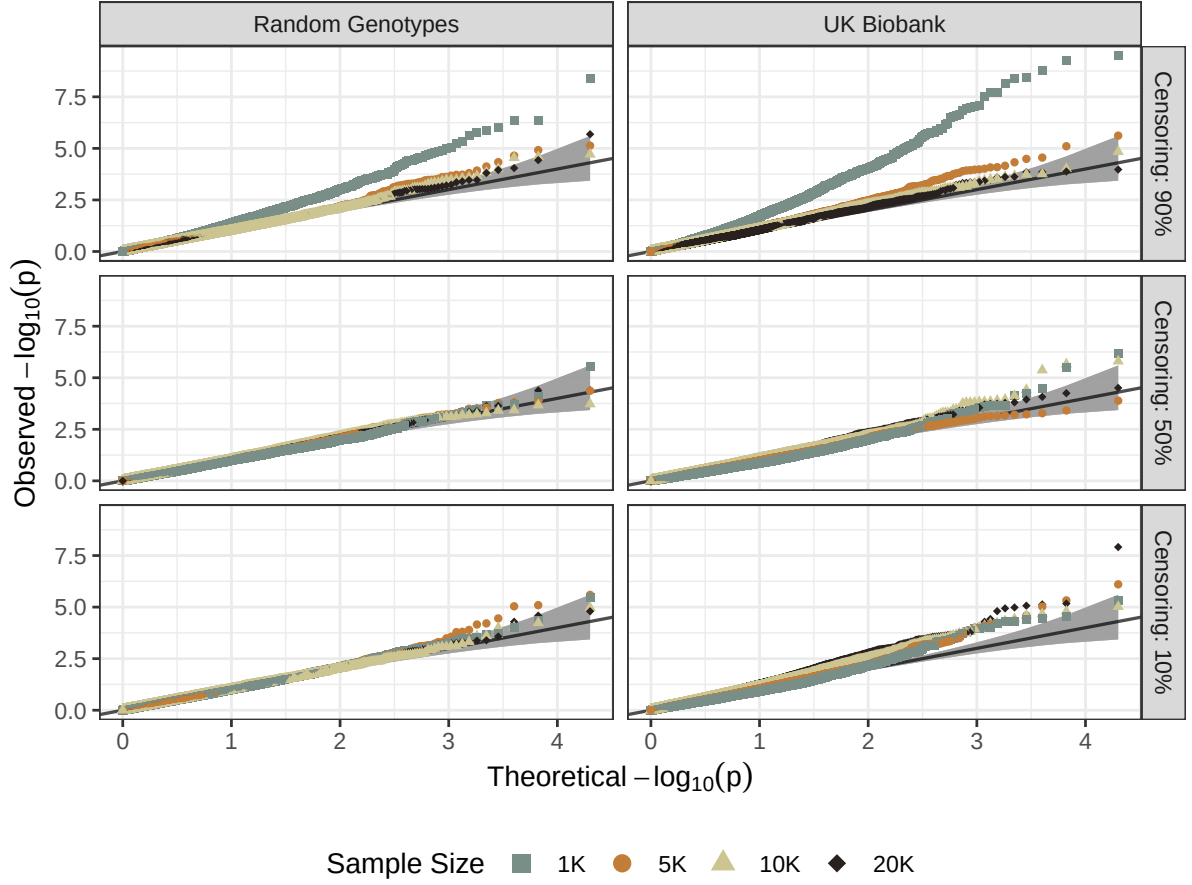


Figure 2.1: **QQ-plots of P -values for CphGxE under the null model.** Synthetic time-to-event traits were simulated with additive genetic, environmental exposure, and covariate effects. Genotype data with 15,000 SNPs were either simulated using synthetic biallelic independent genotypes (Random Genotypes) or taken from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled to sample sizes of 1k, 5k, 10k and 20k individuals. We randomly selected 10% of all variants to be causal with additive effects set at 40% of the frailty variance explained (FVE) for each trait. The frailty variance due to environmental exposures is set to 10% of the total frailty variance. The remainder of the frailty variance is explained by the covariates and noise is introduced by sampling the time variables at random. Data were analyzed using CphGxE under varying percentages of censoring. Under the null hypothesis $H_0 : \sigma^2 = 0$. Results are based on 100 simulated traits per scenario.

2.3.1 CphGxE produces asymptotically calibrated P -values at high censoring rates

We generate synthetic time-to-event data based on a frailty model incorporating covariates, population structure, environmental exposures, and gene-by-environment

interaction effects. The genetic components of the frailty are simulated under two scenarios. In the first scenario, the data are simulated using real genotypes from white British individuals in the UK Biobank. After quality control and LD pruning, we had a data set of 349,411 individuals and 15,000 SNPs (Methods). In the second scenario, data are simulated using synthetic biallelic independent genotypes. Under the null model, we simulate traits consisting of only additive genetic, covariate, and environmental exposure effects. For the genetic frailty, we randomly sample 10% of the SNPs and scale their effect sizes such that they explain 40% of the total frailty variance. For the covariate and environmental exposure frailty, we simulate 3 continuous covariates and 100 continuous environmental exposures. The variance due to environmental exposures is set to 10% of the total frailty variance. The remainder of the frailty variance is explained by the covariates and noise is introduced by randomly sampling the observation and censoring times based on the frailty model. In these simulation experiments, we assess the calibration of CphGxE as a function of the number of individuals being analyzed (varying between 1k, 5k, 10k and 20k randomly subsampled individuals), the censoring rate in the synthetic time-to-event data (varying between 10%, 50%, and 90% censoring), and the genotype data being used (random genotypes or UK Biobank genotypes).

Under the null model, we find that CphGxE produces asymptotically calibrated P -values when the genetic data is simulated using synthetic genotypes (Fig. 2.1 and Tabs. 2.1, B.1, B.2). When the data is simulated using UK Biobank genotypes, we find that CphGxE produces asymptotically calibrated P -values at high censoring rates. However, at low censoring rate, the calibration of the P -values deteriorates with increasing sample size. Overall, this translates to CphGxE asymptotically preserving empirical type I error rates estimated at significance levels $\alpha = 0.05, 0.01, \text{ and } 0.001$, respectively (Tabs. 2.1, B.1, B.2) for most relevant scenarios, as censoring rates tend to be high in real data (Dey et al. 2022).

Table 2.1: **Empirical type I error rates for CphGxE with 90% censoring.** Synthetic traits were simulated with additive genetic, environmental exposure, and covariate effects. Genotype data with 15,000 SNPs were either simulated using synthetic biallelic independent genotypes (Random Genotypes) or taken from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled to sample sizes of 1k, 5k, 10k and 20k individuals. We randomly selected 10% of all variants to be causal with additive effects set at 40% of the frailty variance explained for each trait. The frailty variance due to environmental exposures is set to 10% of the total frailty variance. The remainder of the frailty variance is explained by the covariates and noise is introduced by randomly sampling the observation and censoring times based on the frailty model. Data were analyzed using CphGxE under varying percentages of censoring. Under the null hypothesis $H_0 : \sigma^2 = 0$. 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.

Dataset	Sample Size	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
Random Genotypes	1k	0.110 (0.313)	0.039 (0.193)	0.01 (0.1)
	5k	0.059 (0.235)	0.014 (0.115)	0.003 (0.058)
	10k	0.052 (0.222)	0.012 (0.108)	0.002 (0.047)
	20k	0.058 (0.235)	0.015 (0.121)	0.002 (0.046)
UK Biobank	1k	0.157 (0.364)	0.075 (0.264)	0.028 (0.164)
	5k	0.080 (0.271)	0.022 (0.146)	0.004 (0.066)
	10k	0.063 (0.244)	0.018 (0.131)	0.002 (0.046)
	20k	0.061 (0.240)	0.014 (0.119)	0.002 (0.044)

2.3.2 CphGxE estimates of heritable frailty variance and fixed effects are asymptotically unbiased

Under the Cox proportional hazards model, the variance of the random frailty term represents the unobserved heterogeneity in risk across individuals and mathematically exhibits similarity to heritability in the context of survival traits. Specifically, the frailty variance quantifies the proportion of variability in the underlying hazard that is attributable to genetic or other shared features, similar to the proportion of phenotypic variance explained in linear mixed models. Because the frailty model in CphGxE explicitly

partitions the frailty into genetic relatedness and similarity in environmental exposures as well as fixed effect sizes of covariates, it provides a natural framework for estimating this “frailty heritability.” As such, fitting the null model in CphGxE offers an efficient and principled way to assess heritability in time-to-event data.

We find that CphGxE produces asymptotically unbiased variance component estimates of the frailty due to genetic relatedness when the genetic data is simulated using synthetic genotypes (Fig. B.1). When the data is simulated using UK Biobank genotypes, the variance component of genetic relatedness estimates appear asymptotically unbiased but the variance component of environmental exposure similarity estimates are slightly negatively biased (Fig. B.2). The fixed effect estimates are always unbiased (Fig. B.3).

2.3.3 CphGxE is well powered and correctly prioritizes gene-by-environment interaction variants

To assess the power of CphGxE, we again generate synthetic time-to-event data based on a frailty model incorporating covariates, population structure, environmental exposures, and gene-by-environment interaction effects. For the genetic frailty, we randomly sample 10% of the SNPs and scale their effect sizes such that they explain 30% of the total frailty variance. For the covariate and environmental exposure frailty, we simulate 3 continuous covariates and 100 continuous environmental exposures. The variance due to environmental exposures is set to 10% of the total frailty variance. Under the alternative hypothesis, we simulate 10 SNPs to have gene-by-environment interaction effects with 20 environmental exposure features and we vary the frailty variance they explain individually (varying between 1%, 2%, and 5% of the total frailty variance). The remainder of the frailty variance is caused by the covariates and noise is again introduced by sampling the time variables at random. We assess the power of CphGxE as a function of the number of individuals being analyzed (varying between 1k, 5k, 10k and 20k randomly subsampled individuals), the censoring rate in the synthetic time-to-event data (varying between 10%,

50%, and 90% censoring), and the genotype data being used (random genotypes or UK Biobank genotypes).

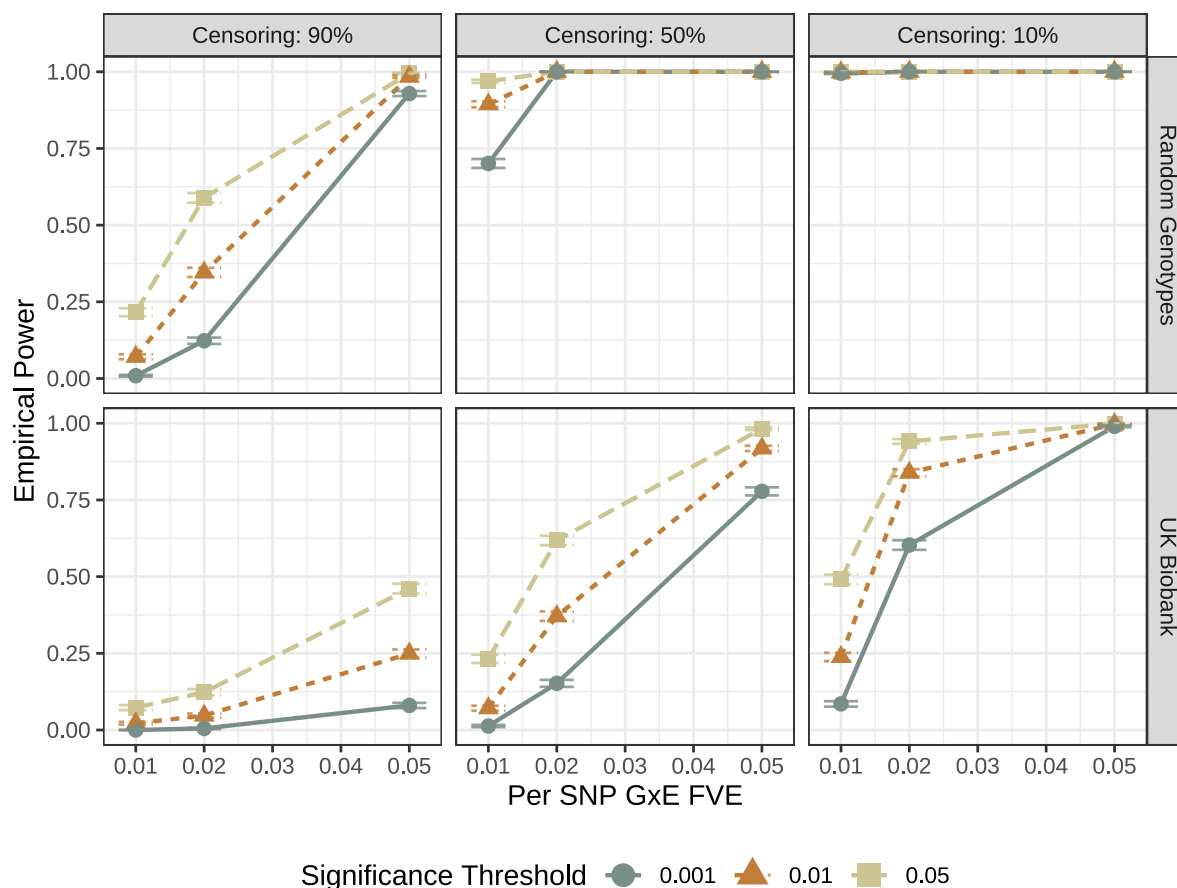


Figure 2.2: **Empirical power as function of per SNP $G \times E$ frailty variance explained.** Synthetic time-to-event traits were simulated with additive genetic, environmental exposure, covariate, and gene-by-environment interaction effects. Genotype data with 15,000 SNPs were either simulated using synthetic biallelic independent genotypes (Random Genotypes) or taken from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled to 20k individuals. We randomly selected 10% of all variants to be causal with additive effects set at 30% of the frailty variance explained (FVE) for each trait. The frailty variance due to environmental exposures is set to 10% of the total frailty variance. The empirical power is shown as function of the per SNP gene-by-environment interaction variance to the frailty (1%, 2%, and 5%). The remainder of the frailty variance is explained by the covariates and noise is introduced by sampling the time variables at random. Data were analyzed using CphGxE under varying percentages of censoring (10%, 50%, and 90% censoring). The analyses used significance thresholds of $\alpha = 0.05$, 0.01, and 0.001. Results are based on 100 simulations per scenario, with error bars representing the standard deviation across replicates.

The results show that when the genetic data is simulated using synthetic genotypes,

CphGxE is well powered to detect variants with gene-by-environment interaction effects even when the variants explain a small amount of the frailty variance (Fig. 2.2). When the data is simulated using UK Biobank genotypes, we find that the power of CphGxE is diminished, but correctly prioritizes gene-by-environment interaction variants at all censoring rates (Fig. 2.3, Tab. B.3). As expected, the power and ability to correctly prioritize gene-by-environment interaction variants of CphGxE increases with increasing sample size and decreasing censoring rate (Figs. B.4 - B.6). In simulations based on UK Biobank genotypes with 20k individuals and 10% censoring, CphGxE identifies at least 49.1% of the SNPs with gene-by-environment interaction effects ($AUC = 0.828$) even when they contribute as little as 1% to the frailty variance (at significance level $\alpha = 0.05$). By contrast, in simulations using synthetic genotypes, CphGxE identifies 100% of the SNPs with gene-by-environment interaction effects ($AUC = 1.000$). When genetic variants have larger marginal gene-by-environment effects and individually account for 5% of the frailty variance, the empirical power of CphGxE remains high despite censoring rates of 90% with an empirical power of 46.1% ($AUC = 0.797$) in simulations based on UK Biobank genotypes and 99.5% ($AUC = 0.999$) in simulations using synthetic genotypes.

2.3.4 The iterative null fit in CphGxE is sensitive to censoring rates and initial parameter values

Fitting a REML model can be challenging and fitting algorithms that work in one setting may not work in another (Harville 1977). The AI algorithm relies on iterative updates of variance component estimates based on the average of the observed and expected information matrices. Convergence can be slow or fail entirely when initial values are chosen poorly (Lindstrom et al. 1988). Moreover, the repeated computation and inversion of matrices is computationally demanding for large datasets.

We simulated data for 10k individuals at varying censoring rates (10%, 50%, and 90%) using both synthetic and UK Biobank genotypes. The frailty variance due to genetic

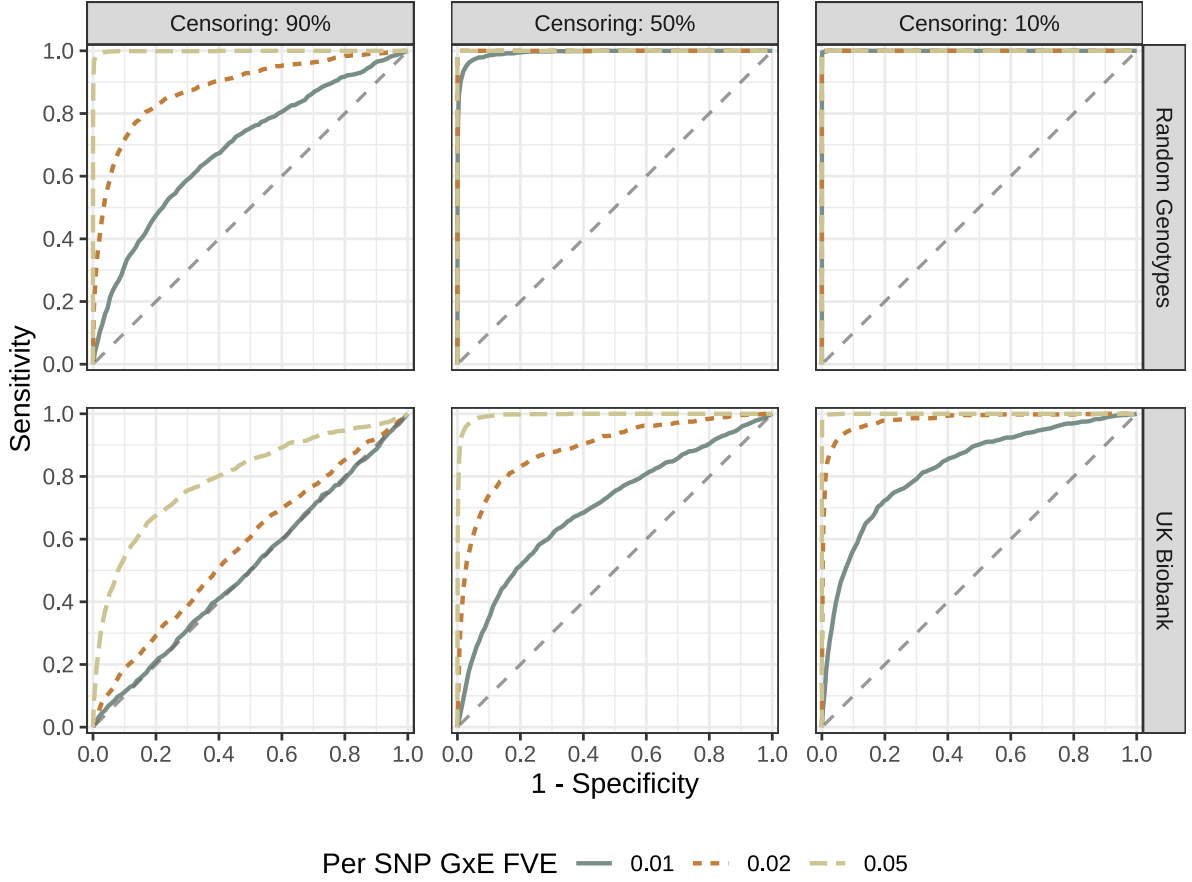


Figure 2.3: **Receiver Operating Characteristic (ROC) curves comparing the ability of CphGxE to correctly prioritize gene-by-environment interaction variants under the alternative model.** Synthetic time-to-event traits were simulated with additive genetic, environmental exposure, covariate, and gene-by-environment interaction effects. Genotype data with 15,000 SNPs were either simulated using synthetic biallelic independent genotypes (Random Genotypes) or taken from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled to 20k individuals. We randomly selected 10% of all variants to be causal with additive effects set at 30% of the frailty variance explained (FVE) for each trait. The frailty variance due to environmental exposures is set to 10% of the total frailty variance. ROC curves are computed under simulations as function of the per SNP gene-by-environment interaction variance of the frailty (1%, 2%, and 5%). The remainder of the frailty variance is explained by the covariates and noise is introduced by sampling the time variables at random. Data were analyzed using CphGxE under varying percentages of censoring (10%, 50%, and 90% censoring). Results are based on 100 simulations per scenario.

relatedness and environmental exposure are set to 40% and 10% of the total frailty variance respectively. The remainder of the frailty variance is explained by the covariates and noise is introduced by sampling the time variables at random. We then fit the null model using

CphGxE with varying initial parameter values $\theta = (h^2, \tau^2)$ and recorded the variance component estimates at each step of the iterative fitting algorithm. We found that the iterative null fit in CphGxE is sensitive to censoring rates and initial parameter values (Fig. B.7).

When the initial parameters are close to the true variance component values ($\theta_0 = (0, 0)$ and $\theta_0 = (1, 1)$), the implemented algorithm is numerically stable (Figs. B.7A and B), however, when the censoring rate is high (90%), it does not reliably converge to the true variance component value of the genetic relatedness frailty (Figs. B.1, B.7), especially when using UK Biobank genotypes. When the initial parameter values are far from the true variance component values ($\theta_0 = (10, 10)$), we observe that the fitting process diverges at low censoring rates for synthetic genotypes (Fig. B.7C) despite the stabilizing measures implemented in the algorithm.

2.3.5 Caching of the null fit improves the runtime of the $G \times E$ test

We assessed the runtime of CphGxE as a function of the number of individuals being analyzed (varying between 1k, 5k, 10k, and 20k individuals). In its current implementation, the computation is implemented as naive matrix operations, which is computationally demanding. The slowest part is the null fit (Fig. B.8), scaling polynomially with the number of individuals and requiring on the order of 20 hours for 20k individuals. Fortunately, the null fit needs to be performed only once and partial results are cached and reused for the $G \times E$ test. This makes the variant test much faster (Fig. B.8), requiring on the order of 10 seconds for 20k individuals per variant.

2.4 Discussion

The presented method is an approach to gene-by-environment interaction analysis of censored time-to-event phenotypes. It is a method developed in the line of marginal epistasis frameworks (Crawford et al. 2017; Stamp et al. 2023; Stamp et al. 2025), i.e. it measures the aggregate effect of interactions. In deviation from the marginal epistasis frameworks, it looks for marginal interactions a focal SNP has with environmental exposures instead of other SNPs. This framework results in more powerful inference compared to methods looking for interactions between a focal SNP and a single environmental exposure at the cost of not knowing which environmental exposures are involved in the interaction. Marginal interaction tests have been implemented mostly for quantitative traits (Zhou et al. 2025). The present method extends this approach to the analysis of censored time-to-event phenotypes by assuming that genes and environmental exposures are the source of the frailty that is shared between individuals. The feature space of environmental exposure features is sparse compared to the number of SNPs, making estimators efficient (Stamp et al. 2025).

In simulation studies, we show that CphGxE controls type I error rates and produces asymptotically calibrated P -values at high censoring rates. We also show that CphGxE is powered to detect SNPs involved in gene-by-environment interaction even when they explain small fractions of the frailty variance.

The presented implementation of CphGxE is a naive implementation of the method, offering many opportunities for future improvements. For example, the null fit can be implemented with the ability to scale to biobank data (Dey et al. 2022; Kim et al. 2025). By implementing an efficient caching mechanism of the partial results of the null fit, CphGxE can be made to scale to biobank data. This approach requires efficient implementations of stochastic traces with shared random vectors and genotype and exposure streaming like in the sparse marginal epistasis test (Stamp et al. 2025).

The simulation studies show that not only sample size and censoring rate are crucial for the calibration and power of the $G \times E$ test. The test is only calibrated when the null fit converges reliably and scalability depends on fast convergence. The iterative Newton-Raphson based algorithm converges relatively fast when the initial parameter values are close to the true variance component values, but has issues with high censoring rates and relatedness in the data. Investigating alternative algorithms for the null fit is a promising avenue for future work.

Lastly, the presented work is not yet applied to real data. In a revised version of this manuscript, we will apply the method to the UK Biobank data to illustrate the utility of the method and generate summary statistics for the $G \times E$ test.

2.5 Data and software availability

Source code and tutorials for implementing the "Cox proportional hazards gene-by-environment interaction test" (CphGxE) are available open source as an R package which is available online on GitHub (<https://github.com/lcrawlab/cphgxe>).

2.6 Web resources

- CphGxE, <https://github.com/lcrawlab/CphGxE>
- UK Biobank, <http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=1967>

CHAPTER 3

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

The text in this chapter is an adaptation of the material as it appears in Stamp et al. (2025). *Sparse modeling of interactions enables fast detection of genome-wide epistasis in biobank-scale studies*. The American Journal of Human Genetics. This article is licensed under the Creative Commons Attribution (CC BY NC ND <https://creativecommons.org/licenses/by-nc-nd/4.0/>) licence. The coauthors listed in this publication are Julian Stamp, Samuel Pattillo Smith, Daniel Weinreich, and Lorin Crawford. Lorin Crawford and Daniel Weinreich supervised the research which forms the basis for the dissertation.

3.1 Introduction

Genome-wide association studies (GWAS) have identified thousands of genetic loci linked with various complex traits and common diseases, offering valuable insights into the genetic foundations of phenotypic variation (Abdellaoui et al. 2023). As of late, there have been many efforts to estimate proportions of genetic variance beyond what is attributable to additive effects (Crawford et al. 2017; Stamp et al. 2023; Fu et al. 2023; Pattillo Smith et al. 2024; Balvert et al. 2024). Epistasis, which refers to interactions between genetic loci, is thought to play a key role in constituting the genetic basis of evolution (Weinreich et al. 2006; Fröhlich et al. 2024). While many studies have shown epistasis to be pervasive in model organisms (Mackay et al. 2024; Weatherly et al. 2022; Zwarts et al. 2011), controversies remain with respect to its role in humans (Polderman et al. 2015). For example, some epistatic interactions identified in association mapping studies can be explained by additive effects of unobserved variants (Hemani et al. 2021). Though previous studies have shown that genetic variance is mainly additive (Hivert et al. 2021; Mackay et al. 2024), these conclusions have recently been challenged (Pattillo Smith et al. 2024).

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 (Purcell et al. 2007; Schüpbach et al. 2010; Prabhu et al. 2012; Wan et al. 2010). With advancements in sequencing technologies, many contemporary genome-wide association studies (GWAS) are conducted on biobank-scale data sets, comprising hundreds of thousands of individuals genotyped at millions of markers and phenotyped for thousands of traits (Abdellaoui et al. 2023; Bycroft et al. 2018; Nagai et al. 2017). This is crucial as the effect of epistatic interactions is hypothesized to be small for many traits (Hivert et al. 2021; Polderman et al. 2015), and traditional search algorithms are known to be most powered when large

training data sets are available (Purcell et al. 2007; Hivert et al. 2021). 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 data set. The “Marginal Epistasis Test” (MAPIT) evaluates each SNP individually and identifies candidates involved in epistasis without requiring the identification of their exact interacting partners (Crawford et al. 2017). Recently, the concept of marginal epistasis has been leveraged to estimate the contribution of nonadditive heritability in complex traits using GWAS summary statistics (Pattillo Smith et al. 2024). It has also been extended to explore the importance of genetic interactions across multiple traits simultaneously (Stamp et al. 2023). 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 (Haseman et al. 1972; Zhou 2017). Although MAPIT mitigates the reduction of power due to multiple testing burden, its implementation on data sets with large sample sizes remains computationally intensive (Crawford et al. 2017). 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 (Wu et al. 2018; Mbatchou et al. 2021). Efforts have been made to address this limitation, such as the “fast marginal epistasis test” (FAME) (Fu et al. 2023), which leverages a stochastic MoM framework and introduces both computationally efficient stochastic trace estimators (Hutchinson 1989) and innovative methods to expedite matrix multiplication (Liberty et al. 2009). However, despite these advancements, further

work is necessary to scale the method to genome-wide applications.

This work introduces the sparse marginal epistasis test, which focuses on searching for epistasis in regions of the genome with known functional enrichment (Maurano et al. 2012) 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 (Finucane et al. 2015). Consequently, the remaining heritability is attributed to regions expected to play a regulatory role (Maurano et al. 2012; Cano-Gamez et al. 2020; Finucane et al. 2015) and that are active in trait-specific tissue (Finucane et al. 2018; Boix et al. 2021). Second, the sparse nature of this approach leads to more efficient estimators for model parameters (Yang et al. 2010; Zhou 2017) 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 (Georgolopoulos et al. 2021) and GWAS summary statistics, we use SME to analyze complex traits in individuals from the UK Biobank (Bycroft et al. 2018) and identify genetic interactions associated with regulatory genomic elements.

3.2 Material and methods

3.2.1 The sparse marginal epistasis test

SME performs a genome-wide search for SNPs involved in genetic interactions while conditioning on information derived from functional genomic data (Fig. 3.1A). Consider a genome-wide association study (GWAS) with N individuals who have been genotyped for J single nucleotide polymorphisms (SNPs) encoded as $\{0, 1, 2\}$ copies of a reference allele at each locus. Also assume that we have access to an external reference $\mathcal{S}$ that

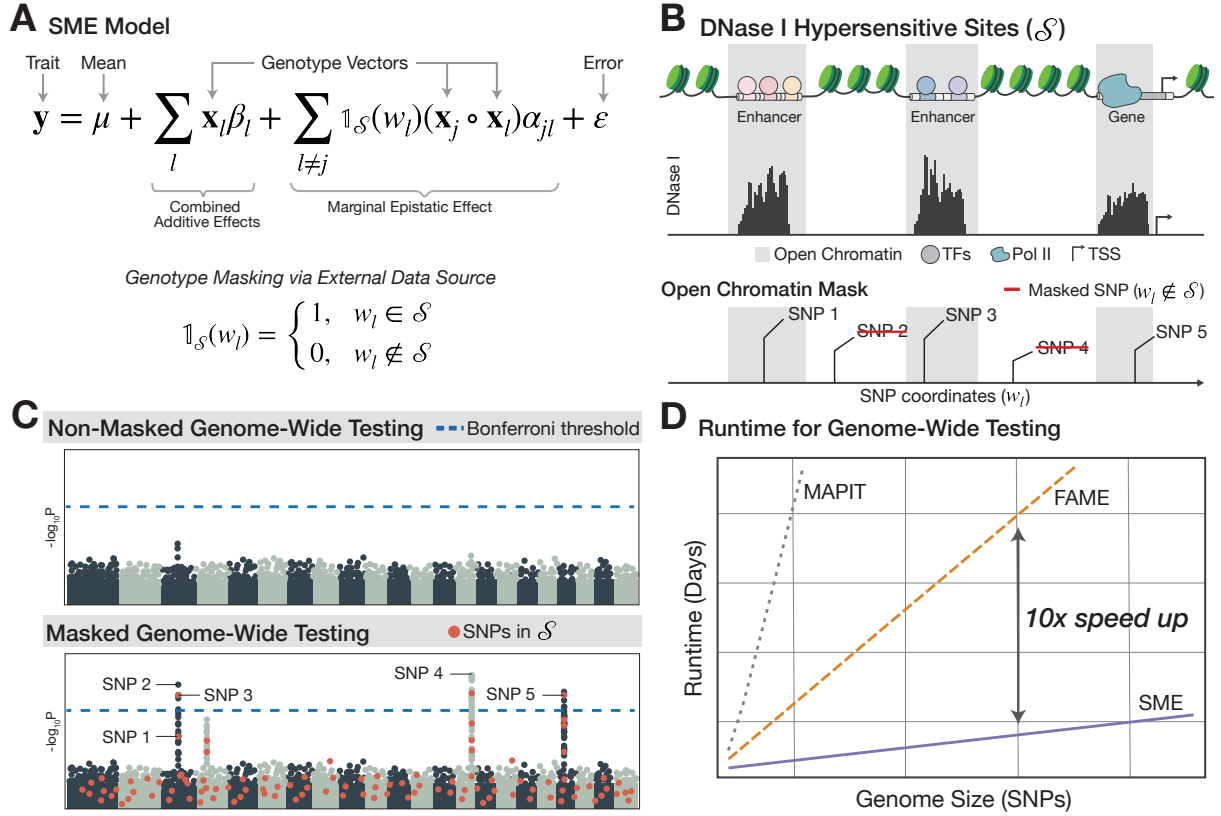


Figure 3.1: Schematic overview of the sparse marginal epistasis (SME) test. (A) SME examines one SNP at a time and estimates marginal epistatic effects—the combined pairwise interaction effects between a j -th focal SNP and other variants on the genome (indexed by $l \neq j$). The key to SME is that it incorporates genomic data $\mathcal{S}$ through a binary indicator function $\mathbb{1}_{\mathcal{S}}(w_l)$ 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 data sets. It achieves runtimes 10× to 90× faster than current state-of-the-art methods.

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 (Crawford et al. 2017). By

examining one SNP at a time (indexed by j), SME fits the following linear mixed model

$$\mathbf{y} = \mu + \sum_l \mathbf{x}_l \beta_l + \sum_{l \neq j} (\mathbf{x}_j \circ \mathbf{x}_l) \alpha_l \cdot \mathbb{K}_{\mathcal{S}}(w_l) + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \tau^2 \mathbf{I}) \quad (3.1)$$

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

$$\mathbb{K}_{\mathcal{S}}(w_l) = \begin{cases} 1 & \text{if } w_l \in \mathcal{S} \\ 0 & \text{if } w_l \notin \mathcal{S} \end{cases} \quad (3.2)$$

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 (Fig. 3.1B). In this case, $\mathcal{S}$ could be a set of genomic regions for which DNase-seq implicates chromatin accessibility and w_l could encode the physical location of the l -th SNP on the genome. Here, $\mathbb{K}_{\mathcal{S}}(w_l) = 1$ 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 Eq. (3.2) are considered.

The benefits of this sparse approach are twofold. 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 (Fig. 3.1C). 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 (Fig. 3.1D). Specifically, SME introduces an approximation to the efficient stochastic trace estimator (Hutchinson 1989; Wu et al. 2018; Fu et al. 2023; Wu et al. 2022) which allows the algorithm to avoid repeating costly matrix computations when estimating model parameters across each SNP that is being tested (Figs. C.1-C.3).

3.2.2 Variance component model formulation

For biobank-scale data, there are often more SNPs than individuals. To overcome an undetermined system in Eq. (3.1), 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 $J^* = \sum_{l \neq j} \mathbb{1}_{\mathcal{S}}(w_l)$ representing the number of interactions considered in the model (Bulik-Sullivan et al. 2015; Yang et al. 2010; Wu et al. 2011; Zhou et al. 2013; Crawford et al. 2017). Assuming that the phenotype has been mean centered and scaled, these normal assumptions allow Eq. (3.1) to be rewritten as

$$\mathbf{y} = \mathbf{m} + \mathbf{g}_j + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \tau^2 \mathbf{I}) \quad (3.3)$$

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

Probabilistically, Eq. (3.3) translates to SME assuming that $\mathbf{m} \sim \mathcal{N}(\mathbf{0}, \omega^2 \mathbf{K})$ where the covariance matrix $\mathbf{K} = \mathbf{X} \mathbf{X}^\top / J$ accounts for the relatedness between individuals in the data and the corresponding component ω^2 models the phenotypic variance explained by additive effects. The second term can be written as $\mathbf{g}_j \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{G}_j)$ where $\mathbf{G}_j = \mathbf{D}_j \mathbf{X}_{-j} \mathbf{W}_j \mathbf{X}_{-j}^\top \mathbf{D}_j / J^*$ 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, $\mathbf{W}_j = \text{diag}[\mathbb{1}_{\mathcal{S}}(w_1), \dots, \mathbb{1}_{\mathcal{S}}(w_{J-1})]$ 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, 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 $\{\mathbb{I}_{\mathcal{S}}(w_l)\}_{l \neq j}$. The main takeaway from the variance component formulation of SME is that the term σ^2 measures SNP-specific contribution to the nonadditive genetic variance.

3.2.3 Point estimates and hypothesis testing

The model in Eq. (3.3) has three variance components that can be estimated using a computationally efficient method-of-moments (MoM) algorithm (Zhou 2017). In expectation

$$\mathbb{E}[\mathbf{y}^\top \mathbf{A} \mathbf{y}] = \sum_{k=1}^3 \text{tr}(\mathbf{A} \boldsymbol{\Sigma}_{jk}) \delta_k \quad (3.4)$$

with $\mathbf{A}$ being a symmetric and non-negative definite matrix used to create weighted second moments, $\text{tr}(\bullet)$ denotes the trace of a matrix, and we use shorthand to represent $[\boldsymbol{\Sigma}_{j1}; \boldsymbol{\Sigma}_{j2}; \boldsymbol{\Sigma}_{j3}] = [\mathbf{K}; \mathbf{G}_j; \mathbf{I}]$ and $\boldsymbol{\delta} = (\omega^2, \sigma^2, \tau^2)$, respectively. In practice, we replace the left hand side of Eq. (3.4) with the realized value $\mathbf{y}^\top \mathbf{A} \mathbf{y}$. We also use the realized covariance matrices in place of the arbitrary $\mathbf{A}$. The point estimates for each variance component is then given as

$$\hat{\delta}_{jk} = \mathbf{y}^\top \mathbf{H}_{jk} \mathbf{y} \quad (3.5)$$

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 $(\mathbf{S}_j)_{kt} = \text{tr}(\boldsymbol{\Sigma}_{jk} \boldsymbol{\Sigma}_{jt})$.

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 $\hat{\sigma}_j^2$ using Eq. (3.5) and compute an approximate standard error

$$\mathbb{V}[\hat{\sigma}_j^2] \approx 2 \mathbf{y}^\top \mathbf{H}_j^\top \mathbf{V}_j \mathbf{H}_j \mathbf{y} \quad (3.6)$$

where $\mathbf{V}_j = \hat{\omega}_j^2 \mathbf{K} + \hat{\sigma}_j^2 \mathbf{G}_j + \hat{\tau}_j^2 \mathbf{I}$. Note that the point estimates from Eqs. (3.4) and (3.5) are unbiased and can lead to negative values when the true variance component is zero (Zhou 2017). The one-sided hypothesis test formalizes the constraint that only positive estimates of $\hat{\sigma}_j^2$ can be indicative of marginal epistasis.

3.2.4 Scalable computation via stochastic MoM

The right-hand side of Eq. (3.5) 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 (Hutchinson 1989; Wu et al. 2018; Fu et al. 2023; Wu et al. 2022)

$$\text{tr}(\Sigma_{jr} \Sigma_{js}) \approx \frac{1}{B} \sum_{b=1}^B \mathbf{z}_b^T \Sigma_{jr} \Sigma_{js} \mathbf{z}_b \quad (3.7)$$

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 $\mathbf{z}_b^T \mathbf{X} \mathbf{X}^T \mathbf{X} \mathbf{X}^T \mathbf{z}_b = \|\mathbf{X} \mathbf{X}^T \mathbf{z}_b\|^2$ is reduced to (i) applying $\mathbf{X}^T$ to the vector $\mathbf{z}_b$ and then (ii) applying $\mathbf{X}$ to the resulting vector $\mathbf{X}^T \mathbf{z}_b$ for all B random vectors. Importantly, Eq. (3.7) 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 (Liberty et al. 2009; Wu et al. 2022; Wu et al. 2018; Fu et al. 2023). 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 $\Omega = 3$, $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_J]$ is a matrix where the j -th column contains the sample mean for the j -th SNP, and $\mathbf{Q} = \text{diag}(q_1^2, \dots, q_J^2)$ contains the variance of each SNP as the diagonal entries. With this specification, we can write $\mathbf{X}^T \mathbf{z}_b = \mathbf{Q}^{-1}(\mathbf{A}^T - \mathbf{U}^T)\mathbf{z}_b$. The first term $\mathbf{Q}^{-1}\mathbf{A}^T \mathbf{z}_b$ can be solved in $\mathcal{O}(NJ/\log_3(\max\{N, J\}))$ time. The second term $\mathbf{Q}^{-1}\mathbf{U}^T \mathbf{z}_b$ corresponds to scaling the random N -dimensional vector $\mathbf{z}_b$ which can be computed in time $\mathcal{O}(N + J)$ (Wu et al. 2018; Wu et al. 2022; Fu et al. 2023).

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 (Fig. C.4).

Since the error terms in Eq. (3.3) are assumed to be independent, the only two intermediate products that need to be computed in Eq. (3.7) 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., $\text{tr}(\mathbf{K}^2) \approx \|\mathbf{K}\mathbf{z}_b\|^2$ and $\text{tr}(\mathbf{K}\mathbf{G}_j) \approx \mathbf{z}_b^\top \mathbf{K}\mathbf{G}_j\mathbf{z}_b$). 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 $1/L$. Fig. C.1 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., Figs. C.2 and C.3). 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 $\mathcal{O}(BNJ^*/\log_3(N))$.

3.2.5 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 single nucleotide polymorphisms (SNPs) in the genotype data. Next, all AT/CG SNPs were removed to avoid possible confounding due to sequencing errors. Then, SNPs with minor allele frequency less than 1% were removed using the PLINK 2.0 (Chang et al. 2015) 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 -value $< 10^{-6}$. 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 assess the scale invariance of the signal identified by SME. Here, we rerun 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.

3.2.6 GWAS summary statistics

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 our analyses.

3.2.7 Generating masks from external data sources

Below is a description of the data sets 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 (DHS) measured over 12 days of *ex vivo* erythroid differentiation (Georgolopoulos et al. 2021). The DHS intervals were reported using the hg38 human reference genome. To map correspondence to the UK Biobank data (which uses the hg19 as reference), we performed a lift-over using CrossMap (Zhao et al. 2014). We mapped each SNP in the UK Biobank to the genomic intervals in the DHS data using the R software package GenomicRanges (Lawrence et al. 2013). The resulting mask comprised of $J^* = 4952$ 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 two-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 (Maurano et al. 2012). Second, GWAS summary statistics have been suggested to tag nonadditive genetic effects contributing to trait architecture (Pattillo Smith et al. 2024). 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 (VITD). 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 VITD.

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 (Berisa et al. 2016).

3.2.8 Simulation studies

To characterize the behavior of the sparse marginal epistasis test, we generate quantitative traits using chromosome 1 of the white British cohort from the UK Biobank (Crawford et al. 2017; Pattillo Smith et al. 2024; Stamp et al. 2023). This 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

$$\mathbf{y} = \sum_{a \in \mathcal{A}} \mathbf{x}_a \beta_a + \sum_{g_1 \in \mathcal{G}_1} \sum_{g_2 \in \mathcal{G}_2} (\mathbf{x}_{g_1} \circ \mathbf{x}_{g_2}) \alpha_{g_1 g_2} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \lambda^2 \mathbf{I}) \quad (3.8)$$

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 $H^2 = h_{\mathcal{A}}^2 + h_{\mathcal{G}}^2$. 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 the $\mathcal{G}_2$, 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 phenotypic variance explained by a single SNP are the epistatic heritability $h_{\mathcal{G}}^2$ and the cardinality of the set to which the SNP belongs. For example, a SNP in $\mathcal{G}_1$ will explain on average $h_{\mathcal{G}}^2/|\mathcal{G}_1|$ 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 data set 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 linkage disequilibrium between SNPs. Therefore, to simulate masks which 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, 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.

3.2.9 Software tools and data sources

Software for running the “FAst Marginal Epistasis test” (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 “MArginal ePIstasis Test” (MAPIT) was implemented using the `mvmapit` software package in R and is available both on CRAN (<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 Numbers 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 linkage disequilibrium maps were taken from <https://bitbucket.org/nygcresearch/ldetect-data/>. The DNase I hypersensitive site (DHS) data are available at <https://doi.org/10.5281/zenodo.5291736>. We used `CrossMap` (<https://crossmap.readthedocs.io/>) and `GenomicRanges` (<https://doi.org/doi: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/>.

3.3 Results

3.3.1 SME scales to biobank genome-wide association studies

To compare the expected central processing unit (CPU) computation for conducting a biobank-scale genome-wide analysis using SME, FAME (Fu et al. 2023), and MAPIT (Crawford et al. 2017), we measured the average runtime per SNP for each method on an Intel(R) Xeon(R) 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 (Methods). The memory requirements for MAPIT are prohibitively high, requiring resources on the order of terabytes for biobank-scale data sets 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 gigabytes of memory at biobank-scale data.

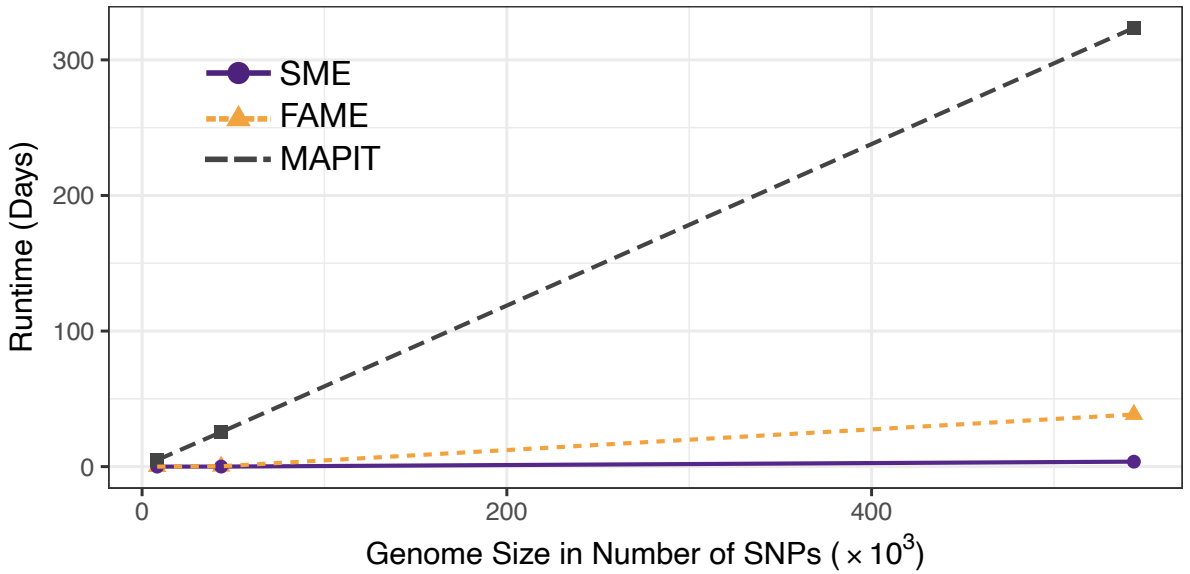


Figure 3.2: **Computational time for running SME and other marginal epistatic approaches on biobank-scale data as a function of the genome size.** The other methods compared include FAME (Fu et al. 2023) and MAPIT (Crawford et al. 2017). Here, we analyze genotype data from a fixed set of 349,411 individuals from the UK Biobank and vary the genome size. All results were computed on a single core of an Intel(R) Xeon(R) 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 data sets of this size. Instead, the runtime for MAPIT was measured on smaller sample sizes (up to 20k 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.

We find that SME performs genome-wide testing $10\times$ faster than FAME and $90\times$

faster than MAPIT (Fig. 3.2). While analyzing the complete data set with a single core, SME requires only 3.7 days of runtime compared to FAME and MAPIT requiring 38.4 and 324 days, respectively. The greatest speedup in SME is achieved by approximating the stochastic trace when estimating model parameters (Fig. C.1). 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 (Figs. C.2 and C.3). The proposed approach enables SME to be effectively applied to the UK Biobank, facilitating genome-wide association studies 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 data set consisting of $N = 349,411$ individuals.

3.3.2 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 (Crawford et al. 2017; Pattillo Smith et al. 2024; Stamp et al. 2023). After quality control, we had a data set of 349,411 individuals and 43,332 SNPs (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 ($\mathcal{S}$) 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 (Fig. C.5). 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 number of individuals being analyzed (varying between 20k, 50k, 100k and 300k randomly subsampled individuals).

Under the null model, we find that SME produces well-calibrated P -values and unbiased variance component estimates (Fig. 3.3 and Table C.1) using a uniformly sparse mask. Specifically, higher levels of sparsity lead to more accurate estimates of the marginal epistatic variance components. We also see 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 3.1). In contrast, FAME produces inflated test statistics as the number of samples in a data set 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. (2017) for assessment of its calibration on small-to-moderately sized data).

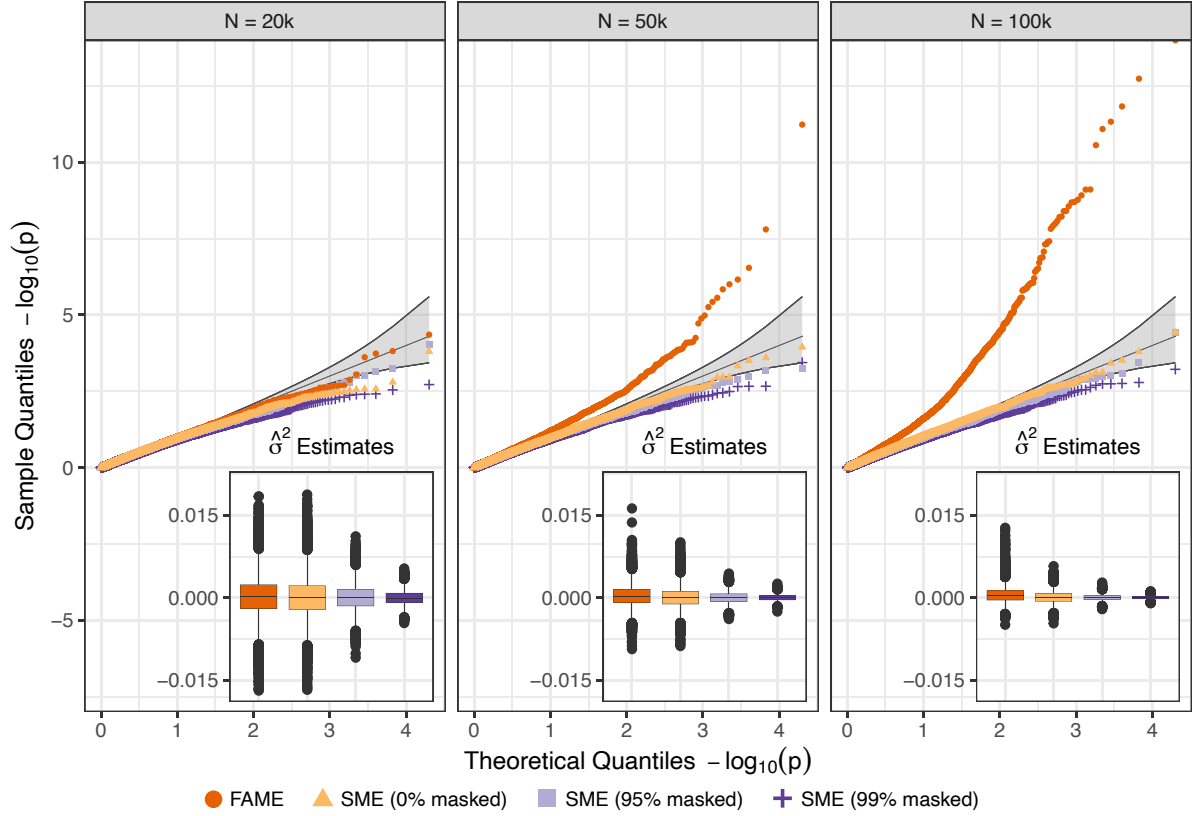


Figure 3.3: **While using a mask that induces uniform sparsity, SME is well-calibrated under the null hypothesis and does not identify epistasis when traits are generated by only additive effects.** Synthetic traits were simulated with only additive effects using chromosome 1 from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled using sample sizes of 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). 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.

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 (Fig. C.6). While type I error control remains conservative (Table C.2), 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.

Table 3.1: **While using a mask that induces uniform sparsity, SME controls 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 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 300k individuals was only analyzed with SME.

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

3.3.3 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 (Crawford et al. 2017; Pattillo Smith et al. 2024; Stamp et al. 2023). These data were subsampled using sample sizes of 50k, 100k, and 300k 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 phenotypic variance explained (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 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 (Fig. C.5A). 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 $P < 5 \times 10^{-8}$ (Fadista et al. 2016).

We find using the uniformly sparse masking scheme significantly enhances the power of SME, with greater levels of sparsity leading to better method performance (Fig. 3.4). When analyzing 300k 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 99.8% power even with relatively small sample sizes (e.g., 50k 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 (Pattillo Smith et al. 2024). Here, we observed that the SME framework protects against the false discovery of nonadditive genetic effects and underestimates the marginal epistatic variance component (σ^2) when causal SNPs involved in pairwise interactions were unobserved (Fig. C.7).

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 resulting in uniform sparsity (Fig. C.8). For example, when analyzing 300k 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 Fig. C.5B). 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 $j = 1, \dots, J$ tested focal SNPs in the data set. 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 (Fig. C.9). For future users of the SME software, we want to note that there is likely an application specific tradeoff between adding SNPs to a mask to reduce potential bias and finding the degree of sparsity needed for a optimally powered

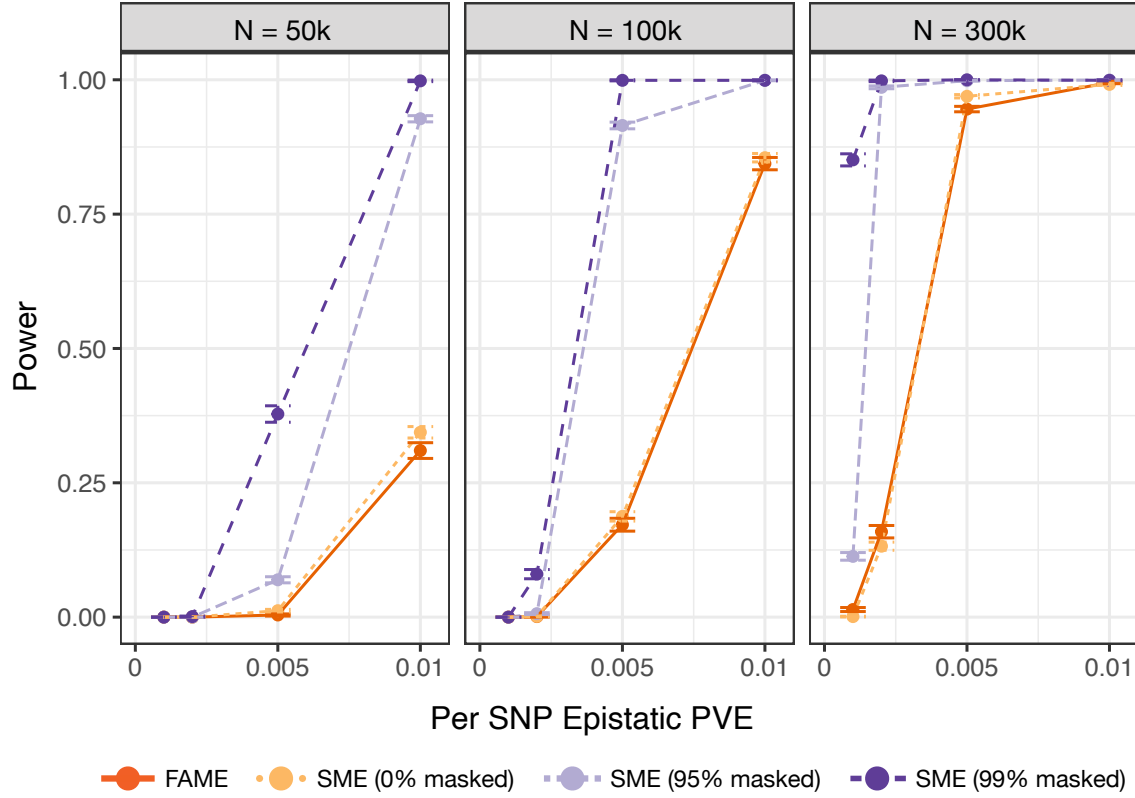


Figure 3.4: **Uniform sparse modeling of interactions enhances the empirical power of SME.** Synthetic traits were simulated with both additive and pairwise epistatic effects using chromosome 1 from individuals of self-identified European ancestry in the UK Biobank. Data were subsampled using sample sizes of 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 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.

test.

3.3.4 SME uses chromatin information to identify epistasis in hematology traits

We apply SME to four hematology traits — mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV), and hematocrit (HCT) — assayed in 349,411 white British individuals in the UK Biobank (Bycroft et al. 2018) genotyped at 543,813 SNPs genome-wide. As an external data source, we leverage DNase I-hypersensitive sites (DHS) data measured over 12 days of *ex vivo* erythroid differentiation (Georgolopoulos et al. 2021). Of the quality controlled SNPs in our data, 4,932 of them are located in DHS regions enriched for transcriptional activity (Maurano et al. 2012). Since previous GWAS results have found genes associated with MCH, MCHC, and MCV to also be implicated in erythroid differentiation (Ding et al. 2012), 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 (Bhoopalan et al. 2020), 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 (Figs. 3.5 and C.10-C.12). 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 nonadditive gene action related to erythropoiesis and red blood cell traits (Table 3.2).

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 secretagomin (*SCGN*). SCGN

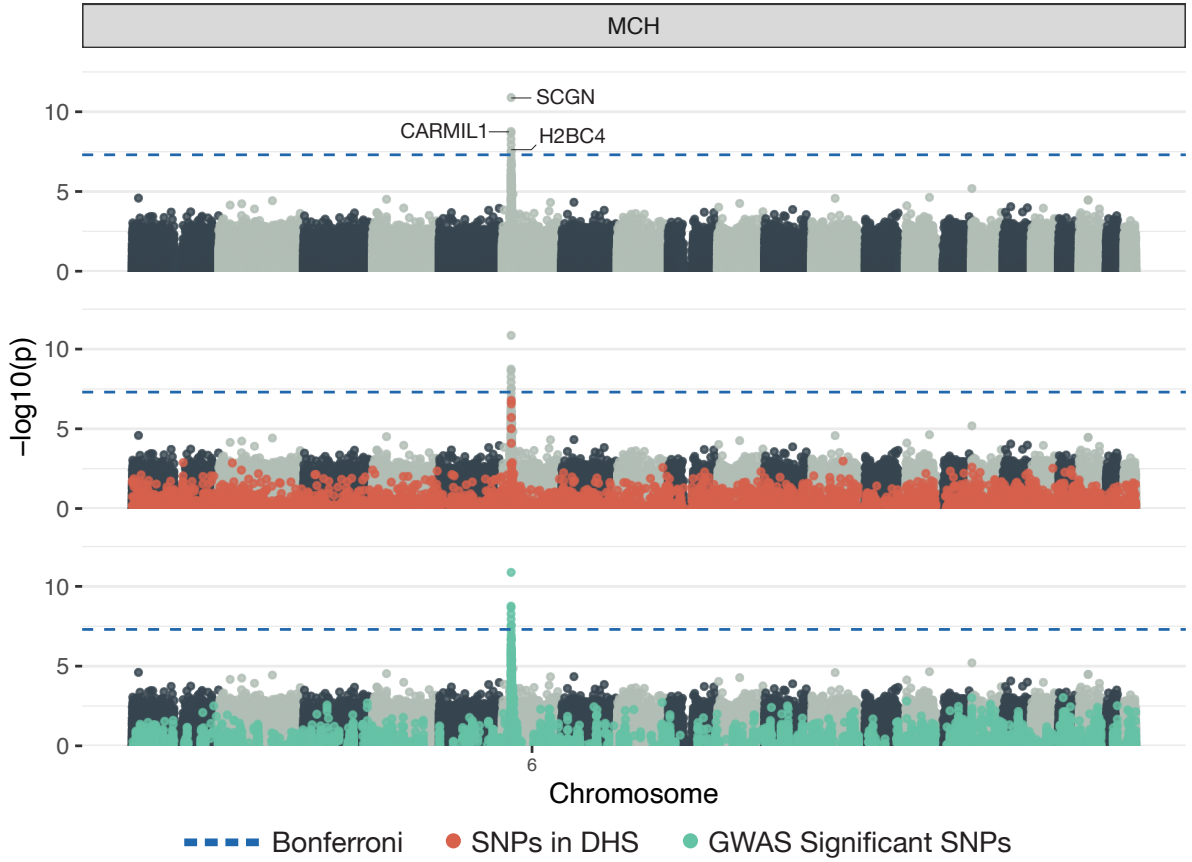


Figure 3.5: **Manhattan plots of a genome-wide interaction analysis using SME to study mean corpuscular hemoglobin (MCH) assayed in individuals in the UK Biobank.** As a mask in this study, we leveraged DNase I-hypersensitive sites (DHS) data measured over 12 days of *ex vivo* erythroid differentiation (Georgolopoulos et al. 2021; Maurano et al. 2012). 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 are also found to have a significant (additive) association with the trait according to a GWAS (Methods).

regulates exocytosis by interacting with two soluble NSF adaptor proteins (SNAP-25 and SNAP-23) and is critical for cell growth in some tissues (Qin et al. 2020). 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 (Ray et al. 2022). Specifically, CARMIL proteins regulate actin dynamics by regulating the activity of CP (Jung et al. 2022; Stark et al. 2017). Erythropoiesis leads to modifications in the expression of membrane and cytoskeletal proteins, whose interactions impact cell structure and function (Nigra et al. 2020; Gokhin et al. 2016). Both *SCGN* and *CARMIL1* have previously been associated with hemoglobin concentration (Timoteo et al. 2023; Ding et al. 2012). A complete list of the results for all traits are listed in Tables C.4-C.7. 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 10k individuals. Importantly, neither baseline identified any genome-wide significant associations (Table 3.2).

3.3.5 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 (Bycroft et al. 2018) and genotyped at the 543,813 SNPs genome-wide. These traits include: body height, mean corpuscular hemoglobin (MCH), uric acid (which we refer to as urate), and vitamin D levels (VITD). 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 VITD (547 SNPs). For each trait, we again use Manhattan plots to visually display the variant-level mapping results, with chromosomes shown in alternating colors for clarity (Figs. C.13-C.16). The nearest genes mapped to the lead SNPs of peaks are highlighted.

Table 3.2: SME identifies marginal epistasis in hematology 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: mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV), and hematocrit (HCT). As a mask, we leveraged DNase I-hypersensitive sites (DHS) data measured over 12 days of *ex vivo* erythroid differentiation (Georgiopoulos et al. 2021; Maurano et al. 2012). 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 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 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 only applied to a random subset of 10k individuals. Coordinates and references in Tab. C.13.

Trait	ID	P -Value	PVE	P -Value (SME 0% Masked)	P -Value (MAPIT)	Gene
MCH	rs4711092	1.41×10^{-11}	0.007	4.78×10^{-4}	0.47	<i>SCGN</i>
MCH	rs9366624	1.80×10^{-9}	0.011	5.75×10^{-7}	9.37×10^{-3}	<i>CARMIL1</i>
MCH	rs9461167	2.34×10^{-9}	0.007	1.40×10^{-5}	3.05×10^{-3}	<i>CARMIL1</i>
MCH	rs9379764	5.53×10^{-9}	0.012	4.58×10^{-6}	0.21	<i>CARMIL1</i>
MCH	rs441460	1.20×10^{-8}	0.008	3.10×10^{-6}	0.36	<i>CARMIL1</i>
MCH	rs198834	2.77×10^{-8}	0.008	2.17×10^{-6}	1.83×10^{-3}	<i>H2BC4</i>
MCH	rs13203202	3.17×10^{-8}	0.012	1.90×10^{-4}	0.13	<i>CARMIL1</i>
MCV	rs9276	9.09×10^{-10}	0.002	7.27×10^{-1}	0.18	<i>HLA-DPB1</i>
MCV	rs9366624	1.86×10^{-8}	0.008	2.63×10^{-7}	0.16	<i>CARMIL1</i>

Importantly, even when conditioning on GWAS summary statistics, SME identifies significant marginal epistatic variants for both height and urate (see Table C.3). 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 (Yengo et al. 2022). A complete list of the results for all traits are listed in Tables C.8-C.11. Again, as 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 10k individuals. Neither baseline had enough power to identify any significant associations at the genome-wide threshold (Table C.3).

3.3.6 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, 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 (Bycroft et al. 2018) 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 (Figs. C.17-C.20).

Next, for all significant marginal epistasis associations, we apply two additional variations of SME and FAME (Table C.12). 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 (Wood et al. 2014). Lastly, despite its inflated observed type I error, FAME does not find significant marginal epistasis at any of these loci.

3.4 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 nonadditive genetic effects only explain a small portion of the overall trait variance (Crawford et al. 2017; Stamp et al. 2023). SME overcomes these limitations by inducing sparsity, essentially limiting the combined interaction for 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 data sets 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 have also found interaction pathways. Split-half analyses on distinct subsets of the UK Biobank also show that the nonadditive 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 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 to be 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 tradeoff within the software.

An important consideration when mapping epistasis in real data is that statistically inferred interactions in genome-wide association studies may arise from same-locus additive effects (Wood et al. 2014). 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. (2021) that an initial set of signals pointing towards evidence of genetic interactions were better explained using linear models of unobserved variants in the same haplotype (Pattillo Smith et al. 2024). 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 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 I/O 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.

3.5 Data and software availability

Source code and tutorials for implementing the "sparse marginal epistasis test" (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 are publicly available at <https://doi.org/10.5281/zenodo.14607997>.

3.6 Web resources

- SME, <https://github.com/lcrawlab/sme>
- MAPIT, <https://github.com/lcrawlab/mvMAPIT>
- FAME, <https://github.com/sriramlab/FAME>
- FAME (GitHub commit hash prefix cfdd03f; date May 7, 2024), <https://doi.org/10.5281/zenodo.14607997>
- UK Biobank, <http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=1967>
- UKB GWAS summary statistics, <https://www.nealelab.is/uk-biobank>
- Linkage disequilibrium maps, <https://bitbucket.org/nygcresearch/ldetect-data/>
- DNase I hypersensitive site (DHS) data, <https://doi.org/10.5281/zenodo.5291736>
- Chain file for genome liftover, <http://hgdownload.soe.ucsc.edu/goldenPath/hg38/liftOver/>

3.7 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. Crawford. 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 Numbers 14649. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of any of the funders.

Conclusion

In this dissertation, I have contributed several advances to the field of statistical genetics. Across three chapters, I developed and characterized statistical models to study gene-by-gene and gene-by-environment interactions, unlock the scalability of compute intensive methods, and provide an approach to a-priori integrate multi-omic data as biological priors for association studies. Beyond the theoretical contributions, I have published implementations of the methods as open-source software that are available to the community.

In Chapter 1, I introduced a multivariate marginal epistasis test which is designed to detect variants involved in gene-by-gene interactions with effects on multiple traits. It leverages genetic correlations to partition covariance components and detects marginal epistasis. In extensive simulations, I demonstrated that this approach outperforms standard univariate methods. Applied to hematological traits assayed in mice, it identified novel variants involved in gene-by-gene interactions.

In Chapter 2, I developed a frailty approach to test for gene-by-environment interactions. It builds on the marginal interaction framework but adapts the framework to time-to-event

traits. In order to handle the linear mixed model setup required by the framework, the Cox proportional hazards ansatz was reformulated as a Poisson generalized linear mixed model. This method partitions variation of frailty in e.g. survival or age-at-onset traits into genetic, environmental exposure, and gene-by-environment interaction components. As a starting point for making this method scale to biobank-scale data, I proposed to cache compute intensive quantities in the null model fit and employ an efficient score test for association testing. In simulations, I demonstrated that this method is well-powered and correctly prioritizes variants involved in gene-by-environment interactions.

In Chapter 3, I presented a marginal epistasis test which incorporates external biological data as priors into genome-wide interaction analyses. The immediate payoff of this approach are twofold. First, I showed in extensive simulations that using appropriate priors significantly improves the efficiency of the estimators resulting in a substantial boost in power. Second, I showed that the resulting sparse modeling of interactions together with algorithmic improvements unlocks the scalability of the marginal interaction framework. Both in extensive simulations and in an application to hematological traits from the UK Biobank, I demonstrated that this method vastly outperforms existing methods in terms of scalability and power. Using DNase-seq data from ex vivo erythroid differentiation, the method identified statistical epistasis in genes associated with protein interactions during erythroid differentiation.

The role that nonlinear contributions to phenotypic variation play in human genetics has been a topic of debate for a long time (Hill et al. [2008](#)). So far, modeling genotype to phenotype mappings beyond the linear approximation have incurred substantial statistical and computational challenges, casting uncertainty on the role gene-by-gene and gene-by-environment interactions play for the genetic architecture of complex traits. While the present dissertation is not sufficient to give a conclusive answer about whether or not statistically searching for interactions will lead to new discoveries, let alone a more profound understanding of the genetic architecture of complex traits, it makes multiple contributions

that will help shed light on the role of gene interactions. Most importantly, it introduces computational and statistical modeling advances that make it possible to falsify the theoretical predictions about the role of interactions by enabling genome-wide association studies of nonlinear contributions to complex traits at biobank-scale. Longstanding conceptual debates about interactions become empirically testable questions. Beyond the immediate application of the proposed methods, the framework developed here serves as a scalable example for integrating multi-omics and environmental data into rigorous statistical inference thereby narrowing the divide between statistical associations and their underlying biological mechanisms. I believe that the methodological and computational innovations introduced in this dissertation will improve our understanding of the nonlinear contributions to phenotypic variation.

Having layed out the contributions of this dissertation, I would like to reflect on thoughts that I did not report or emphasize in the chapters, either because they are conjectures or merely unexpected observations I have not had the chance to explore. There is a possibility that the explanations will be obvious for a reader more familiar with the field.

I was surprised by the pattern that emerged in the results of applying mvMAPIT to hematology traits of heterogeneous stock mice data from the Wellcome Trust Centre for Human Genetics. None of the univariate tests—except for ALY and LIC¹—showed a signal of marginal epistasis. However, the covariance component statistic suggests significant marginal epistasis for many trait pairs. Notably, the signal in the covariance component statistic is independent of the correlation between traits. The results that include the covariance highlight new associations for traits whose measurements are both highly correlated (e.g. hematocrit and hemoglobin) or not correlated at all (e.g. basophils and platelets). A reasonable thought would be to question whether the covariance model is calibrated. However, in simulations I can show that the covariance statistic is conservative,

¹I suspect the signal in ALY and LIC may be driven by outliers in the trait measurements.

with QQ-plots indicating a deflation of the type I error rate. More likely, it shows that trait covariance and covariance of trait variance components are independent. It appears that minor differences in traits can have a great impact on the covariance component statistic. The univariate statistic and the covariance statistic converge when the traits are identical—therefore, how can there be such a strong signal in the covariance component statistic when the univariate test shows no signal and the traits are highly correlated? How important for this result is it that the mouse genotypes are from an experimental cross and not from a natural population? The traits hematocrit and hemoglobin are not just highly correlated, but are related measurements on the same blood cells. This makes me wonder about the impact phenotype definitions may have on studying complex traits. Could we go as far as computing additional traits from one another to boost discovery in multivariate association studies?

Working with genetic data from natural populations in chapters 2 and 3, I repeatedly had to grapple with the circumstance that the structure of the feature space severely impacts the statistical inference. While generally this may be expected, it surprised me to what extent the here presented methods are affected, in particular because the feature space is not used explicitly but used to compute covariance matrices. With unstructured data, CphGxE is well calibrated and efficiently prioritizes gene-by-environment interaction variants. Using real genotypes, even after LD pruning to minimize correlation between variants, the performance of the method suffers. The calibration becomes dependent on censoring rate, the error in the frailty variance estimate of the polygenic background increases, and the power and discriminatory ability decreases. In fact, this trend amplifies for the interaction covariance matrix. In SME, localized sparsity i.e. mostly high LD features, negatively bias the variance component estimate of interactions. Is there a trade-off necessary between a feature space that produces well calibrated tests and our ability to capture nonlinear contributions to phenotypic variation? How can we go about quantitatively characterizing what aspects of the feature space impact the statistical

inference?

Lastly, I would like to outline a few future directions for this work. I believe that much of the future progress in the field will come from scaling the available genetic data sets. To be able to leverage these data efficiently, scalability of methods is crucial.

There are a number of algorithmic improvements that could be made to the methods I have presented in this work, some of which I believe could be beneficial for other methods in the field as well. In order to improve scalability of CphGxE, the implementation of the null fit should build on the ideas introduced in SME. Specifically, it should employ column-wise streaming of large data such as genotypes, and using a block wise computation of quadratic forms wherever possible. This approach significantly improves the efficient use of compute resources. While computing clusters often have many nodes and CPUs to distribute load, managing the memory requirements is crucial to maximize the use of available compute resources. The most time consuming part of CphGxE is the null fit, however that part has been shown to scale to biobank data in previous work. In order to make the efficient score statistic scale, it would be great to store the projection matrix computed in the null fit in a way that it as well can be streamed and reused for the score statistic. Doing so will make it possible to scale the computation of the average information matrix as well as the score statistic. In general, I believe a lot of computational efficiency can be gained in statistical methods by partially precomputing quantities. It has been shown repeatedly that there are diminishing returns when computing the genetic relatedness matrix from more and more variants. I believe it could be beneficial to develop software to make it efficient and easy to precompute and stream measures of genetic relatedness. Many mixed model methods, in particular in the marginal interaction framework as they tend to have to repeatedly recompute the similar quantities.

One angle to advance computational efficiency for statistical genetics methods that I would like to pursue further, is developing a C++ library that makes it easy to use the mailman algorithm. Conceptually, the mailman algorithm is a more efficient way to

compute matrix-vector products when the matrix is over a finite alphabet. This makes it a great tool for genomic data analysis as often data is represented as a finite set of counts. In the current implementation of the mailman algorithm in SME, there is a pre-processing step that resembles encoding of a genotype block. Over a genome-wide test, substantial time is spent on this encoding step even though it is deterministic and in theory is the same for every trait. I envision that the C++ library implementation of the mailman algorithm could provide a preprocessing function and file format for efficiently storing and streaming the encoded genotype block data ready for mailman matrix-vector products.

In terms of the future of methods that detect nonlinear contributions to phenotypic variation, I believe that summary statistics from genome-wide association studies hold a lot of potential. Marginal effects do not only capture the true additive effects of variants in LD with the test variant. In theory, the marginal effect size equally captures interaction effects (Mäki-Tanila et al. [2014](#)). While the relative contribution of interaction effects may be much smaller, especially because they are weighted by allele frequency, the sheer number of interactions that may be captured makes me wonder whether summary statistics from genome-wide association studies could be used to detect nonlinear contributions to phenotypic variation. The key to this approach is recognizing that the portion of interaction effects captured by the marginal effect size will depend on the allele frequency. In future work, I would like to explore whether summary statistics from genome-wide association studies on multiple populations with different allele frequencies could be used to detect nonlinear contributions to phenotypic variation.

APPENDIX A

Supplementary information for Chapter 1

A.1 Supplementary figures

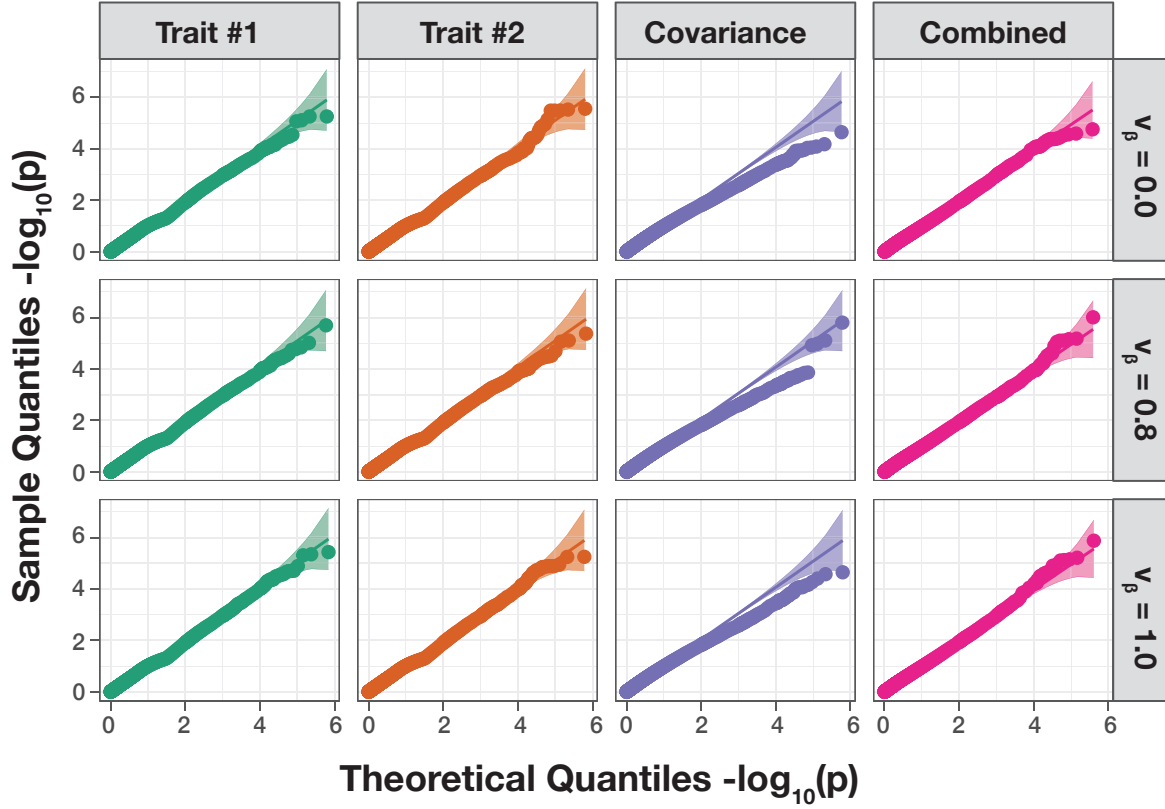


Figure A.1: The mvMAPIT framework using Fisher’s method produces well-calibrated P-values when traits are generated by only additive effects (sample size $N = 1,000$ individuals). In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row of quantile-quantile (QQ) plots corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent additive effects ($v_\beta = 0$), traits with highly correlated additive effects ($v_\beta = 0.8$), and traits with perfectly correlated additive effects ($v_\beta = 1$). The first two columns show P-values resulting from the univariate MAPIT test on “trait #1” and “trait #2”, respectively. The third column depicts the “covariance” P-values which corresponds to assessing the pairwise interactions affecting both traits is. Lastly, the fourth column shows the final “combined” P-value which combines the P-values from the first three columns using Fisher’s method. The 95% confidence interval for the null hypothesis of no marginal epistatic effects is shown in grey. Each plot combines results from 100 simulated replicates.

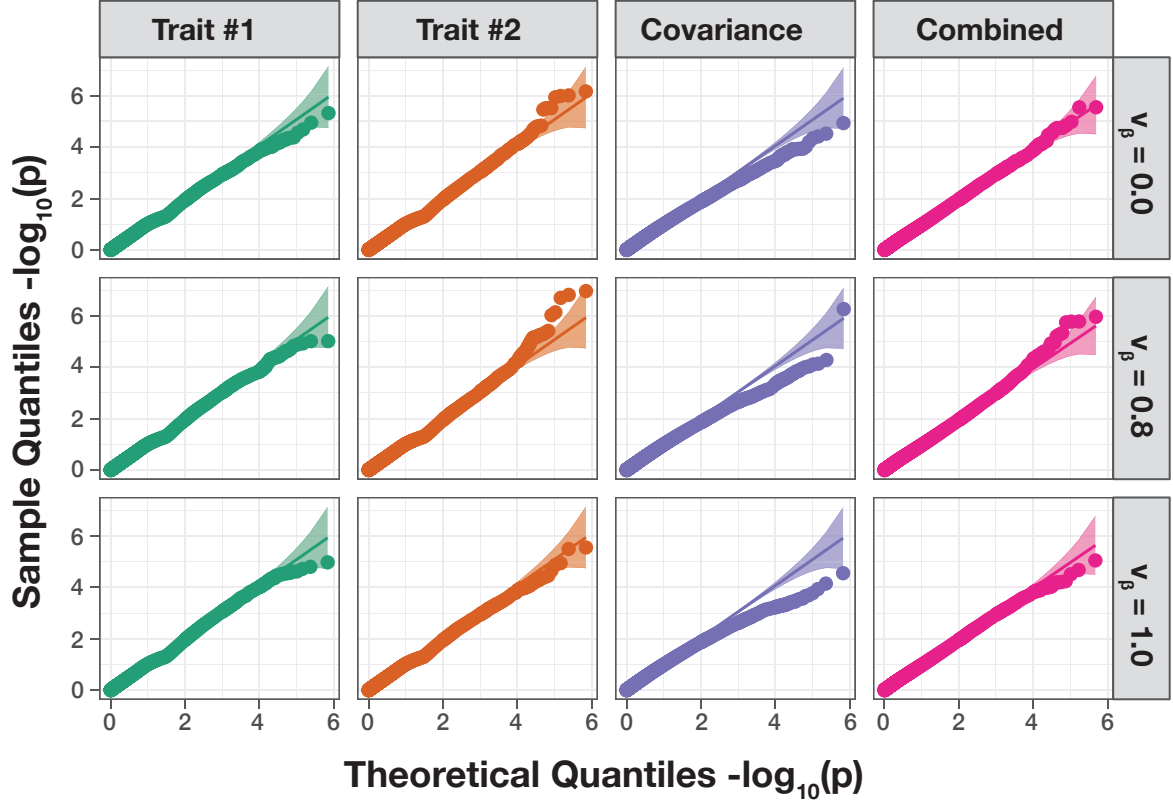


Figure A.2: The mvMAPIT framework using Fisher’s method produces well-calibrated P -values when traits are generated by only additive effects (sample size $N = 1,750$ individuals). In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row of quantile-quantile (QQ) plots corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent additive effects ($v_\beta = 0$), traits with highly correlated additive effects ($v_\beta = 0.8$), and traits with perfectly correlated additive effects ($v_\beta = 1$). The first two columns show P -values resulting from the univariate MAPIT test on “trait #1” and “trait #2”, respectively. The third column depicts the “covariance” P -values which corresponds to assessing the pairwise interactions affecting both traits is. Lastly, the fourth column shows the final “combined” P -value which combines the P -values from the first three columns using Fisher’s method. The 95% confidence interval for the null hypothesis of no marginal epistatic effects is shown in grey. Each plot combines results from 100 simulated replicates.

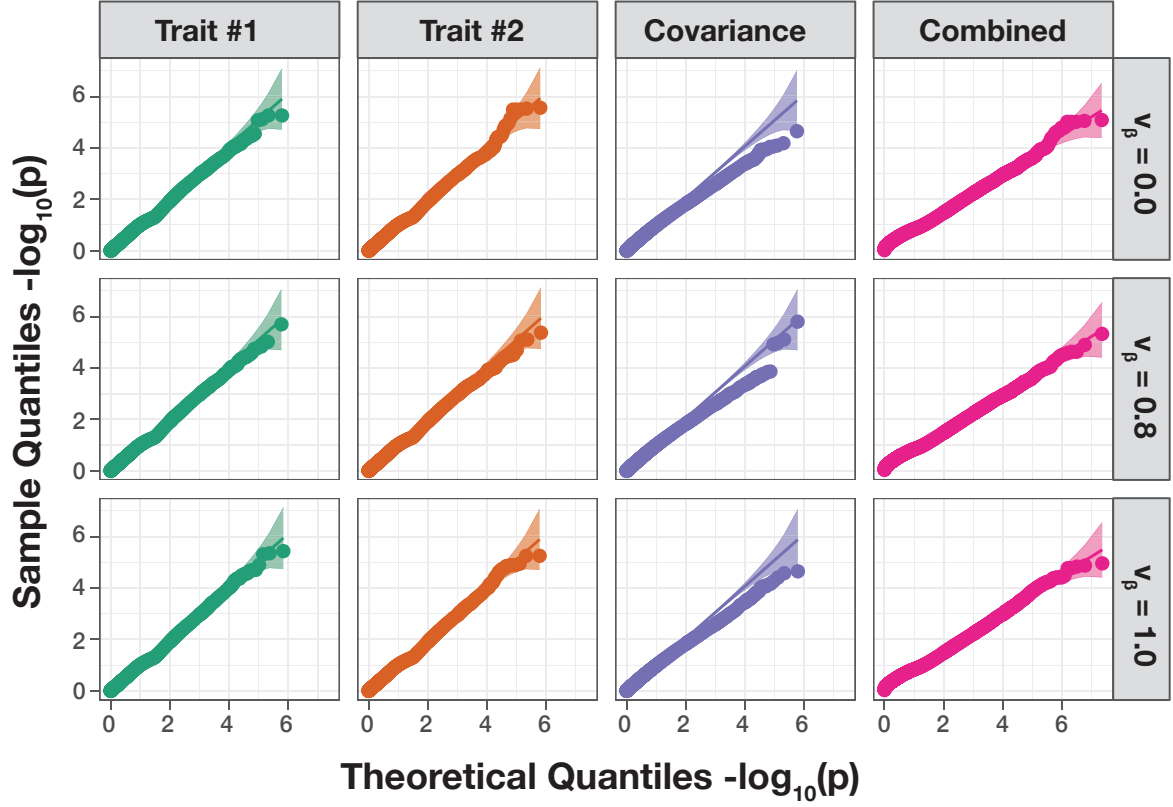


Figure A.3: **The mvMAPIT framework using the harmonic mean produces well-calibrated P-values when traits are generated by only additive effects (sample size $N = 1,000$ individuals).** In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row of quantile-quantile (QQ) plots corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent additive effects ($v_\beta = 0$), traits with highly correlated additive effects ($v_\beta = 0.8$), and traits with perfectly correlated additive effects ($v_\beta = 1$). The first two columns show P -values resulting from the univariate MAPIT test on “trait #1” and “trait #2”, respectively. The third column depicts the “covariance” P -values which corresponds to assessing the pairwise interactions affecting both traits is. Lastly, the fourth column shows the final “combined” P -value which combines the P -values from the first three columns using Fisher’s method. The 95% confidence interval for the null hypothesis of no marginal epistatic effects is shown in grey. Each plot combines results from 100 simulated replicates.

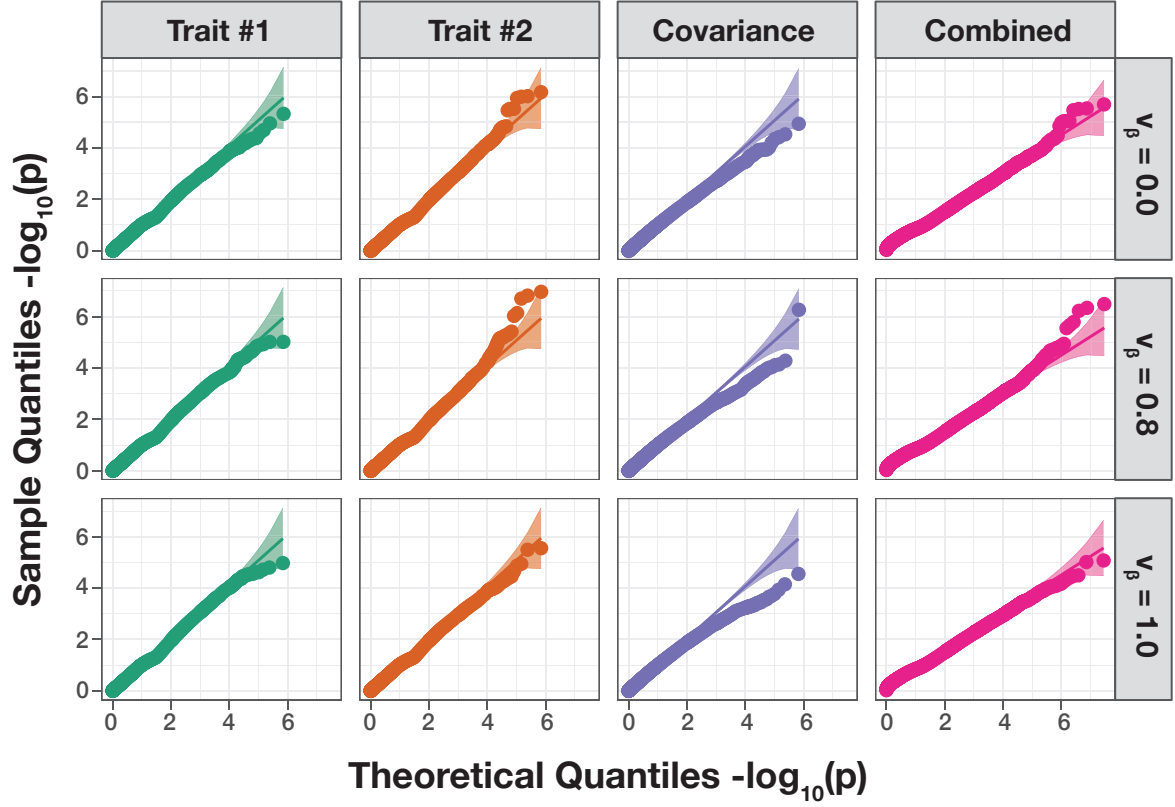


Figure A.4: The mvMAPIT framework using the harmonic mean produces well-calibrated P-values when traits are generated by only additive effects (sample size $N = 1,750$ individuals). In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row of quantile-quantile (QQ) plots corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent additive effects ($v_\beta = 0$), traits with highly correlated additive effects ($v_\beta = 0.8$), and traits with perfectly correlated additive effects ($v_\beta = 1$). The first two columns show P-values resulting from the univariate MAPIT test on “trait #1” and “trait #2”, respectively. The third column depicts the “covariance” P-values which corresponds to assessing the pairwise interactions affecting both traits is. Lastly, the fourth column shows the final “combined” P-value which combines the P-values from the first three columns using Fisher’s method. The 95% confidence interval for the null hypothesis of no marginal epistatic effects is shown in grey. Each plot combines results from 100 simulated replicates.

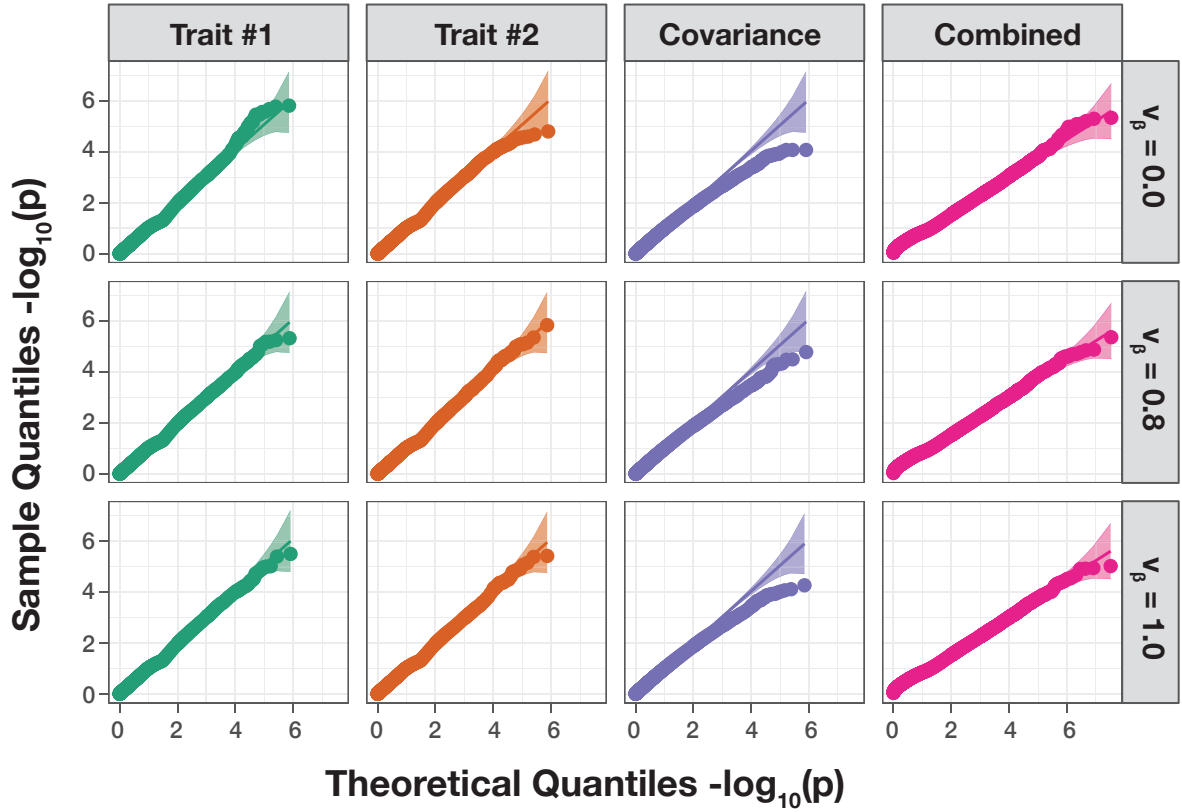


Figure A.5: The mvMAPIT framework using the harmonic mean produces well-calibrated P-values when traits are generated by only additive effects (sample size $N = 2,500$ individuals). In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row of quantile-quantile (QQ) plots corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent additive effects ($v_\beta = 0$), traits with highly correlated additive effects ($v_\beta = 0.8$), and traits with perfectly correlated additive effects ($v_\beta = 1$). The first two columns show P-values resulting from the univariate MAPIT test on “trait #1” and “trait #2”, respectively. The third column depicts the “covariance” P-values which corresponds to assessing the pairwise interactions affecting both traits is. Lastly, the fourth column shows the final “combined” P-value which combines the P-values from the first three columns using Fisher’s method. The 95% confidence interval for the null hypothesis of no marginal epistatic effects is shown in grey. Each plot combines results from 100 simulated replicates.

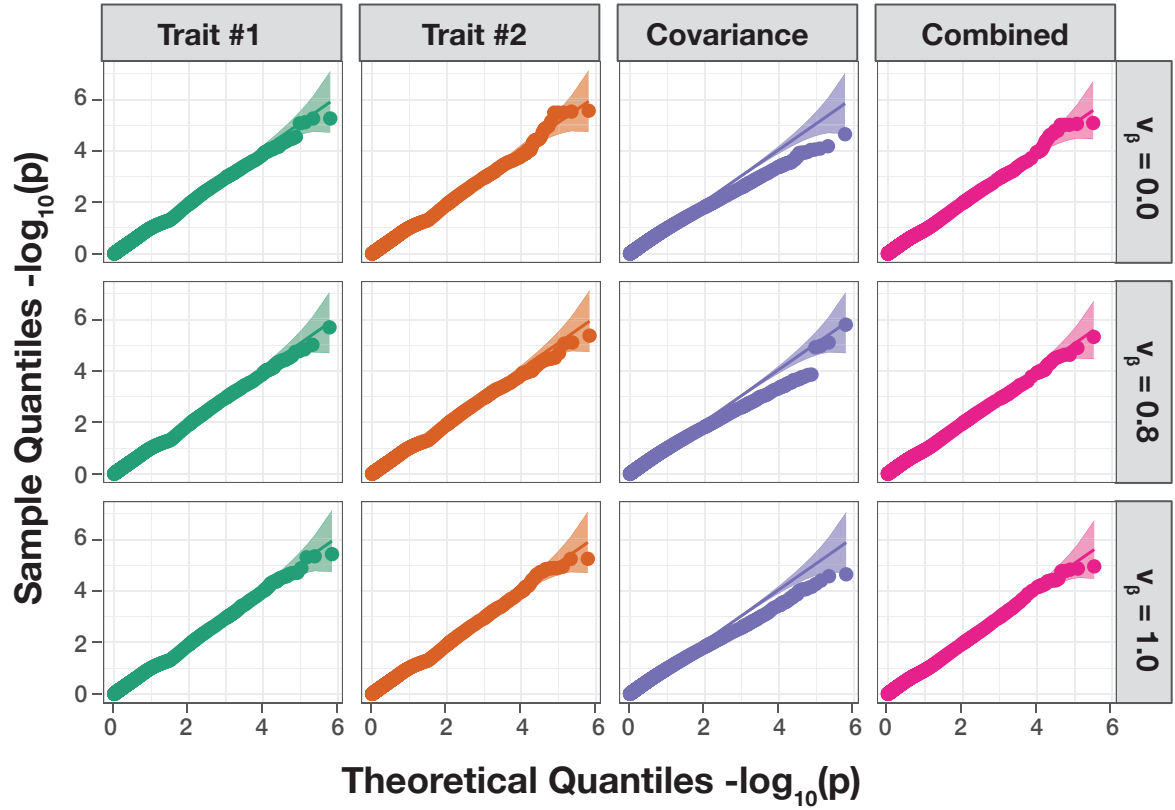


Figure A.6: The mvMAPIT framework using the Cauchy combined test produces well-calibrated P -values when traits are generated by only additive effects (sample size $N = 1,000$ individuals). In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row of quantile-quantile (QQ) plots corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent additive effects ($v_\beta = 0$), traits with highly correlated additive effects ($v_\beta = 0.8$), and traits with perfectly correlated additive effects ($v_\beta = 1$). The first two columns show P -values resulting from the univariate MAPIT test on “trait #1” and “trait #2”, respectively. The third column depicts the “covariance” P -values which corresponds to assessing the pairwise interactions affecting both traits is. Lastly, the fourth column shows the final “combined” P -value which combines the P -values from the first three columns using Fisher’s method. The 95% confidence interval for the null hypothesis of no marginal epistatic effects is shown in grey. Each plot combines results from 100 simulated replicates.

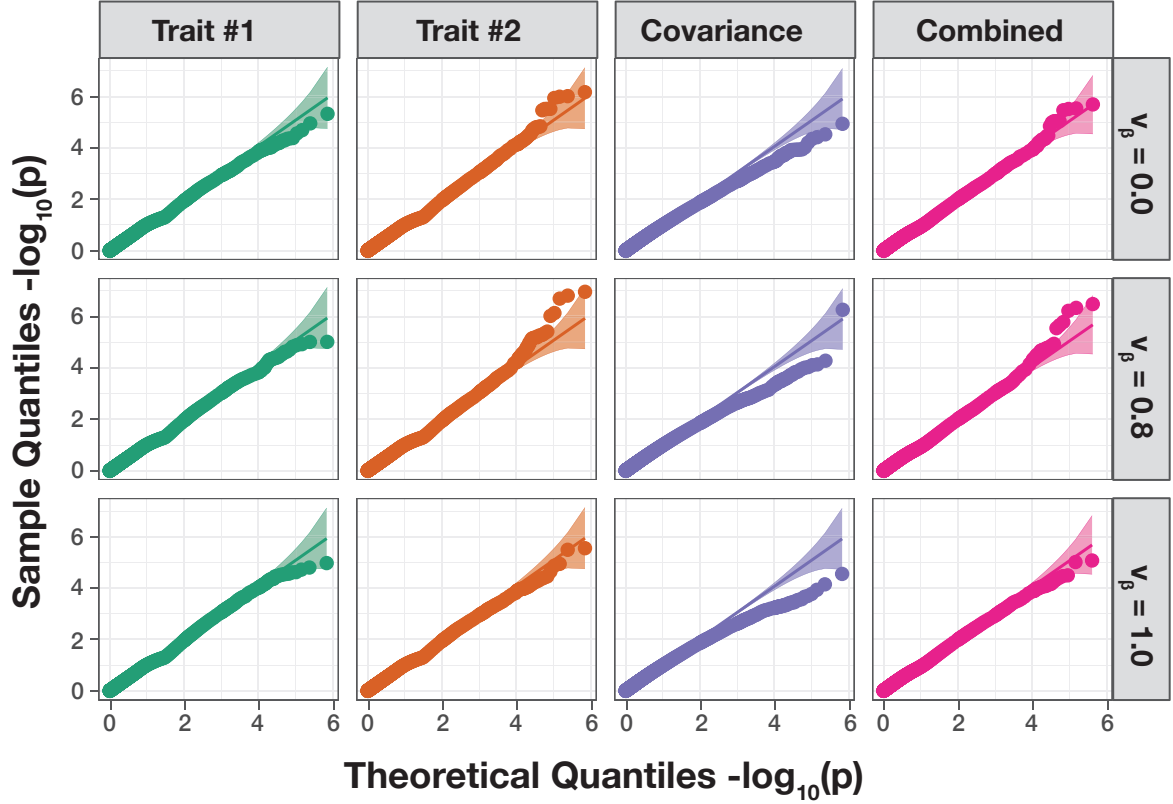


Figure A.7: The mvMAPIT framework using the Cauchy combined test produces well-calibrated P-values when traits are generated by only additive effects (sample size $N = 1,750$ individuals). In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row of quantile-quantile (QQ) plots corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent additive effects ($v_\beta = 0$), traits with highly correlated additive effects ($v_\beta = 0.8$), and traits with perfectly correlated additive effects ($v_\beta = 1$). The first two columns show P-values resulting from the univariate MAPIT test on “trait #1” and “trait #2”, respectively. The third column depicts the “covariance” P-values which corresponds to assessing the pairwise interactions affecting both traits is. Lastly, the fourth column shows the final “combined” P-value which combines the P-values from the first three columns using Fisher’s method. The 95% confidence interval for the null hypothesis of no marginal epistatic effects is shown in grey. Each plot combines results from 100 simulated replicates.

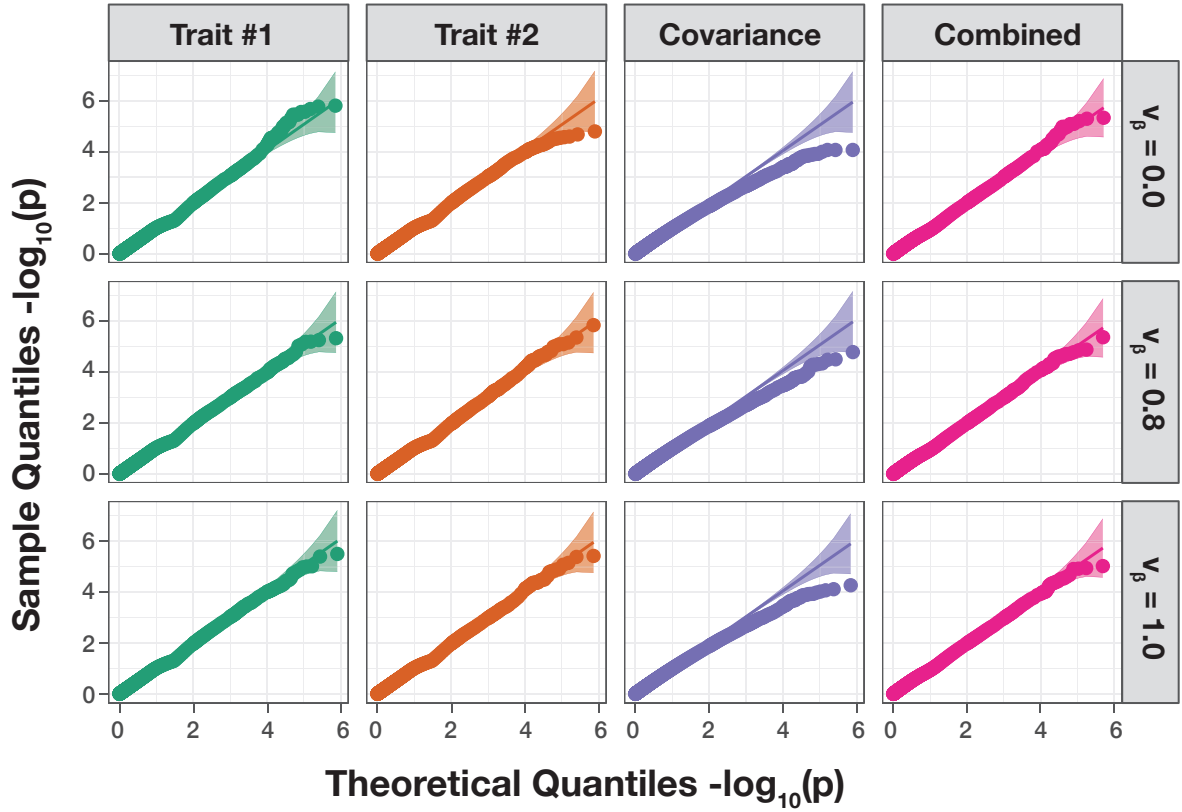


Figure A.8: The mvMAPIT framework using the Cauchy combined test produces well-calibrated P -values when traits are generated by only additive effects (sample size $N = 2,500$ individuals). In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row of quantile-quantile (QQ) plots corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent additive effects ($v_\beta = 0$), traits with highly correlated additive effects ($v_\beta = 0.8$), and traits with perfectly correlated additive effects ($v_\beta = 1$). The first two columns show P -values resulting from the univariate MAPIT test on “trait #1” and “trait #2”, respectively. The third column depicts the “covariance” P -values which corresponds to assessing the pairwise interactions affecting both traits is. Lastly, the fourth column shows the final “combined” P -value which combines the P -values from the first three columns using Fisher’s method. The 95% confidence interval for the null hypothesis of no marginal epistatic effects is shown in grey. Each plot combines results from 100 simulated replicates.

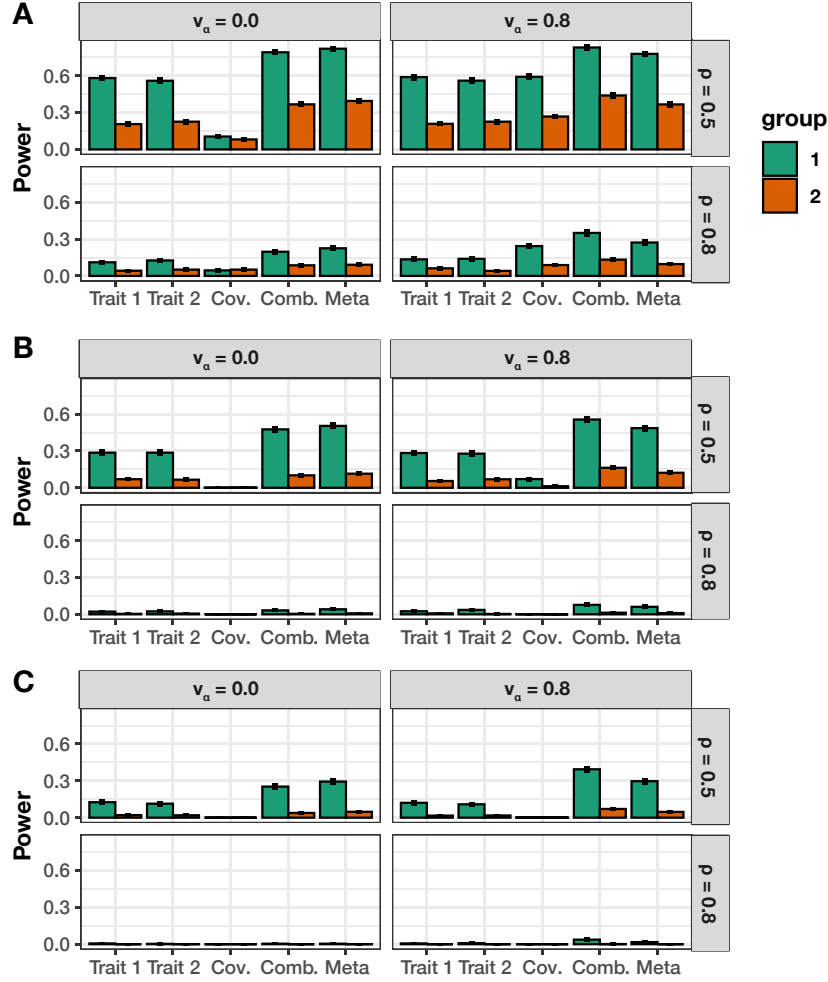


Figure A.9: **Empirical power of mvMAPIT with Fisher's method to detect group #1 (10) and group #2 (20) epistatic variants across complex traits with moderate broad-sense heritability.** In these simulations, both quantitative traits are simulated to have broad-sense heritability $H^2 = 0.6$ with architectures made up of both additive and epistatic effects. The parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to a setting where the epistatic effects for interactive pairs have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent epistatic effects ($v_\alpha = 0$) and traits with highly correlated epistatic effects ($v_\alpha = 0.8$). This plot shows the empirical power of mvMAPIT at significance levels **(A)** $P = 5 \times 10^{-2}$, **(B)** $P = 5 \times 10^{-4}$, and **(C)** $P = 1 \times 10^{-5}$, respectively. Group #1 and #2 causal markers are colored in green and orange, respectively. For comparison, the “trait #1” and “trait #2” bars correspond to the univariate MAPIT model, the “cov” bars corresponds to power contributed by the covariance test, and “comb” details power from the overall association identified by mvMAPIT in the combination approach. Lastly, the “meta” bars shows power from the association identified by combining only the P -values of the two univariate cases “trait #1” and “trait #2”. Results are based on 100 simulations per parameter combination and the horizontal bars represent standard errors.

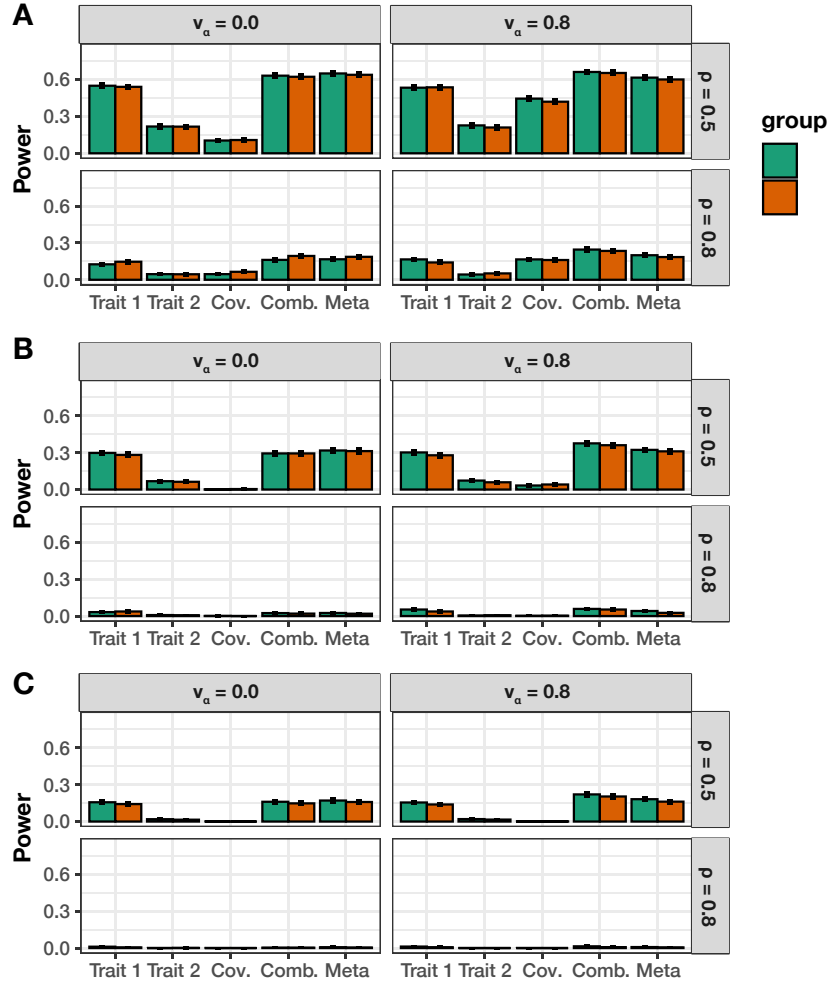


Figure A.10: **Empirical power of mvMAPIT with Fisher's method to detect group #1 (10) and group #2 (10) epistatic variants across complex traits with different levels of broad-sense heritability.** In these simulations, one of the quantitative traits has a moderate broad-sense heritability $H^2 = 0.6$, while the other has heritability $H^2 = 0.3$. Both traits have architectures made up of both additive and epistatic effects. The parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to a setting where the epistatic effects for interactive pairs have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent epistatic effects ($v_\alpha = 0$) and traits with highly correlated epistatic effects ($v_\alpha = 0.8$). This plot shows the empirical power of mvMAPIT at significance levels (A) $P = 5 \times 10^{-2}$, (B) $P = 5 \times 10^{-4}$, and (C) $P = 1 \times 10^{-5}$, respectively. Group #1 and #2 causal markers are colored in green and orange, respectively. For comparison, the “trait #1” and “trait #2” bars correspond to the univariate MAPIT model, the “cov” bars corresponds to power contributed by the covariance test, and “comb” details power from the overall association identified by mvMAPIT in the combination approach. Lastly, the “meta” bars shows power from the association identified by combining only the P -values of the two univariate cases “trait #1” and “trait #2”. Results are based on 100 simulations per parameter combination and the horizontal bars represent standard errors.

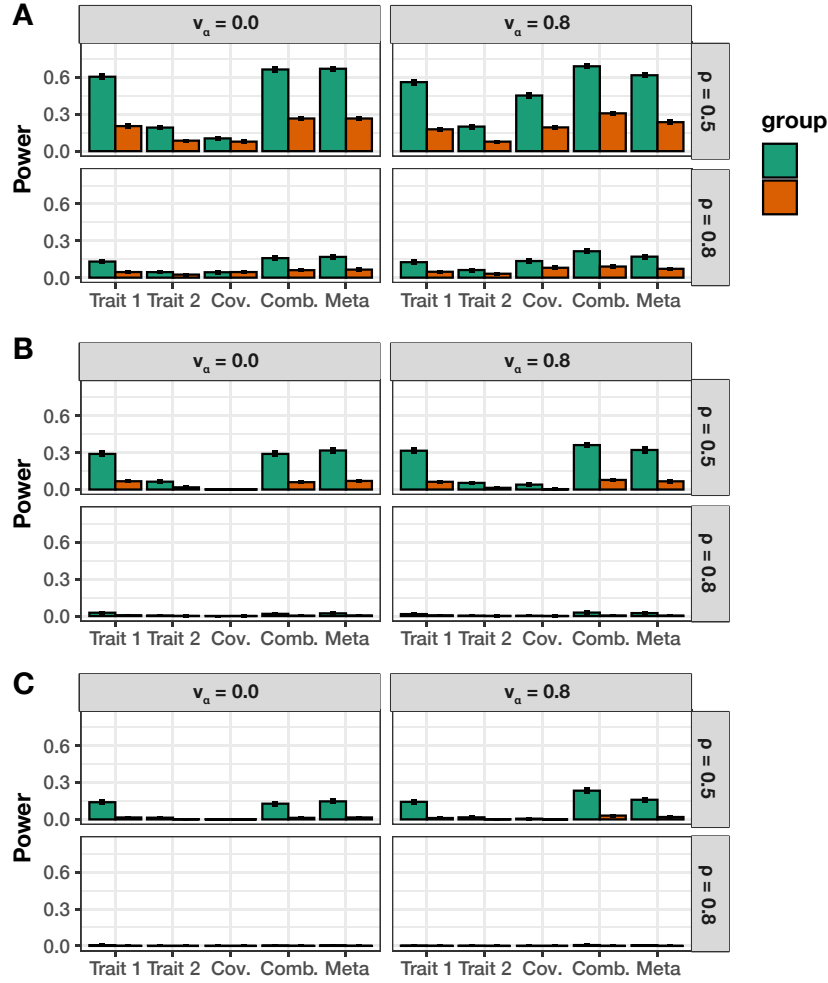


Figure A.11: **Empirical power of mvMAPIT with Fisher's method to detect group #1 (10) and group #2 (20) epistatic variants across complex traits with different levels of broad-sense heritability.** In these simulations, one of the quantitative traits has a moderate broad-sense heritability $H^2 = 0.6$, while the other has heritability $H^2 = 0.3$. Both traits have architectures made up of both additive and epistatic effects. The parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to a setting where the epistatic effects for interactive pairs have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent epistatic effects ($v_\alpha = 0$) and traits with highly correlated epistatic effects ($v_\alpha = 0.8$). This plot shows the empirical power of mvMAPIT at significance levels (A) $P = 5 \times 10^{-2}$, (B) $P = 5 \times 10^{-4}$, and (C) $P = 1 \times 10^{-5}$, respectively. Group #1 and #2 causal markers are colored in green and orange, respectively. For comparison, the “trait #1” and “trait #2” bars correspond to the univariate MAPIT model, the “cov” bars corresponds to power contributed by the covariance test, and “comb” details power from the overall association identified by mvMAPIT in the combination approach. Lastly, the “meta” bars shows power from the association identified by combining only the P -values of the two univariate cases “trait #1” and “trait #2”. Results are based on 100 simulations per parameter combination and the horizontal bars represent standard errors.

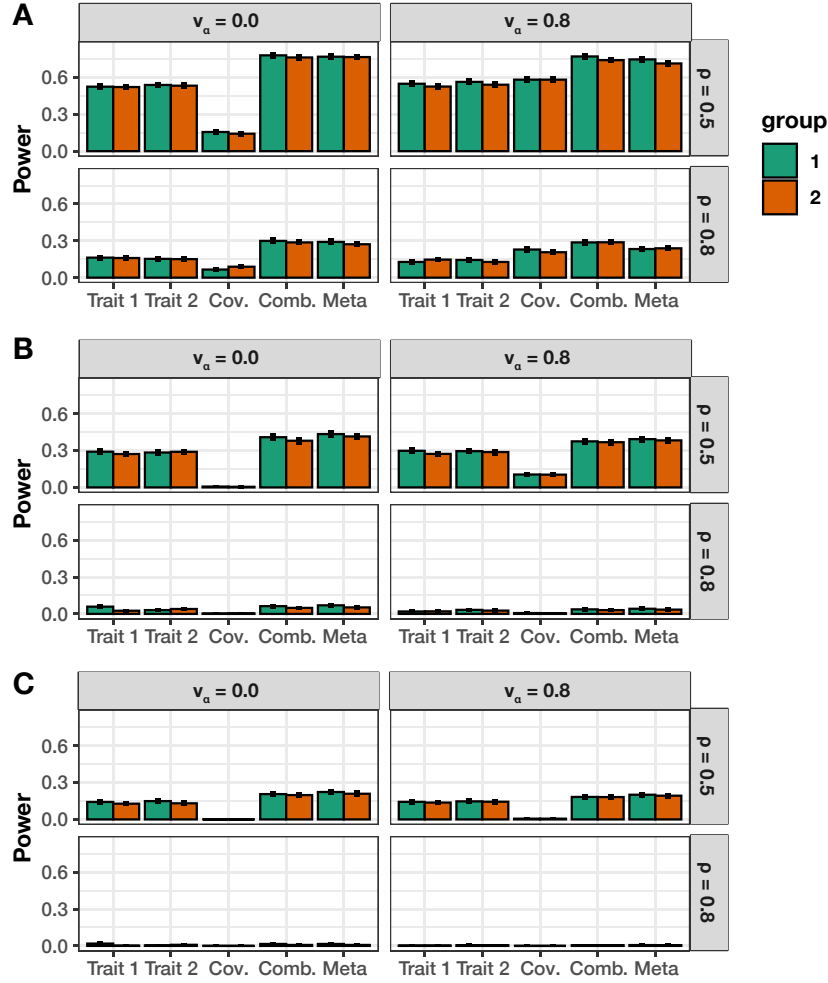


Figure A.12: **Empirical power of mvMAPIT with the harmonic mean combination approach to detect group #1 (10) and group #2 (10) epistatic variants across complex traits with moderate broad-sense heritability.** In these simulations, both quantitative traits are simulated to have broad-sense heritability $H^2 = 0.6$ with architectures made up of both additive and epistatic effects. The parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to a setting where the epistatic effects for interactive pairs have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent epistatic effects ($v_\alpha = 0$) and traits with highly correlated epistatic effects ($v_\alpha = 0.8$). This plot shows the empirical power of mvMAPIT at significance levels (A) $P = 5 \times 10^{-2}$, (B) $P = 5 \times 10^{-4}$, and (C) $P = 1 \times 10^{-5}$, respectively. Group #1 and #2 causal markers are colored in green and orange, respectively. For comparison, the “trait #1” and “trait #2” bars correspond to the univariate MAPIT model, the “cov” bars corresponds to power contributed by the covariance test, and “comb” details power from the overall association identified by mvMAPIT in the combination approach. Lastly, the “meta” bars shows power from the association identified by combining only the P -values of the two univariate cases “trait #1” and “trait #2”. Results are based on 100 simulations per parameter combination and the horizontal bars represent standard errors.

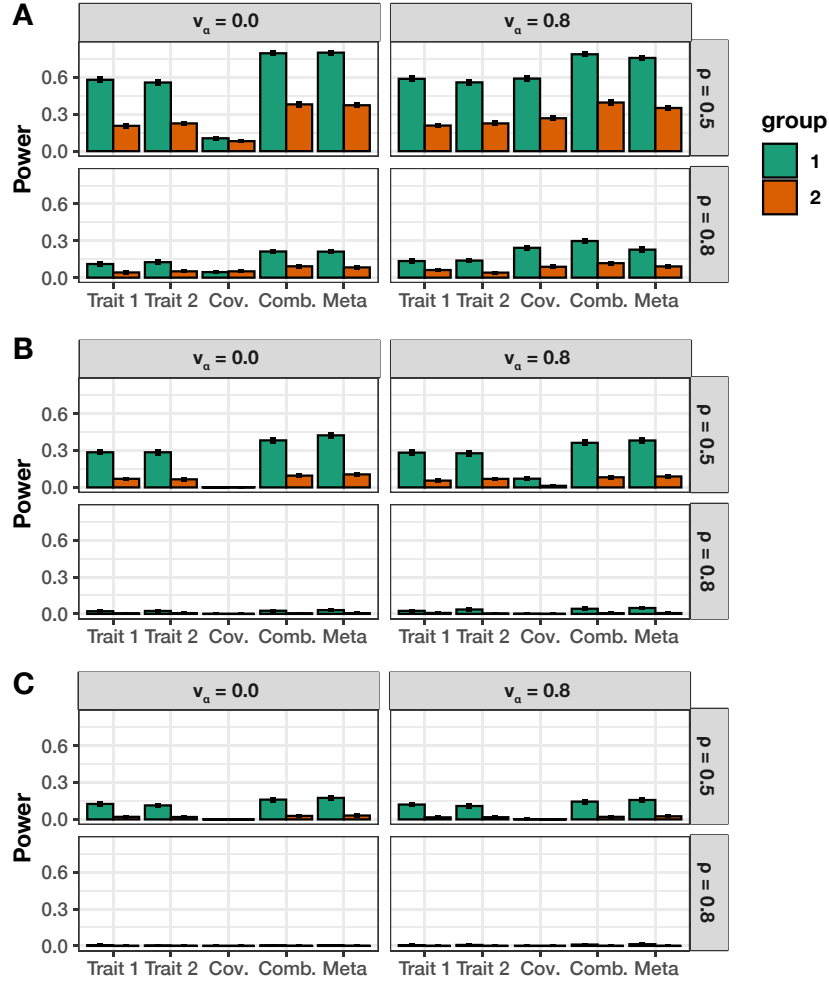


Figure A.13: **Empirical power of mvMAPIT with the harmonic mean combination approach to detect group #1 (10) and group #2 (20) epistatic variants across complex traits with moderate broad-sense heritability.** In these simulations, both quantitative traits are simulated to have broad-sense heritability $H^2 = 0.6$ with architectures made up of both additive and epistatic effects. The parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to a setting where the epistatic effects for interactive pairs have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent epistatic effects ($v_\alpha = 0$) and traits with highly correlated epistatic effects ($v_\alpha = 0.8$). This plot shows the empirical power of mvMAPIT at significance levels (A) $P = 5 \times 10^{-2}$, (B) $P = 5 \times 10^{-4}$, and (C) $P = 1 \times 10^{-5}$, respectively. Group #1 and #2 causal markers are colored in green and orange, respectively. For comparison, the “trait #1” and “trait #2” bars correspond to the univariate MAPIT model, the “cov” bars corresponds to power contributed by the covariance test, and “comb” details power from the overall association identified by mvMAPIT in the combination approach. Lastly, the “meta” bars shows power from the association identified by combining only the P -values of the two univariate cases “trait #1” and “trait #2”. Results are based on 100 simulations per parameter combination and the horizontal bars represent standard errors.

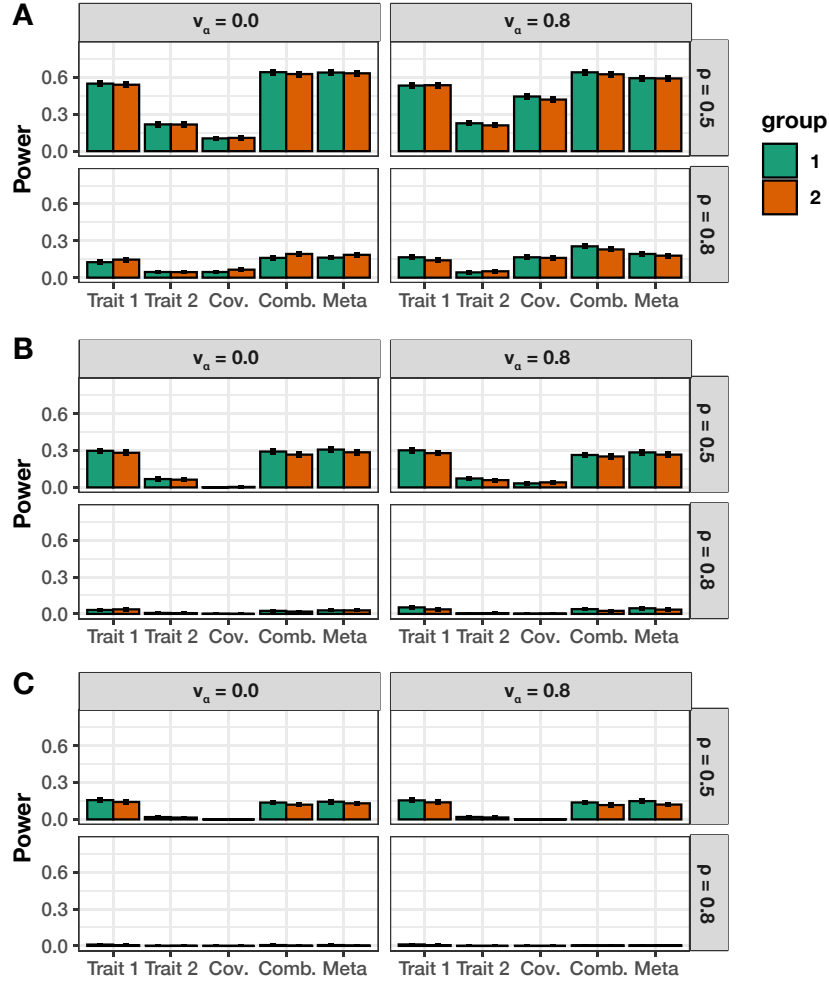


Figure A.14: **Empirical power of mvMAPIT with the harmonic mean combination approach to detect group #1 (10) and group #2 (10) epistatic variants across complex traits with different levels of broad-sense heritability.** In these simulations, one of the quantitative traits has a moderate broad-sense heritability $H^2 = 0.6$, while the other has heritability $H^2 = 0.3$. Both traits have architectures made up of both additive and epistatic effects. The parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to a setting where the epistatic effects for interactive pairs have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent epistatic effects ($v_\alpha = 0$) and traits with highly correlated epistatic effects ($v_\alpha = 0.8$). This plot shows the empirical power of mvMAPIT at significance levels **(A)** $P = 5 \times 10^{-2}$, **(B)** $P = 5 \times 10^{-4}$, and **(C)** $P = 1 \times 10^{-5}$, respectively. Group #1 and #2 causal markers are colored in green and orange, respectively. For comparison, the “trait #1” and “trait #2” bars correspond to the univariate MAPIT model, the “cov” bars corresponds to power contributed by the covariance test, and “comb” details power from the overall association identified by mvMAPIT in the combination approach. Lastly, the “meta” bars shows power from the association identified by combining only the P -values of the two univariate cases “trait #1” and “trait #2”. Results are based on 100 simulations per parameter combination and the horizontal bars represent standard errors.

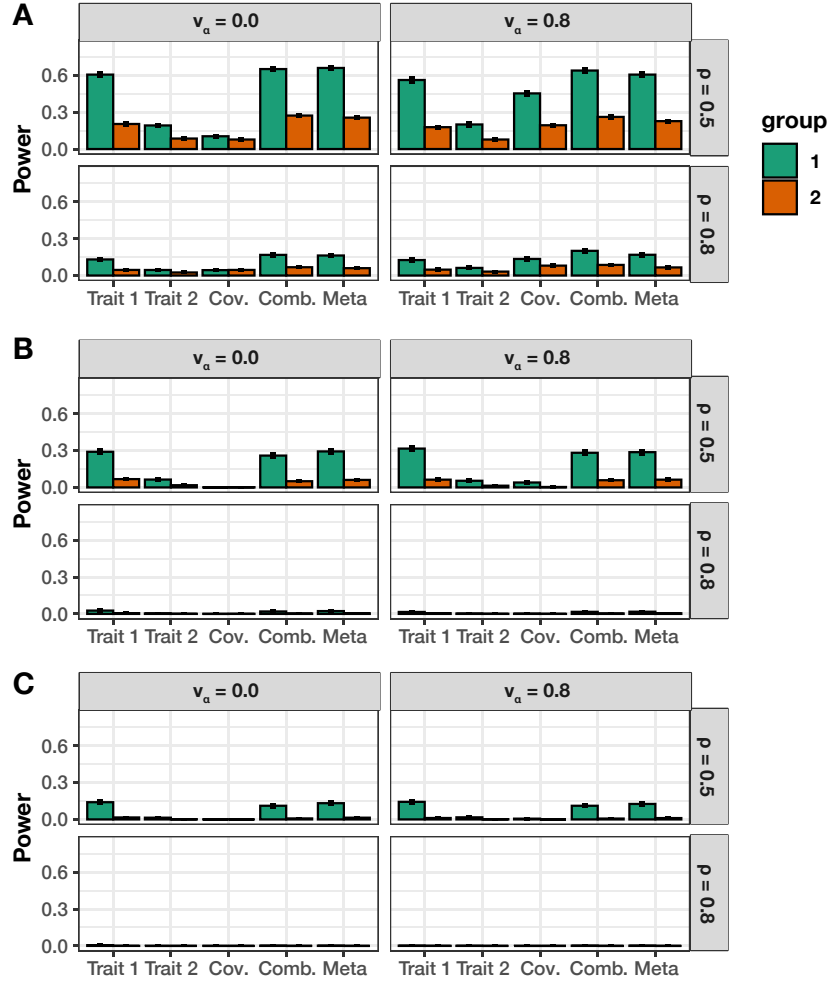


Figure A.15: **Empirical power of mvMAPIT with the harmonic mean combination approach to detect group #1 (10) and group #2 (20) epistatic variants across complex traits with different levels of broad-sense heritability.** In these simulations, one of the quantitative traits has a moderate broad-sense heritability $H^2 = 0.6$, while the other has heritability $H^2 = 0.3$. Both traits have architectures made up of both additive and epistatic effects. The parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to a setting where the epistatic effects for interactive pairs have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent epistatic effects ($v_\alpha = 0$) and traits with highly correlated epistatic effects ($v_\alpha = 0.8$). This plot shows the empirical power of mvMAPIT at significance levels **(A)** $P = 5 \times 10^{-2}$, **(B)** $P = 5 \times 10^{-4}$, and **(C)** $P = 1 \times 10^{-5}$, respectively. Group #1 and #2 causal markers are colored in green and orange, respectively. For comparison, the “trait #1” and “trait #2” bars correspond to the univariate MAPIT model, the “cov” bars corresponds to power contributed by the covariance test, and “comb” details power from the overall association identified by mvMAPIT in the combination approach. Lastly, the “meta” bars shows power from the association identified by combining only the P -values of the two univariate cases “trait #1” and “trait #2”. Results are based on 100 simulations per parameter combination and the horizontal bars represent standard errors.

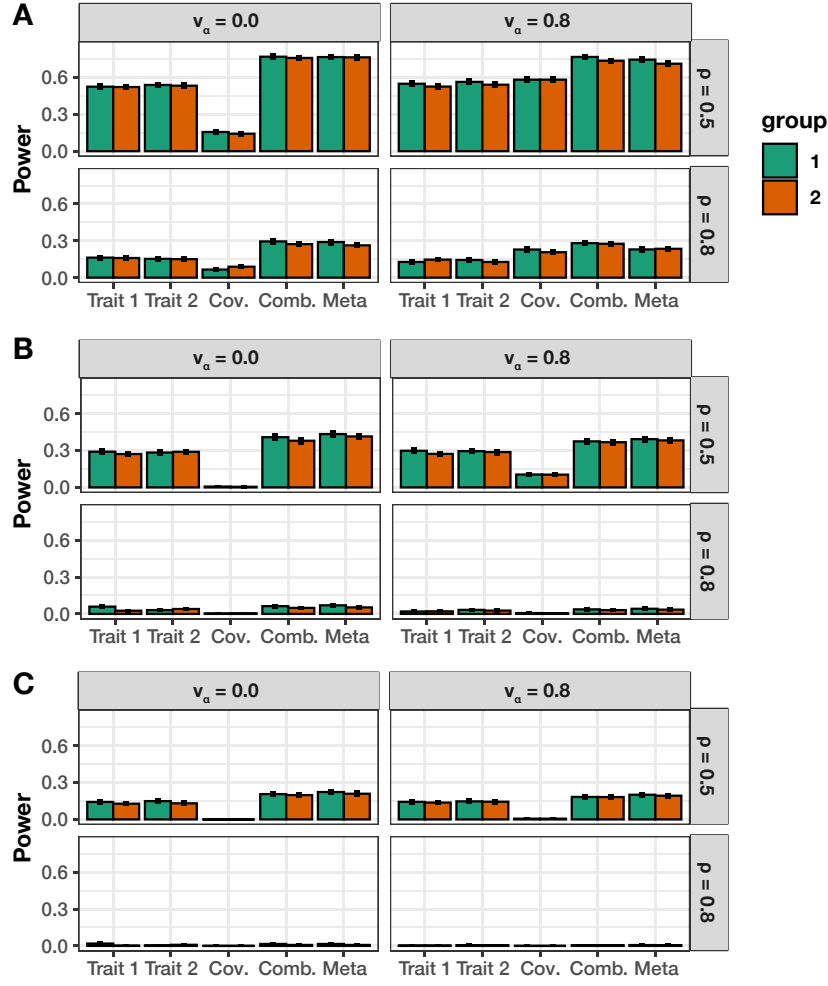


Figure A.16: **Empirical power of mvMAPIT with the Cauchy combination test approach to detect group #1 (10) and group #2 (10) epistatic variants across complex traits with moderate broad-sense heritability.** In these simulations, both quantitative traits are simulated to have broad-sense heritability $H^2 = 0.6$ with architectures made up of both additive and epistatic effects. The parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to a setting where the epistatic effects for interactive pairs have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent epistatic effects ($v_\alpha = 0$) and traits with highly correlated epistatic effects ($v_\alpha = 0.8$). This plot shows the empirical power of mvMAPIT at significance levels **(A)** $P = 5 \times 10^{-2}$, **(B)** $P = 5 \times 10^{-4}$, and **(C)** $P = 1 \times 10^{-5}$, respectively. Group #1 and #2 causal markers are colored in green and orange, respectively. For comparison, the “trait #1” and “trait #2” bars correspond to the univariate MAPIT model, the “cov” bars corresponds to power contributed by the covariance test, and “comb” details power from the overall association identified by mvMAPIT in the combination approach. Lastly, the “meta” bars shows power from the association identified by combining only the P -values of the two univariate cases “trait #1” and “trait #2”. Results are based on 100 simulations per parameter combination and the horizontal bars represent standard errors.

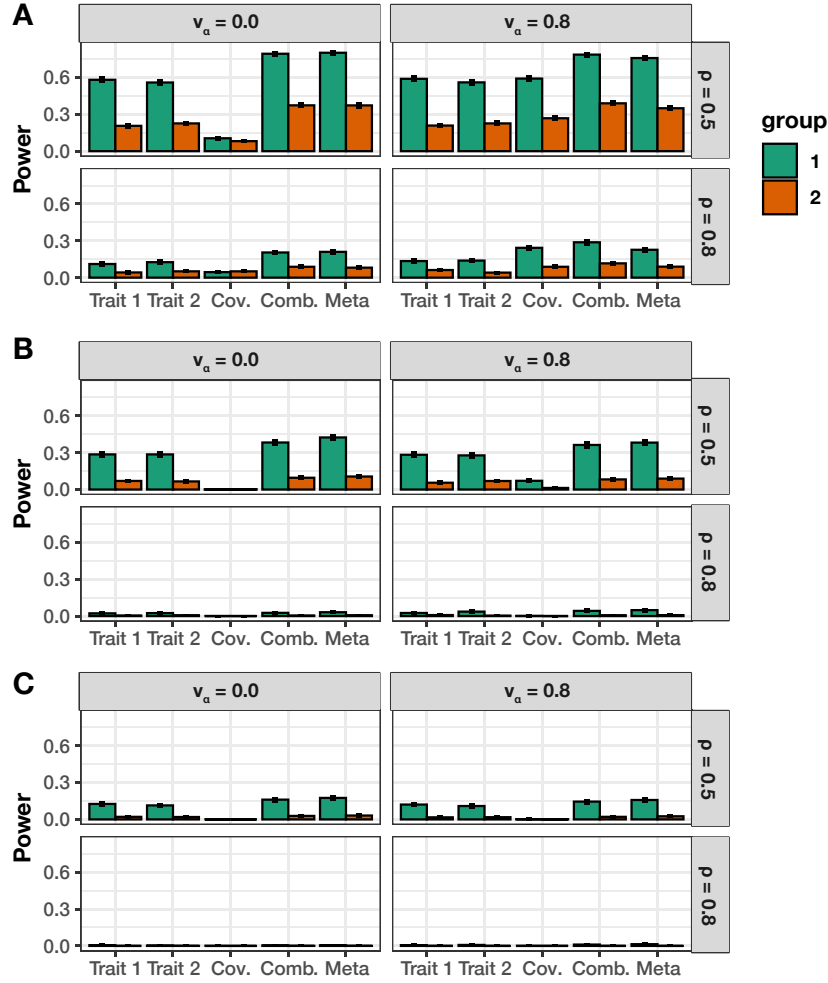


Figure A.17: **Empirical power of mvMAPIT with the Cauchy combination test approach to detect group #1 (10) and group #2 (20) epistatic variants across complex traits with moderate broad-sense heritability.** In these simulations, both quantitative traits are simulated to have broad-sense heritability $H^2 = 0.6$ with architectures made up of both additive and epistatic effects. The parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to a setting where the epistatic effects for interactive pairs have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent epistatic effects ($v_\alpha = 0$) and traits with highly correlated epistatic effects ($v_\alpha = 0.8$). This plot shows the empirical power of mvMAPIT at significance levels **(A)** $P = 5 \times 10^{-2}$, **(B)** $P = 5 \times 10^{-4}$, and **(C)** $P = 1 \times 10^{-5}$, respectively. Group #1 and #2 causal markers are colored in green and orange, respectively. For comparison, the “trait #1” and “trait #2” bars correspond to the univariate MAPIT model, the “cov” bars corresponds to power contributed by the covariance test, and “comb” details power from the overall association identified by mvMAPIT in the combination approach. Lastly, the “meta” bars shows power from the association identified by combining only the P -values of the two univariate cases “trait #1” and “trait #2”. Results are based on 100 simulations per parameter combination and the horizontal bars represent standard errors.

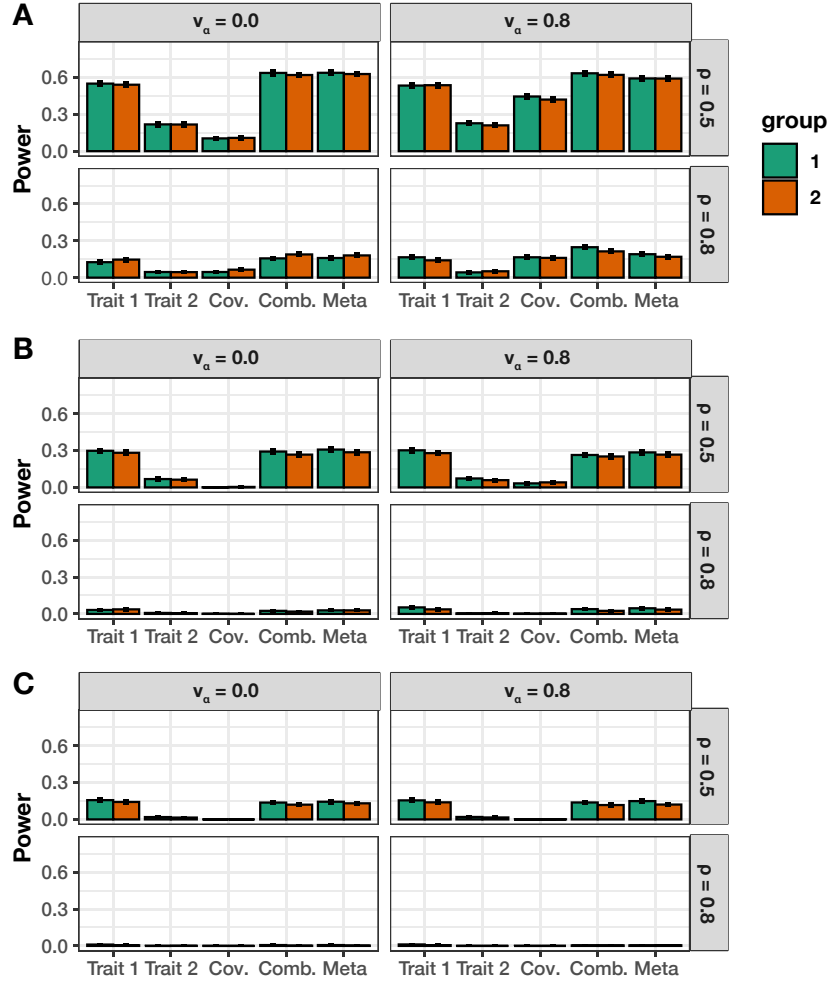


Figure A.18: **Empirical power of mvMAPIT with the Cauchy combination test approach to detect group #1 (10) and group #2 (10) epistatic variants across complex traits with different levels of broad-sense heritability.** In these simulations, one of the quantitative traits has a moderate broad-sense heritability $H^2 = 0.6$, while the other has heritability $H^2 = 0.3$. Both traits have architectures made up of both additive and epistatic effects. The parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to a setting where the epistatic effects for interactive pairs have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent epistatic effects ($v_\alpha = 0$) and traits with highly correlated epistatic effects ($v_\alpha = 0.8$). This plot shows the empirical power of mvMAPIT at significance levels **(A)** $P = 5 \times 10^{-2}$, **(B)** $P = 5 \times 10^{-4}$, and **(C)** $P = 1 \times 10^{-5}$, respectively. Group #1 and #2 causal markers are colored in green and orange, respectively. For comparison, the “trait #1” and “trait #2” bars correspond to the univariate MAPIT model, the “cov” bars corresponds to power contributed by the covariance test, and “comb” details power from the overall association identified by mvMAPIT in the combination approach. Lastly, the “meta” bars shows power from the association identified by combining only the P -values of the two univariate cases “trait #1” and “trait #2”. Results are based on 100 simulations per parameter combination and the horizontal bars represent standard errors.

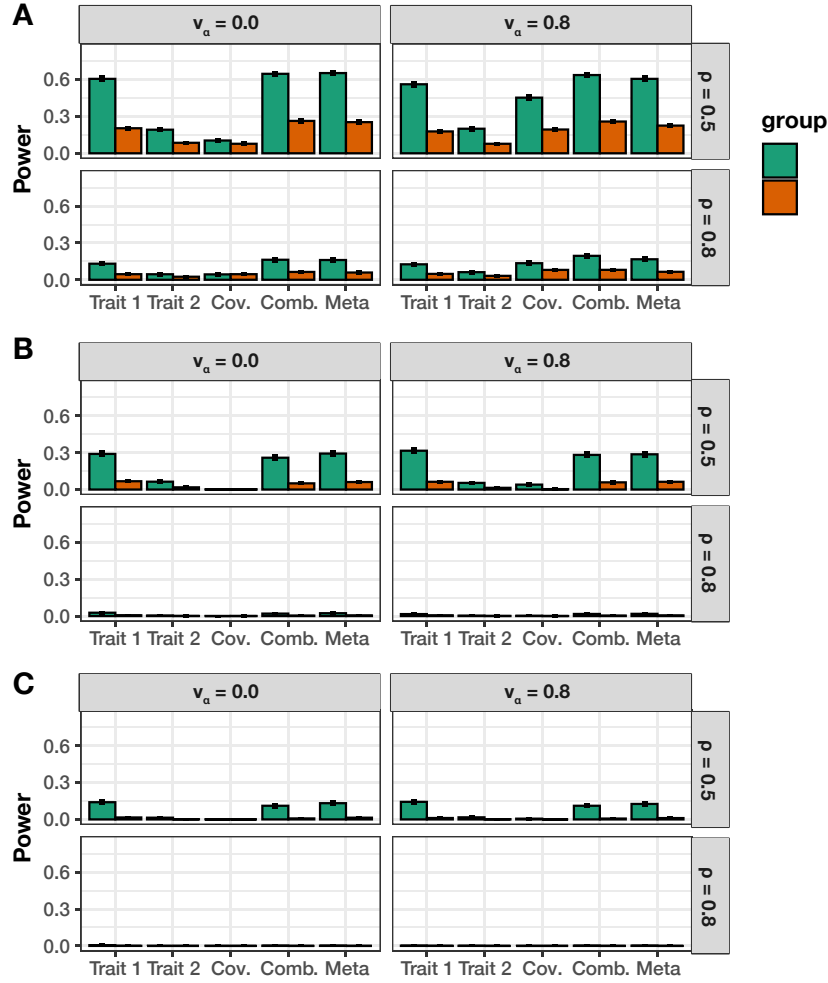


Figure A.19: **Empirical power of mvMAPIT with the Cauchy combination test approach to detect group #1 (10) and group #2 (20) epistatic variants across complex traits with different levels of broad-sense heritability.** In these simulations, one of the quantitative traits has a moderate broad-sense heritability $H^2 = 0.6$, while the other has heritability $H^2 = 0.3$. Both traits have architectures made up of both additive and epistatic effects. The parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to a setting where the epistatic effects for interactive pairs have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent epistatic effects ($v_\alpha = 0$) and traits with highly correlated epistatic effects ($v_\alpha = 0.8$). This plot shows the empirical power of mvMAPIT at significance levels **(A)** $P = 5 \times 10^{-2}$, **(B)** $P = 5 \times 10^{-4}$, and **(C)** $P = 1 \times 10^{-5}$, respectively. Group #1 and #2 causal markers are colored in green and orange, respectively. For comparison, the “trait #1” and “trait #2” bars correspond to the univariate MAPIT model, the “cov” bars corresponds to power contributed by the covariance test, and “comb” details power from the overall association identified by mvMAPIT in the combination approach. Lastly, the “meta” bars shows power from the association identified by combining only the P -values of the two univariate cases “trait #1” and “trait #2”. Results are based on 100 simulations per parameter combination and the horizontal bars represent standard errors.

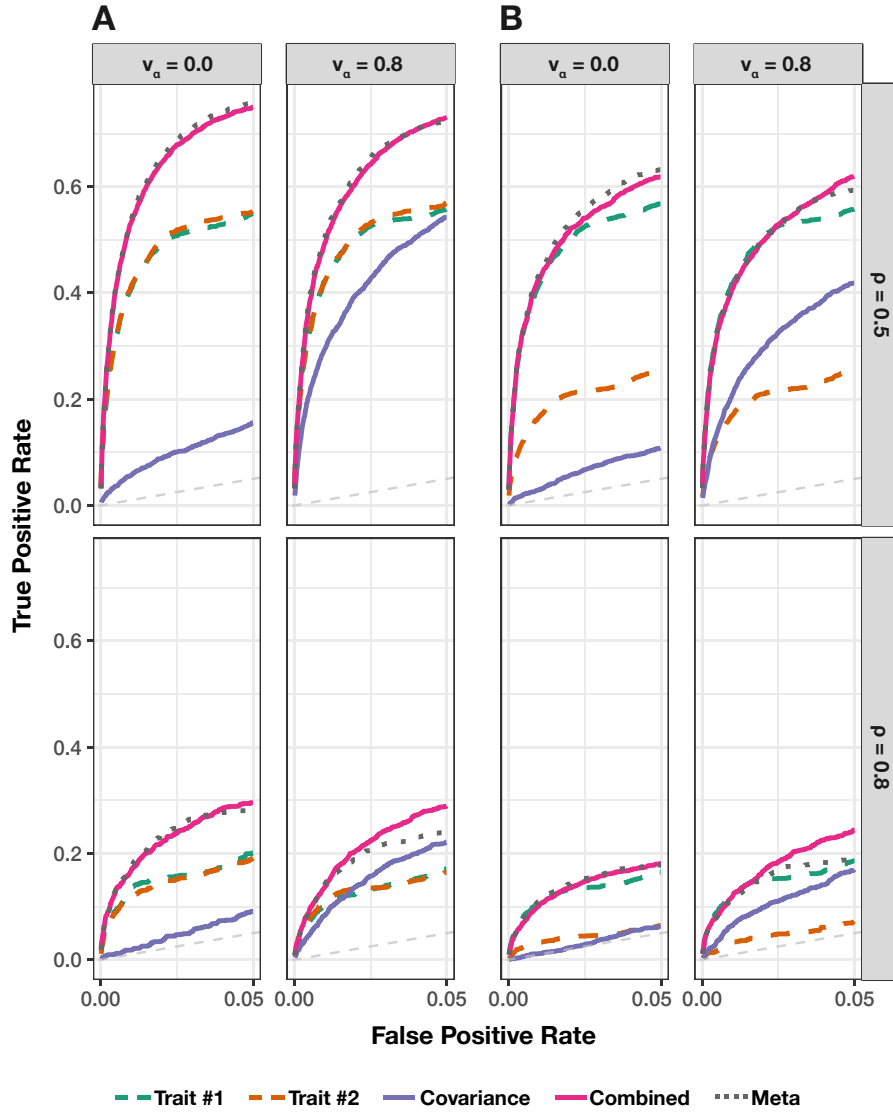


Figure A.20: **Receiver operating characteristic (ROC) curves comparing the ability of mvMAPIT using the harmonic mean to the univariate MAPIT model in detecting group #1 (10) and group #2 (10) epistatic variants across complex traits.** In panel (A) both traits have broad-sense heritability $H^2 = 0.6$; while in panel (B) one of traits has broad-sense heritability $H^2 = 0.6$ and the other has heritability $H^2 = 0.3$. Across the rows, the parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to settings where the epistatic effects across traits are independent ($v_\alpha = 0$) or highly correlated ($v_\alpha = 0.8$). For comparison, the “trait #1” and “trait #2” dashed lines correspond to the univariate MAPIT model. The “covariance” solid purple line corresponds to power contributed by the covariance test. The “combined” pink line shows power from the overall association identified by mvMAPIT in the multivariate approach. Lastly, the “meta” gray dotted line shows power from the association identified by combining only the P -values of the two univariate cases “trait #1” and “trait #2”. Note that the upper limit of the x-axis (i.e., false positive rate) has been truncated at 0.05. All results are based on 100 simulated replicates.

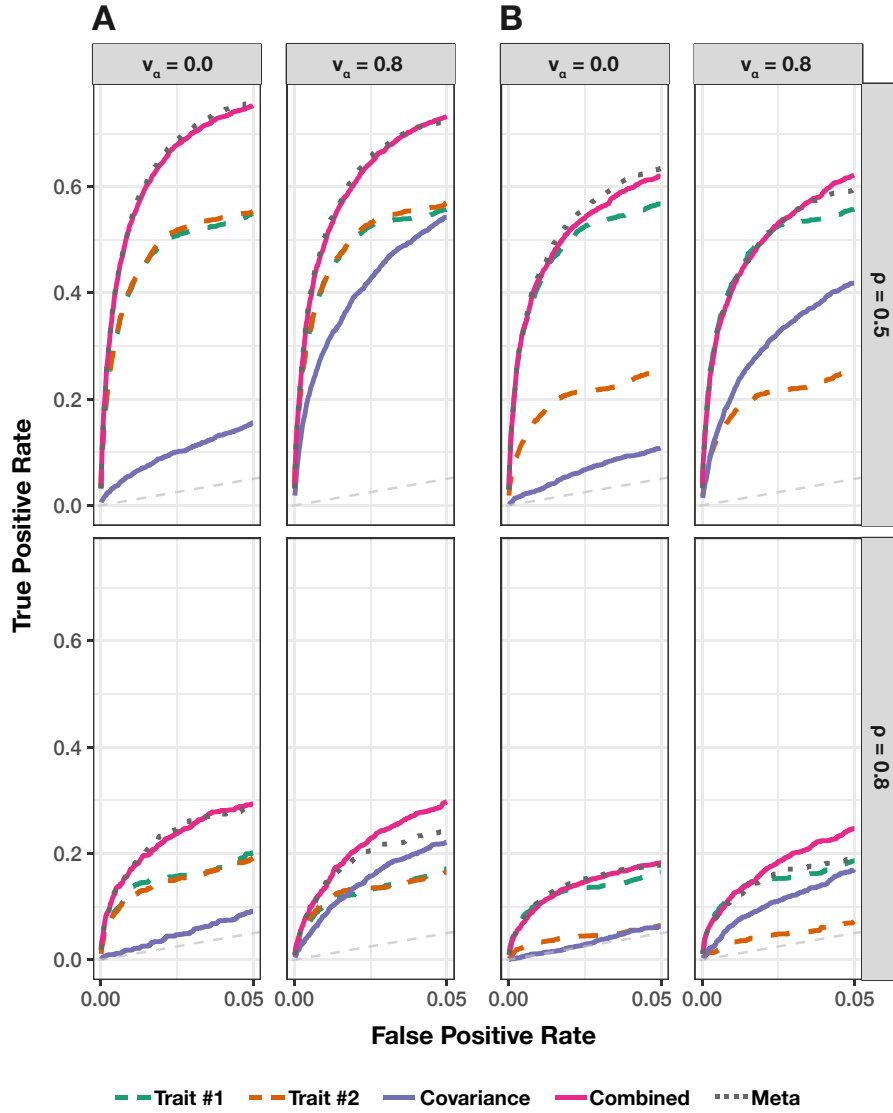


Figure A.21: **Receiver operating characteristic (ROC) curves comparing the ability of mvMAPIT using the Cauchy combination test to the univariate MAPIT model in detecting group #1 (10) and group #2 (10) epistatic variants across complex traits.** In panel (A) both traits have broad-sense heritability $H^2 = 0.6$; while in panel (B) one of traits has broad-sense heritability $H^2 = 0.6$ and the other has heritability $H^2 = 0.3$. Across the rows, the parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to settings where the epistatic effects across traits are independent ($v_\alpha = 0$) or highly correlated ($v_\alpha = 0.8$). For comparison, the “trait #1” and “trait #2” dashed lines correspond to the univariate MAPIT model. The “covariance” solid purple line corresponds to power contributed by the covariance test. The “combined” pink line shows power from the overall association identified by mvMAPIT in the multivariate approach. Lastly, the “meta” gray dotted line shows power from the association identified by combining only the P -values of the two univariate cases “trait #1” and “trait #2”. Note that the upper limit of the x-axis (i.e., false positive rate) has been truncated at 0.05. All results are based on 100 simulated replicates.

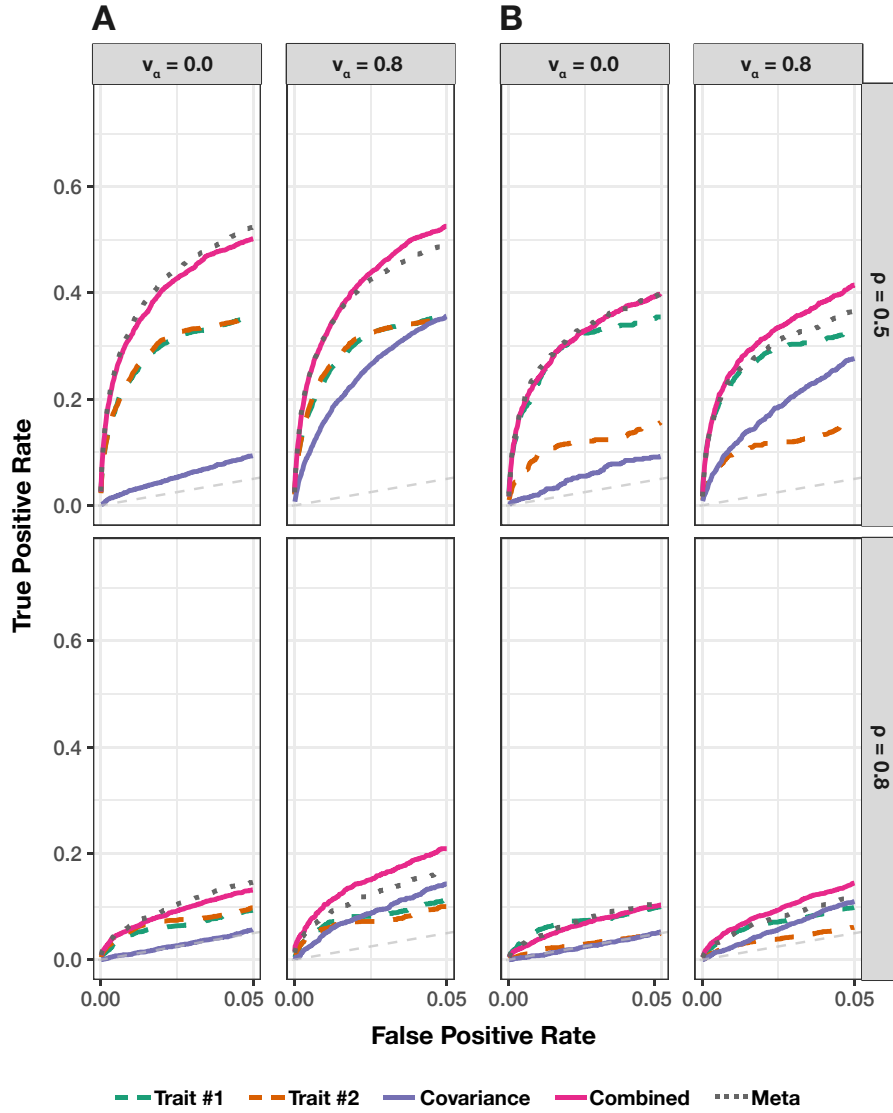


Figure A.22: **Receiver operating characteristic (ROC) curves comparing the ability of mvMAPIT with Fisher's method to the univariate MAPIT model in detecting group #1 (10) and group #2 (20) epistatic variants across complex traits.** In panel (A) both traits have broad-sense heritability $H^2 = 0.6$; while in panel (B) one of traits has broad-sense heritability $H^2 = 0.6$ and the other has heritability $H^2 = 0.3$. Across the rows, the parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to settings where the epistatic effects across traits are independent ($v_\alpha = 0$) or highly correlated ($v_\alpha = 0.8$). For comparison, the “trait #1” and “trait #2” dashed lines correspond to the univariate MAPIT model. The “covariance” solid purple line corresponds to power contributed by the covariance test. The “combined” pink line shows power from the overall association identified by mvMAPIT in the multivariate approach. Lastly, the “meta” gray dotted line shows power from the association identified by combining only the P -values of the two univariate cases “trait #1” and “trait #2”. Note that the upper limit of the x-axis (i.e., false positive rate) has been truncated at 0.05. All results are based on 100 simulated replicates.

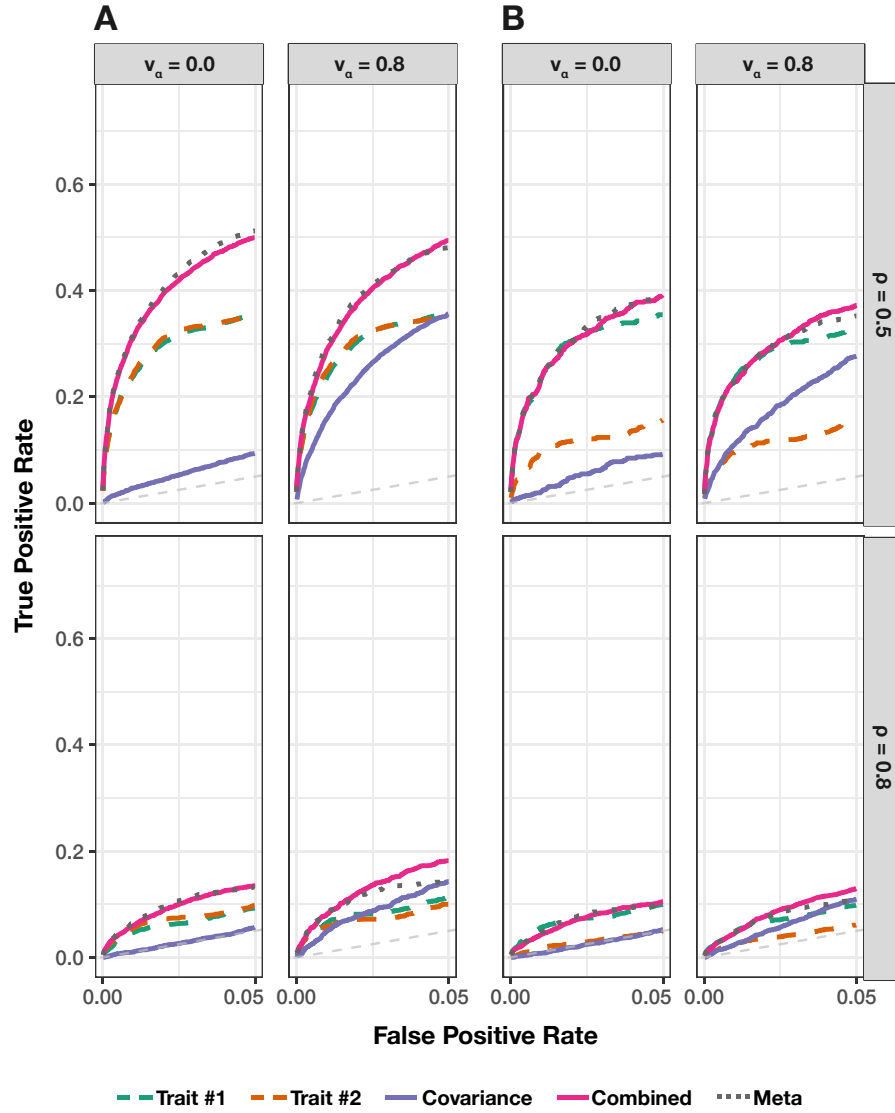


Figure A.23: **Receiver operating characteristic (ROC) curves comparing the ability of mvMAPIT using the harmonic mean to the univariate MAPIT model in detecting group #1 (10) and group #2 (20) epistatic variants across complex traits.** In panel (A) both traits have broad-sense heritability $H^2 = 0.6$; while in panel (B) one of traits has broad-sense heritability $H^2 = 0.6$ and the other has heritability $H^2 = 0.3$. Across the rows, the parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to settings where the epistatic effects across traits are independent ($v_\alpha = 0$) or highly correlated ($v_\alpha = 0.8$). For comparison, the “trait #1” and “trait #2” dashed lines correspond to the univariate MAPIT model. The “covariance” solid purple line corresponds to power contributed by the covariance test. The “combined” pink line shows power from the overall association identified by mvMAPIT in the multivariate approach. Lastly, the “meta” gray dotted line shows power from the association identified by combining only the P -values of the two univariate cases “trait #1” and “trait #2”. Note that the upper limit of the x-axis (i.e., false positive rate) has been truncated at 0.05. All results are based on 100 simulated replicates.

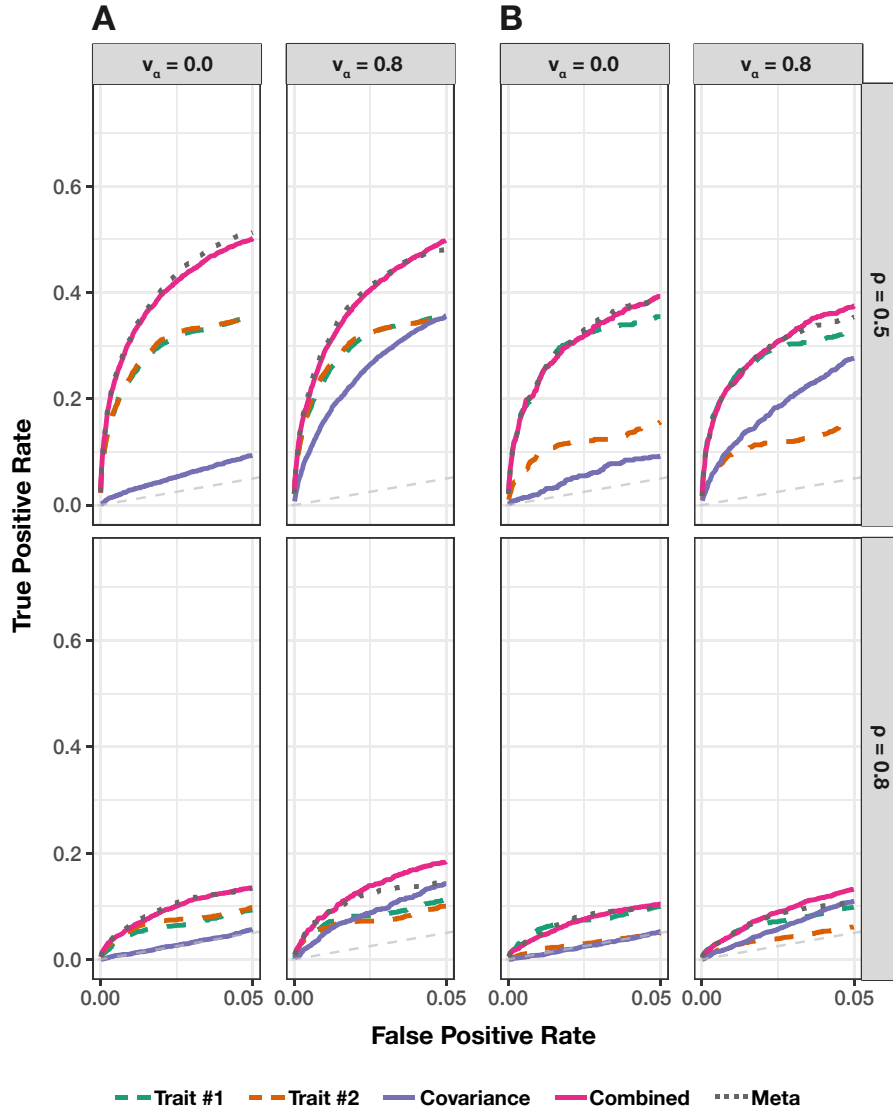


Figure A.24: **Receiver operating characteristic (ROC) curves comparing the ability of mvMAPIT using the Cauchy combination test to the univariate MAPIT model in detecting group #1 (10) and group #2 (20) epistatic variants across complex traits.** In panel (A) both traits have broad-sense heritability $H^2 = 0.6$; while in panel (B) one of traits has broad-sense heritability $H^2 = 0.6$ and the other has heritability $H^2 = 0.3$. Across the rows, the parameter $\rho = \{0.5, 0.8\}$ is used to determine the portion of broad-sense heritability contributed by additive effects. Each column corresponds to settings where the epistatic effects across traits are independent ($v_\alpha = 0.0$) or highly correlated ($v_\alpha = 0.8$). For comparison, the “trait #1” and “trait #2” dashed lines correspond to the univariate MAPIT model. The “covariance” solid purple line corresponds to power contributed by the covariance test. The “combined” pink line shows power from the overall association identified by mvMAPIT in the multivariate approach. Lastly, the “meta” gray dotted line shows power from the association identified by combining only the P -values of the two univariate cases “trait #1” and “trait #2”. Note that the upper limit of the x-axis (i.e., false positive rate) has been truncated at 0.05. All results are based on 100 simulated replicates.

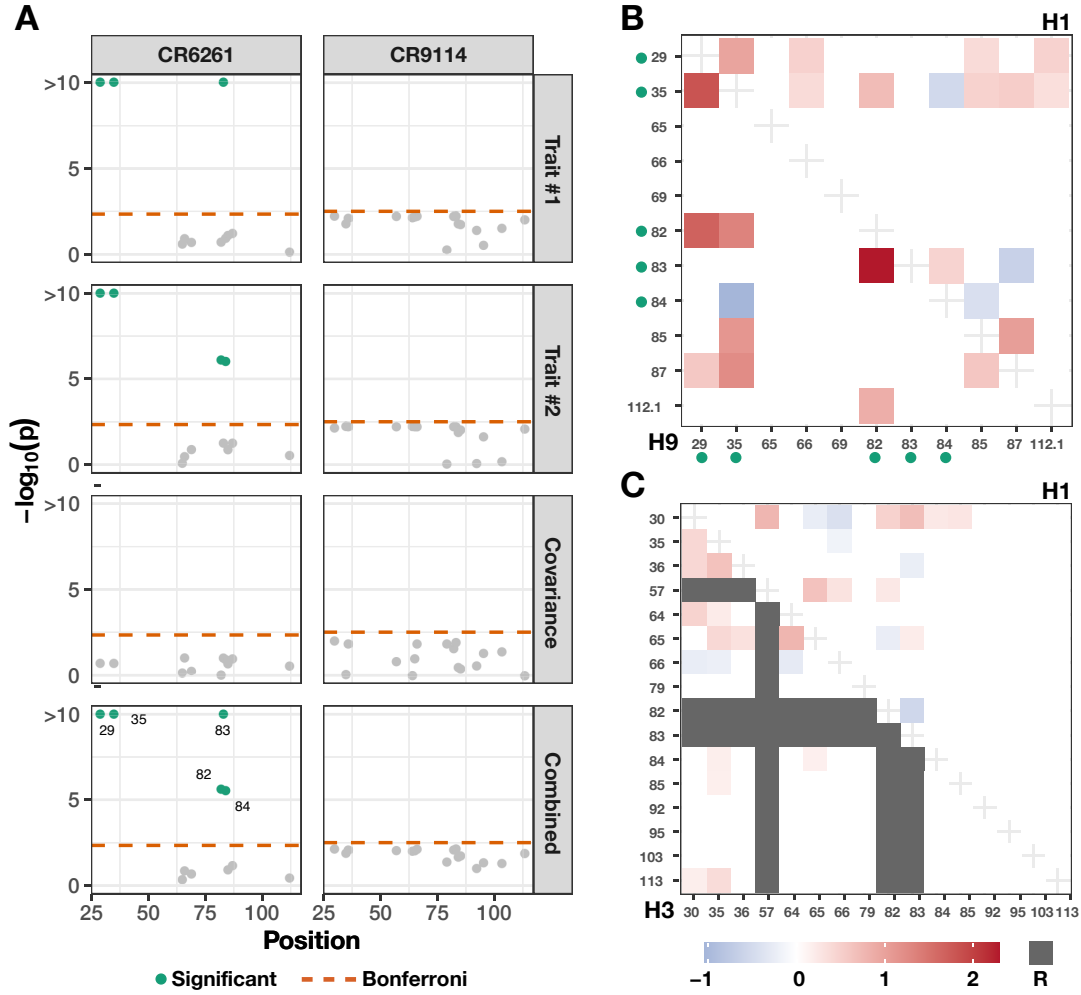


Figure A.25: **Applying mvMAPIT with the harmonic mean to broadly neutralizing antibodies recovers heavy-chain mutations known to be involved in epistatic interactions that affect binding against two influenza strains.** These results are based on protein sequence data from Phillips et al. (2021) who generated a nearly combinatorially complete library for two broadly neutralizing anti-influenza antibodies (bnAbs), CR6261 and CR9114. For each antibody, we assess binding affinity to different influenza strains. For CR6261, traits #1 and #2 are binding measurements to the antigens H_1 and H_9 ; while, for CR9114, we assess the same measurement for H_1 and H_3 . Panel (A) shows Manhattan plots for the different sets of P -values computed during the mvMAPIT analysis. The red horizontal lines indicate a chain-wide Bonferroni corrected significance threshold ($P = 4.55 \times 10^{-3}$ for CR6261 and $P = 3.13 \times 10^{-3}$ for CR9114, respectively). The green colored dots are positions that have significant marginal epistatic effects after multiple correction. Panels (B) and (C) reproduce exhaustive search results originally reported by Phillips et al. (2021). The green dots next to the mutation labels on the axes are the residues that are significant in the multivariate MAPIT analysis and correspond to panel (A). The shaded regions in panel (B) are the regression coefficients for pairwise interactions between positions when assessing binding of CR6261 with H_1 (upper right triangle) and H_9 (lower left triangle). Similarly, panel (C) shows the same information when assessing binding of CR9114 with H_1 (upper right triangle) and H_3 (lower left triangle). Required mutations (indicated by R) are plotted in gray and left out of the analysis (Phillips et al. 2021).

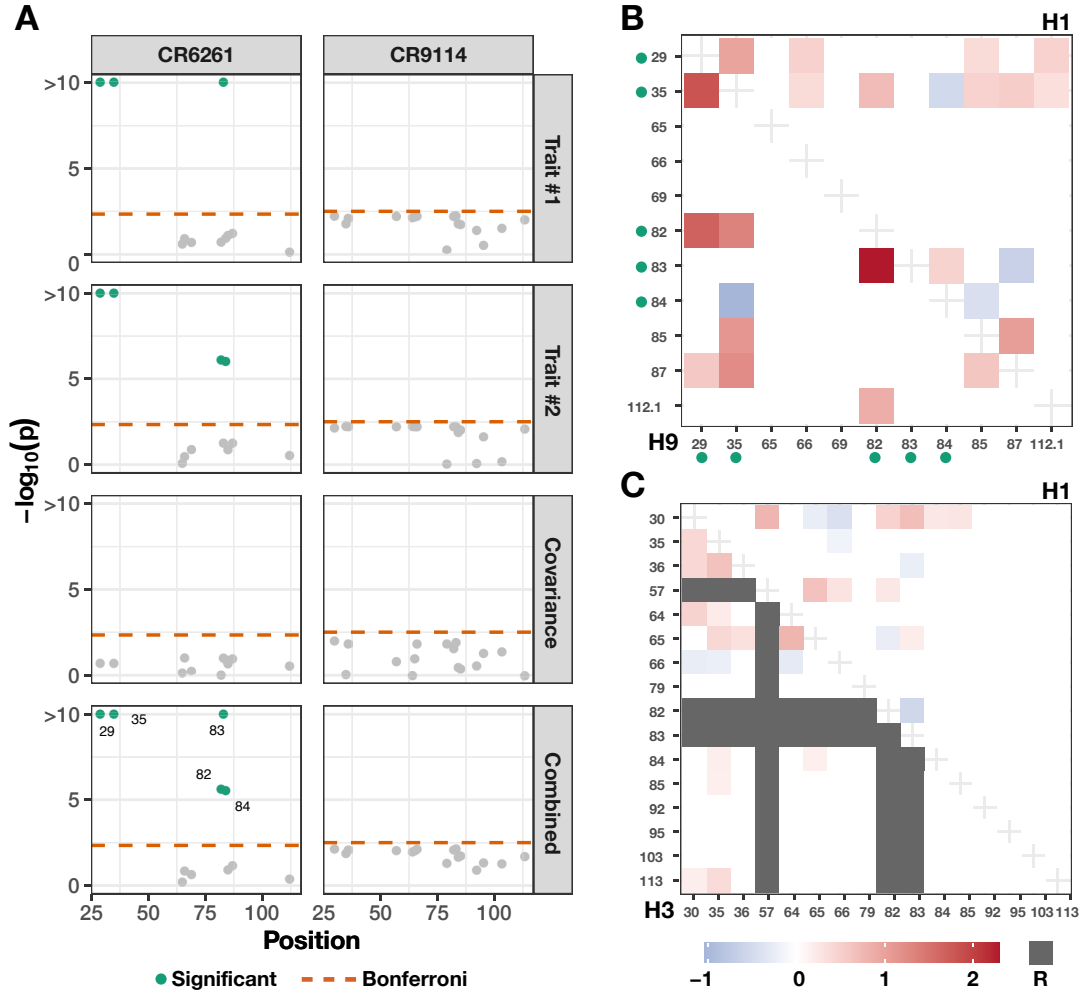


Figure A.26: **Applying mvMAPIT with the Cauchy combination test to broadly neutralizing antibodies recovers heavy-chain mutations known to be involved in epistatic interactions that affect binding against two influenza strains.** These results are based on protein sequence data from Phillips et al. (2021) who generated a nearly combinatorially complete library for two broadly neutralizing anti-influenza antibodies (bnAbs), CR6261 and CR9114. For each antibody, we assess binding affinity to different influenza strains. For CR6261, traits #1 and #2 are binding measurements to the antigens H_1 and H_9 ; while, for CR9114, we assess the same measurement for H_1 and H_3 . Panel (A) shows Manhattan plots for the different sets of P -values computed during the mvMAPIT analysis. The red horizontal lines indicate a chain-wide Bonferroni corrected significance threshold ($P = 4.55 \times 10^{-3}$ for CR6261 and $P = 3.13 \times 10^{-3}$ for CR9114, respectively). The green colored dots are positions that have significant marginal epistatic effects after multiple correction. Panels (B) and (C) reproduce exhaustive search results originally reported by Phillips et al. (2021). The green dots next to the mutation labels on the axes are the residues that are significant in the multivariate MAPIT analysis and correspond to panel (A). The shaded regions in panel (B) are the regression coefficients for pairwise interactions between positions when assessing binding of CR6261 with H_1 (upper right triangle) and H_9 (lower left triangle). Similarly, panel (C) shows the same information when assessing binding of CR9114 with H_1 (upper right triangle) and H_3 (lower left triangle). Required mutations (indicated by R) are plotted in gray and left out of the analysis (Phillips et al. 2021).

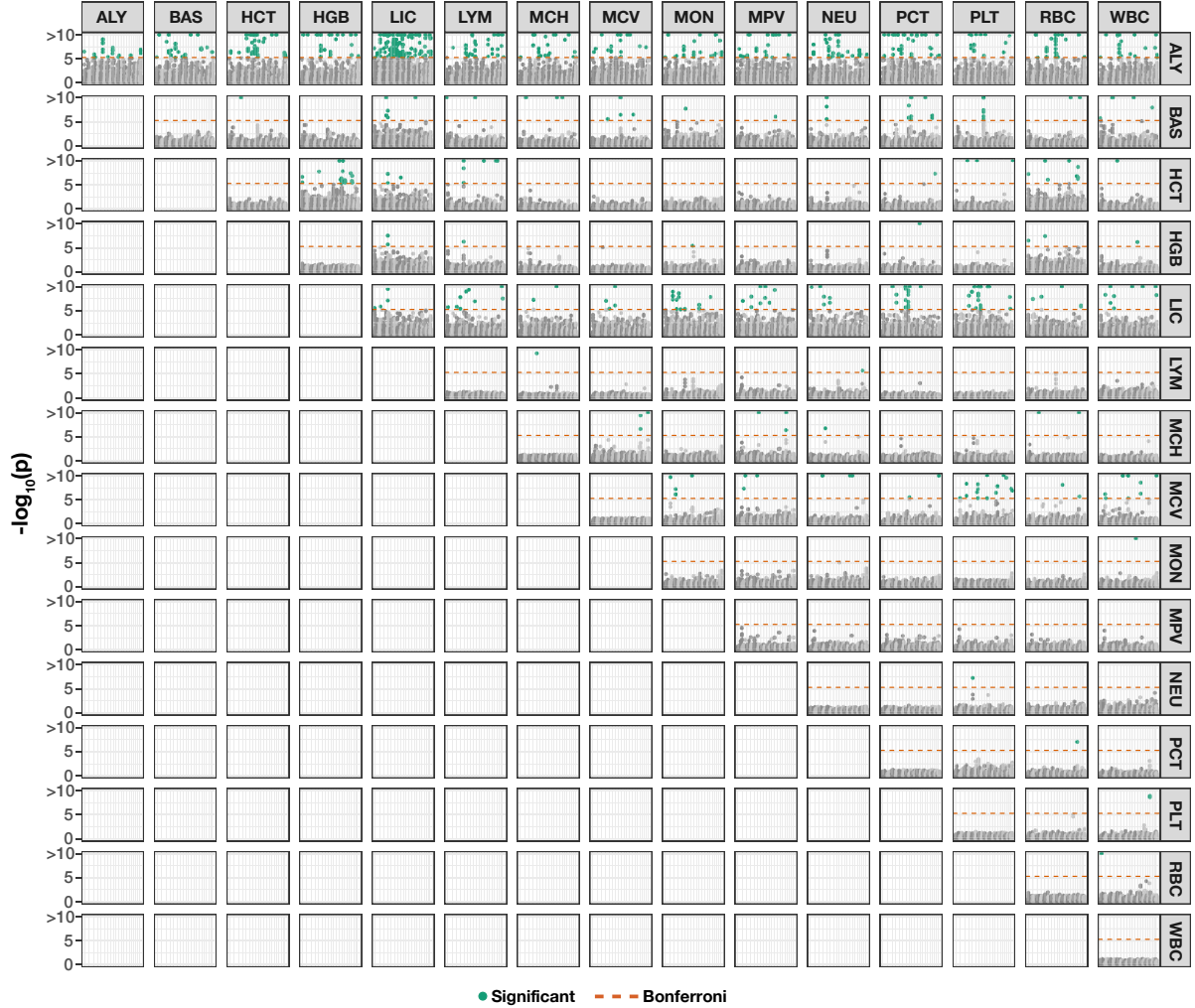


Figure A.27: Manhattan plot of genome-wide interaction study for all trait pairs in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics (Valdar et al. 2006a; Valdar et al. 2006b; *Wellcome Trust Centre for Human Genetics - Mouse Resources 2022*) using mvMAPIT with Fisher's method. The columns correspond to trait #1 in the analysis while the rows denote trait #2. Results on the diagonal correspond to results from running a univariate MAPIT model. The results on the off-diagonals show the combined P -values from mvMAPIT. The red horizontal lines indicate a genome-wide Bonferroni corrected significance threshold ($P = 4.83 \times 10^{-6}$). The green colored dots are SNPs that have significant marginal epistatic effects after multiple correction. Full names for the abbreviations of each trait can be found in the main text (Materials and Methods).

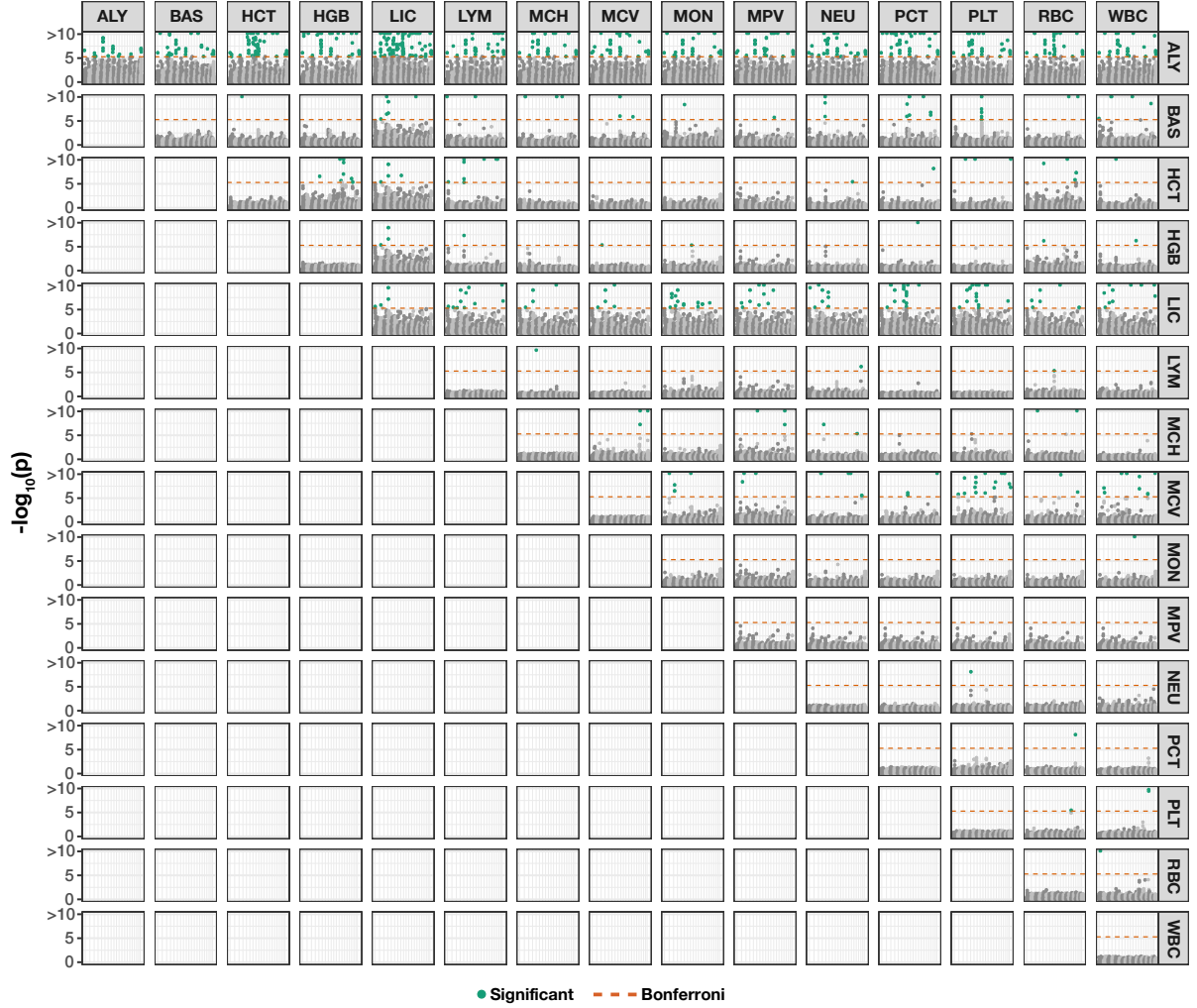
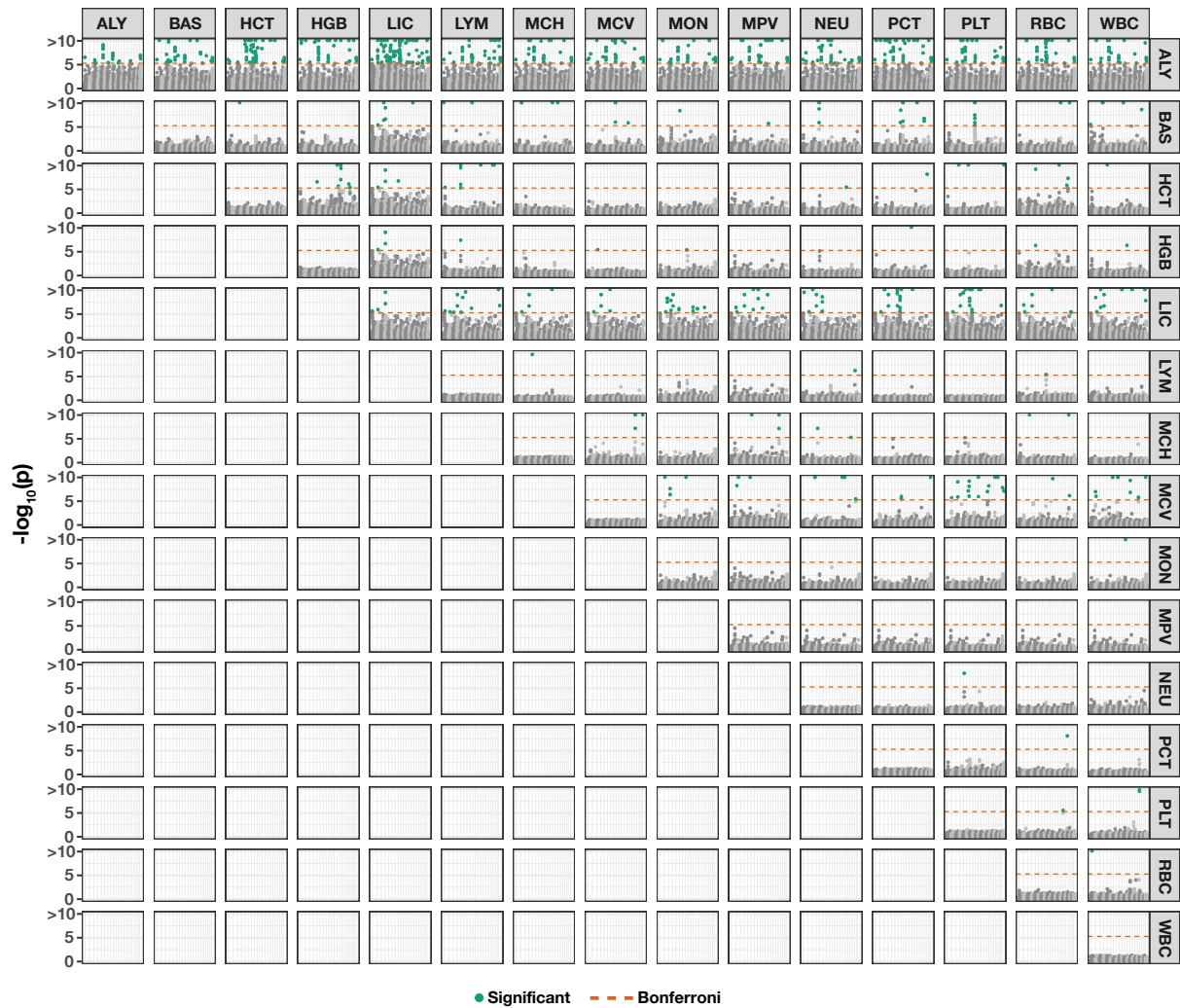
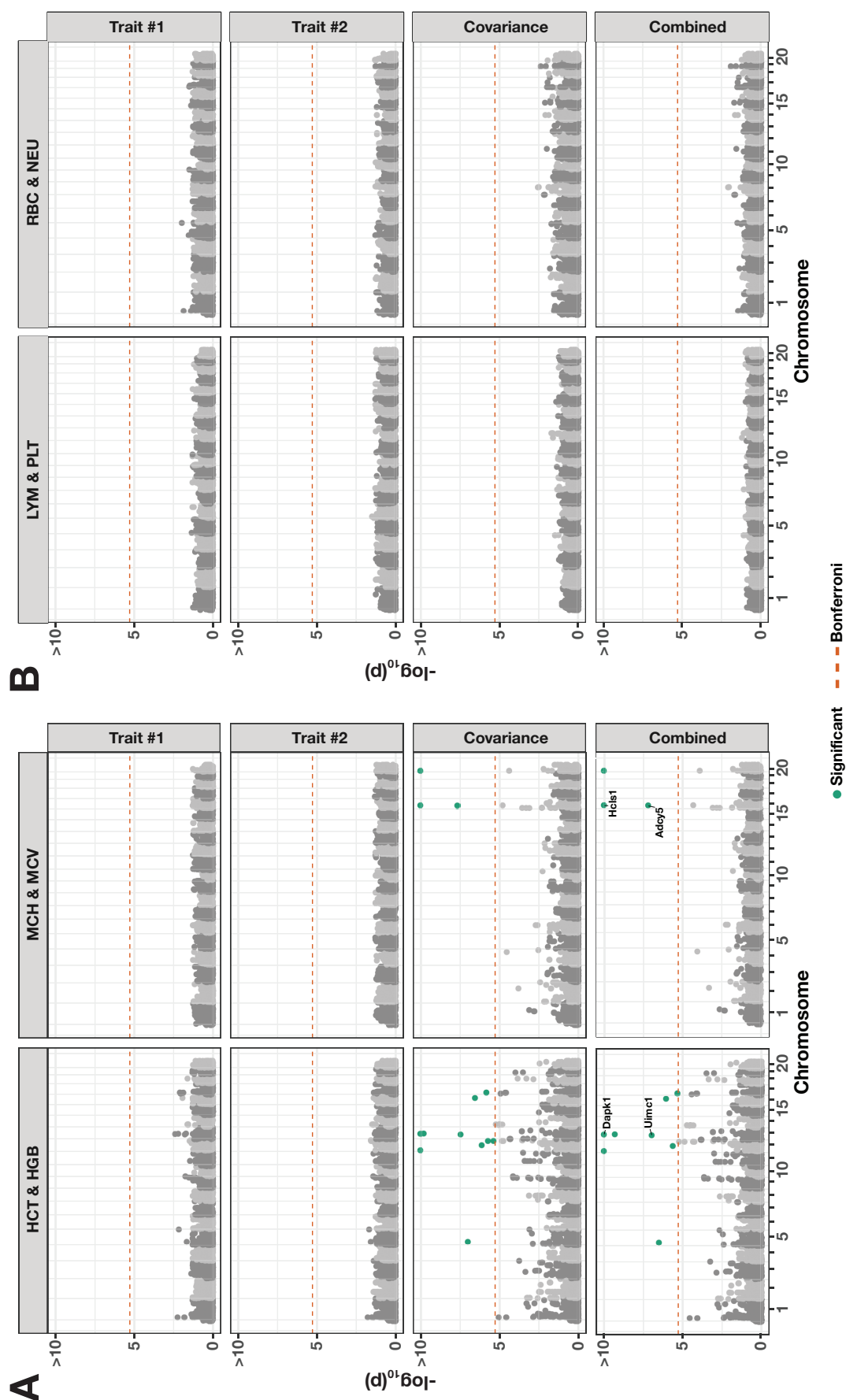


Figure A.28: Manhattan plot of genome-wide interaction study for all trait pairs in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics (Valdar et al. 2006a; Valdar et al. 2006b; *Wellcome Trust Centre for Human Genetics - Mouse Resources 2022*) using mvMAPIT with the harmonic mean. The columns correspond to trait #1 in the analysis while the rows denote trait #2. Results on the diagonal correspond to results from running a univariate MAPIT model. The results on the off-diagonals show the combined P -values from mvMAPIT. The red horizontal lines indicate a genome-wide Bonferroni corrected significance threshold ($P = 4.83 \times 10^{-6}$). The green colored dots are SNPs that have significant marginal epistatic effects after multiple correction. Full names for the abbreviations of each trait can be found in the main text (Materials and Methods).





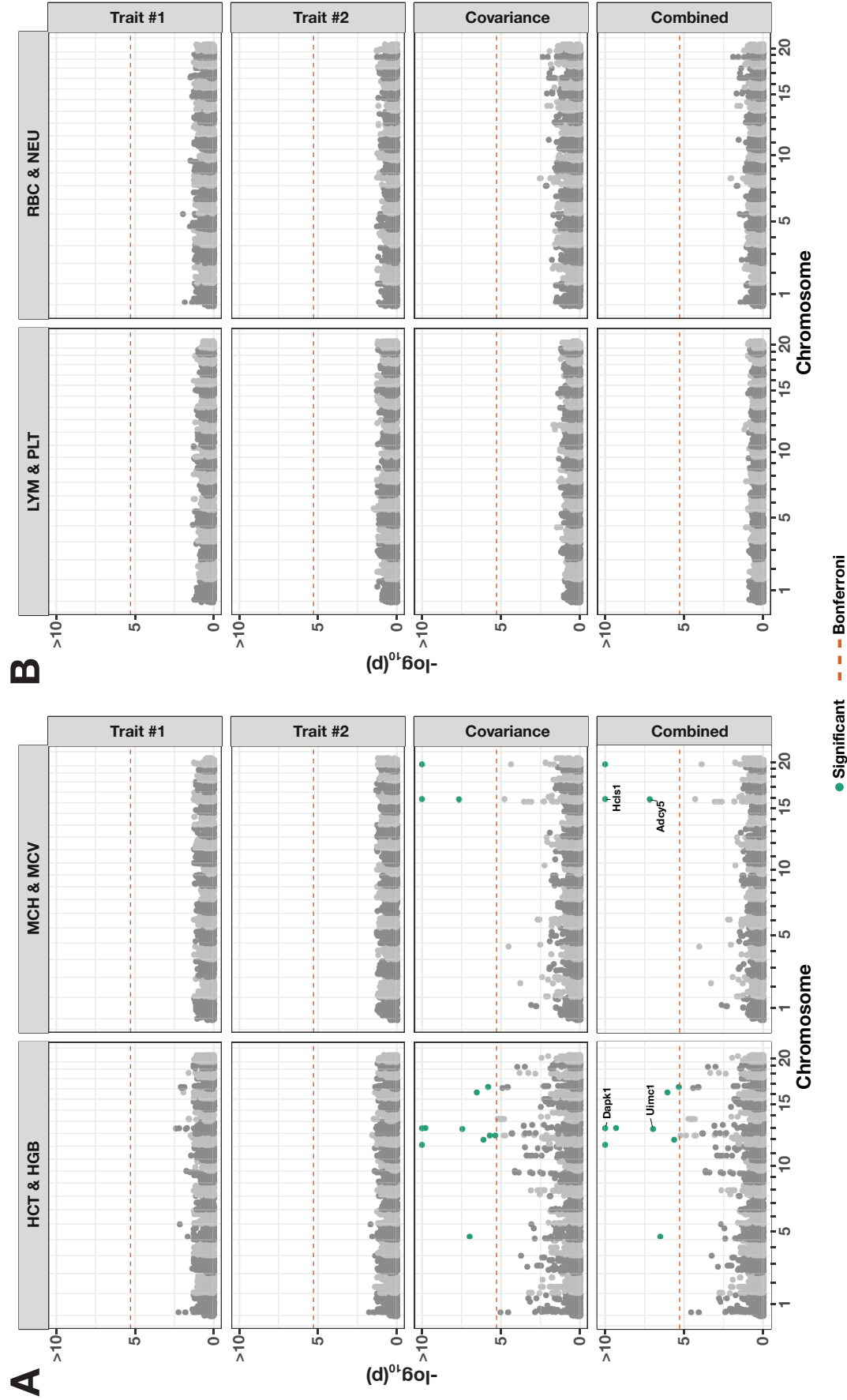


Figure A.31: Manhattan plots of a genome-wide interaction study for two pairs of hematology traits in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics (Valdar et al. 2006a; Valdar et al. 2006b; *Wellcome Trust Centre for Human Genetics - Mouse Resources 2022*) using mvMAPIT with the Cauchy combination test. Panel (A) depicts results for trait pairs hematocrit (HCT) and hemoglobin (HGB) in the left column and mean corpuscular hemoglobin (MCH) and mean corpuscular volume (MCV) in the right column. Panel (B) depicts results for trait pairs lymphocytes (LYM) and platelets (PLT) in the left column and red blood cell count (RBC) and neutrophils (NEU) in the right column. Here, we depict the P -values computed during each step of the mvMAPIT modeling pipeline. The red horizontal lines indicate a genome-wide Bonferroni corrected significance threshold ($P = 4.83 \times 10^{-6}$). The green colored dots are SNPs that have significant marginal epistatic effects after multiple test correction. Significant SNPs were mapped to the closest neighboring genes using the Mouse Genome Informatics database (<http://www.informatics.jax.org>) (Blake et al. 2020; Smith et al. 2009).

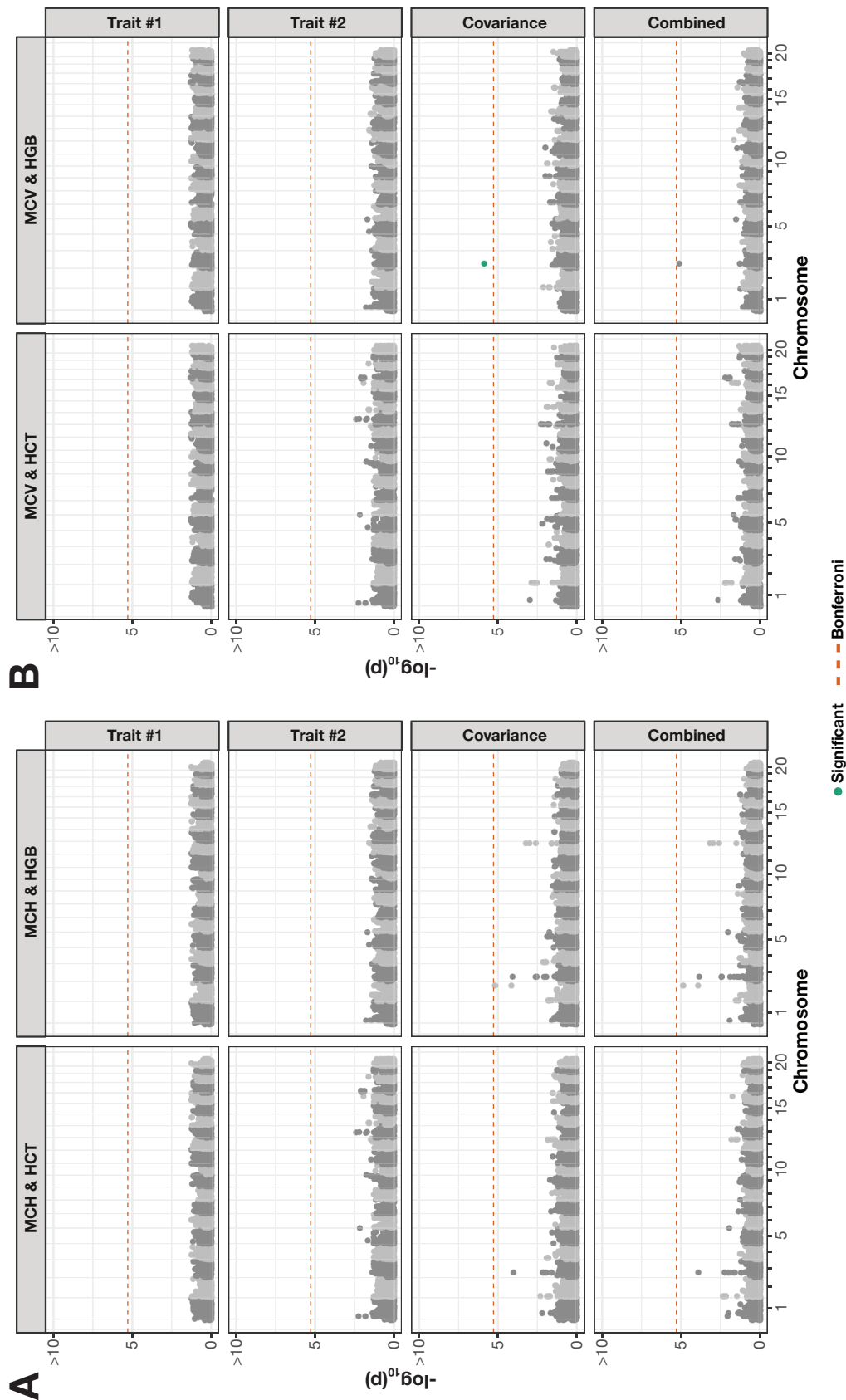
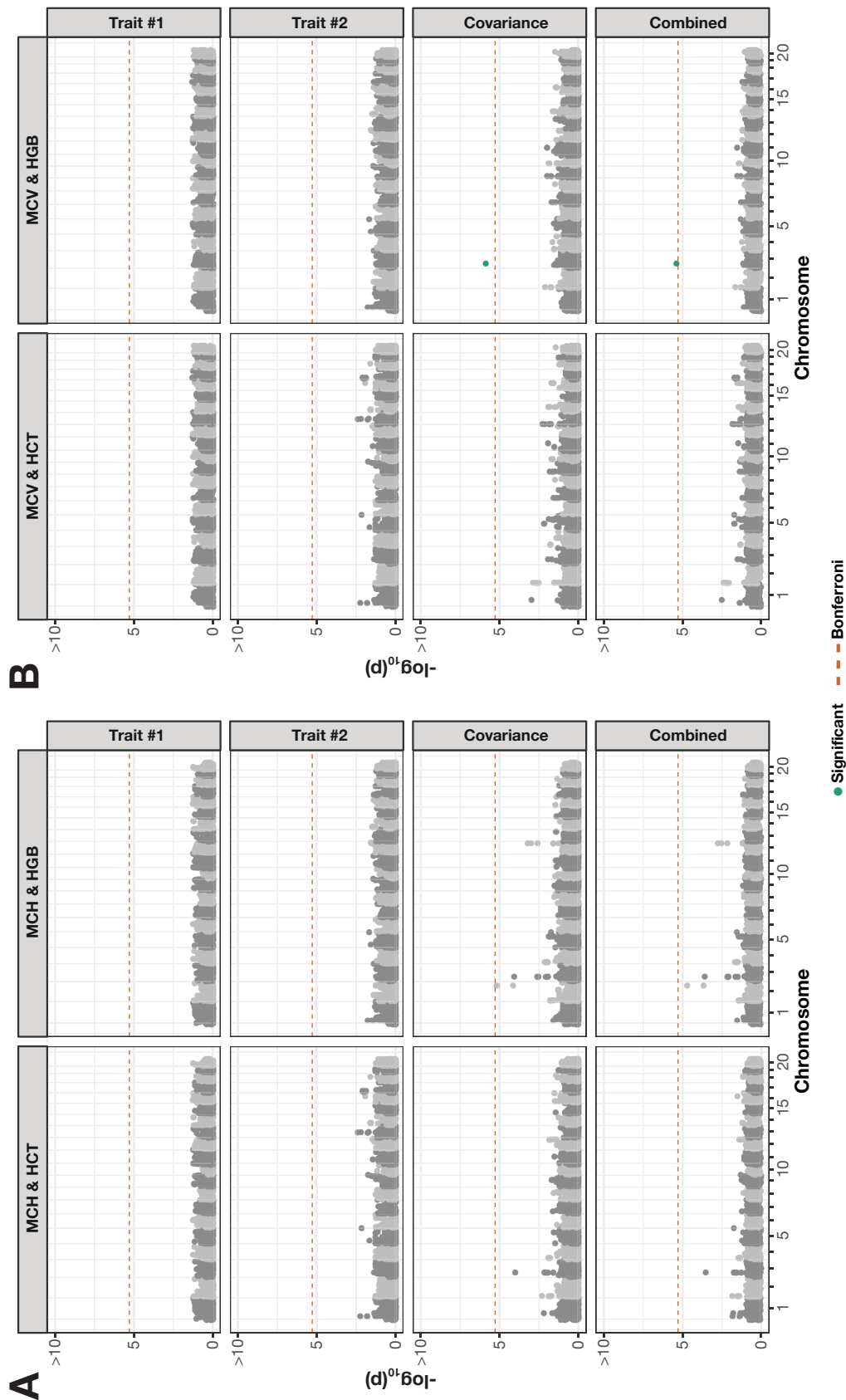


Figure A.32: Manhattan plots of a genome-wide interaction study for two pairs of hematology traits in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics (Valdar et al. 2006a; Valdar et al. 2006b; *Wellcome Trust Centre for Human Genetics - Mouse Resources 2022*) using mvMAPIT with Fisher's method. Panel (A) depicts results for trait pairs mean corpuscular hemoglobin (MCH) and hemoglobin (HGB) in the left column and mean corpuscular volume (MCV) and hematocrit (HCT) in the right column. Panel (B) depicts results for trait pairs mean corpuscular volume (MCV) and hematocrit (HCT) in the left column and mean corpuscular volume (MCV) and hemoglobin (HGB) in the right column. Here, we depict the P -values computed during each step of the mvMAPIT modeling pipeline. The red horizontal lines indicate a genome-wide Bonferroni corrected significance threshold ($P = 4.83 \times 10^{-6}$). The green colored dots are SNPs that have significant marginal epistatic effects after multiple test correction. Significant SNPs were mapped to the closest neighboring genes using the Mouse Genome Informatics database (<http://www.informatics.jax.org>) (Blake et al. 2020; Smith et al. 2009).



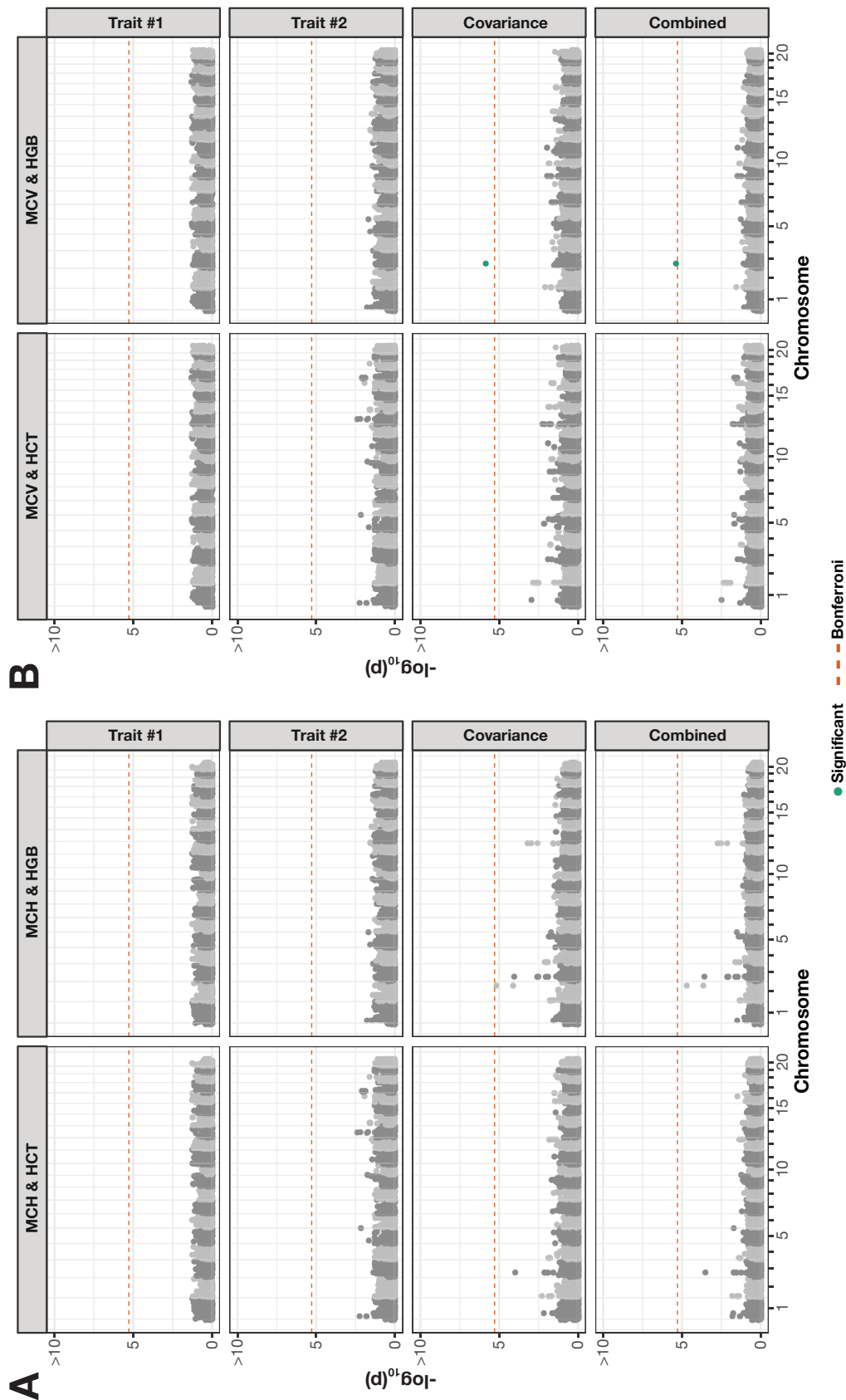


Figure A.34: Manhattan plots of a genome-wide interaction study for two pairs of hematology traits in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics (Valdar et al. 2006a; Valdar et al. 2006b; *Wellcome Trust Centre for Human Genetics - Mouse Resources 2022*) using mvMAPIT with the Cauchy combination test. Panel (A) depicts results for trait pairs mean corpuscular hemoglobin (MCH) and hematocrit (HCT) in the left column and mean corpuscular hemoglobin (HGB) in the right column. Panel (B) depicts results for trait pairs mean corpuscular volume (MCV) and hematocrit (HCT) in the left column and mean corpuscular volume (MCV) and hemoglobin (HGB) in the right column. Here, we depict the P -values computed during each step of the mvMAPIT modeling pipeline. The red horizontal lines indicate a genome-wide Bonferroni corrected significance threshold ($P = 4.83 \times 10^{-6}$). The green colored dots are SNPs that have significant marginal epistatic effects after multiple test correction. Significant SNPs were mapped to the closest neighboring genes using the Mouse Genome Informatics database (<http://www.informatics.jax.org>) (Blake et al. 2020; Smith et al. 2009).

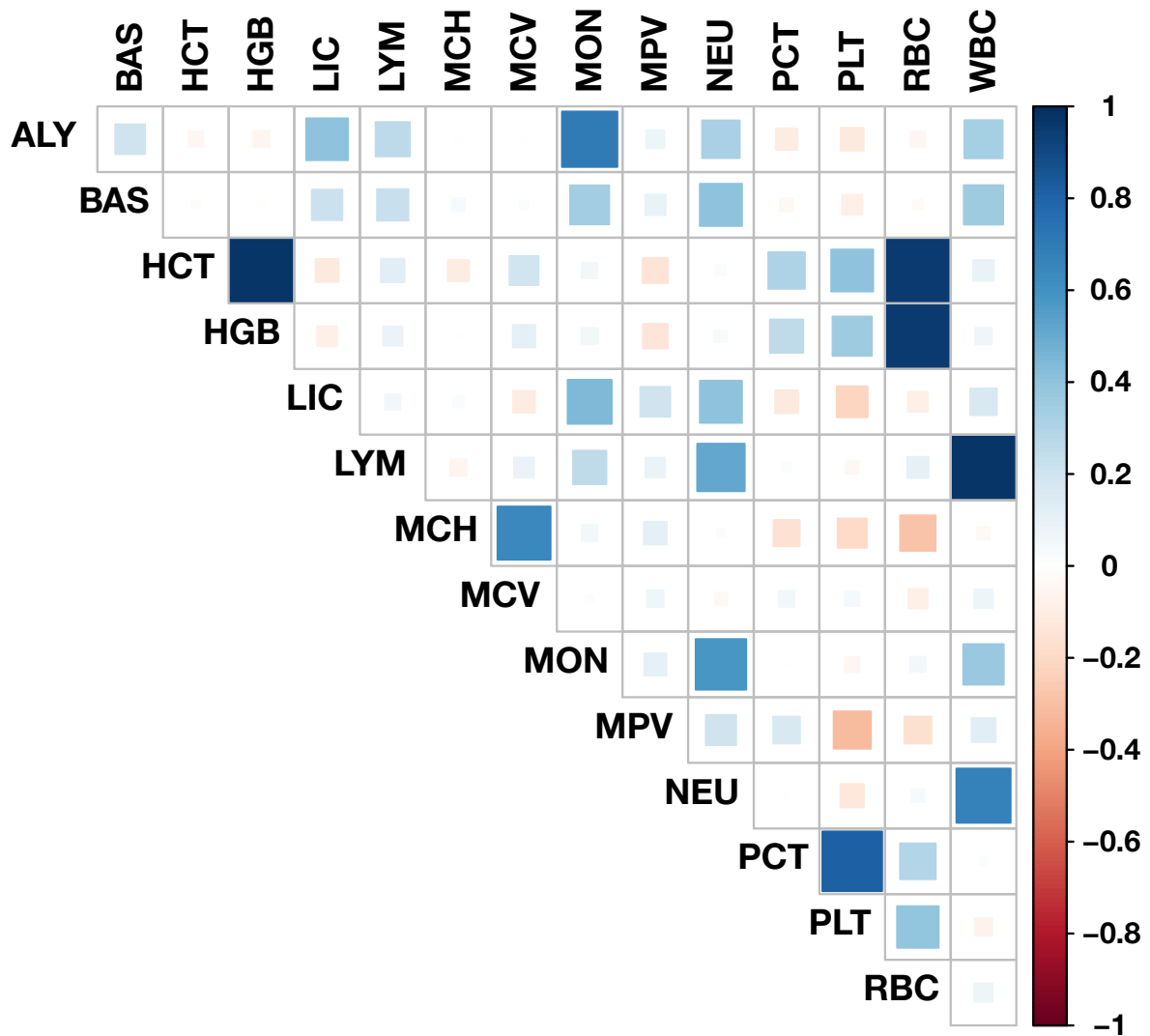


Figure A.35: Empirical correlations for all trait pairs in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics (Valdar et al. 2006a; Valdar et al. 2006b; *Wellcome Trust Centre for Human Genetics - Mouse Resources 2022*). Full names for the abbreviations of each trait can be found in the main text (Materials and Methods).

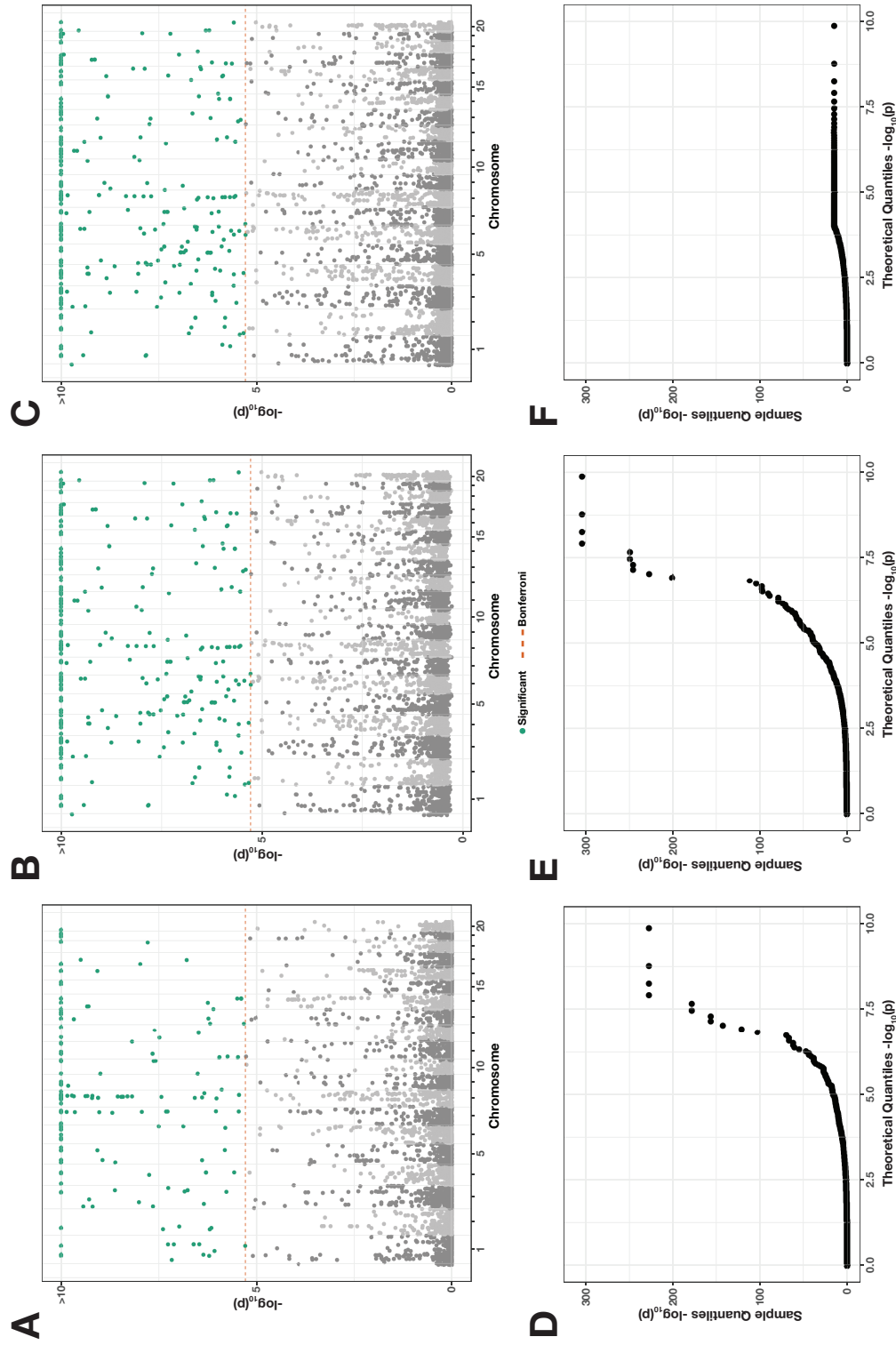


Figure A.36: Manhattan and QQ-plots of a genome-wide interaction study combining all hematology traits in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics (Valdar et al. 2006a; Valdar et al. 2006b; *Wellcome Trust Centre for Human Genetics - Mouse Resources 2022*) using mvMAPIT. We focused on 15 hematological phenotypes including: atypical lymphocytes, basophils, hematocrit, hemoglobin, large immature cells, lymphocytes, mean corpuscular hemoglobin, mean corpuscular volume, monocytes, mean platelet volume, neutrophils, plateletcrit, platelets, red blood cell count, red cell distribution width, and white blood cell count. The red horizontal lines indicate a genome-wide Bonferroni corrected significance threshold ($P = 4.83 \times 10^{-6}$). Panels (A) and (D) correspond to using mvMAPIT with Fisher's method, (B) and (E) correspond to the harmonic mean, and (C) and (F) correspond to the Cauchy combination test. The green colored dots are SNPs that have significant marginal epistatic effects after multiple test correction. Significant SNPs were mapped to the closest neighboring genes using the Mouse Genome Informatics database (<http://www.informatics.jax.org>) (Blake et al. 2020; Smith et al. 2009). The plateau in the P -value distribution displayed in panel (F) is due to numerical limitations of the pcauchy function in the R stats package.

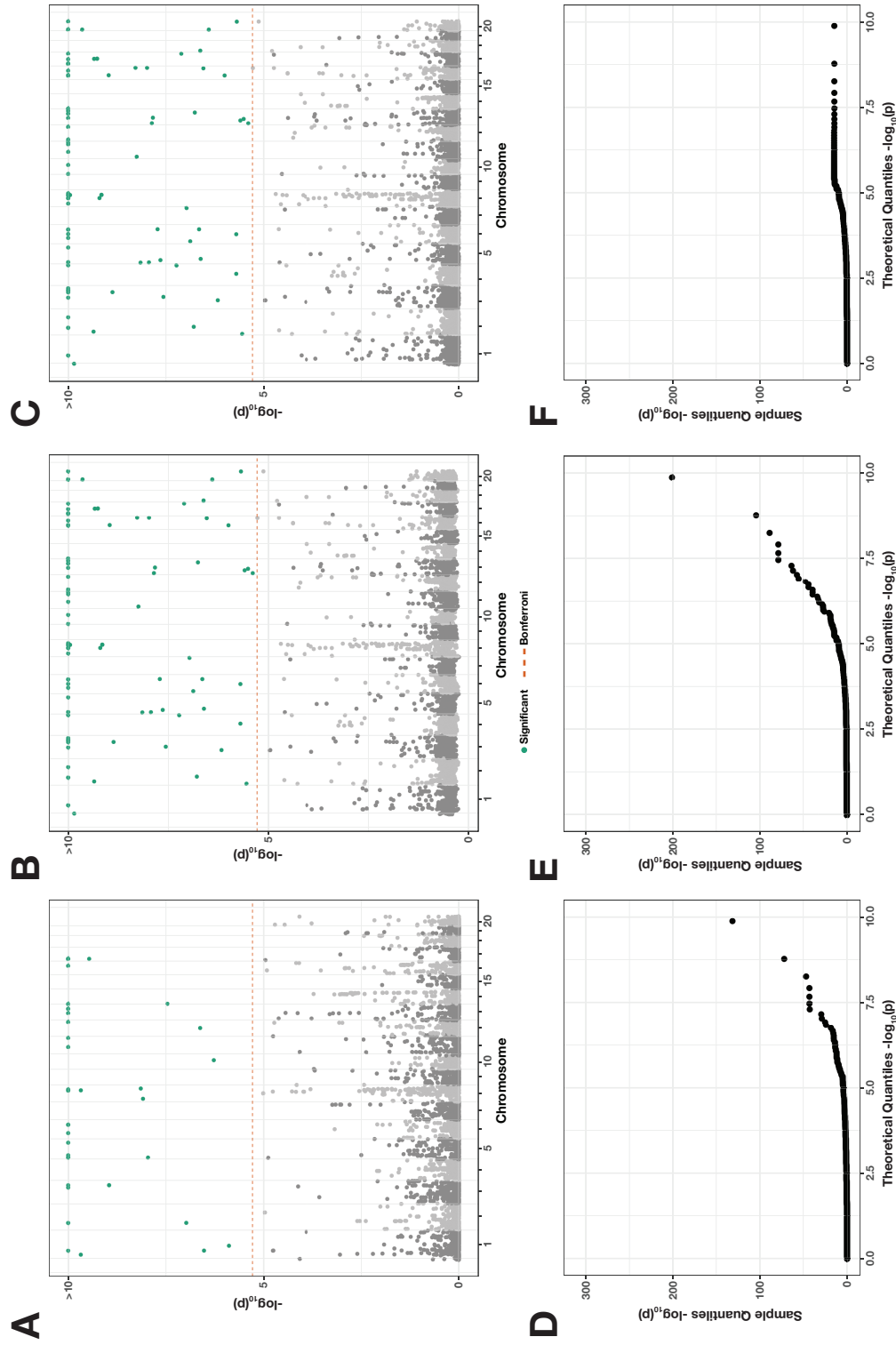


Figure A.37: Manhattan and QQ-plots of a genome-wide interaction study combining many hematology traits in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics (Valdar et al. 2006a; Valdar et al. 2006b; Wellcome Trust Centre for Human Genetics - Mouse Resources 2022) using mvMAPIT. We focused on 13 hematological phenotypes including: basophils, hematocrit, hemoglobin, lymphocytes, mean corpuscular hemoglobin, mean corpuscular volume, monocytes, neutrophils, plateletcrit, platelets, red blood cell count, red cell distribution width, and white blood cell count. The red horizontal lines indicate a genome-wide Bonferroni corrected significance threshold ($P = 4.83 \times 10^{-6}$). Panels (A) and (D) correspond to using mvMAPIT with Fisher's method, (B) and (E) correspond to the harmonic mean, and (C) and (F) correspond to the Cauchy combination test. The green colored dots are SNPs that have significant marginal epistatic effects after multiple test correction. Significant SNPs were mapped to the closest neighboring genes using the Mouse Genome Informatics database (<http://www.informatics.jax.org>) (Blake et al. 2020; Smith et al. 2009). The plateau in the P -value distribution displayed in panel (F) is due to numerical limitations of the `pcauchy` function in the `R` stats package.

A.2 Supplementary tables

Table A.1: **The mvMAPIT framework using Fisher’s method preserves type I error rates under the null model when traits are generated by only additive effects (sample size $N = 1,000$ individuals).** In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent additive effects ($v_\beta = 0$), traits with highly correlated additive effects ($v_\beta = 0.8$), and traits with perfectly correlated additive effects ($v_\beta = 1$). We assess the calibration of the P -values that are produced by mvMAPIT during each of the three key steps in its combinatorial hypothesis testing procedure (see Materials and Methods). We show type I error rates resulting from P -values taken from the “univariate” test on each trait independently, the “covariance” P -values which corresponds to assessing the pairwise interactions affecting both traits, and the final “combined” P -value. Results are summarized over 100 simulated replicates. Values in the parentheses are the standard deviations across replicates.

Step	Add. Effect Corr.	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
Univariate	$v_\beta = 0.0$	0.030 (1×10^{-2})	0.007 (3×10^{-3})	0.0007 (6×10^{-4})
	$v_\beta = 0.8$	0.030 (1×10^{-2})	0.007 (2×10^{-3})	0.0007 (6×10^{-4})
	$v_\beta = 1.0$	0.030 (1×10^{-2})	0.007 (2×10^{-3})	0.0007 (6×10^{-4})
Covariance	$v_\beta = 0.0$	0.040 (1×10^{-2})	0.005 (2×10^{-3})	0.0002 (4×10^{-4})
	$v_\beta = 0.8$	0.040 (1×10^{-2})	0.005 (2×10^{-3})	0.0002 (3×10^{-4})
	$v_\beta = 1.0$	0.040 (1×10^{-2})	0.005 (2×10^{-3})	0.0003 (3×10^{-4})
Combined	$v_\beta = 0.0$	0.030 (1×10^{-2})	0.006 (2×10^{-3})	0.0005 (7×10^{-4})
	$v_\beta = 0.8$	0.040 (1×10^{-2})	0.006 (2×10^{-3})	0.0005 (4×10^{-4})
	$v_\beta = 1.0$	0.040 (1×10^{-2})	0.006 (2×10^{-3})	0.0006 (6×10^{-4})

Table A.2: **The mvMAPIT framework using Fisher’s method preserves type I error rates under the null model when traits are generated by only additive effects (sample size $N = 1,750$ individuals).** In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent additive effects ($v_\beta = 0$), traits with highly correlated additive effects ($v_\beta = 0.8$), and traits with perfectly correlated additive effects ($v_\beta = 1$). We assess the calibration of the P -values that are produced by mvMAPIT during each of the three key steps in its combinatorial hypothesis testing procedure (see Materials and Methods). We show type I error rates resulting from P -values taken from the “univariate” test on each trait independently, the “covariance” P -values which corresponds to assessing the pairwise interactions affecting both traits, and the final “combined” P -value. Results are summarized over 100 simulated replicates. Values in the parentheses are the standard deviations across replicates.

Step	Add. Effect Corr.	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
Univariate	$v_\beta = 0.0$	0.030 (1×10^{-2})	0.008 (2×10^{-3})	0.0007 (5×10^{-4})
	$v_\beta = 0.8$	0.030 (1×10^{-2})	0.008 (2×10^{-3})	0.0009 (7×10^{-4})
	$v_\beta = 1.0$	0.030 (1×10^{-2})	0.008 (3×10^{-3})	0.0009 (9×10^{-4})
Covariance	$v_\beta = 0.0$	0.040 (1×10^{-2})	0.006 (2×10^{-3})	0.0003 (4×10^{-4})
	$v_\beta = 0.8$	0.040 (1×10^{-2})	0.006 (2×10^{-3})	0.0002 (3×10^{-4})
	$v_\beta = 1.0$	0.040 (1×10^{-2})	0.006 (2×10^{-3})	0.0002 (3×10^{-4})
Combined	$v_\beta = 0.0$	0.040 (1×10^{-2})	0.007 (2×10^{-3})	0.0006 (5×10^{-4})
	$v_\beta = 0.8$	0.040 (1×10^{-2})	0.007 (2×10^{-3})	0.0007 (8×10^{-4})
	$v_\beta = 1.0$	0.040 (1×10^{-2})	0.007 (2×10^{-3})	0.0006 (6×10^{-4})

Table A.3: **The mvMAPIT framework using the harmonic mean preserves type I error rates under the null model when traits are generated by only additive effects (sample size $N = 1,000$ individuals).** In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent additive effects ($v_\beta = 0$), traits with highly correlated additive effects ($v_\beta = 0.8$), and traits with perfectly correlated additive effects ($v_\beta = 1$). We assess the calibration of the P -values that are produced by mvMAPIT during each of the three key steps in its combinatorial hypothesis testing procedure (see Materials and Methods). We show type I error rates resulting from P -values taken from the “univariate” test on each trait independently, the “covariance” P -values which corresponds to assessing the pairwise interactions affecting both traits, and the final “combined” P -value. Results are summarized over 100 simulated replicates. Values in the parentheses are the standard deviations across replicates.

Step	Add. Effect Corr.	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
Univariate	$v_\beta = 0.0$	0.030 (1×10^{-2})	0.007 (3×10^{-3})	0.0007 (6×10^{-4})
	$v_\beta = 0.8$	0.030 (1×10^{-2})	0.007 (2×10^{-3})	0.0007 (6×10^{-4})
	$v_\beta = 1.0$	0.030 (1×10^{-2})	0.007 (2×10^{-3})	0.0007 (6×10^{-4})
Covariance	$v_\beta = 0.0$	0.040 (1×10^{-2})	0.005 (2×10^{-3})	0.0002 (4×10^{-4})
	$v_\beta = 0.8$	0.040 (1×10^{-2})	0.005 (2×10^{-3})	0.0002 (3×10^{-4})
	$v_\beta = 1.0$	0.040 (1×10^{-2})	0.005 (2×10^{-3})	0.0003 (3×10^{-4})
Combined	$v_\beta = 0.0$	0.040 (1×10^{-2})	0.006 (2×10^{-3})	0.0005 (5×10^{-4})
	$v_\beta = 0.8$	0.040 (1×10^{-2})	0.006 (2×10^{-3})	0.0004 (4×10^{-4})
	$v_\beta = 1.0$	0.040 (1×10^{-2})	0.006 (2×10^{-3})	0.0005 (5×10^{-4})

Table A.4: **The mvMAPIT framework using the harmonic mean preserves type I error rates under the null model when traits are generated by only additive effects (sample size $N = 1,750$ individuals).** In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent additive effects ($v_\beta = 0$), traits with highly correlated additive effects ($v_\beta = 0.8$), and traits with perfectly correlated additive effects ($v_\beta = 1$). We assess the calibration of the P -values that are produced by mvMAPIT during each of the three key steps in its combinatorial hypothesis testing procedure (see Materials and Methods). We show type I error rates resulting from P -values taken from the “univariate” test on each trait independently, the “covariance” P -values which corresponds to assessing the pairwise interactions affecting both traits, and the final “combined” P -value. Results are summarized over 100 simulated replicates. Values in the parentheses are the standard deviations across replicates.

Step	Add. Effect Corr.	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
Univariate	$v_\beta = 0.0$	0.030 (1×10^{-2})	0.008 (2×10^{-3})	0.0007 (5×10^{-4})
	$v_\beta = 0.8$	0.030 (1×10^{-2})	0.008 (2×10^{-3})	0.0009 (7×10^{-4})
	$v_\beta = 1.0$	0.030 (1×10^{-2})	0.008 (3×10^{-3})	0.0009 (9×10^{-4})
Covariance	$v_\beta = 0.0$	0.040 (1×10^{-2})	0.006 (2×10^{-3})	0.0003 (4×10^{-4})
	$v_\beta = 0.8$	0.040 (1×10^{-2})	0.006 (2×10^{-3})	0.0002 (3×10^{-4})
	$v_\beta = 1.0$	0.040 (1×10^{-2})	0.006 (2×10^{-3})	0.0002 (3×10^{-4})
Combined	$v_\beta = 0.0$	0.040 (1×10^{-2})	0.008 (2×10^{-3})	0.0006 (5×10^{-4})
	$v_\beta = 0.8$	0.040 (1×10^{-2})	0.008 (2×10^{-3})	0.0007 (8×10^{-4})
	$v_\beta = 1.0$	0.040 (1×10^{-2})	0.007 (2×10^{-3})	0.0005 (5×10^{-4})

Table A.5: **The mvMAPIT framework using the harmonic mean preserves type I error rates under the null model when traits are generated by only additive effects (sample size $N = 2,500$ individuals).** In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent additive effects ($v_\beta = 0$), traits with highly correlated additive effects ($v_\beta = 0.8$), and traits with perfectly correlated additive effects ($v_\beta = 1$). We assess the calibration of the P -values that are produced by mvMAPIT during each of the three key steps in its combinatorial hypothesis testing procedure (see Materials and Methods). We show type I error rates resulting from P -values taken from the “univariate” test on each trait independently, the “covariance” P -values which corresponds to assessing the pairwise interactions affecting both traits, and the final “combined” P -value. Results are summarized over 100 simulated replicates. Values in the parentheses are the standard deviations across replicates.

Step	Add. Effect Corr.	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
Univariate	$v_\beta = 0.0$	0.030 (1×10^{-2})	0.009 (2×10^{-3})	0.0010 (9×10^{-4})
	$v_\beta = 0.8$	0.030 (1×10^{-2})	0.009 (2×10^{-3})	0.0009 (7×10^{-4})
	$v_\beta = 1.0$	0.030 (1×10^{-2})	0.009 (3×10^{-3})	0.0009 (7×10^{-4})
Covariance	$v_\beta = 0.0$	0.040 (1×10^{-2})	0.006 (2×10^{-3})	0.0003 (4×10^{-4})
	$v_\beta = 0.8$	0.040 (1×10^{-2})	0.007 (2×10^{-3})	0.0004 (5×10^{-4})
	$v_\beta = 1.0$	0.040 (1×10^{-2})	0.006 (2×10^{-3})	0.0003 (4×10^{-4})
Combined	$v_\beta = 0.0$	0.040 (1×10^{-2})	0.008 (2×10^{-3})	0.0007 (6×10^{-4})
	$v_\beta = 0.8$	0.040 (1×10^{-2})	0.008 (2×10^{-3})	0.0007 (6×10^{-4})
	$v_\beta = 1.0$	0.040 (1×10^{-2})	0.008 (2×10^{-3})	0.0005 (6×10^{-4})

Table A.6: **The mvMAPIT framework using the Cauchy combination test preserves type I error rates under the null model when traits are generated by only additive effects (sample size $N = 1,000$ individuals).** In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent additive effects ($v_\beta = 0$), traits with highly correlated additive effects ($v_\beta = 0.8$), and traits with perfectly correlated additive effects ($v_\beta = 1$). We assess the calibration of the P -values that are produced by mvMAPIT during each of the three key steps in its combinatorial hypothesis testing procedure (see Materials and Methods). We show type I error rates resulting from P -values taken from the “univariate” test on each trait independently, the “covariance” P -values which corresponds to assessing the pairwise interactions affecting both traits, and the final “combined” P -value. Results are summarized over 100 simulated replicates. Values in the parentheses are the standard deviations across replicates.

Step	Add. Effect Corr.	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
Univariate	$v_\beta = 0.0$	0.030 (6×10^{-3})	0.007 (3×10^{-3})	0.0007 (6×10^{-4})
	$v_\beta = 0.8$	0.030 (5×10^{-3})	0.030 (6×10^{-3})	0.0007 (6×10^{-4})
	$v_\beta = 1.0$	0.030 (1×10^{-2})	0.030 (6×10^{-3})	0.0007 (6×10^{-4})
Covariance	$v_\beta = 0.0$	0.040 (6×10^{-3})	0.005 (2×10^{-3})	0.0002 (4×10^{-4})
	$v_\beta = 0.8$	0.040 (5×10^{-3})	0.005 (2×10^{-3})	0.0002 (3×10^{-4})
	$v_\beta = 1.0$	0.040 (6×10^{-3})	0.005 (2×10^{-3})	0.0003 (3×10^{-4})
Combined	$v_\beta = 0.0$	0.030 (5×10^{-3})	0.006 (2×10^{-3})	0.0005 (5×10^{-4})
	$v_\beta = 0.8$	0.030 (4×10^{-3})	0.006 (2×10^{-3})	0.0004 (4×10^{-4})
	$v_\beta = 1.0$	0.030 (5×10^{-3})	0.006 (2×10^{-3})	0.0005 (5×10^{-4})

Table A.7: **The mvMAPIT framework using the Cauchy combination test preserves type I error rates under the null model when traits are generated by only additive effects (sample size $N = 1,750$ individuals).** In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent additive effects ($v_\beta = 0$), traits with highly correlated additive effects ($v_\beta = 0.8$), and traits with perfectly correlated additive effects ($v_\beta = 1$). We assess the calibration of the P -values that are produced by mvMAPIT during each of the three key steps in its combinatorial hypothesis testing procedure (see Materials and Methods). We show type I error rates resulting from P -values taken from the “univariate” test on each trait independently, the “covariance” P -values which corresponds to assessing the pairwise interactions affecting both traits, and the final “combined” P -value. Results are summarized over 100 simulated replicates. Values in the parentheses are the standard deviations across replicates.

Step	Add. Effect Corr.	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
Univariate	$v_\beta = 0.0$	0.030 (6×10^{-3})	0.008 (2×10^{-3})	0.0008 (6×10^{-4})
	$v_\beta = 0.8$	0.030 (5×10^{-3})	0.008 (2×10^{-3})	0.0009 (8×10^{-4})
	$v_\beta = 1.0$	0.030 (6×10^{-3})	0.008 (3×10^{-3})	0.0008 (8×10^{-4})
Covariance	$v_\beta = 0.0$	0.040 (6×10^{-3})	0.006 (2×10^{-3})	0.0003 (4×10^{-4})
	$v_\beta = 0.8$	0.040 (6×10^{-3})	0.006 (2×10^{-3})	0.0002 (3×10^{-4})
	$v_\beta = 1.0$	0.040 (6×10^{-3})	0.006 (2×10^{-3})	0.0002 (3×10^{-4})
Combined	$v_\beta = 0.0$	0.040 (5×10^{-3})	0.007 (2×10^{-3})	0.0006 (5×10^{-4})
	$v_\beta = 0.8$	0.040 (5×10^{-3})	0.007 (2×10^{-3})	0.0007 (8×10^{-4})
	$v_\beta = 1.0$	0.040 (5×10^{-3})	0.007 (2×10^{-3})	0.0005 (5×10^{-4})

Table A.8: **The mvMAPIT framework using the Cauchy combination test preserves type I error rates under the null model when traits are generated by only additive effects (sample size $N = 2,500$ individuals).** In these simulations, quantitative traits are simulated to have narrow-sense heritability $h^2 = 0.6$ with an architecture made up of only additive genetic variation. Each row corresponds to a setting where the additive genetic effects for a causal SNP have different correlation structures across traits. In these simulations, we consider scenarios where we have traits with independent additive effects ($v_\beta = 0$), traits with highly correlated additive effects ($v_\beta = 0.8$), and traits with perfectly correlated additive effects ($v_\beta = 1$). We assess the calibration of the P -values that are produced by mvMAPIT during each of the three key steps in its combinatorial hypothesis testing procedure (see Materials and Methods). We show type I error rates resulting from P -values taken from the “univariate” test on each trait independently, the “covariance” P -values which corresponds to assessing the pairwise interactions affecting both traits, and the final “combined” P -value. Results are summarized over 100 simulated replicates. Values in the parentheses are the standard deviations across replicates.

Step	Add. Effect Corr.	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
Univariate	$v_\beta = 0.0$	0.030 (5×10^{-3})	0.009 (2×10^{-3})	0.001 (9×10^{-4})
	$v_\beta = 0.8$	0.030 (5×10^{-3})	0.009 (2×10^{-3})	0.0009 (7×10^{-4})
	$v_\beta = 1.0$	0.030 (5×10^{-3})	0.009 (3×10^{-3})	0.0009 (7×10^{-4})
Covariance	$v_\beta = 0.0$	0.040 (5×10^{-3})	0.006 (2×10^{-3})	0.0003 (3×10^{-4})
	$v_\beta = 0.8$	0.040 (6×10^{-3})	0.007 (2×10^{-3})	0.0004 (5×10^{-4})
	$v_\beta = 1.0$	0.040 (5×10^{-3})	0.006 (2×10^{-3})	0.0003 (4×10^{-4})
Combined	$v_\beta = 0.0$	0.040 (5×10^{-3})	0.008 (2×10^{-3})	0.0007 (6×10^{-4})
	$v_\beta = 0.8$	0.040 (5×10^{-3})	0.008 (2×10^{-3})	0.0007 (6×10^{-4})
	$v_\beta = 1.0$	0.040 (5×10^{-3})	0.007 (2×10^{-3})	0.0006 (6×10^{-4})

Table A.9: **Complete summary of the marginal epistatic results after applying the mvMAPIT framework to protein sequence data from a nearly combinatorially complete library of two broadly neutralizing anti-influenza antibodies.** Here, data is from Phillips et al. (2021) who generated a nearly combinatorially complete library for two broadly neutralizing anti-influenza antibodies (bnAbs), CR6261 and CR9114. In the first column, we list the antibody being analyzed. In the second column, we give their corresponding residue. In the third and fourth columns, we list all the pairwise antigen combinations done in the analysis. In the remaining columns, we give the results stemming from univariate analyses on antigens #1 and #2, respectively, the covariance (cov) test, and the overall P -values derived by mvMAPIT using Fisher’s method, the harmonic mean, and the Cauchy combination test. Tutorials for how to take these results and recreate the Manhattan plots shown in Figures 1.5, A.25, and A.26 can be found in the mvMAPIT GitHub repository (see Materials and Methods). (XLSX)

Table A.10: **Complete summary of the marginal epistatic results after applying the mvMAPIT framework to 15 hematology traits in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics (Valdar et al. 2006a; Valdar et al. 2006b; *Wellcome Trust Centre for Human Genetics - Mouse Resources* 2022).** In the first column, we list the ID of each SNP. In the second and third columns, we give their corresponding chromosome and basepair according to the mouse assembly NCBI build 34 (accessed from Shifman et al. (2006)). In the fourth and fifth columns, we list all the pairwise trait combinations done in the analysis. In the remaining columns, we give the results stemming from univariate analyses on traits #1 and #2, respectively, the covariance (cov) test, and the overall P -values derived by mvMAPIT using Fisher’s method, the harmonic mean, and the Cauchy combination test. Tutorials for how to take these results and recreate the Manhattan plots shown in Figures A.27-A.34 and A.36-A.37 can be found in the mvMAPIT GitHub repository (see Materials and Methods). (XLSX)

Table A.11: Notable SNPs with marginal epistatic effects after applying the mvMAPIT framework to 15 hematology traits in the heterogeneous stock of mice data set from the Wellcome Trust Centre for Human Genetics (Valdar et al. 2006a; Valdar et al. 2006b; *Wellcome Trust Centre for Human Genetics - Mouse Resources* 2022). In the first two columns, we list SNPs and their genetic location according to the mouse assembly NCBI build 34 (accessed from Shifman et al. (2006)) in the format **Chromosome:Basepair**. The next columns detail the closest neighboring genes found using the Mouse Genome Informatics database (<http://www.informatics.jax.org>) (Blake et al. 2020; Smith et al. 2009), a short summary of the suggested annotated function for those genes, and the reference to the source of the annotation. See Table A.10 for the complete list of SNP and SNP-set level results.

SNP	Location	Genomic Annotation	Reference
rs3699393	2:5887012	anemia and abnormal bone marrow cell development	(Weischenfeldt et al. 2008)
rs13478092	5:3601413	abnormal venous thrombosis	(Berendse et al. 2019)
rs3694887	5:102770070	abnormal B and T cell number and morphology	(Isnard et al. 2000)
rs3694887	5:102770070	abnormal B and T cell number and morphology	(Isnard et al. 2000)
rs13478923	6:99475169	abnormal B cell differentiation, physiology, count	(Feng et al. 2010; Hu et al. 2006)
rs13478924	6:99571626	abnormal B cell differentiation, physiology, count	(Feng et al. 2010; Hu et al. 2006)
rs13478985	6:115245823	decreased bone marrow cell count	(Mortensen et al. 2011; Clarke et al. 2015)
rs3723163	11:103800737	decreased fasting circulating glucose level	(Database 2014)
rs3723163	11:103800737	decreased fasting circulating glucose level	(Database 2014)
rs3724260	12:100163212	anemia	(Raaijmakers et al. 2010)
rs3692165	14:27756640	decreased circulating glucose level	(Database 2014)
rs13482117	14:27614362	decreased circulating glucose level	(Database 2014)
rs13482288	14:81840412	abnormal B cell differentiation and physiology	(Yang et al. 2014)
rs4173870	16:35764290	differentiation of erythrocytes	(Castro-Ochoa et al. 2019)
rs4212102	16:84204704	increased susceptibility to induced thrombosis	(Xu et al. 2017; Souto et al. 2000)
rs4212186	16:84273330	increased susceptibility to induced thrombosis	(Xu et al. 2017; Souto et al. 2000)
rs3711994	19:45078018	abnormal lymphocyte morphology	(Nakayama et al. 2003)

APPENDIX B

Supplementary information for Chapter 2

B.1 The survival and hazard functions

For a continuous random variable T , the survival function $S_T(t)$ is defined as the probability that T exceeds t :

$$S_T(t) = P(T > t) = 1 - F(t) = \int_t^{\infty} f_T(u) du \quad (\text{B.1})$$

for probability density function $f_T(t)$ and cumulative distribution function $F_T(t)$.

The hazard function $\lambda_T(t)$ is defined as the instantaneous rate of failure at time t given survival to that time. It is the limit of the conditional probability of failure in a small interval $(t, t + \Delta t]$ given survival to t as Δt approaches 0:

$$\lambda_T(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (\text{B.2})$$

Following the definition of conditional probability, we see that the conditional probability can be expressed as the ratio of the joint probability of T being in the interval $(t, t + \Delta t]$ and T being greater than t to the probability of T being greater than t . The first probability we recognize as the probability density function $f_T(t)$ and the second probability is the survival function $S_T(t)$. Hence, we can rewrite the hazard function as

$$\lambda_T(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f_T(t)}{S_T(t)} \quad (\text{B.3})$$

B.2 The Breslow estimator

The Breslow estimator is a non-parametric estimator for the cumulative baseline hazard function. It is derived from the partial likelihood of the Cox proportional hazards model and provides a step function estimate that jumps at each observed event time. The estimator takes into account the risk sets at each event time, where the risk set consists

of all subjects who are still at risk (have not experienced the event and are not censored) at that time point. This makes it particularly useful for handling censored survival data while accounting for the effects of covariates through the linear predictor η_i . Without loss of generality, we can order the samples for all N subjects according to the observed time t_i such that $t_1 \leq t_2 \leq \dots \leq t_N$.

We can use the Breslow estimator to get the cumulative baseline hazard.

$$\hat{\Lambda}_0(t) = \sum_{i=1}^n \frac{\delta_i I(t_i \leq t)}{\sum_{j \in R(t_i)} \exp(\eta_j)}$$

Here $R(t_i) = \{j : t_j \geq t_i\}$ is the set of subjects at risk at time t_i .

B.3 Approximation of Gaussian integral using Laplace's method

Laplace's method is a technique for approximating integrals of the form $\int \exp(f(x))dx$ where $f(x)$ has a unique maximum. The method approximates the integrand by a Gaussian function centered at the maximum of $f(x)$. The resulting approximation involves the value of $f(x)$ at its maximum and the determinant of the negative Hessian matrix of $f(x)$ at this point. We use Laplace's method to approximate the integral of the likelihood function in the Cox model with a linear mixed effects model for frailties.

$$\begin{aligned}
L(\lambda_0, \boldsymbol{\beta}, \theta) &= \int_{\mathbb{R}^N} \Pi_{i=1}^N \left[\lambda(t_i; \mathbf{z}_i, \boldsymbol{\beta} | r_{ik})^{\delta_i} S(t_i; \mathbf{z}_i, \boldsymbol{\beta} | r_{ik}) \right] \\
&\quad \times (2\pi)^{-\frac{N}{2}} \frac{1}{\sqrt{|\mathbf{V}(\theta)|}} \exp \left(-\frac{1}{2} \mathbf{r}_k^T (\mathbf{V}(\theta))^{-1} \mathbf{r}_k \right) d\mathbf{r}_k \\
&= (2\pi)^{-\frac{N}{2}} \frac{1}{\sqrt{|\mathbf{V}(\theta)|}} \Pi_{i=1}^N \lambda(t_i; \mathbf{z}_i, \boldsymbol{\beta} | r_{ik})^{\delta_i} \\
&\quad \times \int_{\mathbb{R}^N} \exp \left[\sum_{i=1}^N (\delta_i \{ \mathbf{z}_i^T \boldsymbol{\beta} + r_{ik} \} - \Lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta} + r_{ik})) - \frac{1}{2} \mathbf{r}_k^T (\mathbf{V}(\theta))^{-1} \mathbf{r}_k \right] d\mathbf{r}_k
\end{aligned} \tag{B.4}$$

Writing the expression in the exponential function as

$$f(\mathbf{r}_k) = \sum_{i=1}^n (\delta_i \{ \mathbf{z}_i^T \boldsymbol{\beta} + r_{ik} \} - \Lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta} + r_{ik})) - \frac{1}{2} \mathbf{r}_k^T (\mathbf{V}(\theta))^{-1} \mathbf{r}_k$$

we can approximate the integral with the Laplace method.

Laplace's method approximates the intractable integral by expanding the integrand around the global maximum of the integrand.

$$\int \exp[f(\mathbf{r}_k)] d\mathbf{r}_k \approx (2\pi)^{\frac{n}{2}} | -f''(\hat{\mathbf{r}}_k) |^{-\frac{1}{2}} \cdot \exp(f(\hat{\mathbf{r}}_k))$$

with $\hat{\mathbf{r}}_k = \arg \max_{\mathbf{r}_k} f(\mathbf{r}_k)$. With this approximation, we arrive at the penalized log-likelihood in Eq. [B.5](#).

B.4 Derivation of the penalized log-likelihood

We can write the log-likelihood

$$\begin{aligned}
l(\lambda_0, \boldsymbol{\beta}, \theta) &\approx -\frac{1}{2} \log |\mathbf{V}(\theta)| + \sum_{i=1}^n \delta_i \log \lambda_0(t_i) \\
&\quad - \frac{1}{2} \log |-f''(\hat{\mathbf{r}}_k)| + f(\hat{\mathbf{r}}_k) \\
&= -\frac{1}{2} \log |\mathbf{V}(\theta)| - \frac{1}{2} \log |\hat{\mathbf{W}} + (\mathbf{V}(\theta))^{-1}| \\
&\quad + \sum_{i=1}^n [\delta_i (\log \lambda_0(t_i) + \mathbf{z}_i^T \boldsymbol{\beta} + \hat{r}_{ik}) - \Lambda_0(t_i) \cdot \exp(\mathbf{z}_i^T \boldsymbol{\beta} + \hat{r}_{ik})] \\
&\quad - \frac{1}{2} \hat{\mathbf{r}}_k^T (\mathbf{V}(\theta))^{-1} \hat{\mathbf{r}}_k
\end{aligned} \tag{B.5}$$

This approximation introduces $\hat{\mathbf{r}}_k$ as a maximum likelihood estimate of the genetic frailties and the diagonal matrix $\hat{\mathbf{W}} = \text{diag}(\Lambda_0(t_i) \cdot \exp(\mathbf{z}_i^T \boldsymbol{\beta} + \hat{r}_{ik}))$.

Assuming that W changes slowly as a function of the mean, the expression can further be simplified finding the penalized log-likelihood

$$\begin{aligned}
l_p(\lambda_0, \boldsymbol{\beta}, \theta) &= \sum_{i=1}^n [\delta_i (\log \lambda_0(t_i) + \mathbf{z}_i^T \boldsymbol{\beta} + \hat{r}_{ik}) - \Lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta} + \hat{r}_{ik})] \\
&\quad - \frac{1}{2} \hat{\mathbf{r}}_k^T (\mathbf{V}(\theta))^{-1} \hat{\mathbf{r}}_k
\end{aligned} \tag{B.6}$$

and proceed to find maximum likelihood estimators for the baseline hazard λ_0 and the parameters $\boldsymbol{\beta}$ and θ .

B.5 The permutation test ensures well-calibrated test statistics for small sample sizes

The test statistics from the efficient score test of CphGxE in theory follow a χ^2 -distribution with 1 degree of freedom. However, in the derivation of the method, some approximations have been made to arrive at estimators. As a result, if computed from a χ^2 -distribution, the p -value distribution under the null hypothesis shows type I error

inflation and is only expected to become truly χ^2 for large sample sizes. Calibrated p -values can be computed by performing a permutation test. Permutation tests are a non-parametric statistical method used to assess the significance of test statistics by comparing the observed value against a distribution generated through random permutations of the genetic loci. By permuting the columns of the genotype data, the outcome variable is independent of the molecular data, while the correlation structure between the genetic loci and distributional characteristics of the data can be preserved (Dudoit et al. 2002).

We can estimate distributions of the test statistics $T_1, \dots, T_K$ under the null hypothesis $H_0 : \sigma_k^2 = 0, \forall k \in \{1, \dots, K\}$ by permuting the columns of the input data for genotypes, covariates, survival times, and censoring indicators. The number of total possible permutations can be too large for practical computation. We therefore sample a random subset of B permutations. The permutation test works as follows: For all k , we compute the observed test statistics T_k . For the b -th permutation, $b \in 1, \dots, B$, (2) permute the data. and (3) compute test statistics $T_{1,b}, \dots, T_{K,b}$ for each hypothesis. The permutation distribution of the test statistic T_k for hypothesis H_k , $k = 1, \dots, K$, is given by the empirical distribution of $T_{k,1}, \dots, T_{k,B}$. For one-sided alternative hypotheses of the k -th test of virus-host interactions, the permutation p -value for hypothesis H_k is

$$p_{\text{permute}} = \frac{1}{B} \sum_{b=1}^B \mathbb{1}(T_{k,b} \geq T_k^{\text{observed}})$$

where $\mathbb{1}(\cdot) = 1$ if the condition in parenthesis is true, and 0 otherwise.

This approach generates a true null distribution of test statistics under the assumption that there is no association between the genetic variants and survival outcomes. The empirical p -value represents the proportion of permuted test statistics that are as extreme or more extreme than the observed test statistic. It is important to note that this test can produce p values equal to zero. However, the maximum accuracy of the test is limited by the number of permutations, i.e. a p -value of zero should be understood as $p < B^{-1}$.

B.6 Supplementary figures

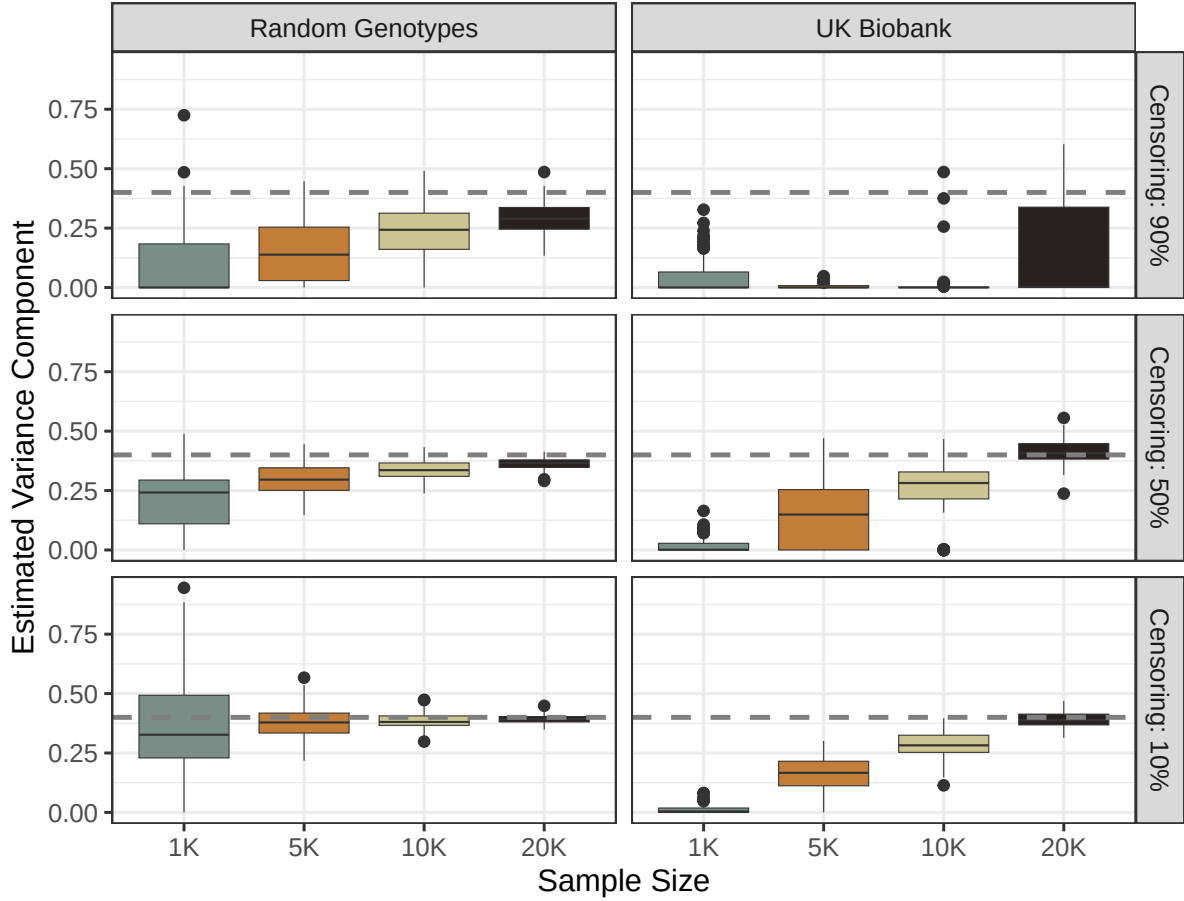


Figure B.1: **Frailty variance component estimates due to genetic relatedness are asymptotically unbiased.** Synthetic traits were simulated with additive genetic, environmental exposure, and covariate effects. Genotype data with 15,000 SNPs were either simulated using synthetic biallelic independent genotypes (Random Genotypes) or taken from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled to sample sizes of 1k, 5k, 10k and 20k individuals. We randomly selected 10% of all variants to be causal with additive effects and we assume that they explain 40% of the frailty variance for each trait as indicated by the gray dashed line. The frailty variance due to environmental exposures is set to 10% of the total frailty variance. The remainder of the frailty variance is explained by the covariates and noise is introduced by sampling the time variables at random. Data were analyzed using CphGxE under varying percentages of censoring. Under the null hypothesis $H_0 : \sigma^2 = 0$. 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.

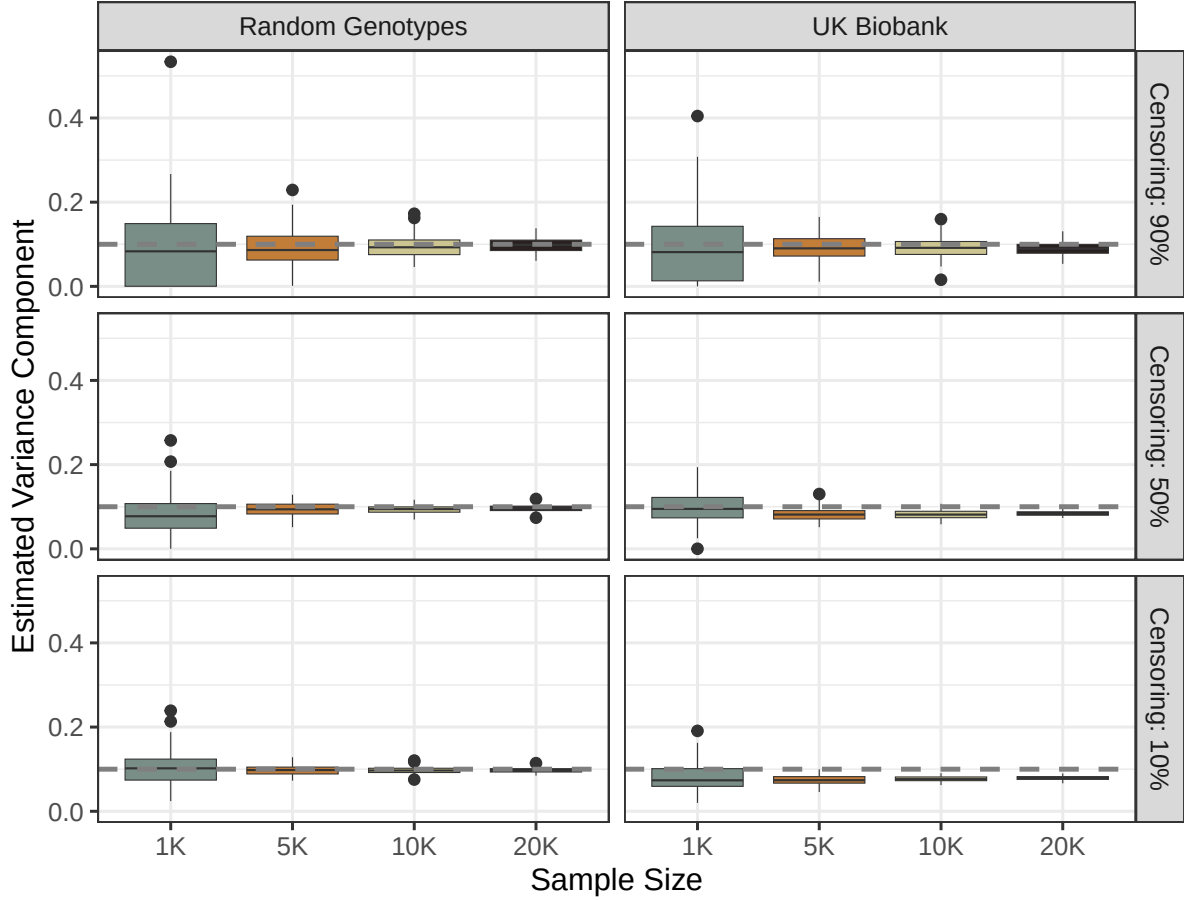


Figure B.2: **Frailty variance component estimates due to environmental exposures are unbiased.** Synthetic traits were simulated with additive genetic, environmental exposure, and covariate effects. Genotype data with 15,000 SNPs were either simulated using synthetic biallelic independent genotypes (Random Genotypes) or taken from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled to sample sizes of 1k, 5k, 10k and 20k individuals. We randomly selected 10% of all variants to be causal with additive effects and we assume that they explain 40% of the frailty variance for each trait. The frailty variance due to environmental exposures is set to 10% of the total frailty variance as indicated by the gray dashed line. The remainder of the frailty variance is explained by the covariates and noise is introduced by sampling the time variables at random. Data were analyzed using CphGxE under varying percentages of censoring. Under the null hypothesis $H_0 : \sigma^2 = 0$. 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.

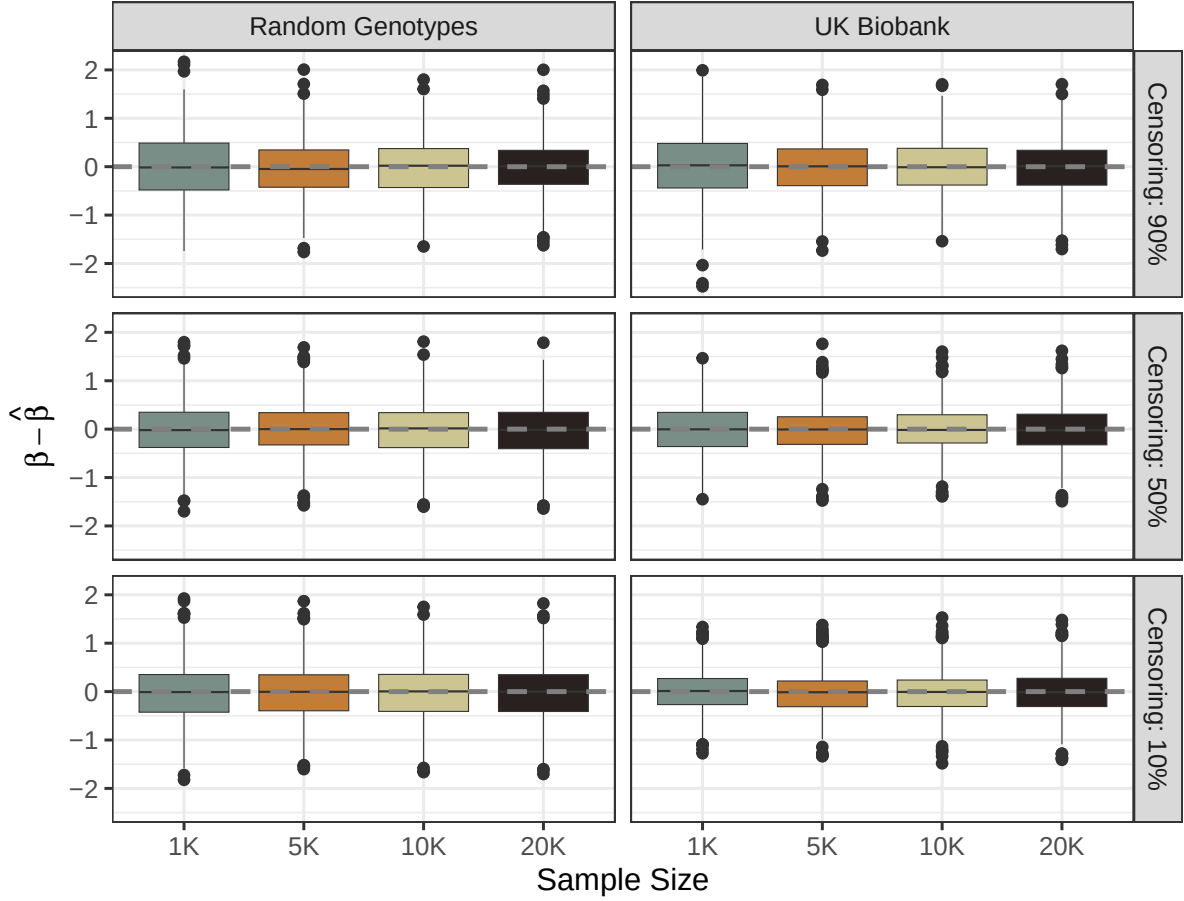


Figure B.3: **Frailty covariate fixed effect estimates are unbiased.** Synthetic traits were simulated with additive genetic, environmental exposure, and covariate effects. Genotype data with 15,000 SNPs were either simulated using synthetic biallelic independent genotypes (Random Genotypes) or taken from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled to sample sizes of 1k, 5k, 10k and 20k individuals. We randomly selected 10% of all variants to be causal with additive effects and we assume that they explain 40% of the frailty variance for each trait. The frailty variance due to environmental exposures is set to 10% of the total frailty variance. The remainder of the frailty variance of 50% is explained by the 3 synthetic covariates and noise is introduced by sampling the time variables at random. Data were analyzed using CphGxE under varying percentages of censoring. Under the null hypothesis $H_0 : \sigma^2 = 0$. 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.

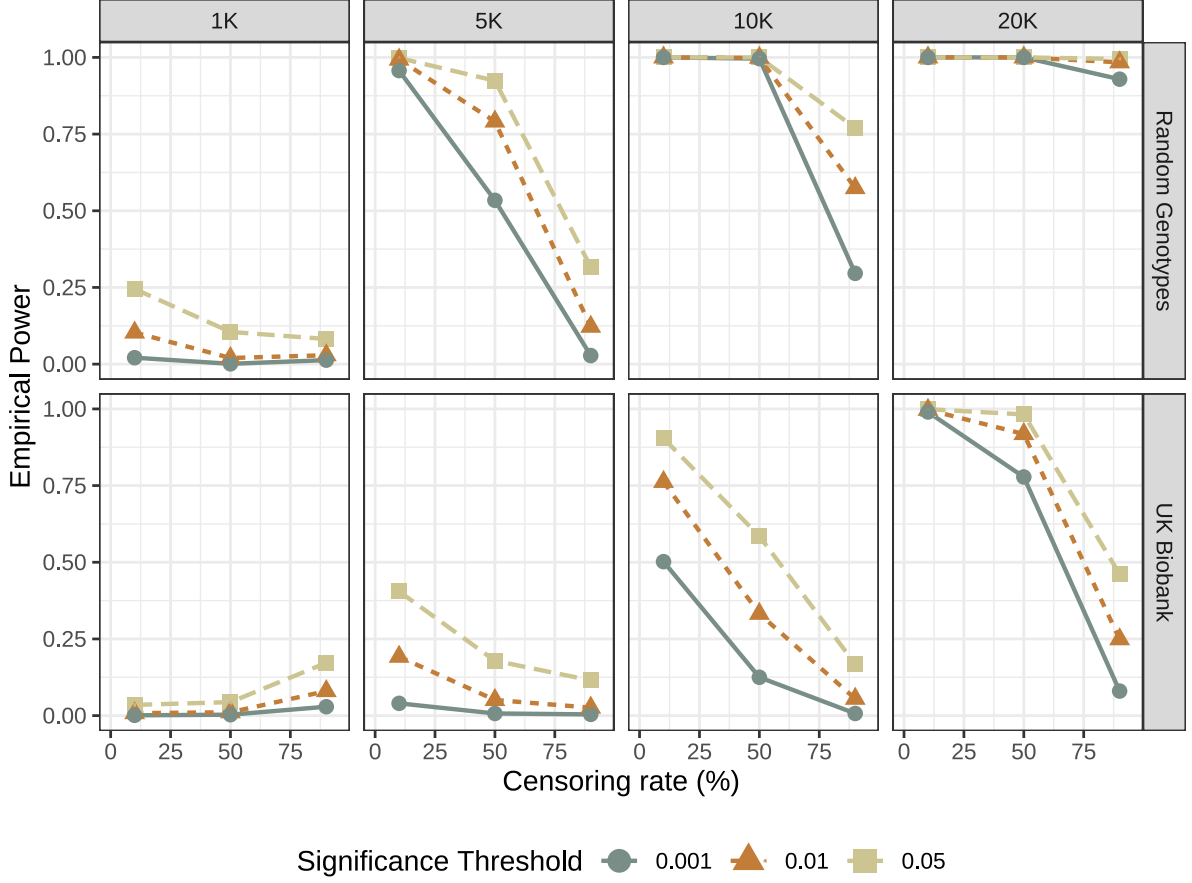


Figure B.4: **Empirical power of CphGxE as function of the censoring rate.** Synthetic time-to-event traits were simulated with additive genetic, environmental exposure, covariate, and gene-by-environment interaction effects. Genotype data with 15,000 SNPs were either simulated using synthetic biallelic independent genotypes (Random Genotypes) or taken from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled to 20k individuals. We randomly selected 10% of all variants to be causal with additive effects set at 30% of the frailty variance explained (FVE) for each trait. The frailty variance due to environmental exposures is set to 10% of the total frailty variance. The empirical power is computed under simulations in which the per-SNP gene-by-environment interaction explains 5% of the total frailty variance. The remainder of the frailty variance is explained by the covariates and noise is introduced by sampling the time variables at random. Data were analyzed using CphGxE under varying percentages of censoring (10%, 50%, and 90% censoring). The analyses used significance thresholds of $\alpha = 0.05$, 0.01, and 0.001. Results are based on 100 simulations per scenario, with error bars representing the standard deviation across replicates.

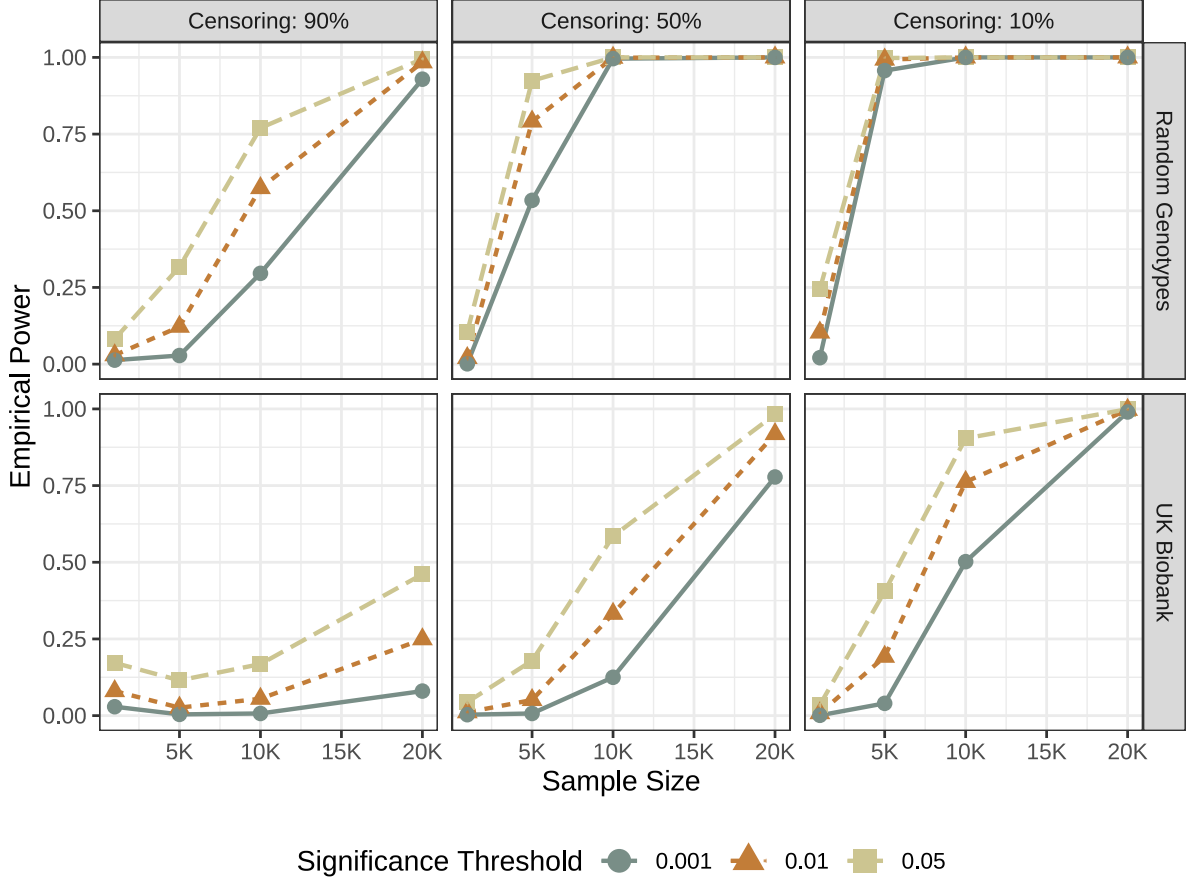


Figure B.5: **Empirical power of CphGxE as function of the censoring rate.** Synthetic time-to-event traits were simulated with additive genetic, environmental exposure, covariate, and gene-by-environment interaction effects. Genotype data with 15,000 SNPs were either simulated using synthetic biallelic independent genotypes (Random Genotypes) or taken from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled to 20k individuals. We randomly selected 10% of all variants to be causal with additive effects set at 30% of the frailty variance explained (FVE) for each trait. The frailty variance due to environmental exposures is set to 10% of the total frailty variance. The empirical power is computed under simulations in which the per-SNP gene-by-environment interaction explains 5% of the total frailty variance. The remainder of the frailty variance is explained by the covariates and noise is introduced by sampling the time variables at random. Data were analyzed using CphGxE under varying percentages of censoring (10%, 50%, and 90% censoring). The analyses used significance thresholds of $\alpha = 0.05$, 0.01, and 0.001. Results are based on 100 simulations per scenario, with error bars representing the standard deviation across replicates.

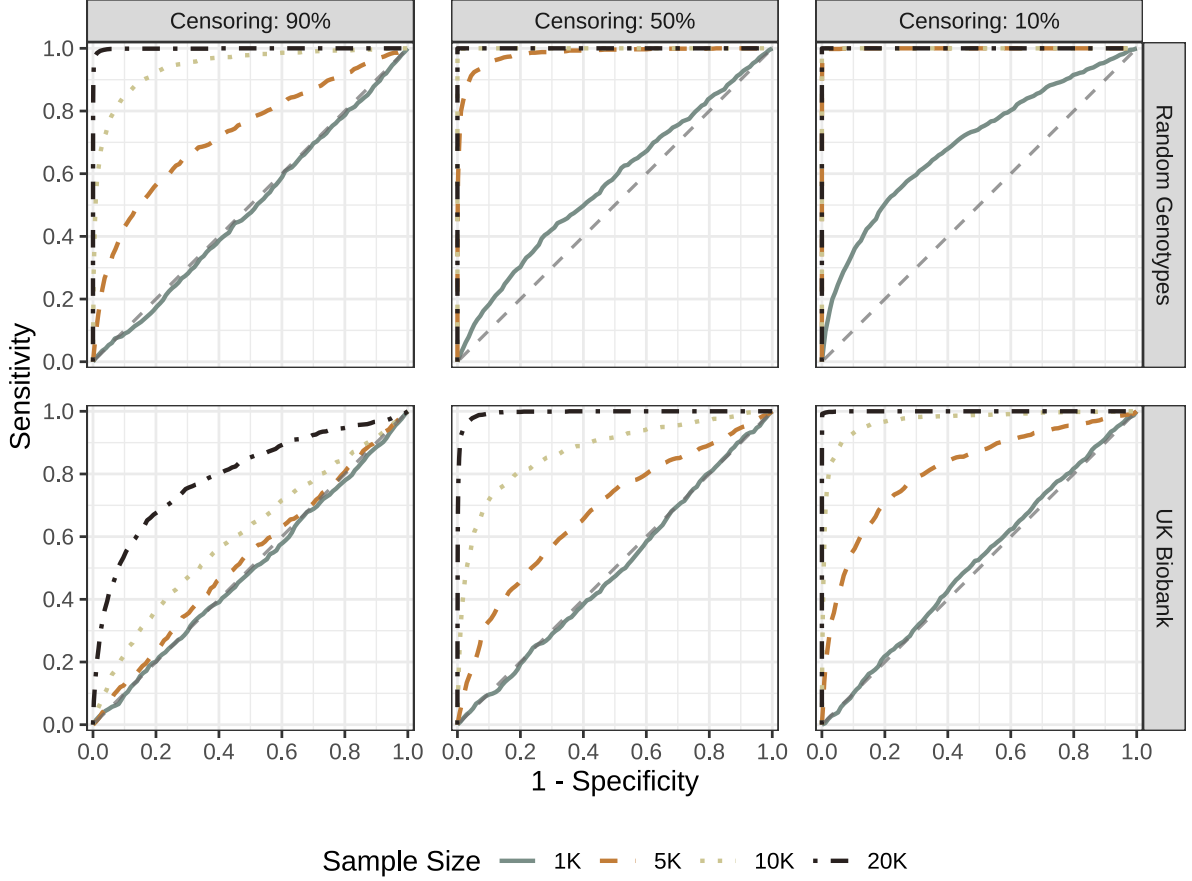


Figure B.6: Receiver Operating Characteristic (ROC) curves comparing the ability of CphGxE to correctly prioritize gene-by-environment interaction variants under the alternative model. Synthetic time-to-event traits were simulated with additive genetic, environmental exposure, covariate, and gene-by-environment interaction effects. Genotype data with 15,000 SNPs were either simulated using synthetic biallelic independent genotypes (Random Genotypes) or taken from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled to 1k, 5k, 10k and 20k individuals. We randomly selected 10% of all variants to be causal with additive effects set at 30% of the frailty variance explained (FVE) for each trait. The frailty variance due to environmental exposures is set to 10% of the total frailty variance. ROC curves are computed under simulations in which the per-SNP gene-by-environment interaction explains 5% of the total frailty variance. The remainder of the frailty variance is explained by the covariates and noise is introduced by sampling the time variables at random. Data were analyzed using CphGxE under varying percentages of censoring (10%, 50%, and 90% censoring). Results are based on 100 simulations per scenario.

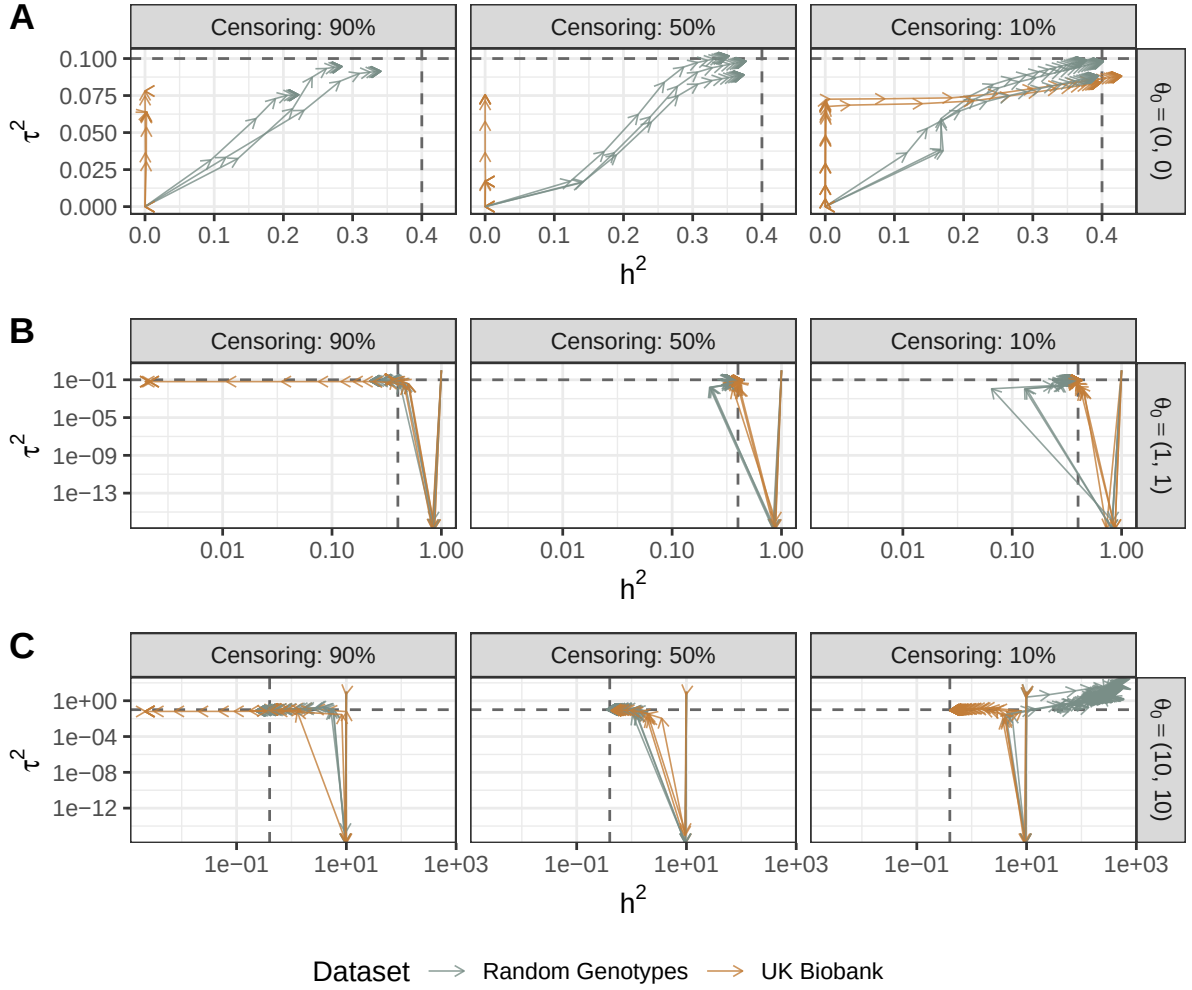


Figure B.7: Null fit convergence is sensitive to initial parameter values and censoring rates. The panels illustrate three examples of the iterative updates of the variance component estimates $\theta = (h^2, \tau^2)$ under the null model for different initial parameter values θ_0 and censoring rates. **(A)** Shows the convergence of the null fit when the initial parameter values are set to $\theta_0 = (0, 0)$. **(B)** Shows the convergence of the null fit when the initial parameter values are set to $\theta_0 = (1, 1)$. **(C)** Shows the convergence of the null fit when the initial parameter values are set to $\theta_0 = (10, 10)$. Synthetic time-to-event traits were simulated with additive genetic, environmental exposure, covariate, and gene-by-environment interaction effects. Genotype data with 15,000 SNPs were either simulated using synthetic biallelic independent genotypes (Random Genotypes) or taken from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled to 10k individuals. We randomly selected 10% of all variants to be causal with additive effects set at 40% of the frailty variance explained (FVE) for each trait. The frailty variance due to environmental exposures is set to 10% of the total frailty variance. The gray dashed lines indicate the true variance component values of the simulated data. The remainder of the frailty variance is explained by the covariates and noise is introduced by sampling the time variables at random. Data were analyzed using CphGxE under varying percentages of censoring (10%, 50%, and 90% censoring).

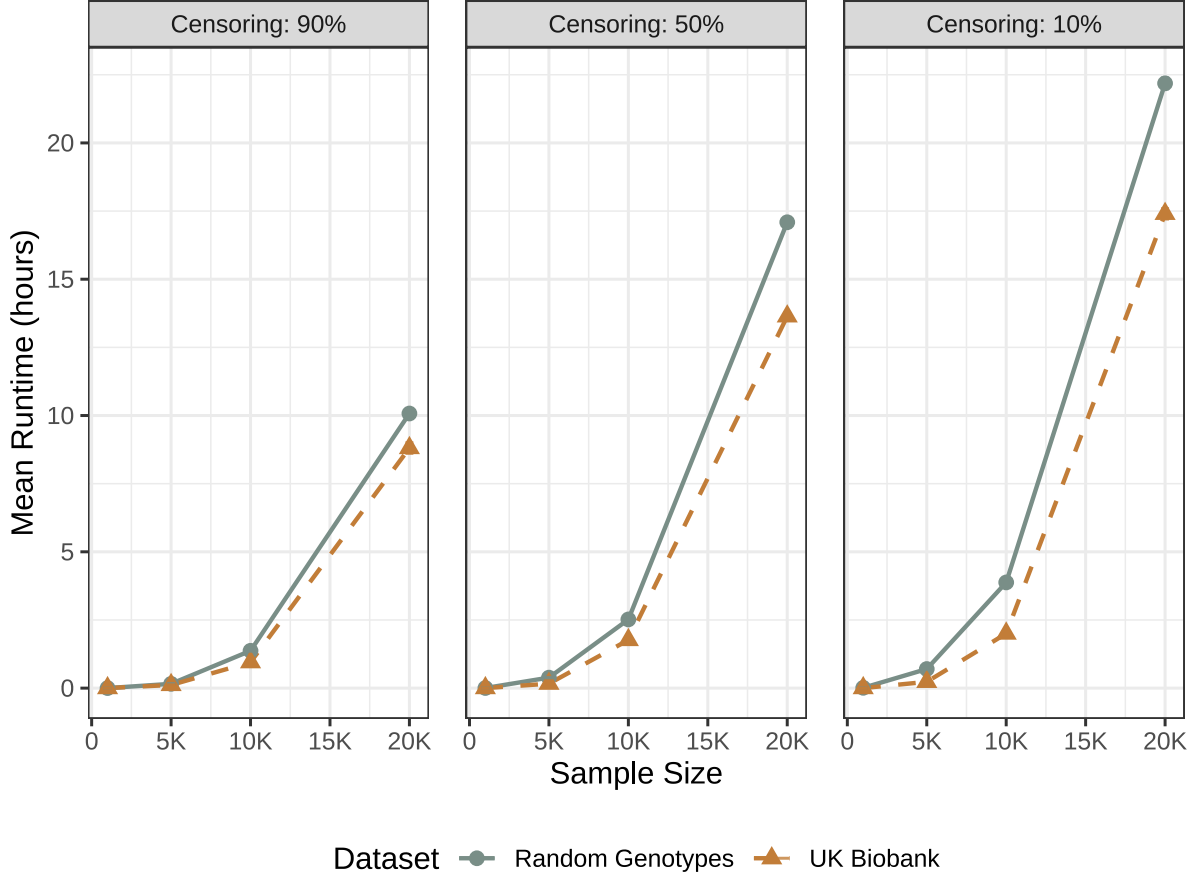


Figure B.8: **Runtime of the null fit in CphGxE as function of sample size in hours.** Synthetic time-to-event traits were simulated with additive genetic, environmental exposure, covariate, and gene-by-environment interaction effects. Genotype data with 15,000 SNPs were either simulated using synthetic biallelic independent genotypes (Random Genotypes) or taken from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled to 1k, 5k, 10k and 20k individuals. We randomly selected 10% of all variants to be causal with additive effects set at 30% of the frailty variance explained (FVE) for each trait. The frailty variance due to environmental exposures is set to 10% of the total frailty variance. The empirical power is computed under simulations in which the per-SNP gene-by-environment interaction explains 5% of the total frailty variance. The remainder of the frailty variance is explained by the covariates and noise is introduced by sampling the time variables at random. Data were analyzed using CphGxE under varying percentages of censoring (10%, 50%, and 90% censoring). Results are based on 100 simulations per scenario, with error bars representing the standard deviation across replicates.

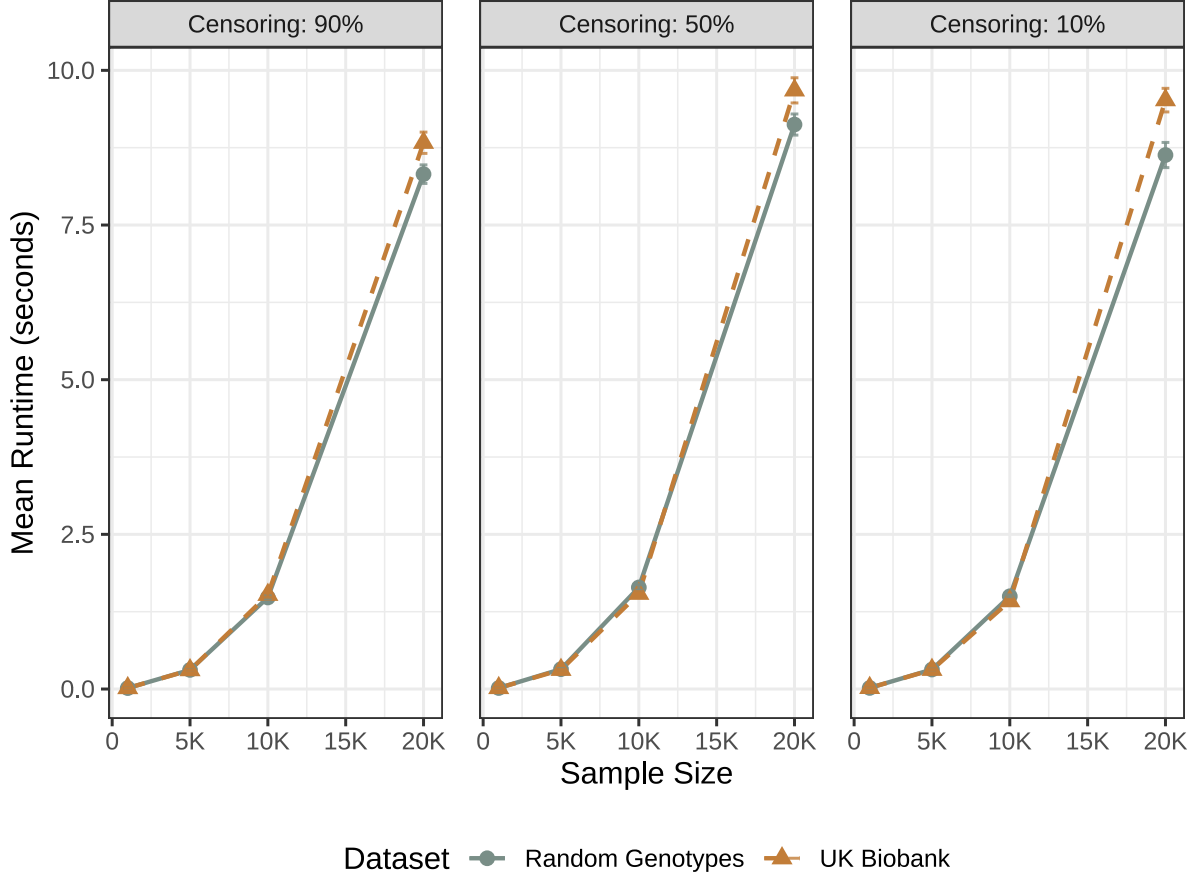


Figure B.9: **Runtime of the $G \times E$ test in CphGxE as function of sample size in seconds per variant.** Synthetic time-to-event traits were simulated with additive genetic, environmental exposure, covariate, and gene-by-environment interaction effects. Genotype data with 15,000 SNPs were either simulated using synthetic biallelic independent genotypes (Random Genotypes) or taken from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled to 1k, 5k, 10k and 20k individuals. We randomly selected 10% of all variants to be causal with additive effects set at 30% of the frailty variance explained (FVE) for each trait. The frailty variance due to environmental exposures is set to 10% of the total frailty variance. The empirical power is computed under simulations in which the per-SNP gene-by-environment interaction explains 5% of the total frailty variance. The remainder of the frailty variance is explained by the covariates and noise is introduced by sampling the time variables at random. Data were analyzed using CphGxE under varying percentages of censoring (10%, 50%, and 90% censoring). Results are based on 100 simulations per scenario, with error bars representing the standard deviation across replicates.

B.7 Supplementary tables

Table B.1: **Empirical type I error rates for CphGxE with 50% censoring.** Synthetic traits were simulated with additive genetic, environmental exposure, and covariate effects. Genotype data with 15,000 SNPs were either simulated using synthetic biallelic independent genotypes (Random Genotypes) or taken from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled to sample sizes of 1k, 5k, 10k and 20k individuals. We randomly selected 10% of all variants to be causal with additive effects and we assume that they explain 40% of the frailty variance for each trait. The frailty variance due to environmental exposures is set to 10% of the total frailty variance. The remainder of the frailty variance is explained by the covariates and noise is introduced by sampling the time variables at random. Data were analyzed using CphGxE under varying percentages of censoring. Under the null hypothesis $H_0 : \sigma^2 = 0$. 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.

Dataset	Sample Size	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
Random Genotypes	1k	0.047 (0.212)	0.001 (0.098)	0.001 (0.037)
	5k	0.054 (0.226)	0.011 (0.108)	0.002 (0.044)
	10k	0.055 (0.227)	0.014 (0.116)	0.002 (0.039)
	20k	0.054 (0.225)	0.012 (0.111)	0.002 (0.040)
UK Biobank	1k	0.038 (0.191)	0.010 (0.101)	0.002 (0.044)
	5k	0.047 (0.211)	0.011 (0.105)	0.001 (0.033)
	10k	0.061 (0.240)	0.015 (0.120)	0.003 (0.053)
	20k	0.069 (0.253)	0.019 (0.135)	0.003 (0.053)

Table B.2: **Empirical type I error rates for CphGxE with 10% censoring.** Synthetic traits were simulated with additive genetic, environmental exposure, and covariate effects. Genotype data with 15,000 SNPs were either simulated using synthetic biallelic independent genotypes (Random Genotypes) or taken from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled to sample sizes of 1k, 5k, 10k and 20k individuals. We randomly selected 10% of all variants to be causal with additive effects and we assume that they explain 40% of the frailty variance for each trait. The frailty variance due to environmental exposures is set to 10% of the total frailty variance. The remainder of the frailty variance is explained by the covariates and noise is introduced by sampling the time variables at random. Data were analyzed using CphGxE under varying percentages of censoring. Under the null hypothesis $H_0 : \sigma^2 = 0$. 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.

Dataset	Sample Size	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
Random Genotypes	1k	0.110 (0.313)	0.039 (0.193)	0.01 (0.1)
	5k	0.059 (0.235)	0.014 (0.115)	0.003 (0.058)
	10k	0.052 (0.222)	0.012 (0.108)	0.002 (0.047)
	20k	0.054 (0.227)	0.014 (0.115)	0.002 (0.036)
UK Biobank	1k	0.157 (0.364)	0.075 (0.264)	0.028 (0.164)
	5k	0.080 (0.271)	0.022 (0.146)	0.004 (0.066)
	10k	0.063 (0.244)	0.018 (0.131)	0.002 (0.046)
	20k	0.097 (0.196)	0.033 (0.178)	0.007 (0.081)

Table B.3: **Area under the Receiver Operating Characteristic (ROC) curves (AUC) comparing the ability of CphGxE to correctly prioritize gene-by-environment interaction variants under the alternative model.** Synthetic time-to-event traits were simulated with additive genetic, environmental exposure, covariate, and gene-by-environment interaction effects. Genotype data with 15,000 SNPs were either simulated using synthetic biallelic independent genotypes (Random Genotypes) or taken from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled to 20k individuals. We randomly selected 10% of all variants to be causal with additive effects set at 30% of the frailty variance explained (FVE) for each trait. The frailty variance due to environmental exposures is set to 10% of the total frailty variance. ROC curves are computed under simulations as function of the per SNP gene-by-environment interaction variance of the frailty $\sigma_{G \times E}^2$ (1%, 2%, and 5%). The remainder of the frailty variance is explained by the covariates and noise is introduced by sampling the time variables at random. Data were analyzed using CphGxE under varying percentages of censoring (10%, 50%, and 90% censoring). Results are based on 100 simulations per scenario. Values correspond to the AUC computed from the aggregated ROC curves used in Figure 2.3.

Dataset	Censoring	$\sigma_{G \times E}^2 = 0.01$	$\sigma_{G \times E}^2 = 0.02$	$\sigma_{G \times E}^2 = 0.05$
Random Genotypes	10%	1.000	1.000	1.000
	50%	0.994	1.000	1.000
	90%	0.689	0.883	0.999
UK Biobank	10%	0.828	0.980	1.000
	50%	0.700	0.890	0.995
	90%	0.504	0.574	0.797

APPENDIX C

Supplementary information for Chapter 3

C.1 Supplementary figures

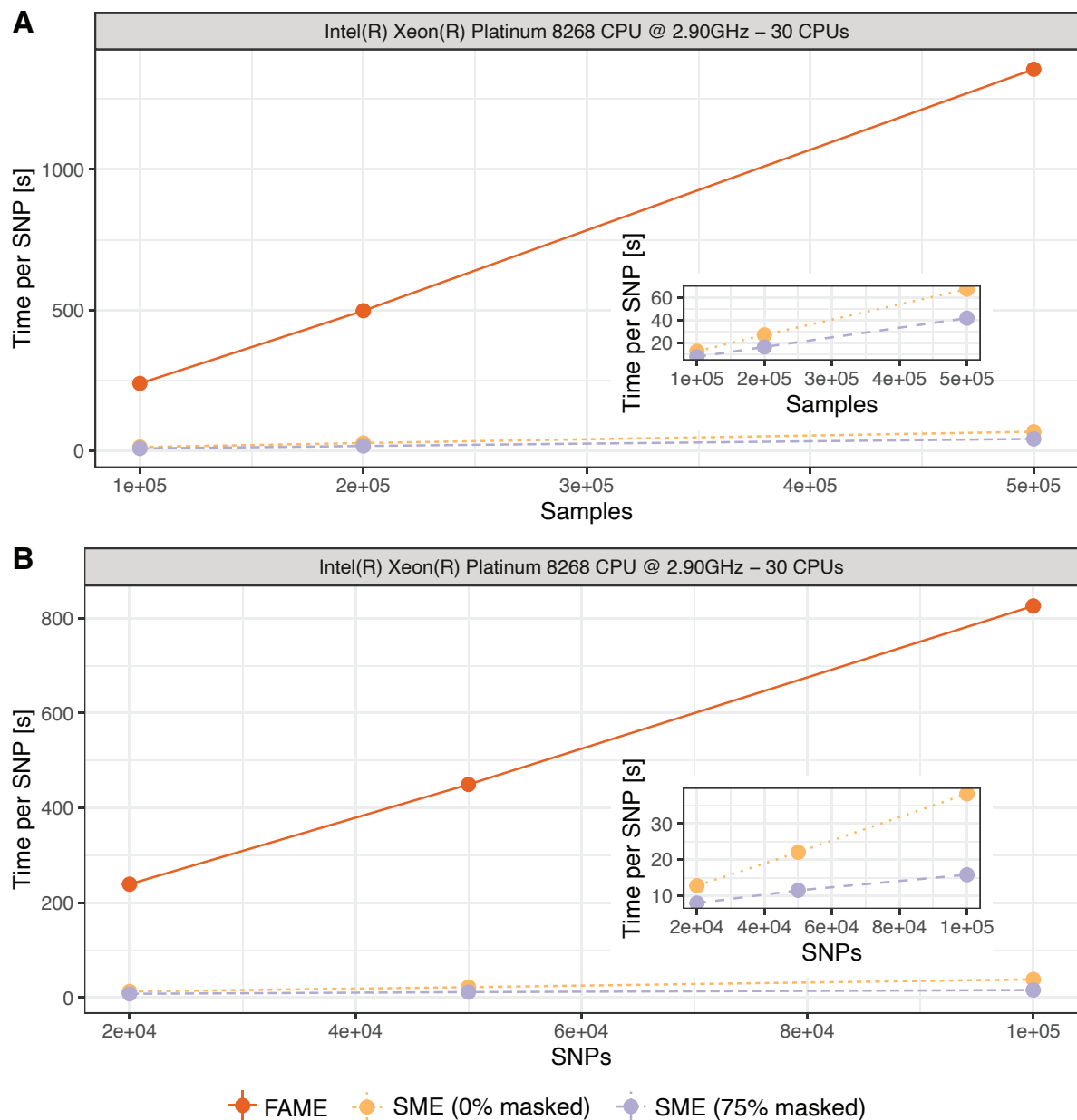


Figure C.2: **Computational time needed to run a single test with SME and FAME (Fu et al. 2023) as a function of sample size and total number of SNPs.** The solid orange line represents the runtime using FAME. The dotted yellow line shows the runtime of SME without a mask. The dashed purple line represents 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.

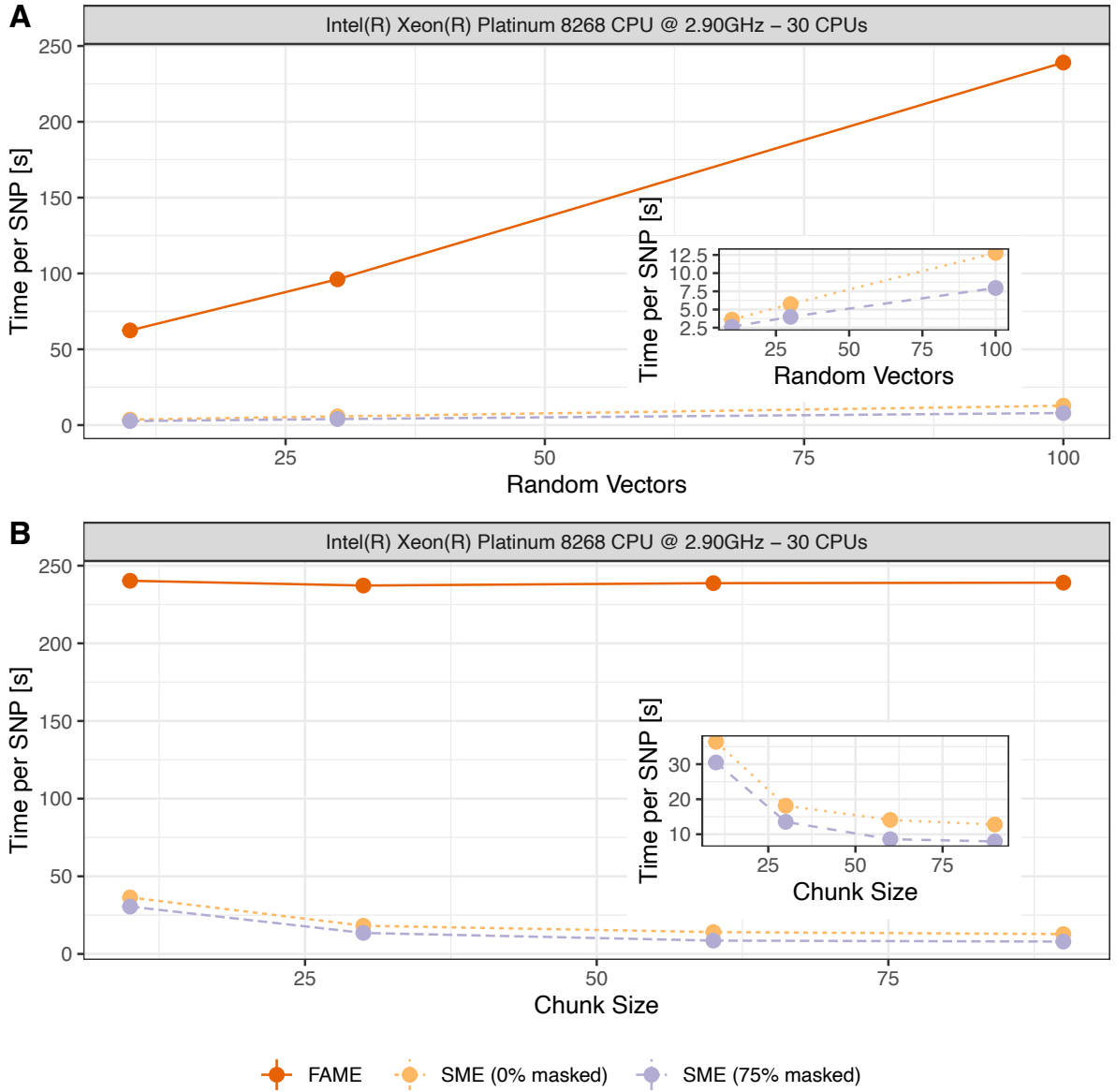


Figure C.3: Computational time needed to run a single test with SME and FAME (Fu et al. 2023) as a function of the number of random vectors and number of SNPs sharing random vectors. The solid orange line represents the runtime using the command-line tool FAME. The dotted yellow line shows the runtime 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.

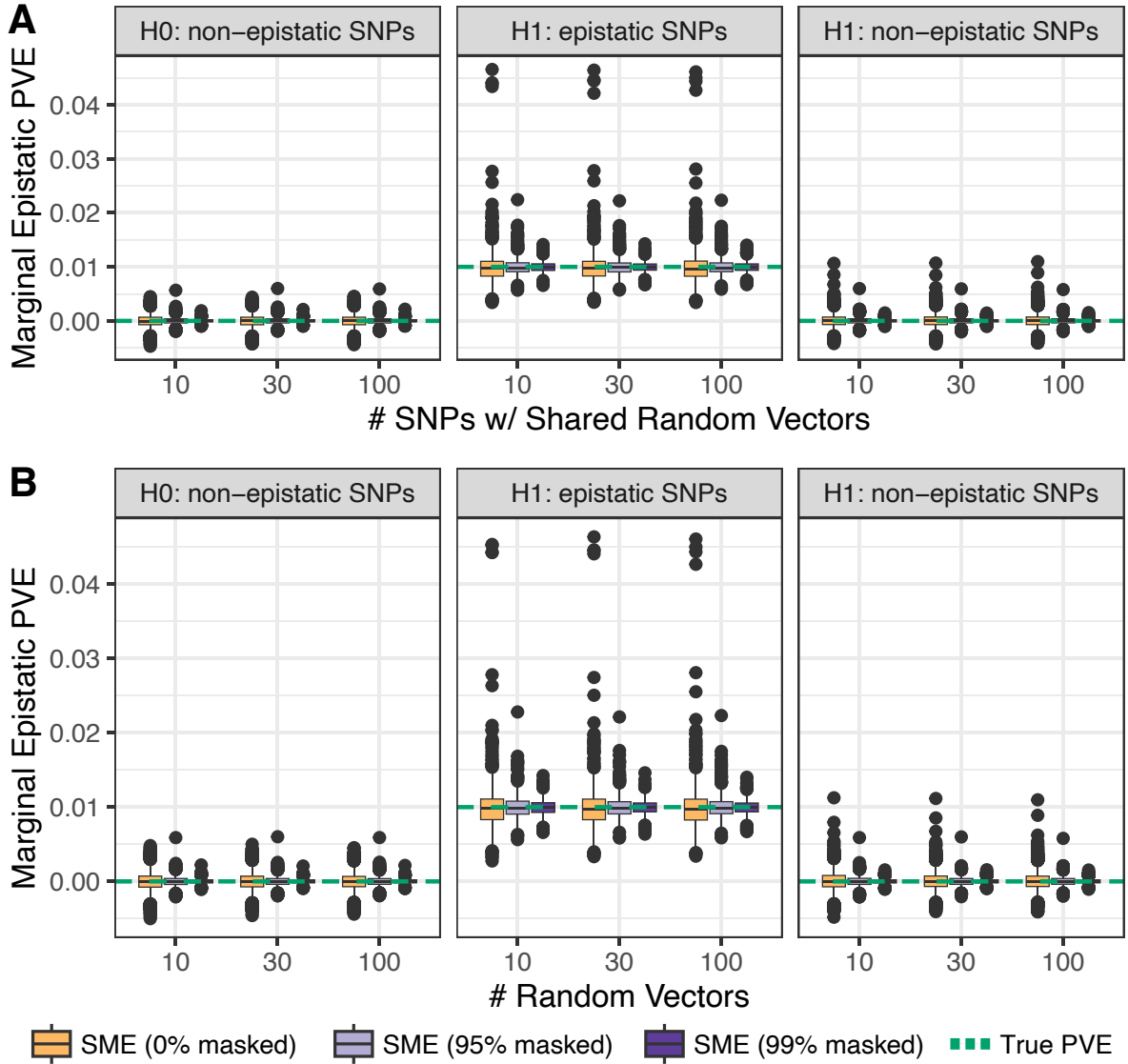


Figure C.4: The SME stochastic method-of-moments algorithm produces estimates that are robust to the choice of number of random vectors and the number of SNPs sharing random vectors. Synthetic traits were simulated under the null hypothesis (H_0) of no epistasis and the alternative hypothesis (H_1) using 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. C.1). 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.

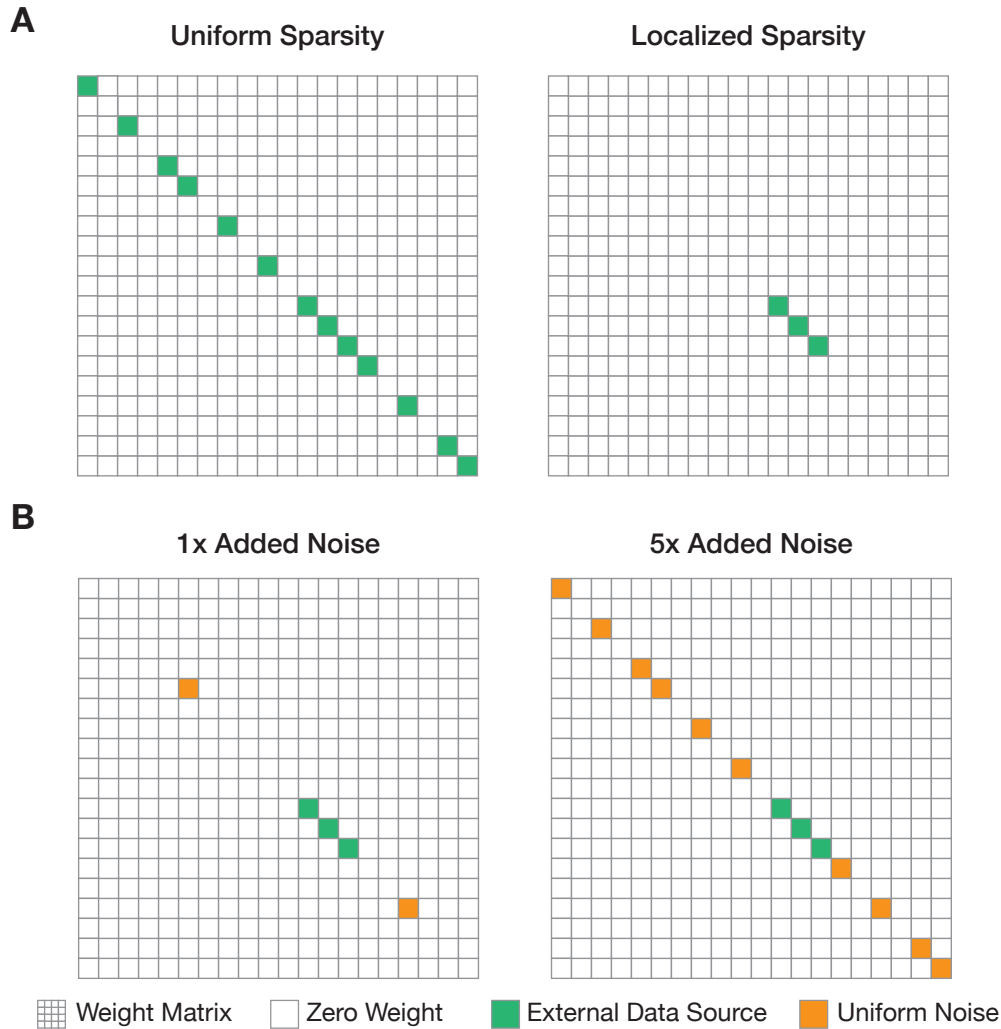


Figure C.5: **Schematic overview illustrating weight matrices with uniform and localized sparsity.** (A) The simulations consider two scenarios that result in different distributions of sparsely modeled gene-interactions. The weight matrices are binary everywhere except for places on the diagonal where 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× and 5× of initially disregarded interactions.

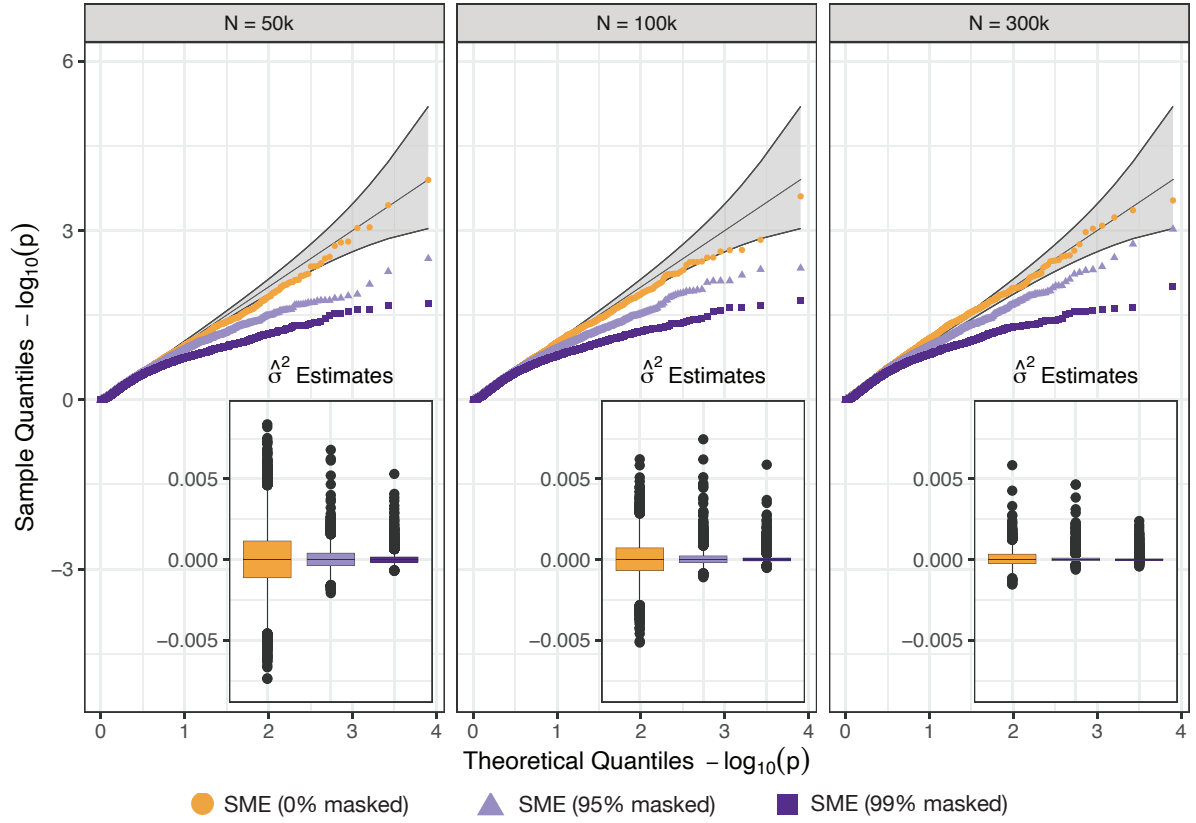


Figure C.6: **While using a mask that induces localized sparsity, SME is negatively biased under the null hypothesis.** Synthetic traits were simulated with only additive effects using chromosome 1 from individuals of self-identified European ancestry in the UK Biobank. These data were then subsampled 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.

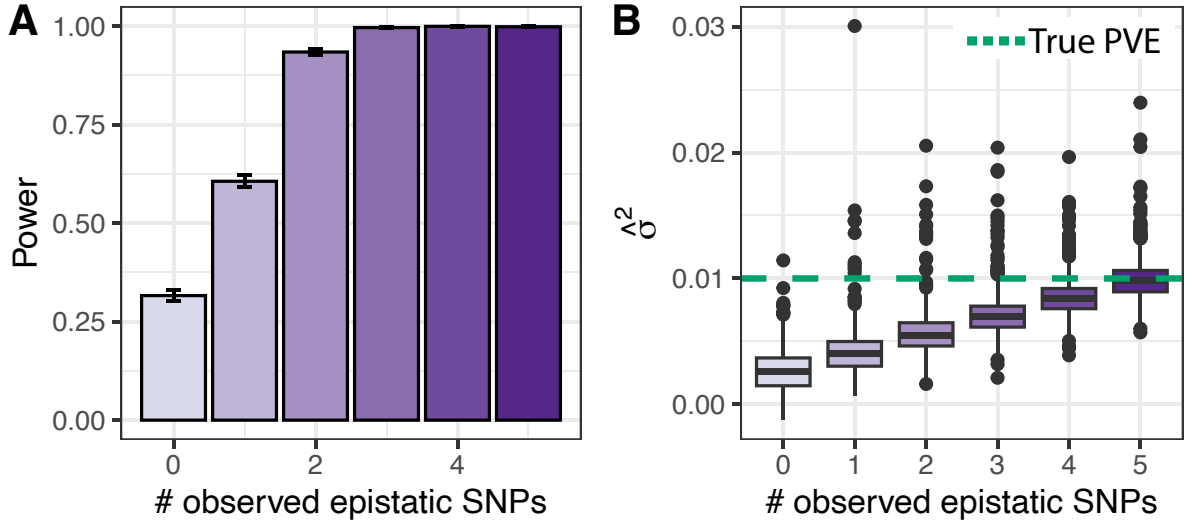


Figure C.7: Consequence of having a misspecified mask on power and variance component estimates in SME. Synthetic traits were simulated using chromosome 1 from individuals of self-identified European ancestry in the UK Biobank. Data were subsampled to 100k individuals. We randomly selected 10% of all 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).

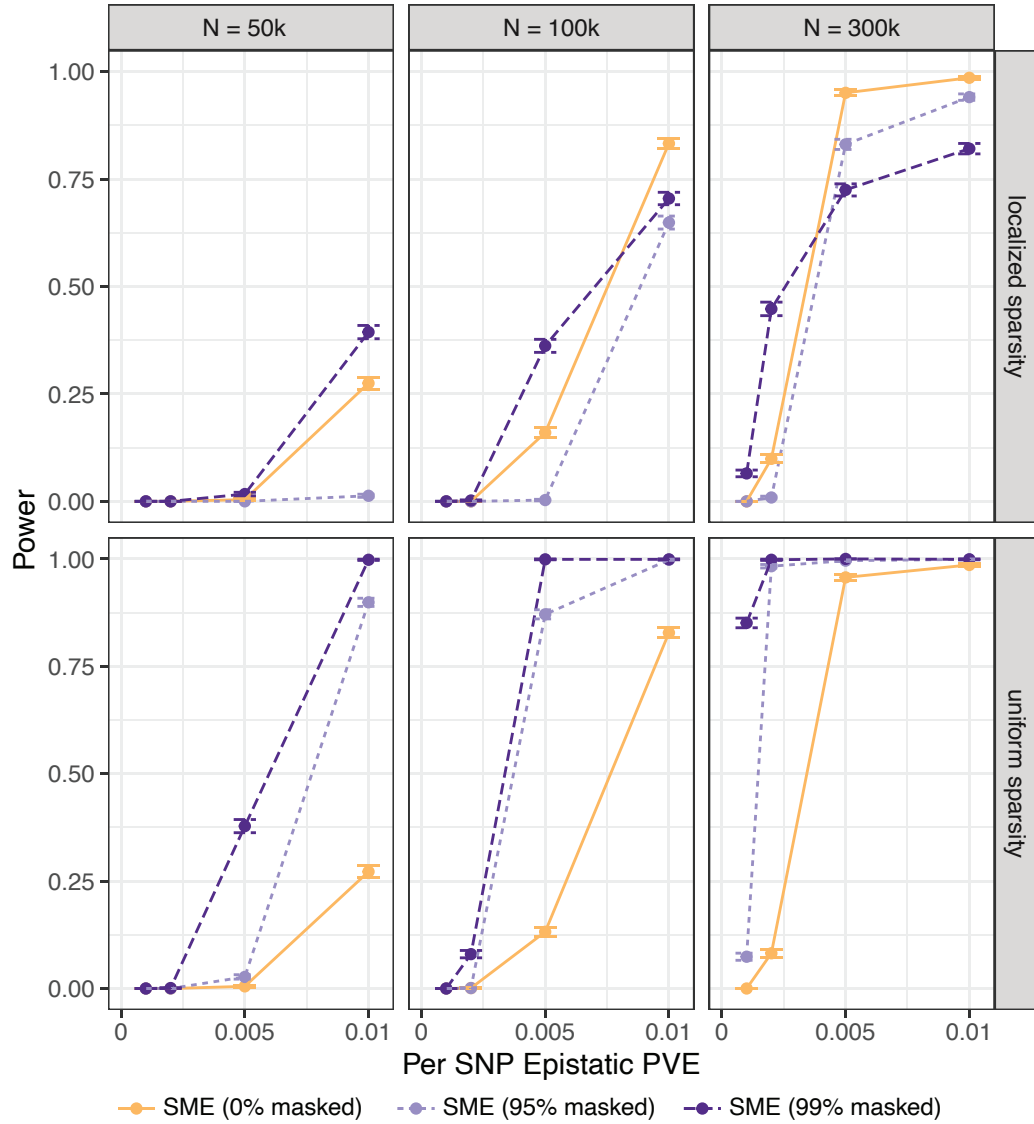


Figure C.8: **SME exhibits less power when using a mask that induces localized sparsity.** Synthetic traits were simulated with both additive and pairwise epistatic effects using chromosome 1 from individuals of self-identified European ancestry in the UK Biobank. Data were subsampled using sample sizes 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.

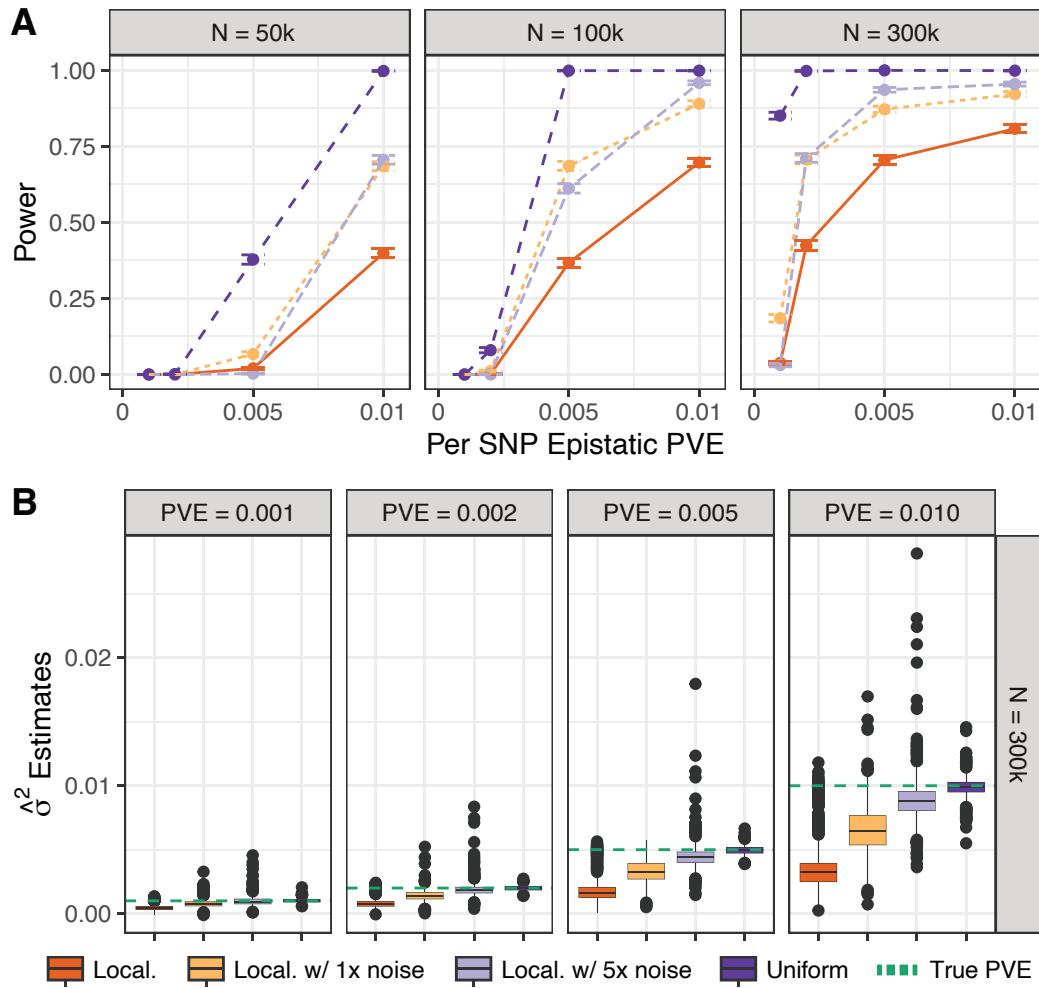


Figure C.9: Adding random noise to external data sources that induce localized sparsity recovers empirical power and reduces the negative bias in variance component estimates. Synthetic traits were simulated with both additive and pairwise epistatic effects using chromosome 1 from individuals of self-identified 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. C.5). **(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.

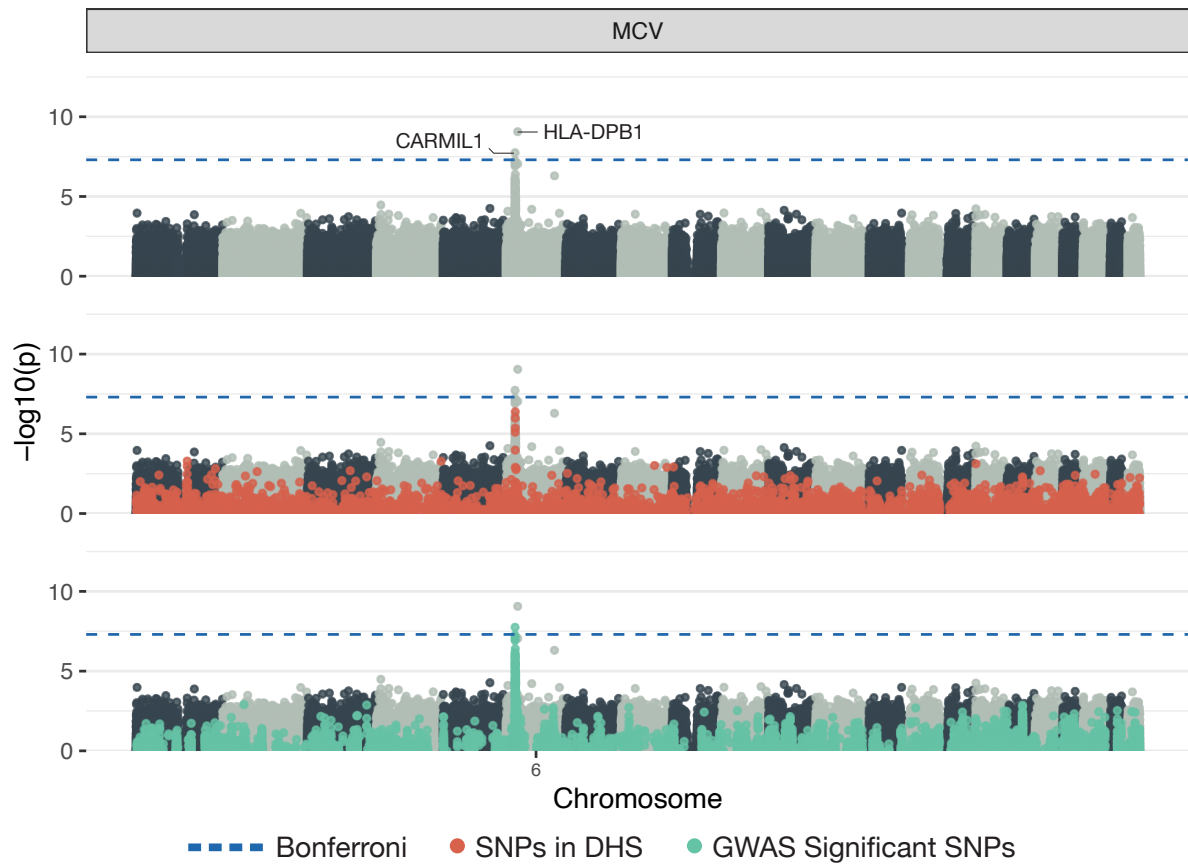


Figure C.10: **Manhattan plots of a genome-wide interaction analysis using SME to study mean corpuscular volume (MCV) assayed in individuals in the UK Biobank.** As a mask in this study, we leveraged DNase I-hypersensitive sites (DHS) data measured over 12 days of *ex vivo* erythroid differentiation (Georgolopoulos et al. 2021; Maurano et al. 2012). 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.

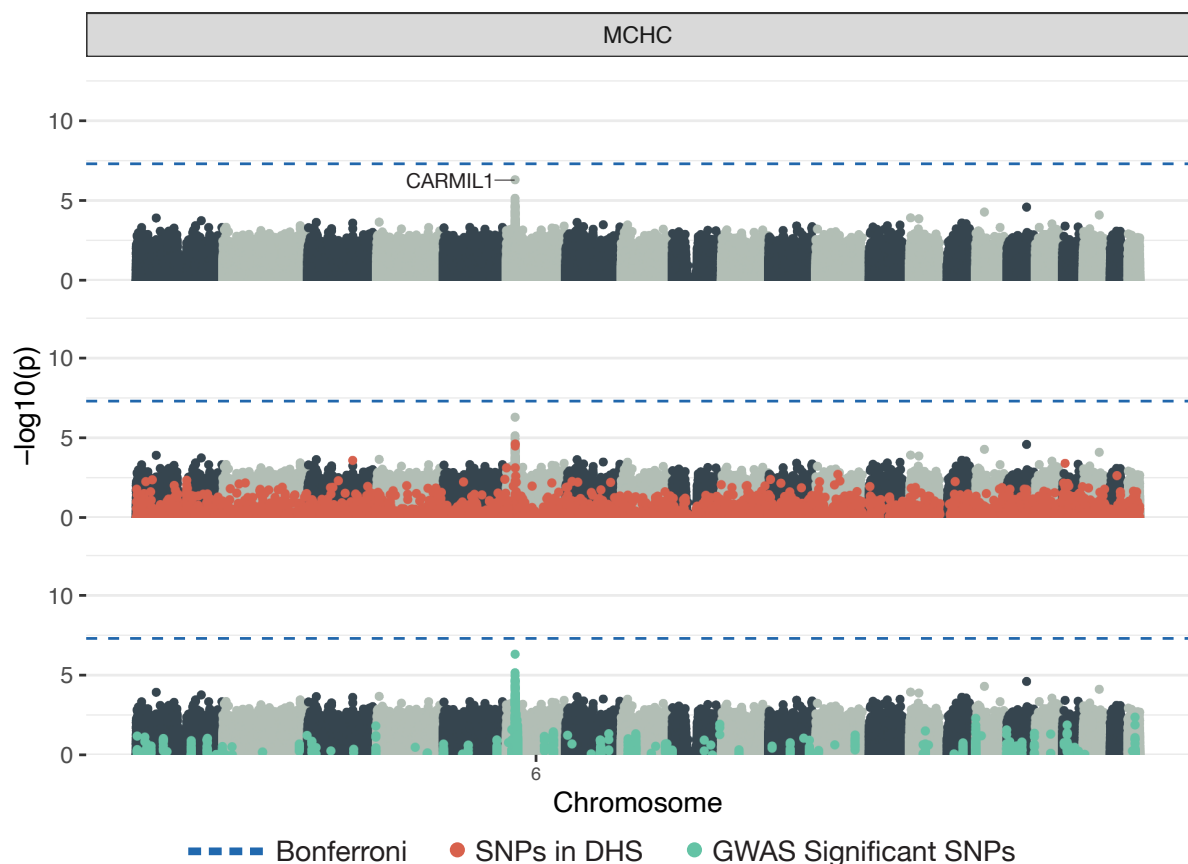


Figure C.11: **Manhattan plots of a genome-wide interaction analysis using SME to study mean corpuscular hemoglobin concentration (MCHC) assayed in individuals in the UK Biobank.** As a mask in this study, we leveraged DNase I-hypersensitive sites (DHS) data measured over 12 days of *ex vivo* erythroid differentiation (Georgolopoulos et al. 2021; Maurano et al. 2012). 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.

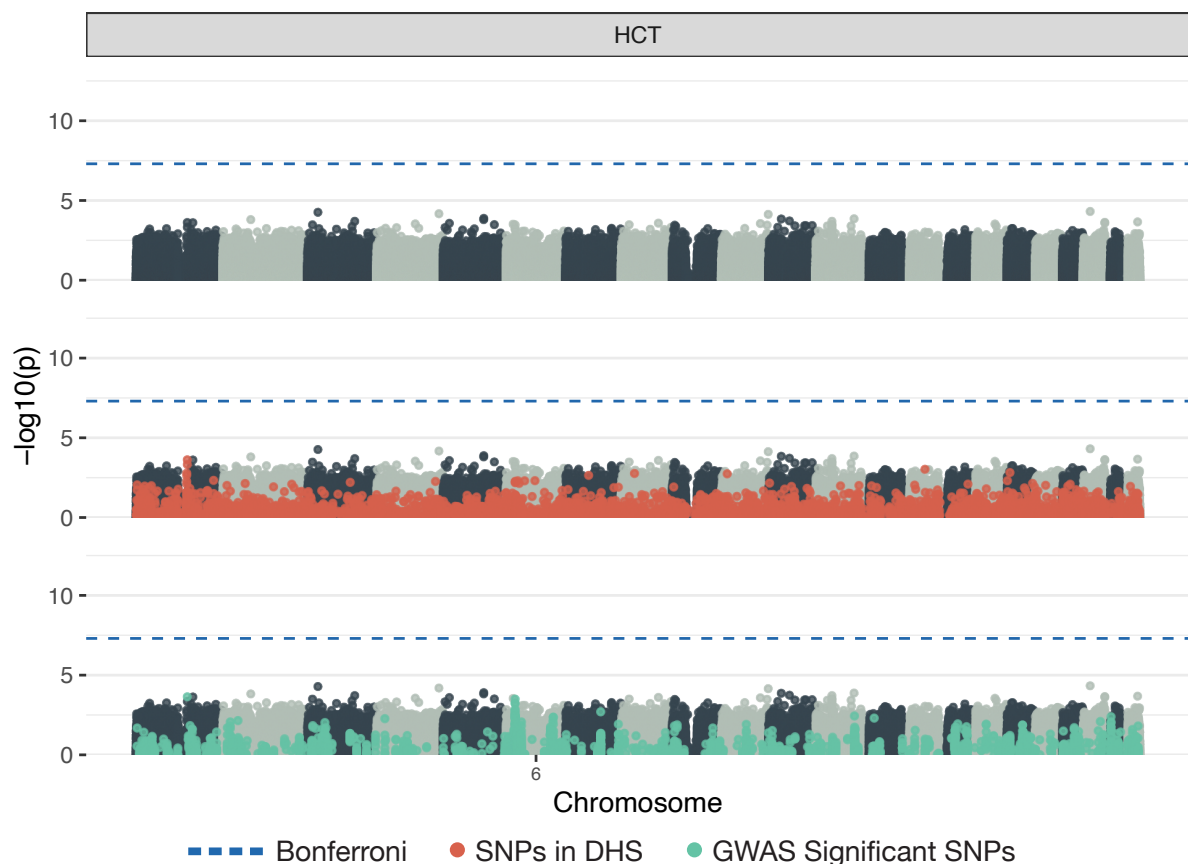


Figure C.12: **Manhattan plots of a genome-wide interaction analysis using SME to study hematocrit (HCT) assayed in individuals in the UK Biobank.** As a mask in this study, we leveraged DNase I-hypersensitive sites (DHS) data measured over 12 days of *ex vivo* erythroid differentiation (Georgolopoulos et al. 2021; Maurano et al. 2012). 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.

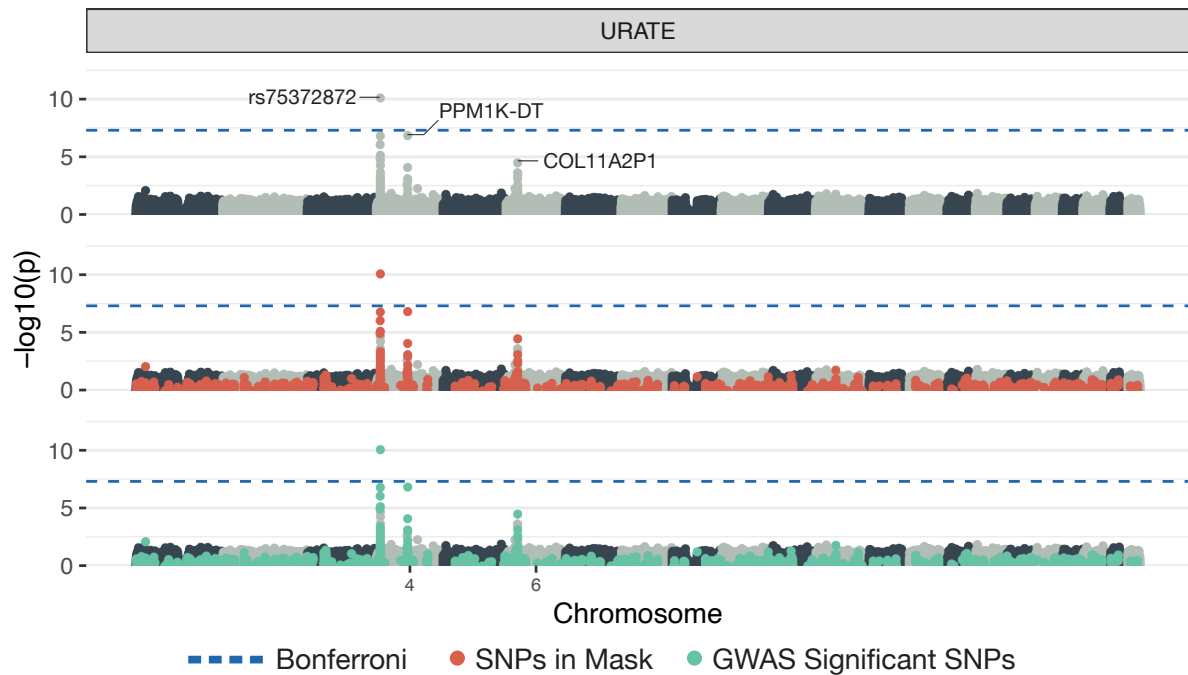


Figure C.13: **Manhattan plots of a genome-wide interaction analysis using SME to study uric acid (urate) assayed in individuals in the UK Biobank.** As a mask in this study, we leveraged 3,536 significant GWAS variants. This means that, while all SNPs are tested for marginal epistasis, only their interactions 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.

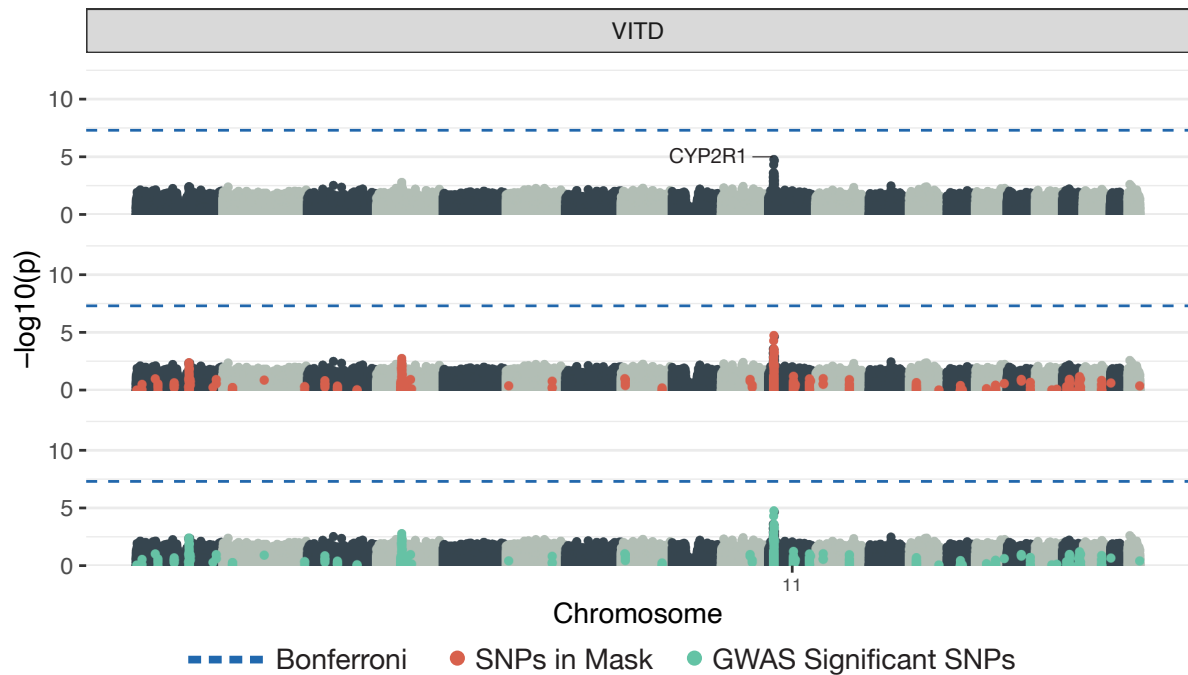


Figure C.14: **Manhattan plots of a genome-wide interaction analysis using SME to study vitamin D levels (VITD) assayed in individuals in the UK Biobank.** As a mask in this study, we leveraged 547 significant GWAS variants. This means that, while all SNPs are tested for marginal epistasis, only their 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.

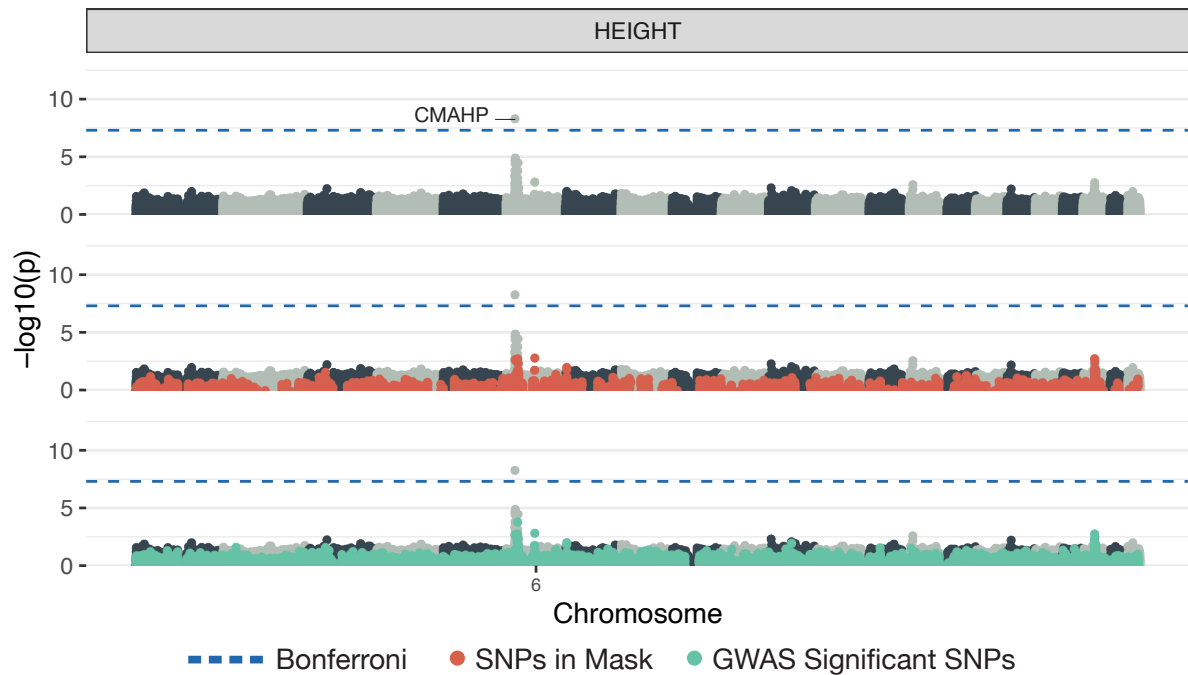


Figure C.15: **Manhattan plots of a genome-wide interaction analysis using SME to study body height assayed in individuals in the UK Biobank.** As a mask in this study, we leveraged the top 5,000 significant GWAS variants ranked by strength of association (i.e., lowest P -values). This means that, while 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.

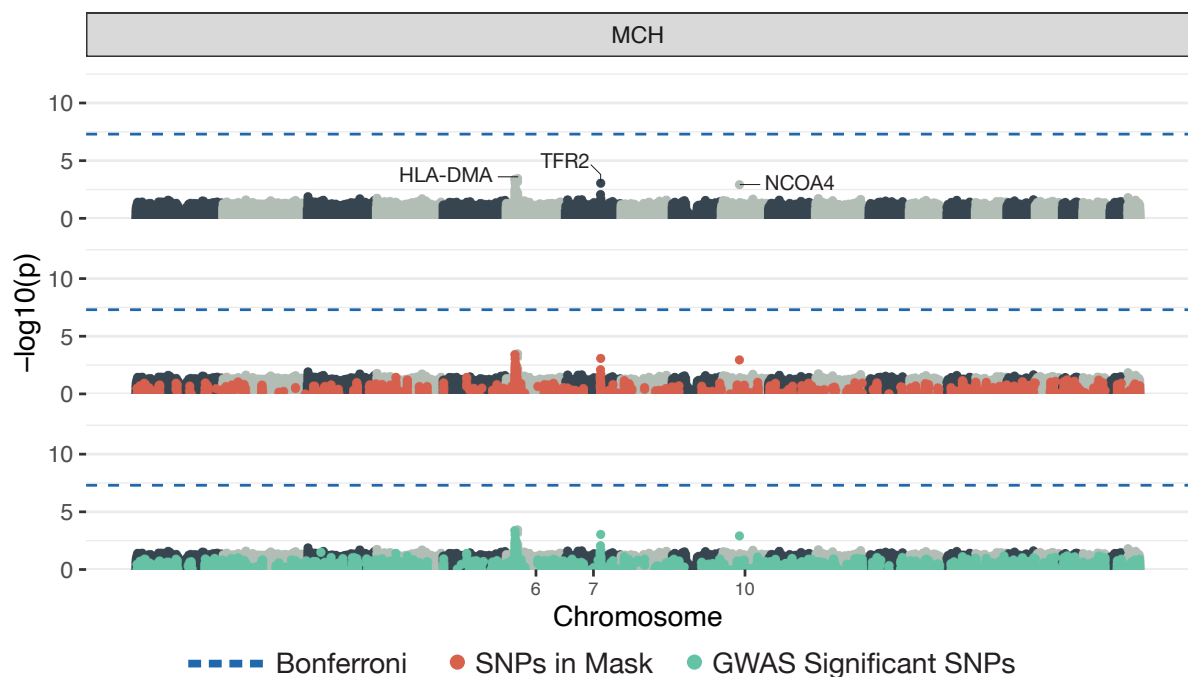


Figure C.16: **Manhattan plots of a genome-wide interaction analysis using SME to study mean corpuscular hemoglobin (MCH) assayed in individuals in the UK Biobank.** As a mask in this study, we leveraged the top 5,000 significant GWAS variants ranked by strength of association (i.e., lowest P -values). 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.

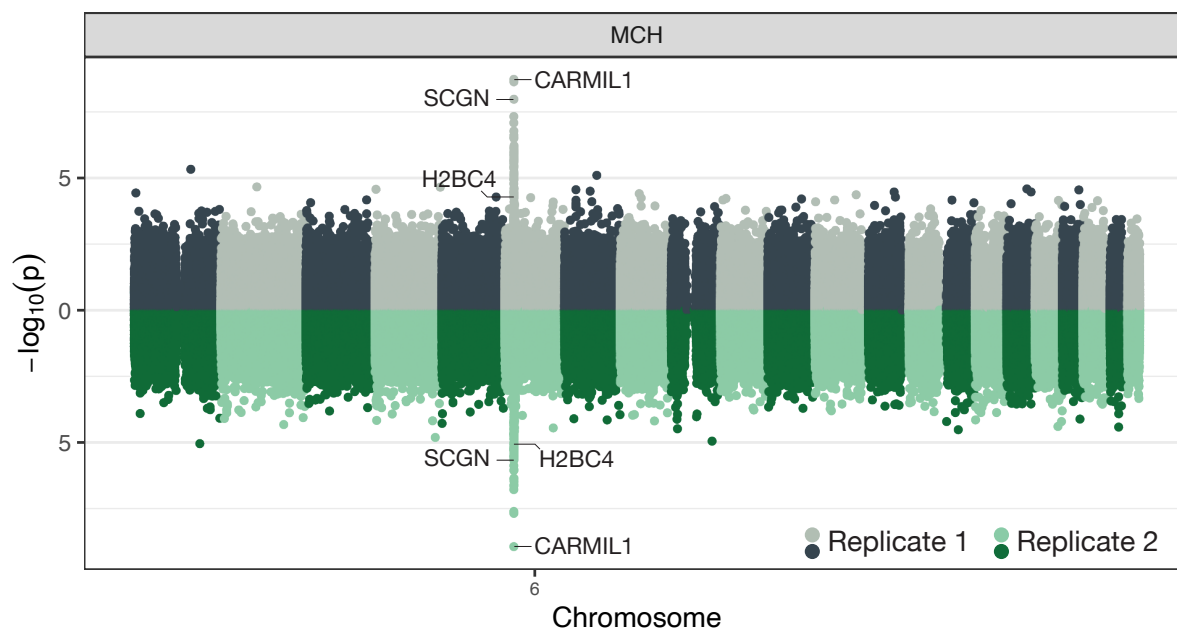


Figure C.17: Manhattan plots of a genome-wide interaction split-half analysis using SME to study mean corpuscular hemoglobin (MCH) assayed in two distinct subsets of 174,705 individuals each taken from the UK Biobank. As a mask in this study, we leveraged DNase I-hypersensitive sites (DHS) data measured over 12 days of *ex vivo* erythroid differentiation (Georgolopoulos et al. 2021; Maurano et al. 2012). 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.

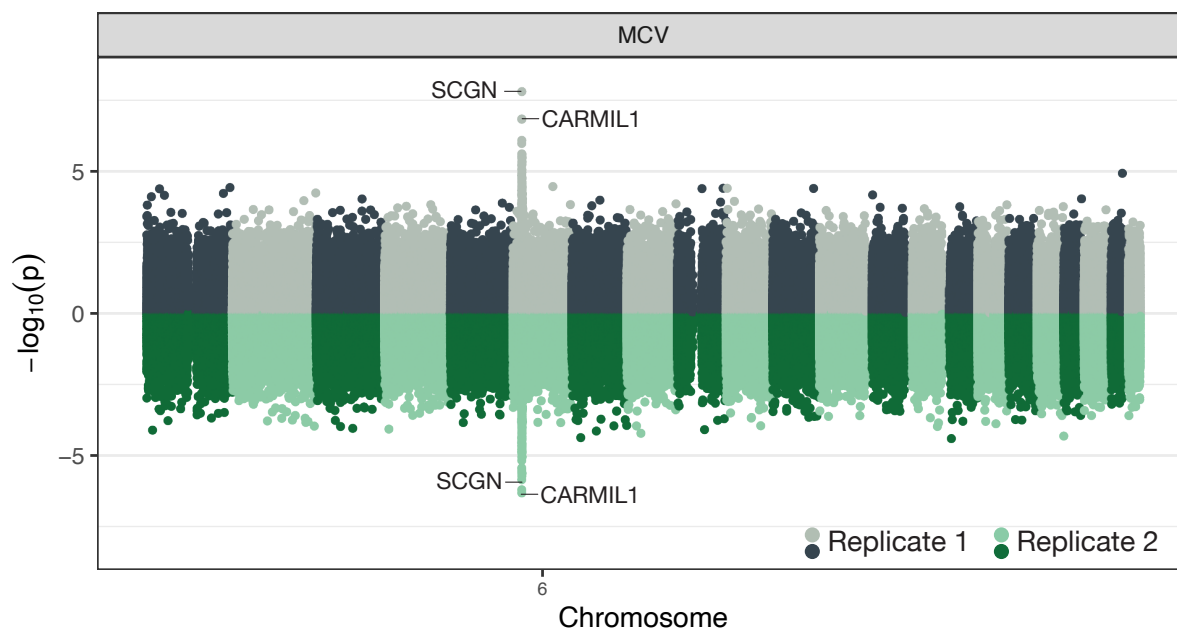


Figure C.18: Manhattan plots of a genome-wide interaction split-half analysis using SME to study mean corpuscular volume (MCV) assayed in two distinct subsets of 174,705 individuals each taken from the UK Biobank. As a mask in this study, we leveraged DNase I-hypersensitive sites (DHS) data measured over 12 days of *ex vivo* erythroid differentiation (Georgolopoulos et al. 2021; Maurano et al. 2012). 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.

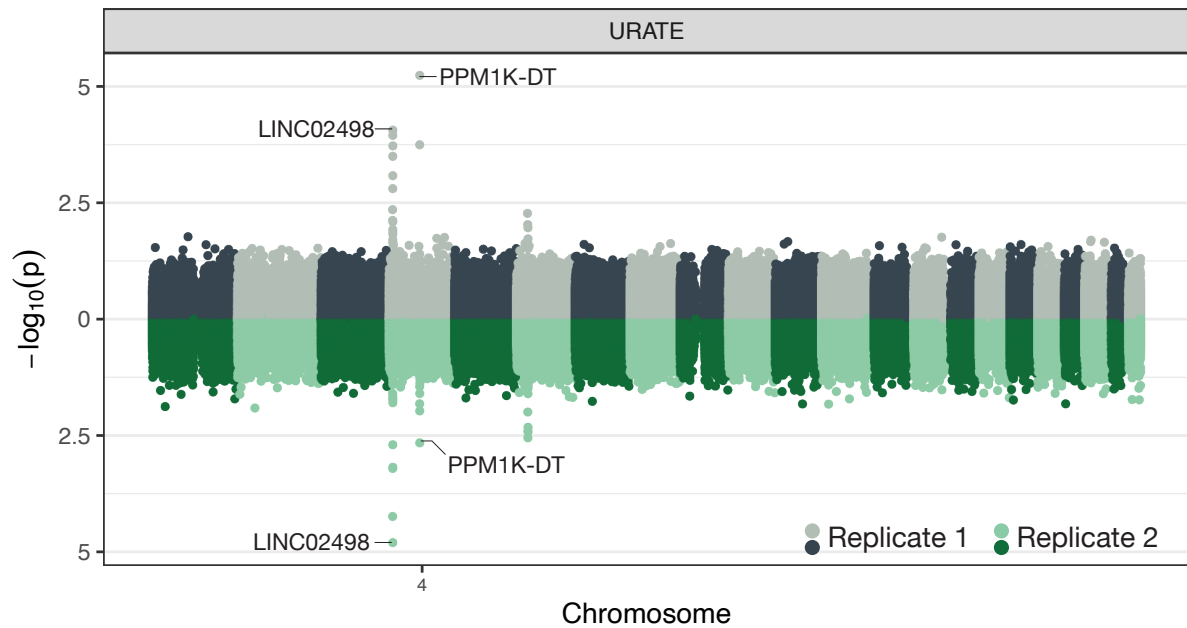


Figure C.19: **Manhattan plots of a genome-wide interaction split-half analysis using SME to study uric acid (urate) assayed in two distinct subsets of 174,705 individuals each taken from the UK Biobank.** As a mask in this study, we leveraged GWAS summary statistics reported in the UK Biobank. This 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.

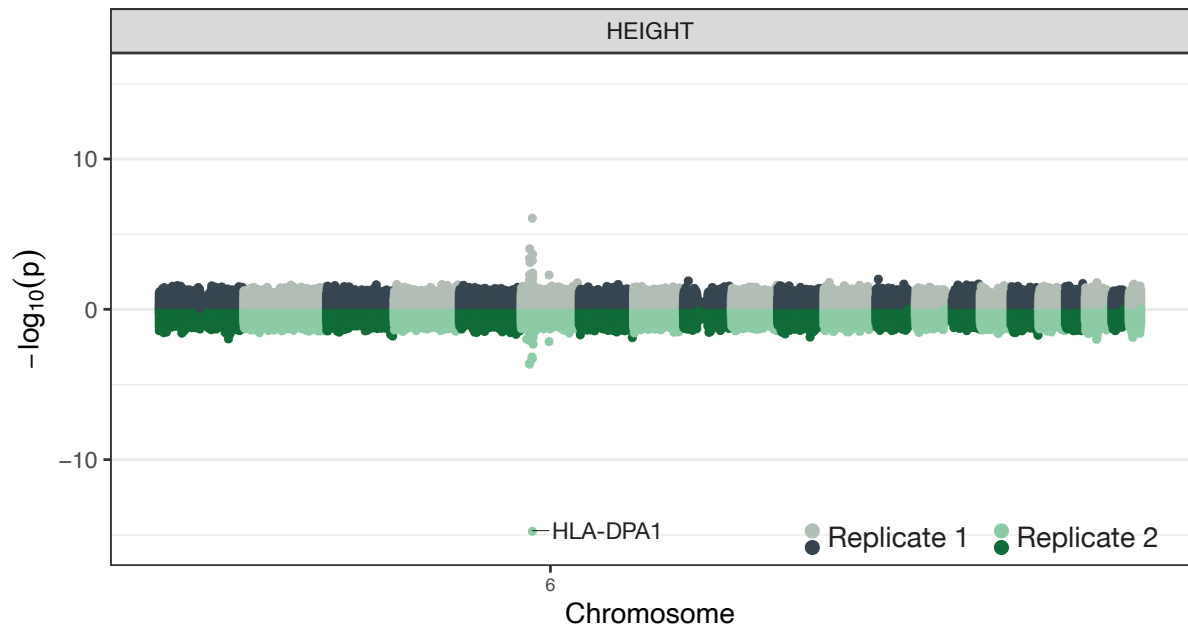


Figure C.20: **Manhattan plots of a genome-wide interaction split-half analysis using SME to study body height assayed in two distinct subsets of 174,705 individuals each taken from the UK Biobank.** As a mask in this study, we leveraged GWAS summary statistics reported in the UK Biobank. This means 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.

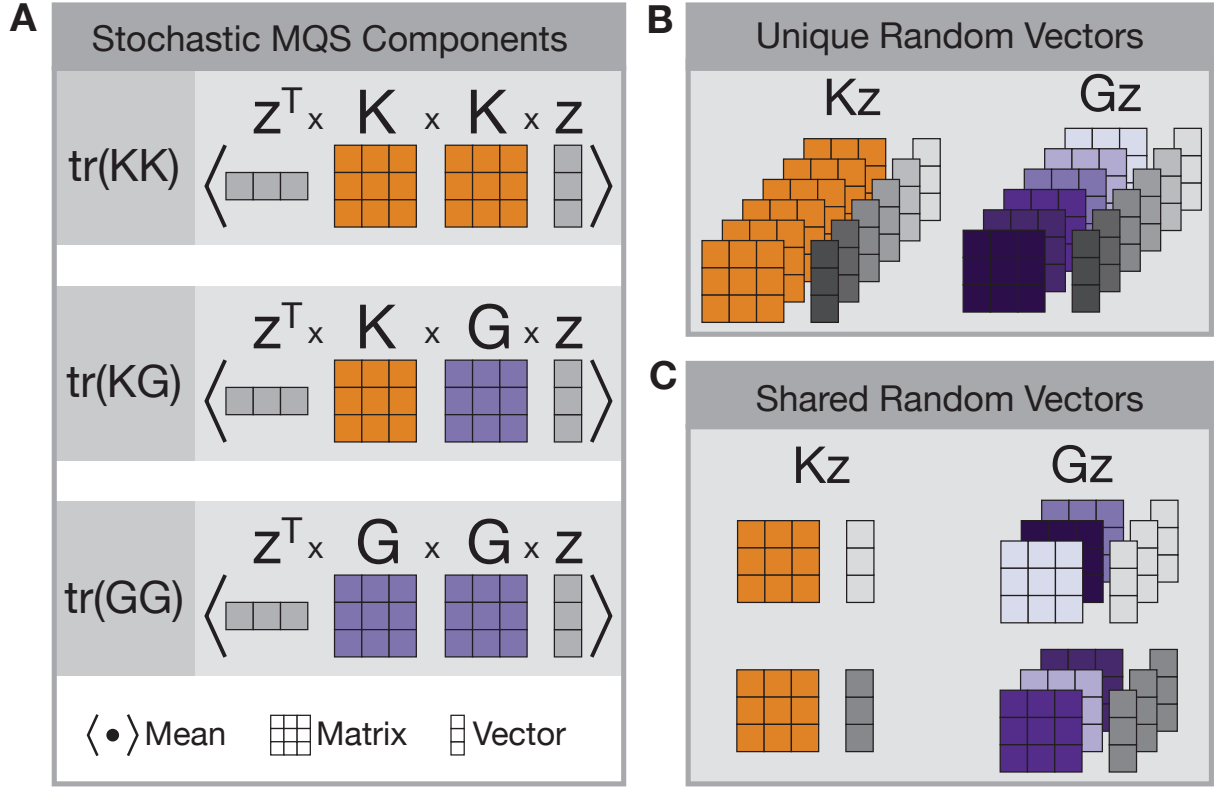


Figure C.1: **Schematic overview illustrating the approximation to the stochastic trace estimator used in SME.** (A) Computing the exact trace of a product for two covariance matrices is computationally infeasible for large studies with many individuals. SME overcomes this limitation by computing point 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 $\mathbf{Kz}$.

C.2 Supplementary tables

Table C.1: **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. 3.3 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 C.2: **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	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
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 C.3: **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	<i>CMAHP</i>	(Yengo et al. 2022)

Table C.4: 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 (Georgolopoulos et al. 2021; Maurano et al. 2012). 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 C.5: 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 (Georgolopoulos et al. 2021; Maurano et al. 2012). 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 C.6: 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 (Georgolopoulos et al. 2021; Maurano et al. 2012). 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 C.7: 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 (Georgolopoulos et al. 2021; Maurano et al. 2012). 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 C.8: 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 C.9: 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 C.10: 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 C.11: 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 C.12: **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 (Georgolopoulos et al. 2021; Maurano et al. 2012). 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

Table C.13: **SME identifies marginal epistasis in hematology 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: mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV), and hematocrit (HCT). As a mask, we leveraged DNase I-hypersensitive sites (DHS) data measured over 12 days of *ex vivo* erythroid differentiation (Georgolopoulos et al. 2021; Maurano et al. 2012). 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 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 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 only applied to a random subset of 10k individuals.

Trait	ID	Coordinates	Ref.
MCH	rs4711092	6:25684405	(Qin et al. 2020; Timoteo et al. 2023; Bauer et al. 2011)
MCH	rs9366624	6:25439492	(Edwards et al. 2013; Yang et al. 2005; Ray et al. 2022; Timoteo et al. 2023; Ding et al. 2012)
MCH	rs9461167	6:25418571	(Edwards et al. 2013; Yang et al. 2005; Ray et al. 2022; Timoteo et al. 2023; Ding et al. 2012)
MCH	rs9379764	6:25414023	(Edwards et al. 2013; Yang et al. 2005; Ray et al. 2022; Timoteo et al. 2023; Ding et al. 2012; Plotnikov et al. 2022; Galesloot et al. 2016)
MCH	rs441460	6:25548288	(Edwards et al. 2013; Yang et al. 2005; Ray et al. 2022; Timoteo et al. 2023; Ding et al. 2012)
MCH	rs198834	6:26114372	(Vuckovic et al. 2020; Zhang et al. 2008)
MCH	rs13203202	6:25582771	(Edwards et al. 2013; Yang et al. 2005; Ray et al. 2022; Timoteo et al. 2023; Ding et al. 2012)
MCV	rs9276	6:33053577	—
MCV	rs9366624	6:25439492	(Edwards et al. 2013; Yang et al. 2005; Ray et al. 2022; Timoteo et al. 2023; Ding et al. 2012)

Bibliography

- Abdellaoui, A. et al. (2023). “15 years of GWAS discovery: Realizing the promise”. In: *The American Journal of Human Genetics* 110.2, pp. 179–194. DOI: [10.1016/j.ajhg.2022.12.011](https://doi.org/10.1016/j.ajhg.2022.12.011).
- All of Us Research Program Investigators et al. (2019). “The ”All of Us” Research Program”. In: *The New England Journal of Medicine* 381.7, pp. 668–676. DOI: [10.1056/NEJMs1809937](https://doi.org/10.1056/NEJMs1809937).
- Balvert, M. et al. (2024). “Considerations in the search for epistasis”. In: *Genome Biology* 25.1, p. 296. DOI: [10.1186/s13059-024-03427-z](https://doi.org/10.1186/s13059-024-03427-z).
- Baranger, D. A. A. et al. (2023). “Multi-omics cannot replace sample size in genome-wide association studies”. In: *Genes, Brain and Behavior* 22.6, e12846. DOI: [10.1111/gbb.12846](https://doi.org/10.1111/gbb.12846).
- Bauer, M. C. et al. (2011). “Identification of a high-affinity network of secretagoin-binding proteins involved in vesicle secretion”. In: *Molecular BioSystems* 7.7, pp. 2196–2204. DOI: [10.1039/COMB00349B](https://doi.org/10.1039/COMB00349B).

- Bayat, A. et al. (2021). “Fast and accurate exhaustive higher-order epistasis search with BitEpi”. In: *Scientific Reports* 11.1, p. 15923. DOI: [10.1038/s41598-021-94959-y](https://doi.org/10.1038/s41598-021-94959-y).
- Berendse, K. et al. (2019). “Liver disease predominates in a mouse model for mild human Zellweger spectrum disorder”. In: *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1865.10, pp. 2774–2787. DOI: [10.1016/j.bbadis.2019.06.013](https://doi.org/10.1016/j.bbadis.2019.06.013).
- Berisa, T. and Pickrell, J. K. (2016). “Approximately independent linkage disequilibrium blocks in human populations”. In: *Bioinformatics* 32.2, pp. 283–285. DOI: [10.1093/bioinformatics/btv546](https://doi.org/10.1093/bioinformatics/btv546).
- 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”. In: *F1000Research* 9, F1000 Faculty Rev. DOI: [10.12688/f1000research.26648.1](https://doi.org/10.12688/f1000research.26648.1).
- Bick, A. G. et al. (2024). “Genomic data in the All of Us Research Program”. In: *Nature* 627.8003, pp. 340–346. DOI: [10.1038/s41586-023-06957-x](https://doi.org/10.1038/s41586-023-06957-x).
- Blake, J. A. et al. (2020). “Mouse Genome Database (MGD): Knowledgebase for mouse–human comparative biology”. In: *Nucleic Acids Research* 49.D1, pp. D981–D987. DOI: [10.1093/nar/gkaa1083](https://doi.org/10.1093/nar/gkaa1083).
- Bloom, J. S. et al. (2013). “Finding the sources of missing heritability in a yeast cross”. In: *Nature* 494.7436, pp. 234–237. DOI: [10.1038/nature11867](https://doi.org/10.1038/nature11867).
- Boix, C. A. et al. (2021). “Regulatory genomic circuitry of human disease loci by integrative epigenomics”. In: *Nature* 590.7845, pp. 300–307. DOI: [10.1038/s41586-020-03145-z](https://doi.org/10.1038/s41586-020-03145-z).
- Boyle, E. A., Li, Y. I., and Pritchard, J. K. (2017). “An Expanded View of Complex Traits: From Polygenic to Omnigenic”. In: *Cell* 169.7, pp. 1177–1186. DOI: [10.1016/j.cell.2017.05.038](https://doi.org/10.1016/j.cell.2017.05.038).
- Brem, R. B. et al. (2005). “Genetic interactions between polymorphisms that affect gene expression in yeast”. In: *Nature* 436.7051, pp. 701–703. DOI: [10.1038/nature03865](https://doi.org/10.1038/nature03865).
- Breslow, N. E. and Clayton, D. G. (1993). “Approximate Inference in Generalized Linear Mixed Models”. In: *Journal of the American Statistical Association* 88.421, pp. 9–25. DOI: [10.1080/01621459.1993.10594284](https://doi.org/10.1080/01621459.1993.10594284).

- Brown, A. A. et al. (2014). “Genetic interactions affecting human gene expression identified by variance association mapping”. In: *eLife* 3, e01381. DOI: [10.7554/eLife.01381](https://doi.org/10.7554/eLife.01381).
- Bulik-Sullivan, B. K. et al. (2015). “LD Score regression distinguishes confounding from polygenicity in genome-wide association studies”. In: *Nature Genetics* 47.3, pp. 291–295. DOI: [10.1038/ng.3211](https://doi.org/10.1038/ng.3211).
- Burton, P. R. et al. (2007). “Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls”. In: *Nature* 447.7145, pp. 661–678. DOI: [10.1038/nature05911](https://doi.org/10.1038/nature05911).
- Bycroft, C. et al. (2018). “The UK Biobank resource with deep phenotyping and genomic data”. In: *Nature* 562.7726, pp. 203–209. DOI: [10.1038/s41586-018-0579-z](https://doi.org/10.1038/s41586-018-0579-z).
- Campbell, M., Walia, H., and Morota, G. (2018). “Utilizing random regression models for genomic prediction of a longitudinal trait derived from high-throughput phenotyping”. In: *Plant Direct* 2.9, e00080. DOI: [10.1002/pld3.80](https://doi.org/10.1002/pld3.80).
- Cano-Gamez, E. and Trynka, G. (2020). “From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases”. In: *Frontiers in Genetics* 11. DOI: [10.3389/fgene.2020.00424](https://doi.org/10.3389/fgene.2020.00424).
- Carlborg, Ö. et al. (2006). “Epistasis and the release of genetic variation during long-term selection”. In: *Nature Genetics* 38.4, pp. 418–420. DOI: [10.1038/ng1761](https://doi.org/10.1038/ng1761).
- Castrignanò, T. et al. (2020). “ELIXIR-IT HPC@CINECA: high performance computing resources for the bioinformatics community”. In: *BMC Bioinformatics* 21.10, p. 352. DOI: [10.1186/s12859-020-03565-8](https://doi.org/10.1186/s12859-020-03565-8).
- Castro-Ochoa, K. F., Guerrero-Fonseca, I. M., and Schnoor, M. (2019). “Hematopoietic cell-specific lyn substrate (HCLS1 or HS1): A versatile actin-binding protein in leukocytes”. In: *Journal of Leukocyte Biology* 105.5, pp. 881–890. DOI: [10.1002/JLB.MR0618-212R](https://doi.org/10.1002/JLB.MR0618-212R).
- Chan, E. K. F. et al. (2011). “Combining Genome-Wide Association Mapping and Transcriptional Networks to Identify Novel Genes Controlling Glucosinolates in *Arabidopsis thaliana*”. In: *PLOS Biology* 9.8, e1001125. DOI: [10.1371/journal.pbio.1001125](https://doi.org/10.1371/journal.pbio.1001125).

- Chang, C. C. et al. (2015). “Second-generation PLINK: rising to the challenge of larger and richer datasets”. In: *GigaScience* 4, p. 7. DOI: [10.1186/s13742-015-0047-8](https://doi.org/10.1186/s13742-015-0047-8).
- Chang, Y.-C. et al. (2020). “GenEpi: gene-based epistasis discovery using machine learning”. In: *BMC Bioinformatics* 21.1, p. 68. DOI: [10.1186/s12859-020-3368-2](https://doi.org/10.1186/s12859-020-3368-2).
- Chari, S. and Dworkin, I. (2013). “The Conditional Nature of Genetic Interactions: The Consequences of Wild-Type Backgrounds on Mutational Interactions in a Genome-Wide Modifier Screen”. In: *PLOS Genetics* 9.8, e1003661. DOI: [10.1371/journal.pgen.1003661](https://doi.org/10.1371/journal.pgen.1003661).
- Chatelain, C. et al. (2025). “Building a human genetic data lake to scale up insights for drug discovery”. In: *Drug Discovery Today* 30.6, p. 104385. DOI: [10.1016/j.drudis.2025.104385](https://doi.org/10.1016/j.drudis.2025.104385).
- Chen, G. et al. (2017). “An association study revealed substantial effects of dominance, epistasis and substance dependence co-morbidity on alcohol dependence symptom count”. In: *Addiction Biology* 22.6, pp. 1475–1485. DOI: [10.1111/adb.12402](https://doi.org/10.1111/adb.12402).
- Chen, H. et al. (2016). “Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models”. In: *The American Journal of Human Genetics* 98.4, pp. 653–666. DOI: [10.1016/j.ajhg.2016.02.012](https://doi.org/10.1016/j.ajhg.2016.02.012).
- Cho, S. and Shin, A. (2025). “Genome-wide interaction study of physical activity and genetic susceptibility on colorectal cancer using UK biobank data”. In: *Scientific Reports* 15.1, p. 30180. DOI: [10.1038/s41598-025-13709-6](https://doi.org/10.1038/s41598-025-13709-6).
- Churchill, G. A. and Doerge, R. W. (2008). “Naive Application of Permutation Testing Leads to Inflated Type I Error Rates”. In: *Genetics* 178.1, pp. 609–610. DOI: [10.1534/genetics.107.074609](https://doi.org/10.1534/genetics.107.074609).
- Clarke, A. J. et al. (2015). “Autophagy is activated in systemic lupus erythematosus and required for plasmablast development”. In: *Annals of the Rheumatic Diseases* 74.5, pp. 912–920. DOI: [10.1136/annrheumdis-2013-204343](https://doi.org/10.1136/annrheumdis-2013-204343).

- Collins, S. R. et al. (2006). “A strategy for extracting and analyzing large-scale quantitative epistatic interaction data”. In: *Genome Biology* 7.7, R63. DOI: [10.1186/gb-2006-7-7-r63](https://doi.org/10.1186/gb-2006-7-7-r63).
- Costanzo, M. et al. (2010). “The Genetic Landscape of a Cell”. In: *Science* 327.5964, pp. 425–431. DOI: [10.1126/science.1180823](https://doi.org/10.1126/science.1180823).
- Cousminer, D. L. et al. (2021). “Genome-wide association study implicates novel loci and reveals candidate effector genes for longitudinal pediatric bone accrual”. In: *Genome Biology* 22.1, p. 1. DOI: [10.1186/s13059-020-02207-9](https://doi.org/10.1186/s13059-020-02207-9).
- Couto Alves, A. et al. (2019). “GWAS on longitudinal growth traits reveals different genetic factors influencing infant, child, and adult BMI”. In: *Science Advances* 5.9, eaaw3095. DOI: [10.1126/sciadv.aaw3095](https://doi.org/10.1126/sciadv.aaw3095).
- Cox, D. R. (1972). “Regression Models and Life-Tables”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 34.2, pp. 187–202. DOI: [10.1111/j.2517-6161.1972.tb00899.x](https://doi.org/10.1111/j.2517-6161.1972.tb00899.x).
- Crawford, L. and Zhou, X. (2018a). *Genome-wide Marginal Epistatic Association Mapping in Case-Control Studies*. Tech. rep. bioRxiv, p. 374983. DOI: [10.1101/374983](https://doi.org/10.1101/374983).
- Crawford, L. et al. (2017). “Detecting epistasis with the marginal epistasis test in genetic mapping studies of quantitative traits”. In: *PLOS Genetics* 13.7, e1006869. DOI: [10.1371/journal.pgen.1006869](https://doi.org/10.1371/journal.pgen.1006869).
- Crawford, L. et al. (2018b). “Bayesian Approximate Kernel Regression With Variable Selection”. In: *Journal of the American Statistical Association* 113.524, pp. 1710–1721. DOI: [10.1080/01621459.2017.1361830](https://doi.org/10.1080/01621459.2017.1361830).
- Crow, J. F. (2010). “On epistasis: why it is unimportant in polygenic directional selection”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 365.1544, pp. 1241–1244. DOI: [10.1098/rstb.2009.0275](https://doi.org/10.1098/rstb.2009.0275).
- Database, M. (2014). “Obtaining and Loading Phenotype Annotations from the International Mouse Phenotyping Consortium (IMPC) Database”. In: *Mouse Genome Informatics and the International Mouse Phenotyping Consortium*.

- Davies, R. B. (1980). “Algorithm AS 155: The Distribution of a Linear Combination of 2 Random Variables”. In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 29.3, pp. 323–333. DOI: [10.2307/2346911](https://doi.org/10.2307/2346911).
- Demmings, E. M. et al. (2019). “Quantitative Trait Locus Analysis of Leaf Morphology Indicates Conserved Shape Loci in Grapevine”. In: *Frontiers in Plant Science* 10.
- Deutschbauer, A. M. and Davis, R. W. (2005). “Quantitative trait loci mapped to single-nucleotide resolution in yeast”. In: *Nature Genetics* 37.12, pp. 1333–1340. DOI: [10.1038/ng1674](https://doi.org/10.1038/ng1674).
- Dewey, M. (2022). *metap: Meta-Analysis of Significance Values*.
- Dey, R. et al. (2022). “Efficient and accurate frailty model approach for genome-wide survival association analysis in large-scale biobanks”. In: *Nature Communications* 13.1, p. 5437. DOI: [10.1038/s41467-022-32885-x](https://doi.org/10.1038/s41467-022-32885-x).
- Ding, K. et al. (2012). “Genetic Loci Implicated in Erythroid Differentiation and Cell Cycle Regulation Are Associated With Red Blood Cell Traits”. In: *Mayo Clinic Proceedings* 87.5, pp. 461–474. DOI: [10.1016/j.mayocp.2012.01.016](https://doi.org/10.1016/j.mayocp.2012.01.016).
- Ding, Y. et al. (2023). “Polygenic scoring accuracy varies across the genetic ancestry continuum”. In: *Nature* 618.7966, pp. 774–781. DOI: [10.1038/s41586-023-06079-4](https://doi.org/10.1038/s41586-023-06079-4).
- Duchesne, P. and Lafaye De Micheaux, P. (2010). “Computing the distribution of quadratic forms: Further comparisons between the Liu–Tang–Zhang approximation and exact methods”. In: *Computational Statistics & Data Analysis* 54.4, pp. 858–862. DOI: [10.1016/j.csda.2009.11.025](https://doi.org/10.1016/j.csda.2009.11.025).
- Dudoit, S., Shaffer, J., and Boldrick, J. (2002). “Multiple Hypothesis Testing in Microarray Experiments”. In: *U.C. Berkeley Division of Biostatistics Working Paper Series*.
- Edwards, M. et al. (2013). “Physiological role of the interaction between CARMIL1 and capping protein”. In: *Molecular Biology of the Cell* 24.19, pp. 3047–3055. DOI: [10.1091/mbc.e13-05-0270](https://doi.org/10.1091/mbc.e13-05-0270).

- Eichler, E. E. et al. (2010). “Missing heritability and strategies for finding the underlying causes of complex disease”. In: *Nature Reviews Genetics* 11.6, pp. 446–450. DOI: [10.1038/nrg2809](https://doi.org/10.1038/nrg2809).
- Ellinghaus, D. et al. (2016). “Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci”. In: *Nature Genetics* 48.5, pp. 510–518. DOI: [10.1038/ng.3528](https://doi.org/10.1038/ng.3528).
- Emani, P. S. et al. (2024). “Single-cell genomics and regulatory networks for 388 human brains”. In: *Science* 384.6698, eadi5199. DOI: [10.1126/science.adi5199](https://doi.org/10.1126/science.adi5199).
- Fadista, J. et al. (2016). “The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants”. In: *European Journal of Human Genetics* 24.8, pp. 1202–1205. DOI: [10.1038/ejhg.2015.269](https://doi.org/10.1038/ejhg.2015.269).
- Fang, G. et al. (2019). “Discovering genetic interactions bridging pathways in genome-wide association studies”. In: *Nature Communications* 10.1, p. 4274. DOI: [10.1038/s41467-019-12131-7](https://doi.org/10.1038/s41467-019-12131-7).
- Feng, X. et al. (2010). “Foxp1 is an essential transcriptional regulator for the generation of quiescent naive T cells during thymocyte development”. In: *Blood* 115.3, pp. 510–518. DOI: [10.1182/blood-2009-07-232694](https://doi.org/10.1182/blood-2009-07-232694).
- Fergus, P. et al. (2020). “Utilizing Deep Learning and Genome Wide Association Studies for Epistatic-Driven Preterm Birth Classification in African-American Women”. In: *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 17.2, pp. 668–678. DOI: [10.1109/TCBB.2018.2868667](https://doi.org/10.1109/TCBB.2018.2868667).
- Finucane, H. K. et al. (2015). “Partitioning heritability by functional annotation using genome-wide association summary statistics”. In: *Nature Genetics* 47.11, pp. 1228–1235. DOI: [10.1038/ng.3404](https://doi.org/10.1038/ng.3404).
- Finucane, H. K. et al. (2018). “Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types”. In: *Nature Genetics* 50.4, pp. 621–629. DOI: [10.1038/s41588-018-0081-4](https://doi.org/10.1038/s41588-018-0081-4).

- Fisher, R. A. (1925). *Statistical Methods for Research Workers*. 5th ed. Biological Monographs and Manuals. Edinburgh: Oliver and Boyd. ISBN: 0-05-002170-2.
- Fiziev, P. P. et al. (2023). “Rare penetrant mutations confer severe risk of common diseases”. In: *Science* 380.6648, eabo1131. DOI: [10.1126/science.abo1131](https://doi.org/10.1126/science.abo1131).
- Flint, J. and Mackay, T. F. (2009). “Genetic architecture of quantitative traits in mice, flies, and humans”. In: *Genome Research* 19.5, pp. 723–733. DOI: [10.1101/gr.086660.108](https://doi.org/10.1101/gr.086660.108).
- Forsberg, S. K. G. et al. (2017). “Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast”. In: *Nature Genetics* 49.4, pp. 497–503. DOI: [10.1038/ng.3800](https://doi.org/10.1038/ng.3800).
- Fröhlich, C. et al. (2024). “Epistasis arises from shifting the rate-limiting step during enzyme evolution of a β -lactamase”. In: *Nature Catalysis* 7.5, pp. 499–509. DOI: [10.1038/s41929-024-01117-4](https://doi.org/10.1038/s41929-024-01117-4).
- Fu, B. et al. (2023). *A biobank-scale test of marginal epistasis reveals genome-wide signals of polygenic epistasis*. DOI: [10.1101/2023.09.10.557084](https://doi.org/10.1101/2023.09.10.557084).
- Fuchsberger, C. et al. (2016). “The genetic architecture of type 2 diabetes”. In: *Nature* 536.7614, pp. 41–47. DOI: [10.1038/nature18642](https://doi.org/10.1038/nature18642).
- Gaertner, B. E. et al. (2012). “More Than the Sum of Its Parts: A Complex Epistatic Network Underlies Natural Variation in Thermal Preference Behavior in *Caenorhabditis elegans*”. In: *Genetics* 192.4, pp. 1533–1542. DOI: [10.1534/genetics.112.142877](https://doi.org/10.1534/genetics.112.142877).
- Galesloot, T. E. et al. (2016). “Meta-GWAS and Meta-Analysis of Exome Array Studies Do Not Reveal Genetic Determinants of Serum Hepcidin”. In: *PLOS ONE* 11.11, e0166628. DOI: [10.1371/journal.pone.0166628](https://doi.org/10.1371/journal.pone.0166628).
- Georgolopoulos, G. et al. (2021). *Discrete regulatory modules instruct hematopoietic lineage commitment and differentiation*. DOI: [10.5281/zenodo.5291737](https://doi.org/10.5281/zenodo.5291737).
- Gerke, J., Lorenz, K., and Cohen, B. (2009). “Genetic Interactions Between Transcription Factors Cause Natural Variation in Yeast”. In: *Science* 323.5913, pp. 498–501. DOI: [10.1126/science.1166426](https://doi.org/10.1126/science.1166426).

- Gilmour, A. R., Thompson, R., and Cullis, B. R. (1995). “Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models”. In: *Biometrics* 51.4, pp. 1440–1450. DOI: [10.2307/2533274](https://doi.org/10.2307/2533274).
- Gokhin, D. S. and Fowler, V. M. (2016). “Feisty filaments: actin dynamics in the red blood cell membrane skeleton”. In: *Current Opinion in Hematology* 23.3, p. 206. DOI: [10.1097/MOH.0000000000000227](https://doi.org/10.1097/MOH.0000000000000227).
- Golan, D., Lander, E. S., and Rosset, S. (2014). “Measuring missing heritability: Inferring the contribution of common variants”. In: *Proceedings of the National Academy of Sciences* 111.49, E5272–E5281. DOI: [10.1073/pnas.1419064111](https://doi.org/10.1073/pnas.1419064111).
- Guo, Y. et al. (2019). “Epi-GTBN: an approach of epistasis mining based on genetic Tabu algorithm and Bayesian network”. In: *BMC Bioinformatics* 20.1, p. 444. DOI: [10.1186/s12859-019-3022-z](https://doi.org/10.1186/s12859-019-3022-z).
- Guo, Y. et al. (2021). “Testing Gene-Gene Interactions Based on a Neighborhood Perspective in Genome-wide Association Studies”. In: *Frontiers in Genetics* 12.
- Gusareva, E. S. and Van Steen, K. (2014a). “Practical aspects of genome-wide association interaction analysis”. In: *Human Genetics* 133.11, pp. 1343–1358. DOI: [10.1007/s00439-014-1480-y](https://doi.org/10.1007/s00439-014-1480-y).
- Gusareva, E. S. et al. (2014b). “Genome-wide association interaction analysis for Alzheimer’s disease”. In: *Neurobiology of Aging* 35.11, pp. 2436–2443. DOI: [10.1016/j.neurobiolaging.2014.05.014](https://doi.org/10.1016/j.neurobiolaging.2014.05.014).
- Gusareva, E. S. et al. (2018). “Male-specific epistasis between WWC1 and TLN2 genes is associated with Alzheimer’s disease”. In: *Neurobiology of Aging* 72, 188.e3–188.e12. DOI: [10.1016/j.neurobiolaging.2018.08.001](https://doi.org/10.1016/j.neurobiolaging.2018.08.001).
- Haar, J. van de et al. (2019). “Identifying Epistasis in Cancer Genomes: A Delicate Affair”. In: *Cell* 177.6, pp. 1375–1383. DOI: [10.1016/j.cell.2019.05.005](https://doi.org/10.1016/j.cell.2019.05.005).
- Harville, D. A. (1977). “Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems”. In: *Journal of the American Statistical Association* 72.358, pp. 320–338. DOI: [10.1080/01621459.1977.10480998](https://doi.org/10.1080/01621459.1977.10480998).

- Haseman, J. K. and Elston, R. C. (1972). “The investigation of linkage between a quantitative trait and a marker locus”. In: *Behavior Genetics* 2.1, pp. 3–19. DOI: [10.1007/BF01066731](https://doi.org/10.1007/BF01066731).
- He, X. et al. (2010). “Prevalent positive epistasis in *Escherichia coli* and *Saccharomyces cerevisiae* metabolic networks”. In: *Nature Genetics* 42.3, pp. 272–276. DOI: [10.1038/ng.524](https://doi.org/10.1038/ng.524).
- Hemani, G. et al. (2011). “EpiGPU: exhaustive pairwise epistasis scans parallelized on consumer level graphics cards”. In: *Bioinformatics* 27.11, pp. 1462–1465. DOI: [10.1093/bioinformatics/btr172](https://doi.org/10.1093/bioinformatics/btr172).
- Hemani, G. et al. (2021). “Phantom epistasis between unlinked loci”. In: *Nature* 596.7871, E1–E3. DOI: [10.1038/s41586-021-03765-z](https://doi.org/10.1038/s41586-021-03765-z).
- Hill, W. G., Goddard, M. E., and Visscher, P. M. (2008). “Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits”. In: *PLOS Genetics* 4.2, e1000008. DOI: [10.1371/journal.pgen.1000008](https://doi.org/10.1371/journal.pgen.1000008).
- Hivert, V. et al. (2021). “Estimation of non-additive genetic variance in human complex traits from a large sample of unrelated individuals”. In: *The American Journal of Human Genetics* 108.5, pp. 786–798. DOI: [10.1016/j.ajhg.2021.02.014](https://doi.org/10.1016/j.ajhg.2021.02.014).
- Horn, T. et al. (2011). “Mapping of signaling networks through synthetic genetic interaction analysis by RNAi”. In: *Nature Methods* 8.4, pp. 341–346. DOI: [10.1038/nmeth.1581](https://doi.org/10.1038/nmeth.1581).
- Hou, K. et al. (2023). “Causal effects on complex traits are similar for common variants across segments of different continental ancestries within admixed individuals”. In: *Nature Genetics* 55.4, pp. 549–558. DOI: [10.1038/s41588-023-01338-6](https://doi.org/10.1038/s41588-023-01338-6).
- Hu, H. et al. (2006). “Foxp1 is an essential transcriptional regulator of B cell development”. In: *Nature Immunology* 7.8, pp. 819–826. DOI: [10.1038/ni1358](https://doi.org/10.1038/ni1358).
- Hu, S. et al. (2025). “Fine-scale population structure and widespread conservation of genetic effect sizes between human groups across traits”. In: *Nature Genetics* 57.2, pp. 379–389. DOI: [10.1038/s41588-024-02035-8](https://doi.org/10.1038/s41588-024-02035-8).

- Huang, W. and Mackay, T. F. C. (2016). “The Genetic Architecture of Quantitative Traits Cannot Be Inferred from Variance Component Analysis”. In: *PLOS Genetics* 12.11, e1006421. DOI: [10.1371/journal.pgen.1006421](https://doi.org/10.1371/journal.pgen.1006421).
- Hutchinson, M. (1989). “A Stochastic Estimator of the Trace of the Influence Matrix for Laplacian Smoothing Splines”. In: *Communications in Statistics - Simulation and Computation* 18.3, pp. 1059–1076. DOI: [10.1080/03610918908812806](https://doi.org/10.1080/03610918908812806).
- Isnard, P. et al. (2000). “Altered lymphoid development in mice deficient for the mAF4 proto-oncogene”. In: *Blood* 96.2, pp. 705–710. DOI: [10.1182/blood.V96.2.705](https://doi.org/10.1182/blood.V96.2.705).
- Jarvis, J. P. and Cheverud, J. M. (2011). “Mapping the Epistatic Network Underlying Murine Reproductive Fatpad Variation”. In: *Genetics* 187.2, pp. 597–610. DOI: [10.1534/genetics.110.123505](https://doi.org/10.1534/genetics.110.123505).
- Jiang, J. (2007). *Linear and generalized linear mixed models and their applications*. Springer series in statistics. New York ; London: Springer. ISBN: 978-0-387-47941-5 978-0-387-47946-0.
- Jiang, Y. and Reif, J. C. (2015). “Modeling Epistasis in Genomic Selection”. In: *Genetics* 201.2, pp. 759–768. DOI: [10.1534/genetics.115.177907](https://doi.org/10.1534/genetics.115.177907).
- Johnson, R. et al. (2023). “The UCLA ATLAS Community Health Initiative: Promoting precision health research in a diverse biobank”. In: *Cell Genomics* 3.1, p. 100243. DOI: [10.1016/j.xgen.2022.100243](https://doi.org/10.1016/j.xgen.2022.100243).
- Jonghe, J. D. et al. (2024). “scTrends: A living review of commercial single-cell and spatial ’omic technologies”. In: *Cell Genomics* 4.12. DOI: [10.1016/j.xgen.2024.100723](https://doi.org/10.1016/j.xgen.2024.100723).
- Jung, G. et al. (2022). “Dual regulation of the actin cytoskeleton by CARMIL-GAP”. In: *Journal of Cell Science* 135.12, jcs258704. DOI: [10.1242/jcs.258704](https://doi.org/10.1242/jcs.258704).
- Kachuri, L. et al. (2024). “Principles and methods for transferring polygenic risk scores across global populations”. In: *Nature Reviews Genetics* 25.1, pp. 8–25. DOI: [10.1038/s41576-023-00637-2](https://doi.org/10.1038/s41576-023-00637-2).
- Kamimoto, K. et al. (2023). “Dissecting cell identity via network inference and in silico gene perturbation”. In: *Nature* 614.7949, pp. 742–751. DOI: [10.1038/s41586-022-05688-9](https://doi.org/10.1038/s41586-022-05688-9).

- Kerin, M. and Marchini, J. (2020a). “A non-linear regression method for estimation of gene–environment heritability”. In: *Bioinformatics* 36.24, pp. 5632–5639. DOI: [10.1093/bioinformatics/btaa1079](https://doi.org/10.1093/bioinformatics/btaa1079).
- (2020b). “Inferring Gene-by-Environment Interactions with a Bayesian Whole-Genome Regression Model”. In: *The American Journal of Human Genetics* 107.4, pp. 698–713. DOI: [10.1016/j.ajhg.2020.08.009](https://doi.org/10.1016/j.ajhg.2020.08.009).
- Kim, D. H. et al. (2025). *Estimating heritability of survival traits using censored multiple variance component model*. DOI: [10.48550/arXiv.2510.26226](https://doi.org/10.48550/arXiv.2510.26226).
- Kim, M. S. et al. (2024). “Association of genetic risk, lifestyle, and their interaction with obesity and obesity-related morbidities”. In: *Cell Metabolism* 36.7, 1494–1503.e3. DOI: [10.1016/j.cmet.2024.06.004](https://doi.org/10.1016/j.cmet.2024.06.004).
- Kirino, Y. et al. (2013). “Genome-wide association analysis identifies new susceptibility loci for Behçet’s disease and epistasis between HLA-B*51 and ERAP1”. In: *Nature Genetics* 45.2, pp. 202–207. DOI: [10.1038/ng.2520](https://doi.org/10.1038/ng.2520).
- Ko, S. et al. (2022). “GWAS of longitudinal trajectories at biobank scale”. In: *The American Journal of Human Genetics* 109.3, pp. 433–445. DOI: [10.1016/j.ajhg.2022.01.018](https://doi.org/10.1016/j.ajhg.2022.01.018).
- Koppelman, G. H. et al. (2025). “Genetic and environmental risk factors for asthma: towards prevention”. In: *The Lancet. Respiratory Medicine*, S2213–2600(25)00256–5. DOI: [10.1016/S2213-2600\(25\)00256-5](https://doi.org/10.1016/S2213-2600(25)00256-5).
- Kraft, P. et al. (2007). “Exploiting Gene-Environment Interaction to Detect Genetic Associations”. In: *Human Heredity* 63.2, pp. 111–119. DOI: [10.1159/000099183](https://doi.org/10.1159/000099183).
- Kroymann, J. and Mitchell-Olds, T. (2005). “Epistasis and balanced polymorphism influencing complex trait variation”. In: *Nature* 435.7038, pp. 95–98. DOI: [10.1038/nature03480](https://doi.org/10.1038/nature03480).
- Kuonen, D. (1999). “Miscellanea. Saddlepoint approximations for distributions of quadratic forms in normal variables”. In: *Biometrika* 86.4, pp. 929–935. DOI: [10.1093/biomet/86.4.929](https://doi.org/10.1093/biomet/86.4.929).

- Kurki, M. I. et al. (2023). “FinnGen provides genetic insights from a well-phenotyped isolated population”. In: *Nature* 613.7944, pp. 508–518. DOI: [10.1038/s41586-022-05473-8](https://doi.org/10.1038/s41586-022-05473-8).
- Lappalainen, T. et al. (2024). “Genetic and molecular architecture of complex traits”. In: *Cell* 187.5, pp. 1059–1075. DOI: [10.1016/j.cell.2024.01.023](https://doi.org/10.1016/j.cell.2024.01.023).
- Lawrence, M. et al. (2013). “Software for Computing and Annotating Genomic Ranges”. In: *PLOS Computational Biology* 9.8, e1003118. DOI: [10.1371/journal.pcbi.1003118](https://doi.org/10.1371/journal.pcbi.1003118).
- Leamy, L., Gordon, R., and Pomp, D. (2011). “Sex-, Diet-, and Cancer-Dependent Epistatic Effects on Complex Traits in Mice”. In: *Frontiers in Genetics* 2.
- Lee, S. H. et al. (2011). “Estimating Missing Heritability for Disease from Genome-wide Association Studies”. In: *The American Journal of Human Genetics* 88.3, pp. 294–305. DOI: [10.1016/j.ajhg.2011.02.002](https://doi.org/10.1016/j.ajhg.2011.02.002).
- Lee, S. et al. (2014). “Rare-Variant Association Analysis: Study Designs and Statistical Tests”. In: *American Journal of Human Genetics* 95.1, pp. 5–23. DOI: [10.1016/j.ajhg.2014.06.009](https://doi.org/10.1016/j.ajhg.2014.06.009).
- Lee, S. and Mostafavi, H. (2025). “Missing regulatory effects on complex traits: Contribution of distal variants”. In: *Cell Genomics* 5.3, p. 100809. DOI: [10.1016/j.xgen.2025.100809](https://doi.org/10.1016/j.xgen.2025.100809).
- Lee, Y. and Nelder, J. A. (1996). “Hierarchical Generalized Linear Models”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 58.4, pp. 619–678.
- Lehner, B. et al. (2006). “Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways”. In: *Nature Genetics* 38.8, pp. 896–903. DOI: [10.1038/ng1844](https://doi.org/10.1038/ng1844).
- Lennon, N. J. et al. (2024). “Selection, optimization and validation of ten chronic disease polygenic risk scores for clinical implementation in diverse US populations”. In: *Nature Medicine* 30.2, pp. 480–487. DOI: [10.1038/s41591-024-02796-z](https://doi.org/10.1038/s41591-024-02796-z).

- Lewinger, J. P. et al. (2013). “Efficient Two-Step Testing of Gene-Gene Interactions in Genome-Wide Association Studies”. In: *Genetic Epidemiology* 37.5, pp. 440–451. DOI: [10.1002/gepi.21720](https://doi.org/10.1002/gepi.21720).
- Liberty, E. and Zucker, S. W. (2009). “The Mailman algorithm: A note on matrix–vector multiplication”. In: *Information Processing Letters* 109.3, pp. 179–182. DOI: [10.1016/j.ipl.2008.09.028](https://doi.org/10.1016/j.ipl.2008.09.028).
- Lindstrom, M. J. and Bates, D. M. (1988). “Newton—Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data”. In: *Journal of the American Statistical Association* 83.404, pp. 1014–1022. DOI: [10.1080/01621459.1988.10478693](https://doi.org/10.1080/01621459.1988.10478693).
- Lishout, F. V. et al. (2013). “An efficient algorithm to perform multiple testing in epistasis screening”. In: *BMC Bioinformatics* 14.1, p. 138. DOI: [10.1186/1471-2105-14-138](https://doi.org/10.1186/1471-2105-14-138).
- Liu, Y. and Xie, J. (2020). “Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures”. In: *Journal of the American Statistical Association* 115.529, pp. 393–402. DOI: [10.1080/01621459.2018.1554485](https://doi.org/10.1080/01621459.2018.1554485).
- Long, A. D. and Langley, C. H. (1999). “The Power of Association Studies to Detect the Contribution of Candidate Genetic Loci to Variation in Complex Traits”. In: *Genome Research* 9.8, pp. 720–731.
- Loos, R. J. F. (2020). “15 years of genome-wide association studies and no signs of slowing down”. In: *Nature Communications* 11.1, p. 5900. DOI: [10.1038/s41467-020-19653-5](https://doi.org/10.1038/s41467-020-19653-5).
- Ma, L., Clark, A. G., and Keinan, A. (2013). “Gene-Based Testing of Interactions in Association Studies of Quantitative Traits”. In: *PLOS Genetics* 9.2, e1003321. DOI: [10.1371/journal.pgen.1003321](https://doi.org/10.1371/journal.pgen.1003321).
- Mackay, T. F. C. and Anholt, R. R. H. (2024). “Pleiotropy, epistasis and the genetic architecture of quantitative traits”. In: *Nature Reviews Genetics*, pp. 1–19. DOI: [10.1038/s41576-024-00711-3](https://doi.org/10.1038/s41576-024-00711-3).
- Mäki-Tanila, A. and Hill, W. G. (2014). “Influence of Gene Interaction on Complex Trait Variation with Multilocus Models”. In: *Genetics* 198.1, pp. 355–367. DOI: [10.1534/genetics.114.165282](https://doi.org/10.1534/genetics.114.165282).

- Marchini, J., Donnelly, P., and Cardon, L. R. (2005). “Genome-wide strategies for detecting multiple loci that influence complex diseases”. In: *Nature Genetics* 37.4, pp. 413–417. DOI: [10.1038/ng1537](https://doi.org/10.1038/ng1537).
- Maurano, M. T. et al. (2012). “Systematic Localization of Common Disease-Associated Variation in Regulatory DNA”. In: *Science* 337.6099, pp. 1190–1195. DOI: [10.1126/science.1222794](https://doi.org/10.1126/science.1222794).
- Mbatchou, J. et al. (2021). “Computationally efficient whole-genome regression for quantitative and binary traits”. In: *Nature Genetics* 53.7, pp. 1097–1103. DOI: [10.1038/s41588-021-00870-7](https://doi.org/10.1038/s41588-021-00870-7).
- McCulloch, C. E. and Searle, S. R. (2005). *Generalized, linear, and mixed models*. Wiley series in probability and statistics. New York: Wiley. ISBN: 978-0-471-19364-7 978-0-471-72207-6. DOI: [10.1002/0471722073](https://doi.org/10.1002/0471722073).
- McDonagh, E. M. et al. (2024). “Human Genetics and Genomics for Drug Target Identification and Prioritization: Open Targets’ Perspective”. In: *Annual Review of Biomedical Data Science* 7.1, pp. 59–81. DOI: [10.1146/annurev-biodatasci-102523-103838](https://doi.org/10.1146/annurev-biodatasci-102523-103838).
- Minikel, E. V. et al. (2020). “Evaluating drug targets through human loss-of-function genetic variation”. In: *Nature* 581.7809, pp. 459–464. DOI: [10.1038/s41586-020-2267-z](https://doi.org/10.1038/s41586-020-2267-z).
- Minikel, E. V. et al. (2024). “Refining the impact of genetic evidence on clinical success”. In: *Nature* 629.8012, pp. 624–629. DOI: [10.1038/s41586-024-07316-0](https://doi.org/10.1038/s41586-024-07316-0).
- Monnahan, P. J. and Kelly, J. K. (2015). “Epistasis Is a Major Determinant of the Additive Genetic Variance in *Mimulus guttatus*”. In: *PLOS Genetics* 11.5, e1005201. DOI: [10.1371/journal.pgen.1005201](https://doi.org/10.1371/journal.pgen.1005201).
- Moore, R. et al. (2019). “A linear mixed-model approach to study multivariate gene–environment interactions”. In: *Nature Genetics* 51.1, pp. 180–186. DOI: [10.1038/s41588-018-0271-0](https://doi.org/10.1038/s41588-018-0271-0).

- Moreno-Grau, S. et al. (2024). “Polygenic risk score portability for common diseases across genetically diverse populations”. In: *Human Genomics* 18.1, p. 93. DOI: [10.1186/s40246-024-00664-y](https://doi.org/10.1186/s40246-024-00664-y).
- Morris, J. A. et al. (2023). “Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens”. In: *Science* 380.6646, eadh7699. DOI: [10.1126/science.adh7699](https://doi.org/10.1126/science.adh7699).
- Mortensen, M. et al. (2011). “The autophagy protein Atg7 is essential for hematopoietic stem cell maintenance”. In: *Journal of Experimental Medicine* 208.3, pp. 455–467. DOI: [10.1084/jem.20101145](https://doi.org/10.1084/jem.20101145).
- Muñoz, P. R. et al. (2014). “Unraveling Additive from Nonadditive Effects Using Genomic Relationship Matrices”. In: *Genetics* 198.4, pp. 1759–1768. DOI: [10.1534/genetics.114.171322](https://doi.org/10.1534/genetics.114.171322).
- Nagai, A. et al. (2017). “Overview of the BioBank Japan Project: Study design and profile”. In: *Journal of Epidemiology*. SI: Overview of BBJ cohort 27.3, Supplement, S2–S8. DOI: [10.1016/j.je.2016.12.005](https://doi.org/10.1016/j.je.2016.12.005).
- Nakayama, K. et al. (2003). “Impaired degradation of inhibitory subunit of NF- κ B (I κ B) and -catenin as a result of targeted disruption of the -TrCP1 gene”. In: *Proceedings of the National Academy of Sciences* 100.15, pp. 8752–8757. DOI: [10.1073/pnas.1133216100](https://doi.org/10.1073/pnas.1133216100).
- Nigra, A. D., Casale, C. H., and Santander, V. S. (2020). “Human erythrocytes: cytoskeleton and its origin”. In: *Cellular and Molecular Life Sciences* 77.9, pp. 1681–1694. DOI: [10.1007/s00018-019-03346-4](https://doi.org/10.1007/s00018-019-03346-4).
- Norris, J. M. and Rich, S. S. (2012). “Genetics of Glucose Homeostasis”. In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 32.9, pp. 2091–2096. DOI: [10.1161/ATVBAHA.112.255463](https://doi.org/10.1161/ATVBAHA.112.255463).
- Onge, R. P. S. et al. (2007). “Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions”. In: *Nature Genetics* 39.2, pp. 199–206. DOI: [10.1038/ng1948](https://doi.org/10.1038/ng1948).

- Patel, R. A. et al. (2022). “Genetic interactions drive heterogeneity in causal variant effect sizes for gene expression and complex traits”. In: *The American Journal of Human Genetics* 109.7, pp. 1286–1297. DOI: [10.1016/j.ajhg.2022.05.014](https://doi.org/10.1016/j.ajhg.2022.05.014).
- Pattillo Smith, S. et al. (2024). “Discovering non-additive heritability using additive GWAS summary statistics”. In: *eLife* 13, e90459. DOI: [10.7554/eLife.90459](https://doi.org/10.7554/eLife.90459).
- Pazokitoroudi, A. et al. (2020). “Efficient variance components analysis across millions of genomes”. In: *Nature Communications* 11.1, p. 4020. DOI: [10.1038/s41467-020-17576-9](https://doi.org/10.1038/s41467-020-17576-9).
- Pazokitoroudi, A. et al. (2021). “Quantifying the contribution of dominance deviation effects to complex trait variation in biobank-scale data”. In: *The American Journal of Human Genetics* 108.5, pp. 799–808. DOI: [10.1016/j.ajhg.2021.03.018](https://doi.org/10.1016/j.ajhg.2021.03.018).
- Pecanka, J. and Jonker, M. A. (2021). “Two-Stage Testing for Epistasis: Screening and Verification”. In: *Epistasis: Methods and Protocols*. Methods in Molecular Biology. New York, NY: Springer US, pp. 69–92. ISBN: 978-1-07-160947-7. DOI: [10.1007/978-1-0716-0947-7_6](https://doi.org/10.1007/978-1-0716-0947-7_6).
- Pérez, P. and Campos, G. de los (2014). “Genome-Wide Regression and Prediction with the BGLR Statistical Package”. In: *Genetics* 198.2, pp. 483–495. DOI: [10.1534/genetics.114.164442](https://doi.org/10.1534/genetics.114.164442).
- Peripato, A. C. et al. (2004). “Epistasis affecting litter size in mice”. In: *Journal of Evolutionary Biology* 17.3, pp. 593–602. DOI: [10.1111/j.1420-9101.2004.00702.x](https://doi.org/10.1111/j.1420-9101.2004.00702.x).
- Pettersson, M. et al. (2011). “Replication and Explorations of High-Order Epistasis Using a Large Advanced Intercross Line Pedigree”. In: *PLOS Genetics* 7.7, e1002180. DOI: [10.1371/journal.pgen.1002180](https://doi.org/10.1371/journal.pgen.1002180).
- Phillips, A. M. et al. (2021). “Binding affinity landscapes constrain the evolution of broadly neutralizing anti-influenza antibodies”. In: *eLife* 10, e71393. DOI: [10.7554/eLife.71393](https://doi.org/10.7554/eLife.71393).
- Plotnikov, D. et al. (2022). “High Blood Pressure and Intraocular Pressure: A Mendelian Randomization Study”. In: *Investigative Ophthalmology & Visual Science* 63.6, p. 29. DOI: [10.1167/iovs.63.6.29](https://doi.org/10.1167/iovs.63.6.29).

- Polderman, T. J. C. et al. (2015). “Meta-analysis of the heritability of human traits based on fifty years of twin studies”. In: *Nature Genetics* 47.7, pp. 702–709. DOI: [10.1038/ng.3285](https://doi.org/10.1038/ng.3285).
- Prabhu, S. and Pe’er, I. (2012). “Ultrafast genome-wide scan for SNP–SNP interactions in common complex disease”. In: *Genome Research* 22.11, pp. 2230–2240. DOI: [10.1101/gr.137885.112](https://doi.org/10.1101/gr.137885.112).
- Purcell, S. et al. (2007). “PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses”. In: *American Journal of Human Genetics* 81.3, pp. 559–575.
- Qin, J. et al. (2020). “Structural and mechanistic insights into secretagogin-mediated exocytosis”. In: *Proceedings of the National Academy of Sciences* 117.12, pp. 6559–6570. DOI: [10.1073/pnas.1919698117](https://doi.org/10.1073/pnas.1919698117).
- Raaijmakers, M. H. G. P. et al. (2010). “Bone progenitor dysfunction induces myelodysplasia and secondary leukaemia”. In: *Nature* 464.7290, pp. 852–857. DOI: [10.1038/nature08851](https://doi.org/10.1038/nature08851).
- Ray, S. et al. (2022). “Functional requirements for a Samd14-capping protein complex in stress erythropoiesis”. In: *eLife* 11, e76497. DOI: [10.7554/eLife.76497](https://doi.org/10.7554/eLife.76497).
- Ripke, S. et al. (2014). “Biological insights from 108 schizophrenia-associated genetic loci”. In: *Nature* 511.7510, pp. 421–427. DOI: [10.1038/nature13595](https://doi.org/10.1038/nature13595).
- Robinson, T., Harkin, J., and Shukla, P. (2021). “Hardware acceleration of genomics data analysis: challenges and opportunities”. In: *Bioinformatics* 37.13, pp. 1785–1795. DOI: [10.1093/bioinformatics/btab017](https://doi.org/10.1093/bioinformatics/btab017).
- Runcie, D. E. et al. (2021). “MegaLMM: Mega-scale linear mixed models for genomic predictions with thousands of traits”. In: *Genome Biology* 22.1, p. 213. DOI: [10.1186/s13059-021-02416-w](https://doi.org/10.1186/s13059-021-02416-w).
- Satterthwaite, F. E. (1946). “An Approximate Distribution of Estimates of Variance Components”. In: *Biometrics Bulletin* 2.6, pp. 110–114. DOI: [10.2307/3002019](https://doi.org/10.2307/3002019).

- Schüpbach, T. et al. (2010). “FastEpistasis: a high performance computing solution for quantitative trait epistasis”. In: *Bioinformatics* 26.11, pp. 1468–1469. DOI: [10.1093/bioinformatics/btq147](https://doi.org/10.1093/bioinformatics/btq147).
- Searle, S. R. (1979). “Estimating Multivariate Variance and Covariance Components Using Quadratic and Bilinear Forms”. In: *Biometrical Journal* 21.4, pp. 389–398. DOI: [10.1002/bimj.4710210407](https://doi.org/10.1002/bimj.4710210407).
- Shao, H. et al. (2008). “Genetic architecture of complex traits: Large phenotypic effects and pervasive epistasis”. In: *Proceedings of the National Academy of Sciences* 105.50, pp. 19910–19914. DOI: [10.1073/pnas.0810388105](https://doi.org/10.1073/pnas.0810388105).
- Sheppard, B. et al. (2021). “A model and test for coordinated polygenic epistasis in complex traits”. In: *Proceedings of the National Academy of Sciences* 118.15, e1922305118. DOI: [10.1073/pnas.1922305118](https://doi.org/10.1073/pnas.1922305118).
- Shifman, S. et al. (2006). “A High-Resolution Single Nucleotide Polymorphism Genetic Map of the Mouse Genome”. In: *PLOS Biology* 4.12, e395. DOI: [10.1371/journal.pbio.0040395](https://doi.org/10.1371/journal.pbio.0040395).
- Smith, C. L. and Eppig, J. T. (2009). “The mammalian phenotype ontology: enabling robust annotation and comparative analysis”. In: *Wiley Interdisciplinary Reviews. Systems Biology and Medicine* 1.3, pp. 390–399. DOI: [10.1002/wsbm.44](https://doi.org/10.1002/wsbm.44).
- Souto, J. C. et al. (2000). “Genetic Susceptibility to Thrombosis and Its Relationship to Physiological Risk Factors: The GAIT Study”. In: *The American Journal of Human Genetics* 67.6, pp. 1452–1459. DOI: [10.1086/316903](https://doi.org/10.1086/316903).
- Stamp, J. et al. (2022). *Leveraging the Genetic Correlation between Traits Improves the Detection of Epistasis in Genome-wide Association Studies*. DOI: [10.7910/DVN/WPFIGU](https://doi.org/10.7910/DVN/WPFIGU).
- (2023). “Leveraging the Genetic Correlation between Traits Improves the Detection of Epistasis in Genome-wide Association Studies”. In: *G3 Genes/Genomes/Genetics*, jkad118. DOI: [10.1093/g3journal/jkad118](https://doi.org/10.1093/g3journal/jkad118).

- Stamp, J. et al. (2025). “Sparse modeling of interactions enables fast detection of genome-wide epistasis in biobank-scale studies”. In: *The American Journal of Human Genetics* 0.0. DOI: [10.1016/j.ajhg.2025.07.004](https://doi.org/10.1016/j.ajhg.2025.07.004).
- Stark, B. C., Lanier, M. H., and Cooper, J. A. (2017). “CARMIL family proteins as multidomain regulators of actin-based motility”. In: *Molecular Biology of the Cell* 28.13, pp. 1713–1723. DOI: [10.1091/mbc.E17-01-0019](https://doi.org/10.1091/mbc.E17-01-0019).
- Stroup, W. W. (2012). *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. 1st. CRC Press.
- Szappanos, B. et al. (2011). “An integrated approach to characterize genetic interaction networks in yeast metabolism”. In: *Nature Genetics* 43.7, pp. 656–662. DOI: [10.1038/ng.846](https://doi.org/10.1038/ng.846).
- Tang, W. et al. (2009). “Epistatic Module Detection for Case-Control Studies: A Bayesian Model with a Gibbs Sampling Strategy”. In: *PLOS Genetics* 5.5, e1000464. DOI: [10.1371/journal.pgen.1000464](https://doi.org/10.1371/journal.pgen.1000464).
- Tao, S. et al. (2012). “Genome-wide two-locus epistasis scans in prostate cancer using two European populations”. In: *Human Genetics* 131.7, pp. 1225–1234. DOI: [10.1007/s00439-012-1148-4](https://doi.org/10.1007/s00439-012-1148-4).
- Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Statistics for Biology and Health. New York, NY: Springer New York. ISBN: 978-1-4419-3161-0 978-1-4757-3294-8. DOI: [10.1007/978-1-4757-3294-8](https://doi.org/10.1007/978-1-4757-3294-8).
- Timoteo, V. J. et al. (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”. In: *The Journal of Nutritional Biochemistry* 111, p. 109126. DOI: [10.1016/j.jnutbio.2022.109126](https://doi.org/10.1016/j.jnutbio.2022.109126).
- Tong, A. H. Y. et al. (2004). “Global Mapping of the Yeast Genetic Interaction Network”. In: *Science* 303.5659, pp. 808–813. DOI: [10.1126/science.1091317](https://doi.org/10.1126/science.1091317).
- Turchin, M. C. et al. (2020). *Pathway Analysis within Multiple Human Ancestries Reveals Novel Signals for Epistasis in Complex Traits*. DOI: [10.1101/2020.09.24.312421](https://doi.org/10.1101/2020.09.24.312421).

- Valdar, W., Flint, J., and Mott, R. (2006a). “Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice”. In: *Genetics* 172.3, pp. 1783–1797. DOI: [10.1534/genetics.104.039313](https://doi.org/10.1534/genetics.104.039313).
- Valdar, W. et al. (2006b). “Genome-wide genetic association of complex traits in heterogeneous stock mice”. In: *Nature Genetics* 38.8, pp. 879–887. DOI: [10.1038/ng1840](https://doi.org/10.1038/ng1840).
- Visscher, P. M. et al. (2012). “Five Years of GWAS Discovery”. In: *American Journal of Human Genetics* 90.1, pp. 7–24. DOI: [10.1016/j.ajhg.2011.11.029](https://doi.org/10.1016/j.ajhg.2011.11.029).
- Visscher, P. M. et al. (2017). “10 Years of GWAS Discovery: Biology, Function, and Translation”. In: *The American Journal of Human Genetics* 101.1, pp. 5–22. DOI: [10.1016/j.ajhg.2017.06.005](https://doi.org/10.1016/j.ajhg.2017.06.005).
- Visscher, P. M. et al. (2025). “Heritable polygenic editing: the next frontier in genomic medicine?” In: *Nature* 637.8046, pp. 637–645. DOI: [10.1038/s41586-024-08300-4](https://doi.org/10.1038/s41586-024-08300-4).
- Vuckovic, D. et al. (2020). “The Polygenic and Monogenic Basis of Blood Traits and Diseases”. In: *Cell* 182.5, 1214–1231.e11. DOI: [10.1016/j.cell.2020.08.008](https://doi.org/10.1016/j.cell.2020.08.008).
- Wainschein, P. et al. (2022). “Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data”. In: *Nature Genetics* 54.3, pp. 263–273. DOI: [10.1038/s41588-021-00997-7](https://doi.org/10.1038/s41588-021-00997-7).
- Wan, X. et al. (2010). “BOOST: A Fast Approach to Detecting Gene-Gene Interactions in Genome-wide Case-Control Studies”. In: *American Journal of Human Genetics* 87.3, pp. 325–340. DOI: [10.1016/j.ajhg.2010.07.021](https://doi.org/10.1016/j.ajhg.2010.07.021).
- Wanstrat, A. and Wakeland, E. (2001). “The genetics of complex autoimmune diseases: non-MHC susceptibility genes”. In: *Nature Immunology* 2.9, pp. 802–809. DOI: [10.1038/ni0901-802](https://doi.org/10.1038/ni0901-802).
- Warburton, P. E. and Sebra, R. P. (2023). “Long-Read DNA Sequencing: Recent Advances and Remaining Challenges”. In: *Annual Review of Genomics and Human Genetics* 24. Volume 24, 2023, pp. 109–132. DOI: [10.1146/annurev-genom-101722-103045](https://doi.org/10.1146/annurev-genom-101722-103045).

- Weatherly, S. M. et al. (2022). “Identification of *Arhgef12* and *Prkci* as genetic modifiers of retinal dysplasia in the *Crb1rd8* mouse model”. In: *PLOS Genetics* 18.6, e1009798. DOI: [10.1371/journal.pgen.1009798](https://doi.org/10.1371/journal.pgen.1009798).
- Wei, W.-H. et al. (2012). “Genome-wide analysis of epistasis in body mass index using multiple human populations”. In: *European journal of human genetics: EJHG* 20.8, pp. 857–862. DOI: [10.1038/ejhg.2012.17](https://doi.org/10.1038/ejhg.2012.17).
- Weinreich, D. M. et al. (2006). “Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins”. In: *Science* 312.5770, pp. 111–114. DOI: [10.1126/science.1123539](https://doi.org/10.1126/science.1123539).
- Weischenfeldt, J. et al. (2008). “NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements”. In: *Genes & Development* 22.10, pp. 1381–1396. DOI: [10.1101/gad.468808](https://doi.org/10.1101/gad.468808).
- Wellcome Trust Centre for Human Genetics - Mouse Resources (2022).
- Wentzell, A. M. et al. (2007). “Linking Metabolic QTLs with Network and cis-eQTLs Controlling Biosynthetic Pathways”. In: *PLOS Genetics* 3.9, e162. DOI: [10.1371/journal.pgen.0030162](https://doi.org/10.1371/journal.pgen.0030162).
- Whitehead, J. (1980). “Fitting Cox’s Regression Model to Survival Data using GLIM”. In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 29.3, pp. 268–275. DOI: [10.2307/2346901](https://doi.org/10.2307/2346901).
- Wilson, D. J. (2019). “The harmonic mean p-value for combining dependent tests”. In: *Proceedings of the National Academy of Sciences* 116.4, pp. 1195–1200. DOI: [10.1073/pnas.1814092116](https://doi.org/10.1073/pnas.1814092116).
- Wood, A. R. et al. (2014). “Another explanation for apparent epistasis”. In: *Nature* 514.7520, E3–E5. DOI: [10.1038/nature13691](https://doi.org/10.1038/nature13691).
- Wu, M. C. et al. (2011). “Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test”. In: *American Journal of Human Genetics* 89.1, pp. 82–93. DOI: [10.1016/j.ajhg.2011.05.029](https://doi.org/10.1016/j.ajhg.2011.05.029).

- Wu, W. et al. (2019). “Retrospective Association Analysis of Longitudinal Binary Traits Identifies Important Loci and Pathways in Cocaine Use”. In: *Genetics* 213.4, pp. 1225–1236. DOI: [10.1534/genetics.119.302598](https://doi.org/10.1534/genetics.119.302598).
- Wu, Y. and Sankararaman, S. (2018). “A scalable estimator of SNP heritability for biobank-scale data”. In: *Bioinformatics* 34.13, pp. i187–i194. DOI: [10.1093/bioinformatics/bty253](https://doi.org/10.1093/bioinformatics/bty253).
- Wu, Y. et al. (2022). “Fast estimation of genetic correlation for biobank-scale data”. In: *American Journal of Human Genetics* 109.1, pp. 24–32. DOI: [10.1016/j.ajhg.2021.11.015](https://doi.org/10.1016/j.ajhg.2021.11.015).
- Xu, F. et al. (2017). “Mutation of the Kunitz-type proteinase inhibitor domain in the amyloid -protein precursor abolishes its anti-thrombotic properties in vivo”. In: *Thrombosis Research* 155, pp. 58–64. DOI: [10.1016/j.thromres.2017.05.003](https://doi.org/10.1016/j.thromres.2017.05.003).
- Yang, C. et al. (2005). “Mammalian CARMIL Inhibits Actin Filament Capping by Capping Protein”. In: *Developmental Cell* 9.2, pp. 209–221. DOI: [10.1016/j.devcel.2005.06.008](https://doi.org/10.1016/j.devcel.2005.06.008).
- Yang, J. et al. (2010). “Common SNPs explain a large proportion of the heritability for human height”. In: *Nature Genetics* 42.7, pp. 565–569. DOI: [10.1038/ng.608](https://doi.org/10.1038/ng.608).
- Yang, Y. et al. (2014). “Arginine Methylation Facilitates the Recruitment of TOP3B to Chromatin to Prevent R Loop Accumulation”. In: *Molecular Cell* 53.3, pp. 484–497. DOI: [10.1016/j.molcel.2014.01.011](https://doi.org/10.1016/j.molcel.2014.01.011).
- Yang, Z. et al. (2025). “Limited overlap between genetic effects on disease susceptibility and disease survival”. In: *Nature Genetics* 57.10, pp. 2418–2426. DOI: [10.1038/s41588-025-02342-8](https://doi.org/10.1038/s41588-025-02342-8).
- Yengo, L. et al. (2022). “A saturated map of common genetic variants associated with human height”. In: *Nature* 610.7933, pp. 704–712. DOI: [10.1038/s41586-022-05275-y](https://doi.org/10.1038/s41586-022-05275-y).
- Zhang, W. et al. (2008). “Evaluation of Genetic Variation Contributing to Differences in Gene Expression between Populations”. In: *The American Journal of Human Genetics* 82.3, pp. 631–640. DOI: [10.1016/j.ajhg.2007.12.015](https://doi.org/10.1016/j.ajhg.2007.12.015).

- Zhang, X. et al. (2010). “TEAM: efficient two-locus epistasis tests in human genome-wide association study”. In: *Bioinformatics* 26.12, pp. i217–i227. DOI: [10.1093/bioinformatics/btq186](https://doi.org/10.1093/bioinformatics/btq186).
- Zhang, Y. D. et al. (2020). “Assessment of polygenic architecture and risk prediction based on common variants across fourteen cancers”. In: *Nature Communications* 11.1, p. 3353. DOI: [10.1038/s41467-020-16483-3](https://doi.org/10.1038/s41467-020-16483-3).
- Zhang, Y. and Liu, J. S. (2007). “Bayesian inference of epistatic interactions in case-control studies”. In: *Nature Genetics* 39.9, pp. 1167–1173. DOI: [10.1038/ng2110](https://doi.org/10.1038/ng2110).
- Zhang, Y., Zhang, J., and Liu, J. S. (2011). “Block-based Bayesian epistasis association mapping with application to WTCCC type 1 diabetes data”. In: *The Annals of Applied Statistics* 5.3, pp. 2052–2077. DOI: [10.1214/11-AOAS469](https://doi.org/10.1214/11-AOAS469).
- Zhao, H. et al. (2025). “Multi-ancestry genome-wide association analyses of polycystic ovary syndrome”. In: *Nature Genetics*, pp. 1–13. DOI: [10.1038/s41588-025-02393-x](https://doi.org/10.1038/s41588-025-02393-x).
- Zhao, H. et al. (2014). “CrossMap: a versatile tool for coordinate conversion between genome assemblies”. In: *Bioinformatics* 30.7, pp. 1006–1007. DOI: [10.1093/bioinformatics/btt730](https://doi.org/10.1093/bioinformatics/btt730).
- Zhong, W. et al. (2023). “A fast and powerful linear mixed model approach for genotype-environment interaction tests in large-scale GWAS”. In: *Briefings in Bioinformatics* 24.1, bbac547. DOI: [10.1093/bib/bbac547](https://doi.org/10.1093/bib/bbac547).
- Zhou, J. et al. (2022). “Higher-order epistasis and phenotypic prediction”. In: *Proceedings of the National Academy of Sciences* 119.39, e2204233119. DOI: [10.1073/pnas.2204233119](https://doi.org/10.1073/pnas.2204233119).
- Zhou, X. (2017). “A unified framework for variance component estimation with summary statistics in genome-wide association studies”. In: *Annals of Applied Statistics* 11.4, pp. 2027–2051. DOI: [10.1214/17-AOAS1052](https://doi.org/10.1214/17-AOAS1052).
- Zhou, X., Carbonetto, P., and Stephens, M. (2013). “Polygenic Modeling with Bayesian Sparse Linear Mixed Models”. In: *PLOS Genetics* 9.2, e1003264. DOI: [10.1371/journal.pgen.1003264](https://doi.org/10.1371/journal.pgen.1003264).

- Zhou, X. and Ning, C. (2025). *fastGxE: Powering genome-wide detection of genotype-environment interactions in biobank studies*. DOI: [10.21203/rs.3.rs-5952773/v1](https://doi.org/10.21203/rs.3.rs-5952773/v1).
- Zhou, X. and Stephens, M. (2014). “Efficient multivariate linear mixed model algorithms for genome-wide association studies”. In: *Nature Methods* 11.4, pp. 407–409. DOI: [10.1038/nmeth.2848](https://doi.org/10.1038/nmeth.2848).
- Zhu, C. et al. (2022). *Amplification is the Primary Mode of Gene-by-Sex Interaction in Complex Human Traits*. DOI: [10.1101/2022.05.06.490973](https://doi.org/10.1101/2022.05.06.490973).
- Zhu, H. and Zhou, X. (2020). “Statistical methods for SNP heritability estimation and partition: A review”. In: *Computational and Structural Biotechnology Journal* 18, pp. 1557–1568. DOI: [10.1016/j.csbj.2020.06.011](https://doi.org/10.1016/j.csbj.2020.06.011).
- Zhu, S. and Fang, G. (2018). “MatrixEpistasis: ultrafast, exhaustive epistasis scan for quantitative traits with covariate adjustment”. In: *Bioinformatics* 34.14, pp. 2341–2348. DOI: [10.1093/bioinformatics/bty094](https://doi.org/10.1093/bioinformatics/bty094).
- Zuk, O. et al. (2012). “The mystery of missing heritability: Genetic interactions create phantom heritability”. In: *Proceedings of the National Academy of Sciences* 109.4, pp. 1193–1198. DOI: [10.1073/pnas.1119675109](https://doi.org/10.1073/pnas.1119675109).
- Zwarts, L. et al. (2011). “Complex genetic architecture of *Drosophila* aggressive behavior”. In: *Proceedings of the National Academy of Sciences* 108.41, pp. 17070–17075. DOI: [10.1073/pnas.1113877108](https://doi.org/10.1073/pnas.1113877108).