Statistical Methods and Applications of Transcriptome-wide Association Studies

by

Haihan Zhang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in the University of Michigan
2025

Doctoral Committee:

Professor Xiang Zhou, Co-Chair
Associate Professor Lam C. Tsoi, Co-Chair
Professor Santhi K. Ganesh
Associate Professor Kevin He

Haihan Zhang

hhzhang@umich.edu

ORCID iD: 0000-0002-5902-7289

© Haihan Zhang 2025

Dedication

To my wife,

You made me complete!

Acknowledgements

As I approach the end of my doctoral journey and look back on the path I have traveled, it is abundantly clear that I did not walk it alone. This journey was made possible by a remarkable community of people who provided the encouragement, intellectual stimulation, and support that sustained me through every challenge.

First and foremost, I want to express my deepest gratitude to my advisors, Dr. Kevin He, Dr. Alex Tsoi, and Dr. Xiang Zhou, for their exceptional mentoring and steadfast support throughout my doctoral journey. I still remember reaching out to Dr. He as a first-year master's student, new to the department and transitioning from molecular biology to statistics with limited knowledge and an uncertain future. Dr. He patiently shared his own experiences, encouraging me to embrace both the opportunities and challenges of this new field. It was Dr. He who then referred me to Dr. Tsoi, my first guide into the world of statistical genetics and dermatology. I am grateful for the countless hours in Dr. Tsoi's office, where he would actively and patiently illustrate complex concepts on his whiteboard until they were crystal clear. I learned so much from his mentorship. Following Dr. Tsoi's advice, I reached out to Dr. Zhou to serve as my dissertation advisor. I was immediately impressed by Dr. Zhou's ability to grasp the core of my questions, even when I struggled to articulate them perfectly. He was always generous with his time and brilliant ideas, providing critical direction for my method development whenever I needed it. To all three of you, thank you. Your collective guidance and insightful comments during our bi-weekly meetings were instrumental in the completion of this dissertation.

My heartfelt appreciation also goes to Dr. Santhi K. Ganesh, who also serves in my dissertation committee. I was incredibly grateful for Dr. Ganesh's prompt and gracious acceptance of my invitation. Her immediate willingness to help, followed by her insightful suggestions during my dissertation proposal, were invaluable and significantly improved the quality of this dissertation.

My heartfelt thanks also extend to the many faculty, staff, and students in the Departments of Biostatistics and Dermatology who have supported me. I am particularly grateful to: Dr. James T. Elder and Dr. Johann E. Gudjonsson for their long-standing support and guidance on my GSRA work; Dr. Zheng Li, Dr. Lu Liu, and Dr. Boran Gao for their insightful feedback on my method development; and Rachael Bogle, Dr. Matthew T. Patrick, and Dr. Phillip E. Stuart for their invaluable assistance with data analysis.

Finally, I want to extend my heartfelt thanks to my lab mates and friends for their incredible companionship and support. Our countless conversations and shared joys form the friendships that I will cherish for a lifetime.

Table of Contents

Dedication ..... **ii**

Acknowledgements ..... **iii**

List of Tables ..... **viii**

List of Figures ..... **ix**

List of Appendices ..... **xii**

Abstract ..... **xiii**

Chapter 1 Introduction **1**

1.1 Transcriptome-wide association studies (TWAS) 1

1.2 Statistical fine-mapping in TWAS 2

1.3 Multi-tissue analysis in TWAS 3

1.4 Dissertation outline 4

Chapter 2 FABIO: TWAS Fine-mapping to Prioritize Causal Genes for Binary Traits **6**

2.1 Introduction 6

2.2 Materials and methods 8

2.2.1 TWAS fine-mapping of binary traits 8

2.2.2 Posterior inclusion probability of genes 9

2.2.3 Simulations 10

2.2.4 Compared methods 13

2.2.5 Real datasets 14

2.2.6 Real data application 15

2.2.7 Software overview 17
2.3 Results 17
2.3.1 Method overview and simulation setup 17
2.3.2 FABIO controls false signals in null simulations 19
2.3.3 FABIO improves power for fine-mapping causal genes in alternative simulations 20
2.3.4 Application to analyzing binary traits in UKBB 24
2.4 Discussion 31
2.5 Supplementary tables 34
2.6 Supplementary figures 37
Chapter 3 mFABIO: An Integrative Multi-tissue TWAS Fine-mapping Approach to Prioritize Causal Genes and Tissues underlying Binary Traits 51
3.1 Introduction 51
3.2 Materials and methods 53
3.2.1 Model for multi-tissue TWAS fine-mapping of binary traits 53
3.2.2 Decomposition of gene-tissue pair effects 54
3.2.3 Hierarchical prior for variable selection 55
3.2.4 Scalable variational inference 56
3.2.5 Calculation of posterior inclusion probability (PIP) 57
3.2.6 Simulations 58
3.2.7 Real datasets 62
3.2.8 Real data application 62
3.3 Results 64
3.3.1 Method overview 64
3.3.2 mFABIO controls false signals in null simulations 66
3.3.3 mFABIO produces calibrated PIPs and improves power for fine-mapping causal signals in alternative simulations 66

3.3.4 Application to analyzing binary traits in UKBB 70
3.4 Discussion 76
3.5 Supplementary figures 79
Chapter 4 Multi-cytokine based TWAS for Seven Inflammatory Skin Disorders Identify Candidate Causal Genes in Keratinocytes 86
4.1 Introduction 86
4.2 Materials and methods 87
4.2.1 GWAS and risk loci of seven common skin disorders 87
4.2.2 GReX modeling 87
4.2.3 TWAS analysis 88
4.2.4 Colocalization analysis 88
4.2.5 scRNA-seq and spatial transcriptomics data analysis 89
4.3 Results 89
4.4 Discussion 100
4.5 Supplementary tables 103
4.6 Supplementary figures 108
Chapter 5 Conclusion 114
5.1 Review of projects 114
5.2 Limitations and future directions 116
5.3 Summary 119
Appendices 120
Bibliography 138

List of Tables

Table 2. 1 Summary results of TWAS fine-mapping in UK Biobank. 25

Table S2. 1 Summary results of TWAS fine-mapping in UK Biobank (FABIO input options for π). 34
Table S2. 2 Summary results of TWAS fine-mapping in UK Biobank (five methods). 35
Table S2. 3 Summary results of TWAS fine-mapping in UK Biobank (down-sampled). 36

Table 3. 1 Gene-level results of multi-tissue TWAS fine-mapping in UK Biobank. 71
Table 3. 2 Gene-tissue pair-level results of multi-tissue TWAS fine-mapping in UK Biobank. . 73

Table 4. 1 Multi-condition TWAS have higher power in revealing disease-associated genes. 92
Table 4. 2 Identification of causal gene candidates for different skin immune-mediated disorders. 96

Table S4. 1 Number of single-condition TWAS significant genes identified in each cytokine condition. 103
Table S4. 2 Number of single-condition TWAS significant protein-coding genes identified in each cytokine condition. 104
Table S4. 3 Number of LD loci that contain TWAS signals. 105
Table S4. 4 Multi-condition TWAS highlight cytokine-specific associations. 106
Table S4. 5 Summary of colocalization analysis for different disease-cytokine pairs. 107

List of Figures

Figure 2. 1 Schematic overview of FABIO for TWAS fine-mapping of binary outcomes. 18
Figure 2. 2 True FDR at an estimated FDR threshold of 0.05 under different conditions. 21
Figure 2. 3 Performance of different methods in simulations with the change of case/control ratio or PVE1. 23
Figure 2. 4 Results of real data applications. 28

Figure S2. 1 Performance of different methods in controlling false signals. 37
Figure S2. 2 True FDR at an estimated FDR threshold of 0.05 under different conditions. 38
Figure S2. 3 Performance of different methods under the baseline simulation setting. 39
Figure S2. 4 Performance of different methods in simulations with the change of the number of causal SNP(s). 40
Figure S2. 5 Performance of different methods in simulations with the change of PVE_2. 41
Figure S2. 6 Performance of different methods in simulations with the change of sparsity of causal genes and SNPs. 42
Figure S2. 7 Power comparison for different methods based on an estimated false discovery rate (FDR) threshold of 0.05. 43
Figure S2. 8 Performance of FOCUS and FOGS with GWAS summary statistics generated by either logistic regression or linear regression. 44
Figure S2. 9 Performance of FABIO with default or beta prior on π. 45
Figure S2. 10 Performance of all five methods. 46
Figure S2. 11 TWAS Manhattan plot of asthma. 47
Figure S2. 12 Results of the Gene Ontology and KEGG pathway enrichment analysis. 48
Figure S2. 13 Box plot of pairwise correlation of GReX for four distinct groups of gene pairs. . 50

Figure 3. 1 Schematic overview of mFABIO for multi-tissue TWAS fine-mapping of binary outcomes. 65
Figure 3. 2 Power of methods under the true FDR threshold of 0.05 across different case:control ratios. 67
Figure 3. 3 Power of methods under the true FDR threshold of 0.05 across different numbers of causal tissue(s). 69
Figure 3. 4 Comparison of significant gene-tissue pairs located in relevant tissues. 74
Figure 3. 5 LocusZoom plots of selected gene candidates uniquely identified by mFABIO. 76

Figure S3. 1 Number of false signals in null simulations. 79
Figure S3. 2 The true false discovery rate (FDR) in alternative simulations at the gene-tissue pair level. 80
Figure S3. 3 The true false discovery rate (FDR) in alternative simulations at the gene level. ... 81
Figure S3. 4 Power of methods under the true FDR threshold of 0.05 across different PVE1 and numbers of causal cis-SNPs. 82
Figure S3. 5 Power of methods under the true FDR threshold of 0.05 across different PVE2. ... 83
Figure S3. 6 Statistical power in alternative simulations at the gene-tissue pair level. 84
Figure S3. 7 Statistical power in alternative simulations at the gene level. 85

Figure 4. 1 Workflow of the multi-condition TWAS on a skin disease. 90
Figure 4. 2 Multi-condition TWAS Manhattan plots of signal genes for different diseases. 95
Figure 4. 3 Plots for GReX modeling. 97
Figure 4. 4 Plots for the analysis of spatial transcriptomics. 100

Figure S4. 1 Venn diagrams showing numbers of genes specifically identified by single condition TWAS, multi-condition TWAS, or shared by both methods. 108
Figure S4. 2 TWAS Manhattan plots of signal genes for different diseases. 109
Figure S4. 3 Boxplot of measurements for different gene categories across diseases. 110
Figure S4. 4 Percentages of colocalized genes among TWAS significant (TWAS+) genes and among TWAS non-significant (TWAS-) genes across disease-cytokine pairs. 111

Figure S4. 5 Heatmap of TWAS signal expression across different cell types in scRNA-seq data. 112
Figure S4. 6 Evaluation plots. 113

List of Appendices

Appendix A: Chapter 2 Supplementary Text 120

A.1 Prior specification of $\boldsymbol{\pi}$ and $\boldsymbol{\sigma 2}$ 120

A.2 Posterior sampling 121

A.3 Model specifications for FOCUS 123

A.4 Model specifications for FOGS 124

Appendix B: Chapter 3 Supplementary Text 125

B.1 Variational inference overview 125

B.2 Update of latent variable $\boldsymbol{z}$ 126

B.3 Update of $\boldsymbol{\mu}$ 127

B.4 Update of $\boldsymbol{\alpha}$ and $\boldsymbol{\sigma\alpha 2}$ 127

B.5 Update of $\boldsymbol{\beta l}$, $\boldsymbol{\sigma\beta 2}$, and $\boldsymbol{\gamma l}$ 128

B.6 Update of $\boldsymbol{\pi G}$ and $\boldsymbol{\pi T}|*$ 132

B.7 Summary of variational inference 134

Appendix C: Chapter 4 Supplementary Text 135

C.1 Cell culture 135

C.2 Bayesian Sparse Linear Mixed Model (BSLMM) 135

C.3 Evaluation of the BSLMM GReX models 136

C.4 TWAS analysis 136

C.5 Cytokine-specific associations 137

Abstract

Genome-wide association studies (GWAS) have successfully identified thousands of genetic variants, primarily single nucleotide polymorphisms (SNPs), associated with complex human diseases and traits. However, many of these associated SNPs reside in non-coding regions of the genome, making it difficult to pinpoint the causal genes and understand the underlying biological mechanisms. To bridge this interpretive gap, transcriptome-wide association studies (TWAS) were developed. This thesis advances the TWAS framework by first proposing novel TWAS fine-mapping methods. Subsequently, we apply both conventional marginal TWAS and our newly developed methods in a large-scale analysis of inflammatory skin disorders, successfully nominating candidate genes and providing deeper insight into the genetic architecture of these conditions.

In Chapter 1, we lay the groundwork for this dissertation, beginning with the fundamental concepts of transcriptome-wide association studies (TWAS). We first review how this framework leverages reference transcriptome data to link genetically predicted gene expression with complex traits. Then, we explore the expansion of these methods into multi-tissue analyses, discussing how both marginal and fine-mapping approaches address the challenge of tissue-specific genetic effects. Finally, we present a comprehensive roadmap for the remaining chapters of this work.

In Chapter 2, we developed FABIO, a novel computational method for TWAS fine-mapping specifically designed for binary traits. Unlike existing methods tailored for quantitative traits, FABIO employs a probit model to directly link multiple genetically regulated expressions (GReX) to a binary outcome. To account for complex correlations arising from linkage disequilibrium and co-regulation, FABIO uniquely models all genes jointly on a chromosome-wide scale. We demonstrate through comprehensive simulations and application to six diseases in the UK Biobank that FABIO achieves superior fine-mapping resolution, producing candidate gene sets smaller than those from current methods.

In Chapter 3, we further extended this work with mFABIO, a novel computational method for multi-tissue TWAS fine-mapping of binary traits. mFABIO adapts the probit model to jointly analyze gene-tissue pairs in a locus, while also accounting for horizontal pleiotropic effects and complex correlations in GReX among both genes and tissues. We validated mFABIO through extensive simulations and by analyzing six binary diseases from the UK Biobank. The results highlight a dual advantage: mFABIO increased the discovery of evidence-supported gene-tissue pairs, while simultaneously improving specificity by reducing the candidate set size.

In Chapter 4, we apply our methodologies to dissect the genetic basis of inflammatory skin diseases. We first conducted large-scale marginal TWAS using transcriptomic data from keratinocytes under eight distinct pro-inflammatory cytokine stimulations, covering seven common skin diseases. We identified 138-328 significant genes per disease within 211 independent loci. To demonstrate the advanced fine-mapping capabilities of our new method in more complex regions, we then applied mFABIO to the multi-condition psoriasis data. This successfully prioritized causal genes from the broader set of associations found in the marginal analysis, showcasing mFABIO's utility in resolving gene signals within a real-world, multi-condition biological system.

In Chapter 5, we conclude by summarizing the body of work presented in this dissertation. This final chapter is structured into three main parts: first, a review of the key findings and conclusions from each preceding project. Second, a frank discussion of the limitations of our studies, addressing the scope and boundaries of our interpretations. Third, the forward-looking that addresses the limitations and proposes promising directions for future research that emerge directly from our findings.

Chapter 1 Introduction

1.1 Transcriptome-wide association studies (TWAS)

Genome-wide association studies (GWAS) have successfully identified thousands of genetic loci associated with complex human traits and diseases. However, a significant challenge remains in translating these statistical associations into biological mechanisms. The majority of single nucleotide polymorphisms (SNPs) identified by GWAS reside in non-coding regions of the genome, suggesting their effects on regulating gene expression rather than on altering protein structure (Maurano et al., 2012). Therefore, the transcriptome-wide association studies (TWAS) were developed as a powerful method to bridge this gap by systematically integrating GWAS data with gene expression data to identify genes whose expression levels are associated with a complex trait.

The core principle of TWAS is to test for an association between the genetically predicted component of gene expression and a phenotype of interest. This is achieved in a two-stage process. First, a predictive model of gene expression is built using a reference dataset that contains both genotype and gene expression data for the same individuals, such as the Genotype-Tissue Expression (GTEx) project (Carithers & Moore, 2015). For each gene, its expression level is modeled as a function of the genotypes of SNPs in its local (cis-) region. Elastic net (Zou & Hastie, 2005) or LASSO (Tibshirani, 1996) regression have been proposed, as they can effectively select a sparse set of predictive cis-SNPs from the highly correlated genetic data. The resulting model captures the genetic component of gene expression, often referred to as the genetically regulated expression (GReX). Second, the weights from these prediction models are applied to the genotypes of individuals in a much larger GWAS cohort, for which direct gene expression measurements are not available. This step generates the GReX of all the available genes for each individual in the GWAS dataset. Finally, the association between the GReX and the complex trait is tested, effectively collapsing the effects of multiple SNPs into a single, gene-based test. Several software

packages have been developed to perform TWAS, like PrediXcan (Gamazon et al., 2015) and FUSION (Gusev et al., 2016).

Compared to the traditional single-variant GWAS analysis, TWAS provide the better biological interpretability. By directly implicating genes rather than individual SNPs, TWAS provides a more direct and biologically interpretable link between a genetic locus and a phenotype. This helps to prioritize genes for the downstream functional experiments. Secondly, TWAS can increase statistical power. Through aggregated effects of multiple cis-eQTLs (expression quantitative trait loci), TWAS can boost the statistical power to detect associations, particularly when a gene's expression is regulated by many variants with small individual effects (Gusev et al., 2016). Finally, TWAS significantly reduce the burden of multiple hypothesis testing by shifting from testing millions of SNPs to testing approximately 20,000 protein-coding genes. Through these strengths, TWAS have been successfully applied across a wide range of diseases, including schizophrenia (Gusev et al., 2018), cardiovascular disease (Li et al., 2022; Zhou et al., 2022; Zhuang et al., 2021), and autoimmune disorders (Liu et al., 2023; Wen et al., 2024), identifying novel and functionally relevant disease-associated genes.

1.2 Statistical fine-mapping in TWAS

Despite their utility, conventional TWAS have inherent limitations. A significant association between a gene's predicted expression and a complex trait does not definitively establish a causal relationship. Such an association can be misleading due to confounding genetic effects. For instance, pleiotropy, where a single genetic variant independently influences both gene expression and the trait through distinct biological pathways, can create a spurious link (Solovieff et al., 2013). Similarly, linkage disequilibrium (LD) - the non-random association of alleles at different loci - can induce a significant TWAS signal if the variants used to predict gene expression are merely correlated with the true, unobserved causal variant for the trait (Zhu et al., 2016). These challenges necessitate more advanced statistical methods to dissect TWAS signals and pinpoint the specific genes driving disease risk.

Therefore, methods for statistical fine-mapping in TWAS have been developed, which aim to move beyond single-gene associations to identify the specific gene or set of genes at a given locus that are causally responsible for the association with a trait. This process is crucial for distinguishing truly causal genes from those confounders due to LD or pleiotropy. The field has

evolved rapidly, with several key methods emerging to address this challenge. FOCUS (Fine-mapping Of Causal gene Sets) was a pioneering method in this area (Mancuso et al., 2019), which adapts the principles of genetic fine-mapping to the gene level. It takes summary-level data from GWAS and eQTL studies as input and uses a Bayesian framework to compute a posterior inclusion probability (PIP) for each gene in a locus. The model considers all possible combinations of causal genes within a region and assigns a probability to each gene being part of the true causal set. This allows researchers to prioritize genes with high confidence and construct a "credible set" of genes that likely contains the causal one(s).

Following FOCUS, several advanced TWAS fine-mapping methods were developed later. FOGS (Fine-mapping Of Gene Sets), a frequentist method, relies on a penalty function to model the gene effect sizes on the outcome trait and applies gene set testing for fine-mapping (Wu & Pan, 2020). GIFT (gene-based integrative fine-mapping through conditional TWAS) performs conditional TWAS analysis by explicitly adjusting for the GReX of other local genes to fine-map putatively causal genes (Liu et al., 2024). cTWAS (causal-TWAS) assesses the focal genes' association with a trait through a Bayesian regression model, while controlling for the effects of nearby genes and variants (Zhao et al., 2024). All these fine-mapping methods were tested in the real-world applications, and successfully prioritized gene candidates by analyzing multiple genes simultaneously.

1.3 Multi-tissue analysis in TWAS

The genetic regulation of gene expression is often highly specific to a particular tissue or cell type. For many complex traits and diseases, the biologically relevant tissue where genetic effects manifest is unknown. Performing a standard TWAS using a single, potentially incorrect, tissue (e.g., using the most accessible blood eQTLs to study a psychiatric disorder) can lead to a loss of statistical power and may miss the true causal gene entirely. Multi-tissue TWAS methods were developed to address this challenge by systematically integrating gene expression information from multiple tissues to increase the power of discovery and help identify the most relevant etiological context.

The first generation of multi-tissue TWAS methods focused on aggregating evidence from single-tissue analyses to produce a single, more powerful test of marginal association. These methods are particularly effective when a gene's regulatory architecture is shared across several

tissues. MultiXcan was the first attempt in this field (Barbeira et al., 2019), which serves as a direct extension of PrediXcan. It first performs a TWAS association test for a gene in every available tissue individually, and then combines these single-tissue association statistics into a unified, omnibus test statistic. By accounting for the correlation in gene expression across tissues (e.g., a gene's expression in the liver is often correlated with its expression in the kidney), MultiXcan effectively boosts power to detect associations that are present in multiple tissues without overly penalizing for multiple testing. UTMOST (unified test for molecular signatures) is the other popular method for multi-tissue TWAS, that takes a different approach (Hu et al., 2019). Instead of combining association statistics at the end, it improves the initial gene expression prediction model. It builds a single, cross-tissue model that leverages shared regulatory patterns to predict a gene's expression more accurately than single-tissue models. This enhanced prediction model is then used in a standard TWAS framework to test for association with a trait.

More recently, TGFM (tissue-gene fine-mapping) was developed as a multi-tissue TWAS fine-mapping method that calculates the posterior inclusion probability (PIP) for each specific gene-tissue pair being causally linked to a disease (Strober et al., 2025). TGFM jointly models genes across multiple tissues, accounting for complex correlation structures, including linkage disequilibrium (LD) between variants and the correlated expression of genes across different tissues. Such TWAS fine-mapping methods allow the precise identification of the causal gene, the relevant tissue, and the underlying mechanism driving a GWAS signal.

1.4 Dissertation outline

This dissertation addresses key challenges in the rapidly evolving field of TWAS by advancing its methodologies, with a specific focus on statistical fine-mapping. Furthermore, it demonstrates the translational power of these techniques by applying both conventional and fine-mapping TWAS frameworks to investigate the genetic architecture of inflammatory skin disorders, integrating insights from multi-omics data.

In Chapter 2, we addressed a key gap in TWAS fine-mapping by introducing FABIO, a method tailored for binary traits. While all existing methods are designed for quantitative outcomes, FABIO leverages a probit model to analyze the joint effect of multiple genetically regulated expressions (GReX) on a binary disease state. Uniquely, FABIO models all genes on a chromosome-wide scale to account for confounders from linkage disequilibrium and co-

regulation. We validated this approach with extensive simulations and analyses of six UK Biobank diseases, showing that FABIO provides enhanced fine-mapping resolution and identifies significantly smaller candidate gene sets than competing methods.

In Chapter 3, we addressed the lack of TWAS fine-mapping methods that leverage multi-tissue data for binary outcomes by developing mFABIO. This novel computational method adapts a probit model to jointly analyze gene-tissue pairs, controlling for both horizontal pleiotropy and complex GReX correlations. Through comprehensive simulations and application to six binary diseases, we demonstrate mFABIO's twofold superiority: it boosts the discovery of causal gene-tissue pairs compared to competing methods while simultaneously enhancing specificity, yielding more refined candidate sets than existing single-tissue approaches.

In Chapter 4, we demonstrated the real-world application of TWAS and our methods to inflammatory skin diseases. We began by conducting a broad, marginal TWAS using multi-condition transcriptomic data from keratinocytes, which revealed a wide range of gene-disease associations. To resolve these signals, we then deployed mFABIO for a multi-condition fine-mapping analysis of the psoriasis data. This second step successfully distilled the initial list of candidates into a core set of probable causal genes, showcasing mFABIO's ability to pinpoint key drivers within a complex, real-world biological system.

In summary, this dissertation introduces a novel suite of TWAS fine-mapping methods specifically designed for binary traits. Our core innovation is the use of a probit model to directly link genetically regulated expression (GReX) to disease status. Chapter 2 establishes this framework in a single-tissue context, which is then extended in Chapter 3 to a more complex, multi-tissue setting. Finally, Chapter 4 demonstrates the practical power of this work by applying these methods alongside conventional TWAS to dissect the genetic architecture of complex diseases, showcasing their ability to resolve causal signals for binary outcomes.

Chapter 2 FABIO: TWAS Fine-mapping to Prioritize Causal Genes for Binary Traits

2.1 Introduction

Transcriptome-wide association studies (TWAS) have gained significant popularity in the field of genetics. They integrate gene expression mapping studies with genome-wide association studies (GWAS) to uncover potentially causal associations between genes and traits (Gamazon et al., 2015; Gusev et al., 2016; Zhu et al., 2016). However, the traditional approach of conducting marginal TWAS analyses, which examine one gene at a time, has important limitations. In such analyses, the focus is solely on testing the marginal association between genetically regulated expression (GReX) of a single gene and a trait of interest. This approach overlooks the fact that neighboring genes often exhibit correlated expression levels and share cis-SNPs that are in close linkage disequilibrium (LD) with each other (Mancuso et al., 2019; Wainberg et al., 2019). Consequently, genes detected through marginal TWAS may not necessarily be causal but rather represent tag genes in the region. To overcome the limitation of marginal TWAS analysis, several statistical methods for TWAS fine-mapping, including FOCUS (fine-mapping of causal gene sets) (Mancuso et al., 2019) and FOGS (Fine-mapping Of Gene Sets) (Wu & Pan, 2020), have been developed (Liu et al., 2024; Schaid et al., 2018; Spain & Barrett, 2015). These TWAS fine-mapping methods can prioritize associated genes by jointly modeling multiple genes within a specific region of interest. The TWAS fine-mapping methods begin by constructing GReX prediction models using data from eQTL mapping studies (Gamazon et al., 2015; Zou & Hastie, 2005). The constructed prediction models are then used to impute GReX in GWAS, enabling the association testing between GReX and the trait of interest at the locus level. Different TWAS fine-mapping methods employ distinct modeling frameworks. Specifically, FOCUS, a Bayesian method, specifies a sparsity-inducing prior on gene effect sizes on the outcome trait and uses Markov chain Monte Carlo (MCMC) for modeling inference. Conversely, FOGS, a frequentist method, relies on a penalty function to model the gene effect sizes on the outcome trait and applies

gene set testing for fine-mapping. TWAS fine-mapping analyses have been carried out in many studies, facilitating the discovery of potentially causal gene-trait associations.

Despite their widespread usage, current TWAS fine-mapping methods have two important limitations. The first limitation is that these methods were primarily designed to model quantitative traits, which are traits that can be measured on a continuous scale, such as height or blood pressure. When applied to binary disease traits, both FOCUS and FOGS still apply a linear regression framework to model the association between multiple GReX and the binary outcome trait (details in the Appendix A), relying on input of marginal GWAS summary statistics generated from either linear regression (Hautakangas et al., 2022; Jones et al., 2021) or logistic regression. Such linear modeling approach of TWAS aligns with the common practice in GWAS association analysis, as a linear regression can be viewed as an effective approximation to logistic regression when effect sizes are small (Zhou et al., 2013). However, unlike the SNP effect sizes in GWAS, the GReX effect sizes observed in TWAS analysis can be relatively large. Consequently, employing linear regression as an approximation to logistic regression in TWAS fine-mapping settings may not accurately capture the binary nature of the outcome, potentially leading to a loss of power. The second limitation of existing TWAS fine-mapping methods is, they only consider the correlation among GReX within local genomic regions defined by LD blocks. However, cis-SNPs for genes that are outside the same genomic region may still be in LD with each other, resulting in potential correlation among GReX across genomic regions. In addition, correlated gene expression may be observed for pairs of genes located outside local LD blocks, again leading to potential correlation among GReX across genomic regions. Moreover, focusing solely on GReX of genes located within local genomic regions may be restrictive. Such an approach may miss causal genes located in other genomic regions that also contribute to the trait of interest (Klebanov & Yakovlev, 2007; Morley et al., 2004), potentially leading to unnecessarily increase in residual error variance and subsequent loss of power. Therefore, modeling multiple genes on an entire chromosome may offer advantages. It can capture GReX correlations that are not captured by modeling small genomic regions individually, thereby potentially increasing the power of TWAS fine-mapping.

Here, we introduce a novel TWAS fine-mapping method called FABIO (Fine-mApping of causal genes for BInary Outcomes) that effectively addresses the aforementioned limitations. FABIO relies on a probit model to directly relate multiple GReX to binary outcome in TWAS fine-mapping. Additionally, it jointly models all genes located on a chromosome to account for the

correlation among GReX arising from cis-SNP LD and expression correlation across genomic regions. We demonstrate the effectiveness of FABIO through comprehensive simulations. We also applied FABIO to analyze six binary disease traits in the UK Biobank dataset, where FABIO was able to narrow down a small set of causal genes, with the set size 27.9-36.9% smaller than those yielded by current TWAS fine-mapping methods. Notably, FABIO prioritized multiple candidate genes previously validated for their association with the respective diseases, such as *GATA3* for asthma; *ABCG2* for gout; and *SH2B3* for hypertension. Putative causal genes identified by FABIO were further validated through enrichment analyses, with disease-associated pathways highlighted in asthma and hypertension. These results support the superior power of FABIO in prioritizing putatively causal genes associated with binary traits.

2.2 Materials and methods

2.2.1 TWAS fine-mapping of binary traits

We consider a genome-wide association study (GWAS) performed on n individuals for a binary trait of interest. We denote $\mathbf{y}$ as an n-vector of binary phenotypes measured on the n individuals. Our goal is to identify genes whose genetically regulated expression (GReX) is associated with the outcome trait through TWAS fine-mapping. Different from the existing TWAS fine-mapping methods (Mancuso et al., 2019; Wu & Pan, 2020), we aim to jointly model all genes on the same chromosome while explicitly accounting for the binary nature of the outcome trait. To do so, we examine one chromosome at a time. For the chromosome of focus, we denote $\mathbf{G}$ as an $n \times p$ GReX matrix constructed for p genes on the chromosome for the same set of individuals. These GReX are assumed to have already been constructed using standard software such as PrediXcan (Gamazon et al., 2015). With the constructed GReX, we consider a probit model linking the GReX to the binary outcome:

$$P(y_i = 1|\widehat{\boldsymbol{g}}_i, \boldsymbol{\alpha}) = 1 - P(y_i = 0|\widehat{\boldsymbol{g}}_i, \boldsymbol{\alpha}) = \Phi\left(\mu + \widehat{\boldsymbol{g}}_i{}^T \boldsymbol{\alpha}\right)(i = 1, \cdots, n), \quad (2.1)$$

where y_i is the binary trait for the i'th individual; $\widehat{\boldsymbol{g}}_i$ is a p-vector of GReX for p genes in the i'th individual; $\boldsymbol{\alpha}$ is a p-vector of corresponding gene effect sizes; μ is a scalar representing the intercept; and Φ is the cumulative distribution function (CDF) of the standard normal distribution.

Following (Albert & Chib, 1993), we introduce a p-vector of auxiliary variables $\boldsymbol{z}$ and obtain the equivalent latent variable representation of equation (2.1) as:

$$y_i = \begin{cases} 1, & if z_i > 0 \\ 0, & if z_i \leq 0' \end{cases} \tag{2.2}$$

$$z_i = \mu + \widehat{\boldsymbol{g}}_i{}^T \boldsymbol{\alpha} + \epsilon_i \epsilon_i \sim N(0,1), \tag{2.3}$$

where the residual error ϵ_i follows the standard normal distribution; and z_i is the i'th element of $\boldsymbol{z}$.

Note that the GReX are included as covariates, with the weights in the linear combination of GreX ($\boldsymbol{\alpha}$) automatically inferred from the model to account for their correlations. As a result, the pair-wise GReX correlation matrix is effectively accounted for in the model. To enable TWAS fine-mapping, we specify a sparsity inducing prior on each element of the gene effect sizes $\boldsymbol{\alpha}$ using a point-normal distribution

$$\alpha_j \sim \pi N(0, \sigma^2) + (1 - \pi)\delta_0, \tag{2.4}$$

where α_j is the j'th element of $\boldsymbol{\alpha}$; π is the proportion of non-zero $\boldsymbol{\alpha}$; and δ_0 denotes a point mass at zero. The above equation effectively assumes that, with probability π, the j'th gene has a non-zero effect and its effect size follows a normal distribution with mean zero and variance σ^2. With probability 1-π, the j'th gene has zero effect. Among the two hyper-parameters in the sparsity inducing prior, π controls the proportion of non-zero $\boldsymbol{\alpha}$ values in equation (2.3) while σ^2 controls the expected magnitude of the non-zero $\boldsymbol{\alpha}$ values. Specifically, for π, we assume a uniform prior distribution on $log\ \pi$ by default, with an alternative option of a beta prior on π. The prior specifications for π and σ^2 follow that of (Guan & Stephens, 2011) and are described in detail in the Appendix A.

2.2.2 Posterior inclusion probability of genes

Based on the modeling specification described above, we developed a Markov chain Monte Carlo (MCMC) algorithm to estimate the model parameters and obtain test statistic for each gene effect size α_j. To do so, we first introduce a set of binary indicator variables (γ_j) to represent whether the j'th gene has non-zero effects or not. We then apply a Metropolis-Hastings algorithm within the MCMC framework to obtain posterior samples of the hyperparameters including γ_j. To improve the convergence of the MCMC algorithm, we implement a technique known as "small world proposal" (Guan & Krone, 2007), where we introduce multiple local moves instead of a single move to induce longer-range proposals on the proposal distributions. The number of compounded local moves is uniformly drawn from a uniform distribution ranging from 1 to 20.

Full details of the sampling algorithm are described in the Appendix A. In the present study, we used a total of 26,000 iterations, with the first 6,000 as the burn-in and the rest as the sampling iterations. With the MCMC algorithm, we obtain for each gene its posterior probability of having a non-zero effect, or $\gamma_j = 1$, a quantity commonly referred to as the posterior inclusion probability (PIP). This PIP serves as an important measure of evidence for the gene's association with the outcome trait. We refer to our method as FABIO (Fine-mApping of causal genes for BInary Outcomes).

2.2.3 Simulations

We conducted simulations to assess the performance of FABIO and compare it with existing fine-mapping methods FOCUS and FOGS. To do so, we used real genotype data and simulated both gene expression and the binary outcome trait. Specifically, we randomly sampled 50,000 individuals of European ancestry in UK Biobank to serve as the GWAS data. We used all 373 individuals of European ancestry in GEUVADIS to serve as the eQTL mapping data (more details on UK Biobank and GEUVADIS are provided in the next section). We focused on chr 1 for our simulations, obtained 405,875 SNPs on chr 1 that are shared between the GWAS and eQTL data, and partitioned these SNPs into 133 independent LD blocks defined by LDetect (Berisa & Pickrell, 2015), following the same procedure of (Mancuso et al., 2019; Wu & Pan, 2020). These LD blocks contain 1 - 60 genes per block (mean = 12, median = 8). We retained 88 LD blocks with at least 5 genes and randomly selected 50 LD blocks among them to carry out the simulations. For the selected LD blocks, we obtained the genes inside them (5 - 45 genes per block, mean = 16, median = 13), and then extracted a common set of cis-SNPs that reside within 100 kb upstream of the transcription start site (TSS) and 100 kb downstream of the transcription end site (TES) of each gene following (Liu et al., 2021; Yuan et al., 2020). We kept 770 genes that had at least 10 cis-SNPs and used these SNPs and genes for simulations. These genes contain a medium of 364 SNPs (average=370) per gene, with over 95% of them having more than 100 SNPs.

We simulated the gene expression in both the GWAS and eQTL datasets following the strategy of (Yuan et al., 2020). Specifically, for each gene in turn, we randomly selected k SNPs to have non-zero effects on gene expression, and then simulated its gene expression level as:

$$g_i = \boldsymbol{m}_i^T \boldsymbol{\beta} + \varepsilon_i \varepsilon_i \sim N(0, 1 - PVE_1), \tag{2.5}$$

$$\beta_j \sim N(0, PVE_1/k). \tag{2.6}$$

where g_i is the simulated gene expression level for the i'th individual (in either the GWAS data or eQTL data), $\boldsymbol{m}_i$ is a k-vector of standardized genotypes for the non-zero effect SNPs for the i'th individual, $\boldsymbol{\beta}$ is a k-vector of simulated SNP effect sizes with β_j being the j'th element, and PVE_1 is a scalar that represents the proportion of gene expression variance explained by genetic effects. We will describe the choice of k in a following paragraph.

We then randomly selected n_g genes to have non-causal effects on the trait, and simulated a binary trait based on the simulated gene expression in the GWAS panel using a logistic regression model:

$$P\{y_i = 1|z_i\} = \frac{1}{1+exp\,(-z_i)}, \tag{2.7}$$

$$z_i = u + \boldsymbol{g}_i^T\boldsymbol{\alpha} + \epsilon_i \epsilon_i \sim N\left(0,1 - \left(n_g \cdot PVE_2\right)\right). \tag{2.8}$$

where y_i is the simulated binary trait for the i'th individual; z_i is the latent variable for the corresponding individual; u is an intercept that controls the case/control ratio as it equals to the expectation of $log\ (case\#/control\#)$; $\boldsymbol{g}_i$ is a n_g-vector of simulated gene expression of the selected causal genes; n_g is the number of selected causal genes; $\boldsymbol{\alpha}$ is a n_g-vector of genes' causal effects. Following (Yuan et al., 2020), we set the non-zero effect $\alpha_j = \sqrt{PVE_2/PVE_1}$ for every element of $\boldsymbol{\alpha}$, where PVE_2 is a scalar that denotes the proportion of variance in z_i that is explained by the causal genes. Note that we used a logistic regression instead of a probit model to evaluate the performance of our method under model misspecifications. We will describe the choice of n_g in the following paragraph.

In the simulations, we first examined the calibration of test statistics from different methods under two null simulation settings: one is the complete null setting, where we set $PVE_2 = 0\%$ so that none of the genes were associated with the trait; the other is a masked null setting, where we randomly selected 10 LD blocks, randomly assigned one gene in each block to be causal, and masked the causal genes by excluding them in the fine-mapping analysis. Besides null simulations, we also examined the power of different methods on detecting the causal genes under the alternative simulation settings with non-zero PVE_2. For both complete null and alternative simulations, we started from a baseline simulation setting and then varied one parameter at a time on top of the baseline setting to examine the influence of different parameters. Specifically, for PVE_1, we set it to be 1%, 5%, 10% (baseline setting), or 20% following the same settings of (Mancuso et al., 2019). For the number of SNPs that display non-zero effects on gene expression

(k in equation (2.6)), we set it to be either 1, 2, 1% (baseline setting), or 10% of all cis-SNPs following (Mancuso et al., 2019). For PVE_2, we varied its value to be either 0% (complete null setting), 0.2%, 0.4% (baseline setting), or 0.6% following (Yuan et al., 2020). For n_g, we followed FOCUS and assumed one causal gene in 10 of the 50 LD blocks, with no causal genes in the remaining 40 blocks. In addition, we noticed that the main simulations in FOGS paper assumed one causal gene per LD block, including a non-sparse case of all the cis-SNPs to be causal for each causal gene. Therefore, we considered four distinct settings covering both non-sparse and sparse conditions: one causal gene in 10 of the 50 LD blocks with either 1% or 100% cis-SNPs to be causal; or one causal gene in each of the 50 LD blocks with either 1% or 100% cis-SNPs to be causal, and all other parameters followed the baseline settings. For the binary trait, we examined different case:control ratios by varying the intercept u, using 1:1 as the baseline setting ($u = 0$) while exploring other ratios of 1:10 ($u = log\ (0.1)$), 1:20 ($u = log\ (0.05)$), and 1:50 ($u = log\ (0.02)$).

FABIO places a uniform prior on log π by default. Additionally, it allows the users to assume π following a beta prior distribution: $\pi \sim Beta(a, b)$, and input *a* and *b* based on their prior knowledge to control the expectation and variance of π. Here, we compared the performance of FABIO using default prior against the beta prior on π with different combinations of *a* and *b* in the simulations under the baseline setting. When applying FABIO with the beta prior on π, we first tried three combinations with fixed expectation and different variances: (1) *a*=1.1 and *b*=99; (2) *a*=0.11 and *b*=9.9; (3) *a*=11 and *b*=990, all of which leads to the same prior expectation $E[\pi]$=1.1/100, which is close to the true π in our baseline setting: 10/770. Furthermore, we included two additional parameter combinations to test the influence of prior expectation on method performance: (4) *a*=1.1 and *b*=9.9; (5) *a*=1.1 and *b* =999, which either over-estimated or under-estimated proportion of causal genes as compared to the truth, respectively.

We performed 100 simulation replicates for each simulation setting. In each replicate, we applied FOCUS and FOGS to analyze one LD block at a time and applied FABIO to either analyze one LD block at a time or analyze all 50 LD blocks jointly. We then summarized the results across simulation replicates. In the analysis, we first examined the calibration of test statistics from different methods under the null simulation settings. For the frequentist method FOGS, we examined the number of false discoveries at the Bonferroni adjusted p-value threshold. For the other two Bayesian methods FABIO and FOCUS, we examined the number of false discoveries

using the 95% credible set based on PIPs. Given those methods produce different test statistics, we evaluated the power of different methods using two rank-based criteria: (a) the power under a false discovery rate (FDR) of 0.05; (b) the area under the curve (AUC) based on the receiver operating characteristics (ROC) curves. Since we knew the ground truth in the simulations, we first compared the power of different methods under a true FDR threshold of 0.05. Specifically, to compute the true FDR in the simulations, we first divided 100 simulation replicates into five groups, with 20 replicates per group. In each group, we obtained the test statistics from each method for all genes across the 20 replicates, either in the form of p-values or PIPs. For each method in turn, we sorted the test statistics from the most significant (i.e. the smallest p-value or the largest PIP) to the least significant (i.e. the largest p-value or the smallest PIP). For each ordered test statistic in the sorted list, we counted the number of false positives detected above the threshold defined by this specific test statistic. The number of false positives, divided by the total number of genes called significant at that threshold, is FDR. We then determined threshold that corresponds to an FDR of 0.05, based on which we further calculated power. Power is calculated as the number of true positives above the threshold in the sorted list divided by the total number of causal genes. Afterwards, we obtained the average power across five groups to minimize stochasticity (Liu et al., 2024; Song et al., 2021; Song et al., 2020; Yuan et al., 2020).

Because we can only compute the true FDR in the simulations but not in the real data, we also compared the power of different methods under an estimated FDR threshold of 0.05. We estimated FDR using different approaches: the Benjamini-Hochberg approach (Benjamini & Hochberg, 1995) to estimate FDR using p-values from the frequentist methods, and the local FDR approach (Efron, 2007) to estimate FDR using PIPs from the Bayesian methods. Specifically for the local FDR approach, we first calculated the local FDR for all the genes, which is effectively 1-PIP (Gao & Zhou, 2024; Liu et al., 2024). We then sorted genes based on these local FDRs from small to large. For each gene in turn, we estimated FDR for that gene by aggregating the local FDRs from the first gene up to that gene in the sorted list. Afterwards, we identified the cutoff value of PIP/local FDR that corresponds to a targeted FDR threshold of 0.05. This is the significant threshold we use that corresponds to an estimated FDR of 0.05.

2.2.4 Compared methods

We compared FABIO with two existing TWAS fine-mapping methods FOCUS (Version 0.6.10) and FOGS (Version 2.0), and two recently published methods GIFT (Version 1.0.0) and

cTWAS (Version 0.1.37) (Zhao et al., 2024). FOCUS and cTWAS are Bayesian methods that use a two-stage regression framework for fine-mapping. Both of them first obtain the SNP prediction weights on gene expression from the eQTL mapping study, and then plugs in the estimated SNP weights on expression in GWAS summary statistics for TWAS fine-mapping. FOGS and GIFT are frequentist fine-mapping method that use a two-stage regression framework. FOGS relies on a SNP set test to examine the joint effects of all SNPs within a gene, and controls for the SNP effects of the other genes in the same region; while GIFT examines one genomic region at a time, jointly models the GReX of all genes residing in the focal region, and carries out TWAS conditional analysis in a maximum likelihood framework. In practice, we noticed that both GIFT and cTWAS are computationally demanding: it takes about three days for either method to analyze on a relatively small LD block with <10 genes and over a week on a moderate-sized LD block with >20 genes in the simulations. This is in direct contrast to the other methods: for example, it took FABIO, FOCUS and FOGS an average of 3.1, 1.5 and 9.2 hours to analyze all 50 LD blocks in the simulations, respectively. Due to the heavy computational demand of these two methods, we limited comparisons to the baseline simulation setting, and further down-sampled to 5,000 GWAS individuals to speed up GIFT, as we used individual-level GWAS data for GIFT following the developers' advice to achieve the best performance.

We used the Bayesian sparse linear mixed models (BSLMM) (Zhou et al., 2013) to construct gene expression prediction models for FABIO, FOCUS and GIFT, while we used the elastic net regression (Zou & Hastie, 2005) for FOGS following the recommendation from its developer to improve its performance, also for cTWAS following the developers' suggestions that cTWAS performs best when prediction models are sparse, and the computing time will increase dramatically for cTWAS with increasing density of variants. With the SNP weights, we directly constructed GReX in the GWAS data for FABIO and GIFT, then carried out TWAS fine-mapping analysis. For the other three methods, we first performed marginal association analysis in the GWAS data with logistic regression using PLINK 2.0 (Chang et al., 2015). This step was used to obtain GWAS summary statistics. We then paired the GWAS summary statistics with the SNP weights as inputs for them. For those methods, we calculated the LD matrix using the same set of individuals in UKBB.

2.2.5 Real datasets

2.2.5.1 GEUVADIS data

The GEUVADIS data (Lappalainen et al., 2013) contain gene expression measurements for 462 individuals from five different populations that include CEPH (CEU), Finns (FIN), British (GBR), Toscani (TSI), and Yoruba. We combined samples from the four European ancestries (CEU, FIN, GBR, and TSI) and obtained a final sample size of 373. In the expression data, we only focused on protein-coding genes annotated from GENCODE (Harrow et al., 2012) (release 12), and removed lowly expressed genes that had zero counts in at least half of the individuals to obtain a total of 14,698 protein-coding genes. We performed PEER (Stegle et al., 2012) normalization to remove confounding effects. Following (Zeng & Zhou, 2017) to remove potential population stratification, we first quantile-normalized the gene expression measurements across individuals in each of the four European populations to a standard normal distribution, and then further quantile-normalized the expression measurements to a standard normal distribution across all 373 individuals from all four European populations. In addition to expression data, we also obtained genotype data of the 373 individuals from the 1000 Genomes Project phase 3. We filtered out SNPs that have a Hardy-Weinberg equilibrium (HWE) p-value < 10^{-4}, a genotype call rate < 95%, or a minor allele frequency (MAF) < 0.05. We retained a final set of 5,395,690 SNPs for analysis.

2.2.5.2 UK Biobank (UKBB) data

The UK Biobank data consists of 487,409 individuals and 92,693,895 imputed autosome SNPs (Bycroft et al., 2018). We followed the same sample QC procedure as Neale Lab (https://github.com/Nealelab/UK_Biobank_GWAS/tree/master/imputed-v2-gwas) to retain a total of 337,198 individuals of European ancestry. We performed SNP QC by filtering out SNPs with a minor allele frequency (MAF) < 0.01, with an imputation information score < 0.8, and with a Hardy-Weinberg equilibrium (HWE) test p-value < 10^{-10}. After these QC steps, we retained 6,725,312 SNPs for our analysis.

2.2.6 Real data application

We performed fine-mapping on six binary disease traits through integrating the GEUVADIS data with the UKBB data. The examined disease traits include asthma, breast cancer, gout, hypertension, prostate cancer, and rheumatoid arthritis. These disease traits were selected following (Yang & Zhou, 2022) and have a wide range of prevalence (0.02 to 0.38). For the two

sex-related traits (breast cancer and prostate cancer), we limited our analysis to female individuals (for breast cancer) or male individuals (for prostate cancer). In the analysis, we extracted cis-SNPs that resided within 100 kb upstream of TSS and 100 kb downstream of TES of the gene, and focused on a common set of cis-SNPs among them that were shared between UKBB and GEUVADIS datasets. We retained a total of 14,388 genes that contained at least 10 cis-SNPs for TWAS fine-mapping analysis. For a given trait in UKBB, we kept all individuals of European ancestry with self-reported case status of that trait in the GWAS data. Same as the simulations, we used BSLMM to construct GReX prediction models for FABIO, FOCUS, and GIFT, and used the elastic net regression for FOGS and cTWAS following the recommendation from its developer. For FABIO, we used individual-level genotypes of the GWAS data and analyzed genes on each chromosome jointly. To correct for population structure and relatedness, we first identified the top 10 principal components (PCs) of genotypes using principal component analysis (PCA). We then regressed the predicted expression level of each gene on the top 10 genotype PCs and sex, and used the linear regression residual as the corrected gene expression level as input for FABIO. We applied FABIO using both the default uniform prior on log π, and the alternative beta prior on π. For the alternative beta prior, we fixed the parameter a of the beta prior to be 1.1, and estimated the parameter b for each disease trait by assuming one causal gene per risk region that contains GWAS or TWAS signals (defined in Table 1). For the other four methods, we first performed marginal association analysis in the GWAS data with logistic regression model using PLINK 2.0 (Chang et al., 2015), including the top 10 genotype PCs and sex as covariates. We then paired the GWAS summary statistics with the SNP weights as input, and analyzed 1,433 independent LD blocks (Berisa & Pickrell, 2015) across all 22 autosomal chromosomes one at a time. For GIFT and cTWAS, due to the heavy computational demand mentioned earlier in the simulations, we again restricted the comparison to a subset of LD blocks that contain no more than 20 genes, covering a total of 8,411 genes from 1,259 unique LD blocks. We compared the results of these two methods with genes identified by FABIO, FOCUS, or FOGS from the same subset of the LD blocks. We also performed marginal TWAS analysis using FUSION (Gusev et al., 2016) with the GWAS summary statistics and the SNP weights calculated by BSLMM. For a fair comparison across methods, we used an estimated FDR threshold of 0.05 for all methods to declare significance. We estimated FDR based on p-values using the Benjamini-Hochberg approach for FOGS and GIFT, and estimated FDR based on PIPs using the local FDR approach for FABIO,

FOCUS, and cTWAS. For the significant genes identified by FABIO, FOCUS, and FOGS, we further performed enrichment analysis with the Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using the clusterprofiler R package (Yu et al., 2012). We declared gene set significance based on a q-value threshold of 0.05.

2.2.7 Software overview

FABIO is implemented in an R package and is freely available at https://github.com/superggbond/FABIO/. FABIO requires an individual-level predicted GReX file of the TWAS cohort of study (generated using standard methods like PrediXcan and BSLMM) in either plain text format or gzipped text format, which has m (number of genes) rows and n (sample size) + 1 columns. The gene names serve as the first column, following n columns where each column represents GReX of genes for an individual. If the users only have genotypes from an individual-level GWAS in PLINK format, FABIO also provides a function to create the predicted GReX file using our prebuilt BSLMM gene expression prediction models derived from the GEUVADIS eQTL panel, the same eQTL panel used in our real data applications. FABIO also requires the binary phenotypes for the same cohort of n individuals in a plain text format, using 1 to indicate the case and 0 to indicate the control. The order of the binary phenotypes should be consistent with the column order of the predicted GReX file. A key parameter in the model is π, which represents the proportion of genes with non-zero effects. We present two prior options for π: a uniform prior on log π, which requires no input of π from the users, and the initial π will be the proportion of genes passing the single-gene tests based on the input predicted GReX file; or a beta prior $\pi \sim Beta(a, b)$, where users can choose the values of a and b if they have prior knowledge to determine the prior expectation and the variance of π. At the end, FABIO outputs the TWAS fine-mapping results as a table with three columns: the names of the genes, the PIPs of the genes, and the FDR estimate for each gene corresponding to the PIP value.

2.3 Results

2.3.1 Method overview and simulation setup

FABIO is described in the Materials and methods section, with technical details provided in the Appendix A and a method schematic shown in Figure 2.1.

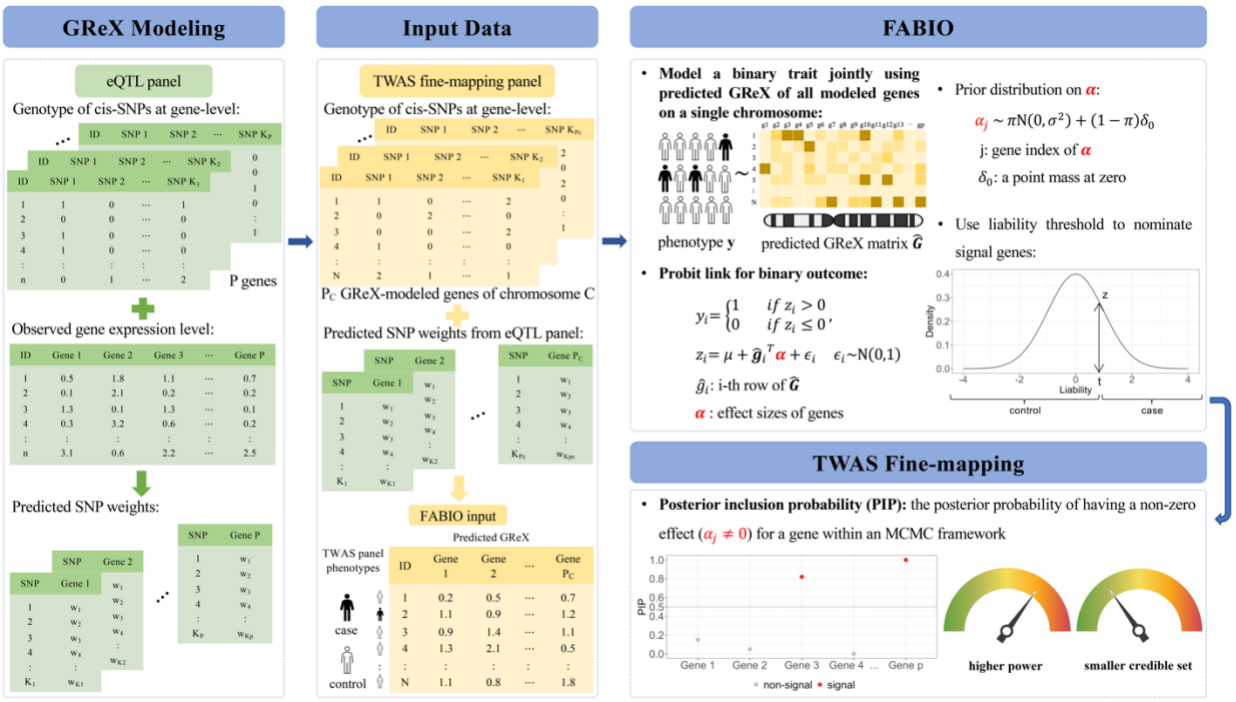

Figure 2. 1 Schematic overview of FABIO for TWAS fine-mapping of binary outcomes.

Same as other existing methods, as a two-step TWAS fine-mapping approach, FABIO requires predicted GReX of the study cohort generated using the standard method like PrediXcan or BSLMM. Shown under the "GReX Modeling" and "Input Data" section, SNP weights are first estimated from the eQTL mapping cohort (sample size = n) with known genotypes and gene expression levels, then the predicted GReX will be generated in the study cohort (sample size = N, usually N > n), using known genotypes and estimated SNP weights. Shown under the "FABIO" section, FABIO explicitly models the binary nature of the outcome trait through a latent variable $\mathbf{z}$ with a sparsity inducing prior on each element of the gene effect sizes $\boldsymbol{\alpha}$. It also simultaneously models all genes on a single chromosome to account for the GReX correlation both within and between LD blocks, through the input of individual-level GReX matrix. We apply MCMC method to estimate the model parameters and obtain test statistic for each gene effect size α_i, and use the posterior inclusion probability (PIP) as the evidence for the gene's association with the binary outcome trait ("TWAS Fine-mapping" section). * Icons used in this figure are from BioRender.com.

Briefly, FABIO is a TWAS fine-mapping method that aims to identify genes whose GReX is associated with a binary trait of interest. Different from existing TWAS fine-mapping approaches, FABIO explicitly models the binary nature of the outcome trait through a probit model, and simultaneously models all genes on an entire chromosome to account for the GReX correlation both within and between LD blocks, which also reduces the residual error variance to enhance the precision of TWAS fine-mapping. We develop an efficient MCMC algorithm for model inference, making FABIO scalable to large biobank scale datasets.

We performed comprehensive simulations to evaluate the performance of FABIO and compared it mainly with two existing TWAS fine-mapping methods FOCUS and FOGS. We also

compared FABIO with two recently published methods GIFT and cTWAS in the baseline simulation due to their heavy computational demand. The simulation details are provided in the Materials and Methods section. Briefly, we obtained genotype data from 373 individuals in GEUVADIS and 50,000 individuals in UK Biobank. We focused on 770 genes residing in 50 randomly selected LD blocks on chromosome 1 and obtained *cis*-SNPs for these genes for simulations. For each gene in turn, we used the genotype data of the *cis*-SNPs from GEUVADIS to simulate the gene expression level in the gene expression mapping study. We also used the genotype data from the same set of *cis*-SNPs in UK Biobank to construct the GReX underlying these genes, selected some genes to be causal, and simulated a binary outcome trait of interest in the GWAS. Afterwards, we applied different methods to perform TWAS fine-mapping to identify the causal genes. In the process, we considered a baseline simulation setting and varied one parameter at a time to examine the influence of different parameters on top of the baseline setting. We examined a total of 11 null settings and 15 alternative settings, with 100 simulation replicates per setting.

2.3.2 FABIO controls false signals in null simulations

First, we examined the performance of different methods under complete null simulation settings where none of the genes is associated with the outcome trait ($PVE_2 = 0\%$). We found that FABIO and FOCUS had reasonably low false positive rates under these settings compared to FOGS (Figure S2.1a). For example, under the baseline setting, FABIO falsely identified an average of 0.22 genes in the 95% credible set while FOCUS falsely identified an average of 0.35 genes per simulation replicate. In contrast, FOGS produced inflated p-values (Figure S2.1b) and falsely identified an average of 3.68 genes per replicate under the Bonferroni corrected threshold. Across settings (Figure S2.1a), FABIO, FOCUS, and FOGS falsely identified an average of 0.25, 0.33, and 4.25 genes per replicate, respectively.

Next, we examined a more challenging null simulation setting where the causal genes were not included in the analysis. Under this setting, because the null genes may display gene expression correlation with the causal genes and may have cis-SNP overlap with them, the null genes will display marginal TWAS signals and could be falsely identified to be causal in the fine-mapping analysis. As expected, we observed much higher false positive rates for all three methods compared to the complete null setting. However, both FABIO and FOCUS still detect much less false signals compared to FOGS. Specifically, with the 95% credible set, FABIO falsely detected

an average of 4.43 genes per replicate while FOCUS falsely detected an average of 4.70 genes (Figure S2.1a). At a Bonferroni corrected threshold, FOGS falsely detected an average of 30.90 genes per replicate (Figure S2.1a), with highly inflated p-values (Figure S2.1c).

2.3.3 FABIO improves power for fine-mapping causal genes in alternative simulations

We examined the performance of different methods in controlling false discovery rate (FDR) under the alternative simulations. Under the baseline setting, the estimated FDR based on PIP for both FABIO and FOCUS are relatively conservative and higher than the true FDR (Figure S2.1d). In contrast, the estimated FDR based on p-values from FOGS are much lower than the truth, which is consistent with its highly inflated p-values observed in the null simulations and can lead to an excessive number of false discoveries if used in practice. Under an estimated FDR of 0.05, the true FDR is 2.3%, 4.0%, and 45.5% in the baseline setting for FABIO, FOCUS, and FOGS, respectively (Figure 2.2a). Similar results were observed under different case/control ratios (Figure 2.2a) and PVE_1 (Figure 2.2b), as well as other simulation settings (Figure S2.2): across all settings, the true FDR that corresponds to the estimated FDR of 0.05 is on average 2.8%, 3.4%, and 39.0% for FABIO, FOCUS, and FOGS, respectively. Note that FOGS achieved calibrated type I error in non-sparse settings (Figure S2.2c), which is consistent with its modeling specifications (details in the Appendix A).

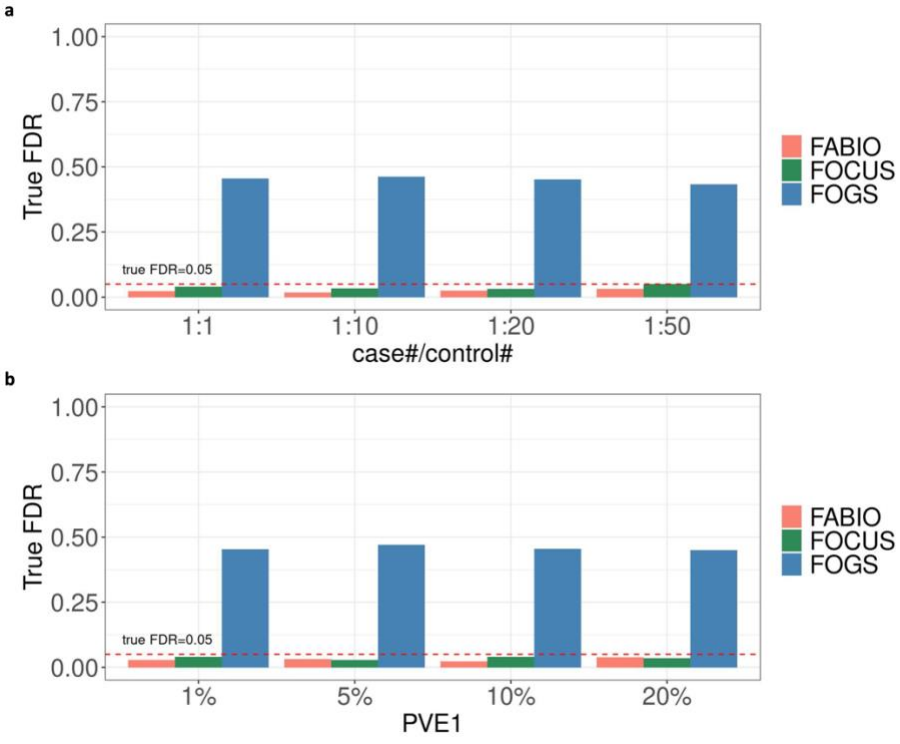

Figure 2. 2 True FDR at an estimated FDR threshold of 0.05 under different conditions.

The red dashed line indicates a true FDR of 0.05. **(a)** Under different case/control ratios. **(b)** Under different percentages of PVE_1.

Next, we evaluated the power of different methods in detecting causal genes under alternative settings. Since different methods use different statistics and have different thresholds (e.g., 95% credible set based on PIPs for FABIO and FOCUS and a Bonferroni adjusted p-value threshold for FOGS), we first compared the size of 95% credible set and the number of true signal genes in the 95% credible set between FABIO and FOCUS. Then for fair comparisons among all three methods, we used two rank-based criteria to evaluate method performance: we either computed power based on a true FDR threshold of 0.05 or measured area under the curve (AUC) using the receiver operating characteristics (ROC) curve. In the simulations, we found that FABIO is more powerful than the other two methods under all alternative simulation settings. For example, in the baseline setting, at a true FDR threshold of 0.05, FABIO, FOCUS, and FOGS achieved a power of 54.6%, 36.5%, and 6.8%, respectively (Figure S2.3a). The power gain brought up by FABIO is partially due to direct modeling of binary traits and partially due to modeling all genes on an entire chromosome together. Specifically, if we restrict FABIO to analyze one LD block at a time same as FOCUS and FOGS, its power is 48.5% (Figure S2.3a), which is at least 32.9% higher than FOCUS or FOGS, supporting the benefits of explicit modeling of binary traits. If we apply FABIO on an entire chromosome, its power will further increase to 54.6% (Figure S2.3a), which is at least 49.6% higher than FOCUS or FOGS, supporting the benefits of analyzing all LD blocks on the same chromosome jointly. Measuring AUCs instead of power at FDR produces similar results. Specifically, in the baseline setting, FABIO achieved an AUC of 82.4% and restricting it to analyzing one LD block at a time reduced its AUC to 80.9%. In comparison, FOCUS achieved an AUC of 73.7% (Bootstrap test p-value = 0.02 for FABIO vs FOCUS) and FOGS achieved 69.0% (p-value = 2.15×10^{-5} for FABIO vs FOGS; Figure S2.3b).

We carefully examined the influence of different simulation parameters on the power of different methods. In terms of case/control ratio, between FABIO and FOCUS, we noticed FABIO can always achieve smaller size of 95% credible set with more true signal genes involved (Figure 2.3a). We found that the power of different methods all reduced with decreasing case/control ratio, though the rank among them did not change. Specifically, when the case/control ratio decreased

from 1:1, 1:10, 1:20, to 1:50, the power of FOGS decreased from 6.8%, 5.2%, 4.2%, to 3.0%; the power of FOCUS decreased from 36.5%, 34.8%, 33.5%, to 23.2%; and the power of FABIO decreased from 54.6%, 50.2%, 49.0%, to 44.8%, which remains higher than the other two methods (Figure 2.3b). The results of AUCs are similar (Figure 2.3c). In terms of SNP effects on gene expression (PVE_1), we had similar observations on the size of 95% credible set between FABIO and FOCUS (Figure 2.3d). We also observed that the performance of all methods improved with increasing PVE_1, though their rank did not change (Figure 2.3e and 2.3f). The power difference between different methods becomes more apparent for relatively small PVE_1 (e.g., 5%), highlighting the benefits of FABIO in more challenging settings. For example, when PVE_1 =5%, FABIO still achieves a power of 55.8%, while the power of FOCUS and FOGS reduces to 26.0% and 6.5%, respectively (Figure 2.3e). In terms of the genetic architecture underlying gene expression, we found that FABIO still identified more true signal genes through 95% credible set than FOCUS (Figure S2.4a), and all methods had better performance with decreased number of SNPs with non-zero effects on gene expression which is equivalent to increased SNP effect sizes, though their rank again did not change (Figure S2.4b and S2.4c). For example, when the number of causal SNPs for each gene increased from 1% to 10%, although the power of both FABIO and FOCUS reduced substantially (from 54.6% to 22.5% for FABIO; and from 36.5% to 11.8% for FOCUS), FABIO still had twice the power of FOCUS. The power of FOGS reduced slightly (from 6.8% to 4.8%) as expected, given that the power was already low when the number of causal SNPs for each gene is 1% (Figure S2.4b). In terms of gene effects on the outcome trait (PVE_2), similar observations on 95% credible sets between FABIO and FOCUS (Figure S2.5a), and when PVE_2 increases from 0.2% to 0.4% and 0.6%, all three methods achieved significant power increase (from 44.0% to 54.6% and 57.2% for FABIO; from 27.5% to 36.5% and 38.8% for FOCUS; and from 2.5% to 6.8% and 7.2% for FOGS) (Figure S2.5b) and better AUC (Figure S2.5c). Finally, the sparsity of causal genes together with causal SNPs also affect method performance. Again, FABIO had better performance than FOCUS in terms of 95% credible sets (Figure S2.6a), and both methods performed better than FOGS in terms of power and AUC in sparse settings (Figure S2.6b and S2.6c). However, FOGS achieved relatively high power (Figure S2.6b) and AUC (Figure S2.6c) in non-sparse settings, consistent with its model specification (details in the Appendix A).

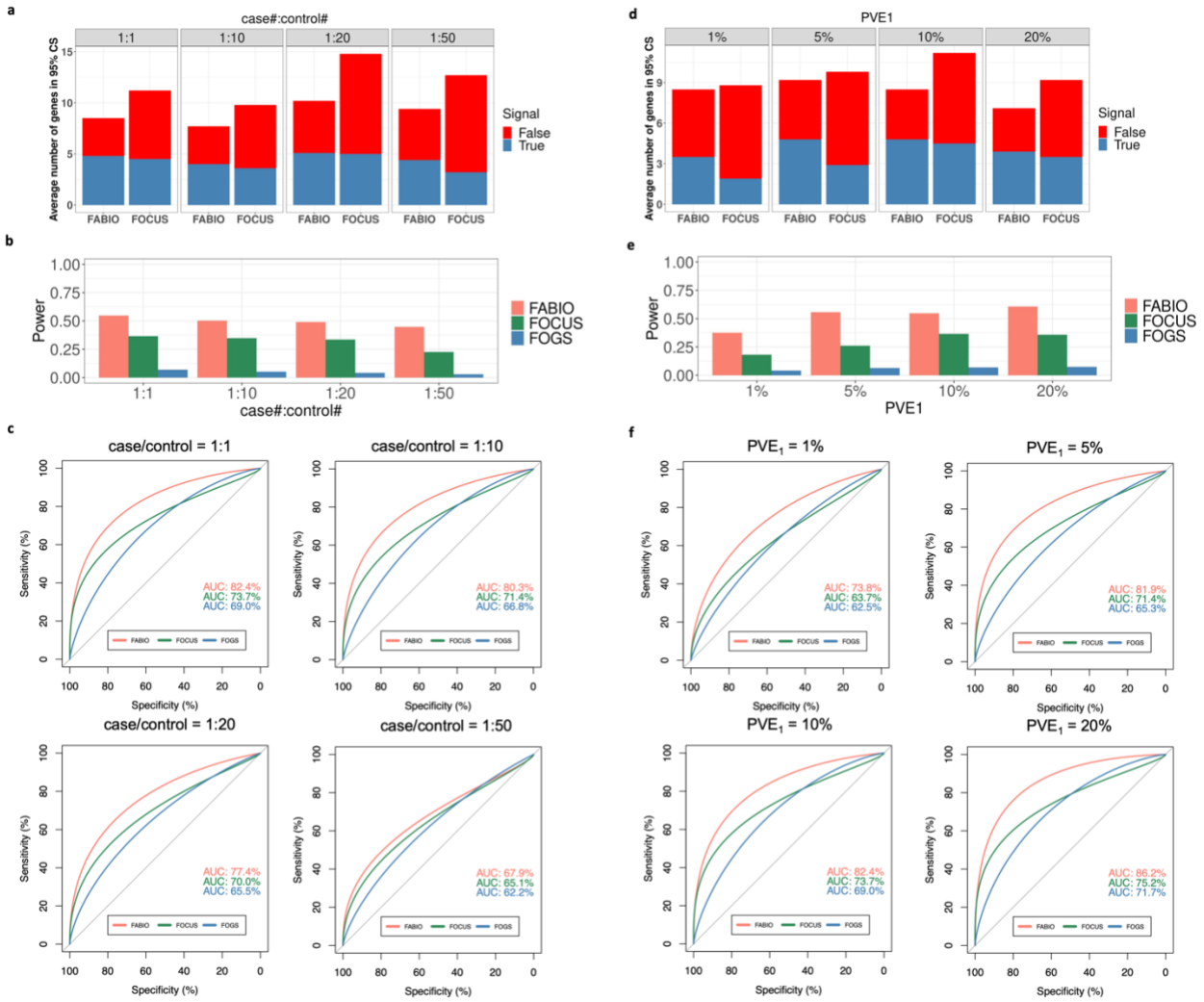

Figure 2. 3 Performance of different methods in simulations with the change of case/control ratio or PVE1.

(a) Average number of genes in 95% credible set (CS) defined by FABIO or FOCUS and the number of true signal genes in 95% CS under different case/control ratios. **(b)** Power comparison for different methods based on a true false discovery rate (FDR) of 0.05 under different case/control ratios. **(c)** ROC curves of different methods with AUCs recorded under different case/control ratios. **(d)** Average number of genes in 95% credible set (CS) defined by FABIO or FOCUS and the number of true signal genes in 95% CS under different PVE1. **(e)** Power comparison for different methods based on a true false discovery rate (FDR) of 0.05 under different PVE1. **(f)** ROC curves of different methods with AUCs recorded under different PVE1.

We have used a true FDR threshold of 0.05 for the power comparisons described above. However, we note that such true FDR can only be calculated in simulations but unavailable for real datasets. Therefore, we estimated FDR using p-values from FOGS and PIPs from FABIO and FOCUS to further compare the power of different methods based on an estimated FDR threshold of 0.05. For FABIO and FOCUS, the estimated FDR under different settings is conservative and

higher than the true FDR; while for FOGS the estimated FDR are lower than the truth, potentially resulting in an excessive number of false discoveries (Figure S2.1d). Nevertheless, at the estimated FDR threshold of 0.05, FABIO retains the highest power across all 15 alternative simulation settings: the power of FABIO, FOCUS, and FOGS in the baseline setting is 58.5%, 37.4%, and 6.6%, respectively (Figure S2.7a), and is on average 50.7%, 31.3%, and 5.5% across all alternative settings, respectively. The conclusion holds regardless of the case/control ratio (Figure S2.7a), the SNP effect on gene expression (PVE_1, Figure S2.7b), the number of causal SNP(s) for each gene (Figure S2.7c), the gene effect on the outcome trait (PVE_2, Figure S2.7d), and the sparsity of causal genes and SNPs (Figure S2.7e).

Next, we explored whether the input summary statistics from GWAS marginal analysis would affect the performance of methods. While we used logistic regression to generate marginal GWAS summary statistics as input for FOCUS and FOGS, we found that using linear regression to generate marginal GWAS summary statistics lead to very similar or slightly reduced performance: at a true FDR threshold of 0.05, the power for FOCUS was 33.3% and 36.5%, and for FOGS, 6.0% and 6.8%, for using linear and logistic regression GWAS summary statistics, respectively (Figure S2.8a). Similar observations are held in terms of AUCs (Figure S2.8b). In addition, we found that FABIO with the default prior on π achieved higher power than using a beta prior, which is influenced by whether prior expectation of π matches the truth (Figure S2.9a and S2.9b). Finally, FABIO also achieved better performance than GIFT and cTWAS in limited comparisons due to their computational burden (Figure S2.10a and S2.10b). These results indicate that FABIO can outperform other existing methods with competitive computing time and retain better performance with smaller GWAS sample size.

2.3.4 Application to analyzing binary traits in UKBB

We performed TWAS fine-mapping analysis on six binary disease traits through integrating the eQTL mapping study of GEUVADIS and the GWAS of UK Biobank (details in Material and Methods). Specifically, we focused on 14,388 genes residing in 1,433 independent LD blocks and examined their associations with each of the six binary traits. The six binary traits include asthma (AS), breast cancer (BRCA), gout (GO), hypertension (HT), prostate cancer (PRCA), and rheumatoid arthritis (RA). Among these traits, we identified 256 (232 unique) GWAS risk regions that contain at least one genome-wide significant SNP ($p < 5\times10^{-8}$) associated with at least one trait. We performed a TWAS marginal analysis using FUSION and identified a

total of 127 (126 unique) significant genes residing in 113 (107 unique) TWAS risk regions across the six traits. Overlapping the results from the two analyses leads to a total of 333 (294 unique) risk regions that contain at least one genome-wide significant SNP or one marginal TWAS significant gene (Table 2.1). These are the likely risk regions that are more likely to contain true signals than the remaining regions.

Table 2. 1 Summary results of TWAS fine-mapping in UK Biobank.

Trait	GWAS risk regions	Sig. TWAS genes	TWAS risk regions	Risk regions with GWAS or TWAS signals	FABIO	FOCUS	FOGS
AS	21	7	6	24	23 (15)	56 (10)	85 (10)
BRCA	15	3	3	18	12 (2)	14 (2)	65 (4)
GO	18	21	18	33	34 (16)	44 (16)	81 (12)
HT	175	85	80	228	254 (159)	227 (128)	231 (117)
PRCA	17	2	2	19	21(7)	18 (1)	70 (0)
RA	10	9	4	11	20 (14)	24 (13)	80 (15)

This table summarizes the number of discoveries for each of the six disease traits (rows) in the TWAS fine-mapping analysis of UK Biobank. A GWAS risk region (1st column) is defined as an LD block that harbors at least one genome-wide significant SNP (p-value $< 5 \times 10^{-8}$). A significant TWAS gene (2nd column) is defined as a gene with a marginal TWAS p-value < 0.05/14,388. A TWAS risk region (3rd column) is defined as an LD block that harbors at least one marginal TWAS significant gene. A risk region with GWAS or TWAS signals (4th column) is defined as an LD block that harbors at least one genome-wide significant SNP or significant TWAS gene. The last three columns list the number of genes discovered by each of the three methods. The number in the bracket is the number of identified genes that are located in a risk region with GWAS or TWAS signals. We used an estimated FDR threshold of 0.05 to declare significance for all methods in the fine-mapping analysis.

Next, we performed TWAS fine-mapping using FABIO, FOCUS and FOGS on all 14,388 genes from the 1,433 independent LD blocks. Same as what we did in the simulations, we estimated FDR using the test statistics generated by different methods and declared significant gene associations based on an estimated FDR threshold of 0.05 for all methods. Note that even though we used the same estimated FDR threshold, such threshold for different methods may still correspond to different levels of false discoveries: as is evident in simulations, the same estimated FDR from FOGS may correspond to a much higher true FDR than that of FABIO or FOCUS. Based on the estimated FDR threshold of 0.05, FABIO identified a total of 364 associated genes

(354 unique) from 248 LD blocks across the six traits, with an average of 1.47 associated genes per LD block. These numbers are reduced when a beta prior is used on π (Table S2.1), since our input parameters of the beta prior may not reflect the unknown truth of the causal gene density. Among the identified genes, 213 of them reside in a risk region with either GWAS or TWAS signals (Table 2.1). As a comparison, FOCUS identified a total of 383 associated genes (350 unique) from 188 LD blocks across six traits, with an average of 2.04 associated genes per LD block. Among the identified genes, 170 of them reside in a risk region with either GWAS or TWAS signals (Table 2.1). FOGS identified 612 genes (439 unique) from 263 LD blocks across six traits, with an average of 2.33 associated genes per LD block. Among the identified genes, 158 of them reside in a risk region with either GWAS or TWAS signals (Table 2.1). In addition, we found that FABIO identified more gene candidates that lied in the GWAS or TWAS risk regions compared to GIFT and cTWAS (Table S2.2). Overall, FABIO was able to narrow down to the smallest set of genes per LD block along with the highest proportion of identified genes residing in likely risk regions in the set (Figure 2.4a), supporting its superior performance.

We further examined the power and robustness of different methods through down-sampling analysis. Specifically, we applied the three TWAS fine-mapping methods using only half of the samples for the same six traits and examined whether the results obtained in such reduced sample analysis is consistent with full sample analysis. We found that FABIO identified more gene candidates that lied in the GWAS or TWAS risk regions than FOCUS and FOGS in the reduced sample analysis (Table S2.3). In addition, FABIO achieved higher consistency between the full sample analysis and reduced sample analysis, with 135 out of 364 (37.1%) genes identified across all six traits in the full sample analysis that were also identified in the reduced sample analysis, more so than FOCUS (86 out of 383; 22.5%) or FOGS (198 out of 612; 32.4%). These observations support the power and robustness of FABIO in the real data application.

We list a few gene examples that are uniquely identified by FABIO. The first example is *GATA3* associated with asthma (PIP=1), a gene that resides in the known risk regions of asthma (Figure S2.11). *GATA3* has been associated with asthma in previous GWAS (Schoettler et al., 2019) and acts as a key regulator of T-cell development, Th2 differentiation, and the Th1/Th2 balance to affect asthma risk (Ray & Cohn, 1999). Transgenic mice with a dominant-negative mutant of GATA3 display decreased expression level of Th2 cytokines IL-4, IL-5, and IL-13,

along with attenuated asthma symptoms, such as airway eosinophilia, mucus production, and IgE synthesis (Zhang et al., 1999), supporting a potentially causal effect of *GATA3* on asthma risks.

The second example is *ABCG2* associated with gout (PIP=1, Figure 2.4b). A common polymorphism (rs2231142, Q141K) in *ABCG2* was identified to be associated with gout and gout related symptoms such as serum urate concentration (Chen et al., 2018; Cleophas et al., 2017; Eckenstaler & Benndorf, 2021; Wrigley et al., 2020) in previous GWAS and our GWAS analysis (Figure 2.4c). ABCG2, the ATP-binding cassette transporter G2, mediates renal and/or extra-renal urate excretion. Consequently, the dysfunction of ABCG2 leads to elevated levels of uric acid, causing hyperuricemia and gout (Matsuo et al., 2013; Matsuo et al., 2009). Indeed, ABCG2-knockout mouse has significantly higher serum uric acid level and significantly lower level of urate excretion from the intestine, which are consistent with the observation in patients with ABCG2 dysfunction, indicating that ABCG2 has an important role in extra-renal urate excretion, especially in intestinal urate excretion (Takada et al., 2014).

The third example is *SH2B3* associated with hypertension (PIP=1, Figure 2.4d). SH2B3, also known as lymphocyte adapter protein (LNK), is an intracellular adaptor protein mainly expressed in hematopoietic and endothelial cells (Dale & Madhur, 2016). A non-synonymous SNP (rs3184504) in exon 3 of *SH2B3* leads to a substitution from arginine to tryptophan in amino acid 262 (R262W), and was associated with hypertension in previous GWAS (Devallière & Charreau, 2011) and in our GWAS analysis (Figure 2.4e). Two different mouse models that disrupt LNK signaling further confirmed that LNK regulates blood pressure and renal inflammation (Rudemiller et al., 2015; Saleh et al., 2015). One is an LNK deleted mouse model, that represents a complete loss of function of LNK, where hypertension was severely exacerbated with enhanced renal and vascular inflammation compared to wild type animals (Saleh et al., 2015). The other is an LNK-mutated mouse model, that represents an alteration of function of LNK, where mutation of the Sh2 domain of LNK led to the attenuation of hypertension, albuminuria, and renal inflammation (Rudemiller et al., 2015). Both animal models supported that modulation of LNK function in immune cells can regulate blood pressure and the associated inflammation and end-organ damage (Dale & Madhur, 2016).

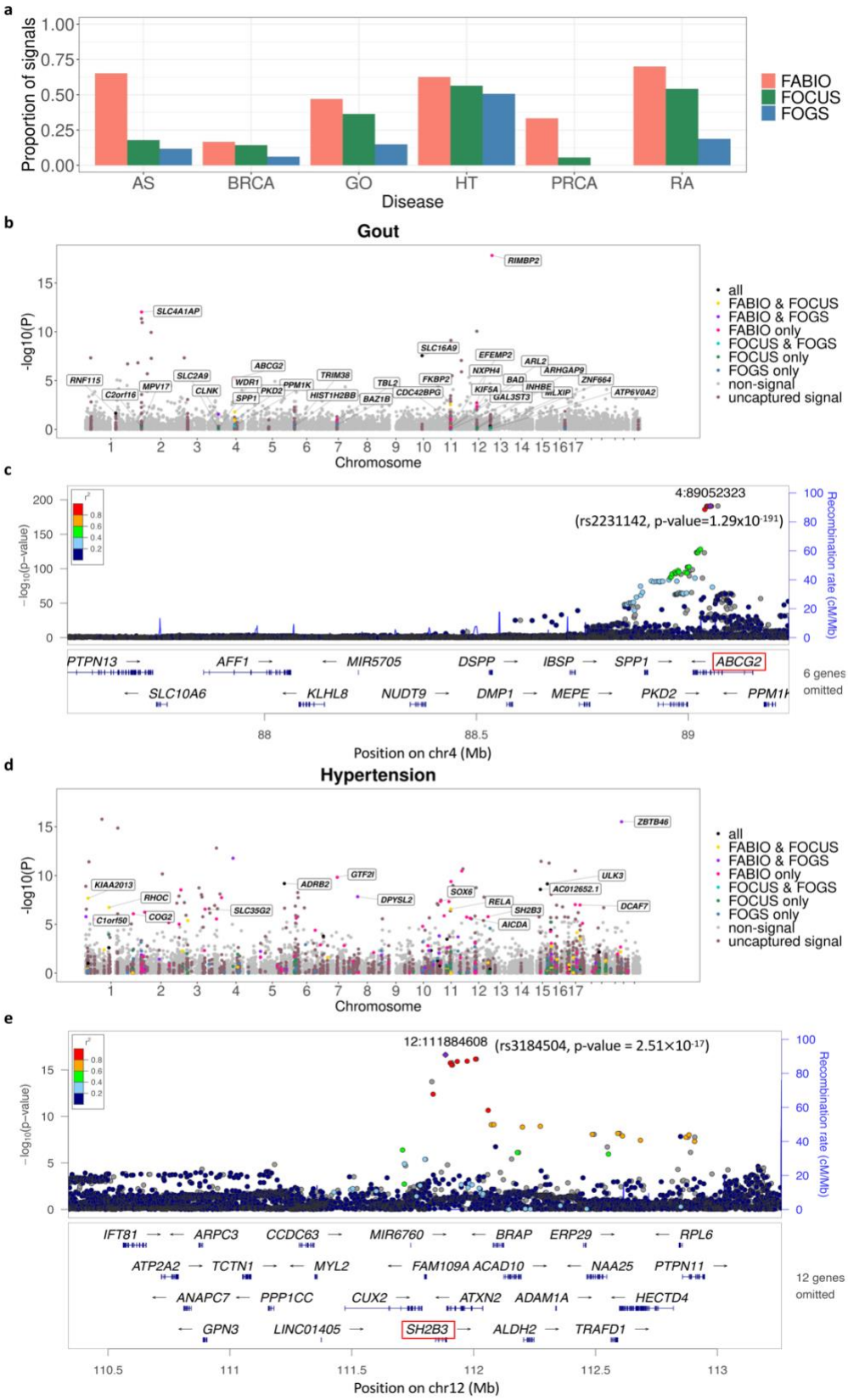

Figure 2. 4 Results of real data applications.

(a) Proportion of identified signal genes that locate in risk regions for the six diseases. Highlighted genes identified in TWAS fine-mapping analysis are labeled in black boxes in **(b)** and **(d)**. Each gene is represented as a dot with x-axis indicating its genomic location and y-axis indicating its -log10 of p-value in the marginal TWAS association test. The dot of each gene is then colored based on different categories: (1) only identified by FABIO; (2) only identified by FOCUS; (3) only identified by FOGS; (4) identified by both FABIO and FOCUS; (5) identified by both FABIO and FOGS; (6) identified by both FOCUS and FOGS; (7) identified by all three methods; (8) located in a known risk region but missed by all three methods (uncaptured signal); (9) a gene shown no significance in both TWAS and GWAS analyses (non-signal). **(b)** TWAS Manhattan plot of gout. **(c)** LocusZoom plot of the corresponding LD block with the GWAS p-value of the SNP rs2231142 in ABCG2 for gout. **(d)** TWAS Manhattan plot of hypertension. **(e)** LocusZoom plot of the corresponding LD block with the GWAS p-value of the SNP rs3184504 in SH2B3 for hypertension.

We further systematically examined the biological relevance of the genes identified by different methods using Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. For asthma, the candidate genes identified by FABIO are significantly enriched in five pathways, with four from Gene Ontology (Figure S2.12a) and one from KEGG (Figure S2.12b). Three of them point to interleukin-13 (IL-13) production: positive regulation of interleukin-13 production (q-value=0.015), interleukin-13 production (q-value=0.015), and regulation of interleukin-13 production (q-value=0.015); and three FABIO significant genes are involved in these pathways, including *TSLP*, *IL4*, and *GATA3* (FABIO PIP=0.92, 0.76, and 1, respectively). The role of IL-13 signaling in asthma has been recognized for a long time in multiple studies (Corren, 2013; Rael & Lockey, 2011; Wills-Karp & Chiaramonte, 2003; Wills-Karp, 2004), and IL-13 is a pleiotropic TH2 cytokine that has been considered as a key to the pathogenesis of asthma through its regulatory role in IgE synthesis, mucus hypersecretion, airway hyperresponsiveness, and fibrosis (Rael & Lockey, 2011). A mast cell related pathway, positive regulation of mast cell activation (q-value=0.015), is also enriched among the candidate genes identified by FABIO, and the involved FABIO significant genes are *SNX4*, *TSLP*, and *IL4* (FABIO PIP=0.69, 0.92, and 0.76, respectively). Mast cells play key roles in asthma through secretion of mediators with pro-inflammatory and airway-constrictive effects, such as histamine and bioactive lipids (Méndez-Enríquez & Hallgren, 2019). By releasing a variety of such mediators, mast cells can trigger bronchoconstriction and alter the permeability of the bronchial barrier that further lead to asthma symptoms (Banafea et al., 2022). The inflammatory bowel disease is highlighted (q-value=0.015) through KEGG enrichment analysis using FABIO significant genes, and the involved genes are *IL18RAP*, *IL4*, *GATA3*, and *SMAD3* (FABIO PIP=1, 0.76, 1, and 0.75, respectively). A systematic review has shown that asthma is associated with both Crohn's disease and ulcerative colitis, the two types of the inflammatory bowel disease (Kuenzig

et al., 2018). As a comparison, only two significant pathways (cytokine-cytokine receptor interaction, q-value=0.006; Thiamine metabolism, q-value=0.006) were identified from KEGG enrichment analysis using significant genes identified by FOCUS (Figure S2.12c), and none for FOGS.

We also identified significantly enriched pathways for hypertension. A total of 11 pathways are enriched using the candidate genes identified by FABIO: three from Gene Ontology (Figure S2.12d) and eight from KEGG (Figure S2.12e). Two of them are directly associated with blood pressure: regulation of systemic arterial blood pressure (q-value=0.017) and regulation of blood pressure (q-value=0.030). Four of them are cardio-related: positive regulation of heart rate (q-value=0.033), adrenergic signaling in cardiomyocytes (q-value=0.012), vascular smooth muscle contraction (q-value=0.025), and dilated cardiomyopathy (q-value=0.030). Renin secretion (q-value=0.044) is also enriched, where renin is a well-known regulator in blood pressure through the renin–angiotensin–aldosterone system (RAAS) (Nakagawa et al., 2020). Using the candidate genes identified by FOCUS, 17 pathways are enriched, with 14 from Gene Ontology (Figure S2.12f) and three from KEGG (Figure S2.12g). However, 10 of them are not directly associated with cardio or blood pressure, such as the top three enriched pathways: positive regulation of adenylate cyclase activity (q-value=0.024), positive regulation of cyclase activity (q-value=0.024), and positive regulation of lyase activity (q-value=0.024). For FOGS, we have two significantly enriched pathways from Gene Ontology (Figure S2.12h), and one of them is directly related to blood pressure: negative regulation of systemic arterial blood pressure (q-value=0.026).

Finally, we note that FABIO is reasonably computationally efficient despite the necessary computational cost incurred for direct modeling of binary outcomes. Specifically, in the fine-mapping of UK Biobank data with 337,198 individuals, using eight CPU threads, it took FABIO an average of 16.14 hours to analyze each chromosome per trait. As a comparison, with the same computational resource, it took an average of 11.84 hours for PLINK to compute the GWAS summary statistics to serve as input for FOCUS and FOGS, which incurred an additional 2.52 and 25.08 hours respective for fine-mapping each chromosome. In terms of physical memory, FABIO uses ~52 Gb memory while PLINK uses ~64 Gb memory to compute the necessary GWAS summary statistics for FOCUS or FOGS, which then incurred minimal memory for additional fine-mapping analysis.

2.4 Discussion

We consider a genome-wide association study (GWAS) performed on n individuals for a binary trait of interest. We denote $\mathbf{y}$ as an n-vector of binary phenotypes measured on the n individuals. Our goal is to identify genes whose genetically regulated expression (GReX) is associated with the outcome trait through TWAS fine-mapping. In this study, we have presented FABIO, a novel TWAS fine-mapping method tailored for binary outcomes. Unlike existing approaches, FABIO not only incorporates a probit model to capture the binary nature of disease traits but also jointly models all genes on the same chromosome to account for GReX correlation across genes in different LD blocks. As a result, FABIO produces reasonably calibrated test statistics and achieves high power. Through comprehensive simulations and real-world applications to six disease traits in the UK Biobank, we have illustrated the benefits of FABIO.

FABIO employs a probit link function to model the relationship between GReX and binary outcome traits. The probit link function introduces a latent variable known as the liability score, which determines the disease status of each individual. Specifically, when the liability score exceeds a certain threshold, then the individual is classified as having the disease. The liability score is then related to a set of explanatory variables for association analysis. A common alternative approach to modeling binary outcomes is to use the logistic link function within generalized linear models. The logistic link function models the probability of disease for each individual based on the explanatory variables. Consequently, the disease probability varies across individuals depending on the underlying explanation variables. Although the interpretation and inference algorithms differ between these two binary outcome modeling approaches, they often produce similar association results in most application settings (Czado & Munk, 2000). Indeed, in our simulations, we verified that FABIO retains reasonable power even when there are model misspecifications, such as using a logistic regression to generate binary outcomes. The results suggest that FABIO can effectively capture the relationship between GReX and binary outcomes, even when the assumptions of the probit link function are not precisely met.

FABIO accounts for gene expression correlation outside local LD blocks by jointly modeling all genes on the entire chromosome. The constructed GReX indeed displays correlation not only within LD blocks but also between neighboring LD blocks, though the correlations decrease with increasing distance. To illustrate this, we focus on 1,568 genes from 133 distinct LD blocks on chromosome 1. First, we calculated the pairwise correlation of GReX for four distinct

groups of gene pairs: those within each LD block, between two adjacent LD blocks, between two distant LD blocks, and between one LD block on chromosome 1 and one on chromosome 2, with the latter serving as a baseline. We observed a clear decreasing trend in GReX correlation across these groups. Notably, the correlations between two adjacent LD blocks and between two distant LD blocks remained substantial - 3.88 and 1.47 times higher, respectively, than the baseline correlation between LD blocks on different chromosomes (Figure S2.13). This suggests that GReX correlation extends beyond individual LD blocks, although with diminishing magnitude over increasing distance. Second, for all six binary traits examined in our real data analysis, we computed the proportion of phenotype variance (R^2) explained by GReX using linear regression in two ways. First way, we included the GReX for all genes across 133 LD blocks on chromosome 1 in the model and computed the R^2. Second way, we included the GReX for all genes within each LD block, computed the R^2 for each block, and then summed the R^2 values across blocks. Intuitively, if GReX correlation between LD blocks is negligible, the R^2 values obtained from these two approaches should be identical. However, we found that the average R^2 computed in the first way is 1.38% (±0.49%) while that the average R^2 computed in the second way is 1.74% (±0.52%), with a 26.1% increase due to GReX correlation between LD blocks. This highlights the importance of modeling GReX correlation between LD blocks, as is done in FABIO.

Furthermore, FABIO employs a well-established Markov chain Monte Carlo (MCMC) algorithm to estimate model parameters and compute the test statistic (PIP) for each gene. This approach avoids the high computational cost associated with directly modeling binary outcomes using a probit link, while maintaining computational efficiency. Compared with other existing linear model-based methods, FABIO outperforms approaches such as FOGS, GIFT, and cTWAS, which require an additional step to generate GWAS summary statistics using external software (e.g., PLINK). Even compared to the fastest existing method, FOCUS, FABIO achieves comparable running times, as mentioned earlier in the Results section.

FABIO is not without limitations. Firstly, due to the probit link function, FABIO requires individual-level GWAS data as input and cannot directly model GWAS summary statistics. In particular, FABIO needs to infer a latent liability variable for each individual during the MCMC inference process, which is only feasible with individual-level GWAS data. While requiring individual-level data may restrict the application of FABIO, this requirement offers an important advantage. It helps mitigate biases commonly observed in methods that make use of only GWAS

summary statistics, particularly when there is a LD matrix mismatch between the reference panel and the analyzed GWAS data (Benner et al., 2017; Yang et al., 2012). Secondly, similar to FOCUS and FOGS, FABIO adopts a two-stage fitting strategy for TWAS fine-mapping: it relies on a reference eQTL mapping study to construct the gene expression prediction model in the first stage, and applies the SNP weights inferred in the eQTL mapping study to construct GReX in the GWAS data for TWAS fine-mapping in the second stage. While the two-stage fitting strategy is convenient and effective, it overlooks the uncertainty in the gene expression modeling in the first stage, potentially leading to power loss (Liu et al., 2021; Yuan et al., 2020). An important future direction is to extend FABIO towards joint likelihood-based inference for TWAS fine-mapping, thus accounting for the inference uncertainty from the first stage. Finally, in our real-world applications, we utilized GEUVADIS data as the eQTL mapping study to fine map the six disease traits. It is important to acknowledge that GEUVADIS data is collected from whole blood, which may not necessary be the disease relevant tissue for the outcome disease. Future extensions of FABIO should consider integrating tissue-specific eQTL mapping studies to reveal tissue specificity underlying the disease trait. Utilizing multiple eQTL panels from various tissues to construct SNP weights has been proposed to leverage shared eQTL across tissues to enhance the power of TWAS (Hu et al., 2019). Alternatively, employing tissue-specific references tailored to each disease may also yield more effective results.

2.5 Supplementary tables

Table S2. 1 Summary results of TWAS fine-mapping in UK Biobank (FABIO input options for π).

Trait	Risk regions with GWAS or TWAS signals	default prior on π	beta prior on π
AS	24	23 (15)	8 (3)
BRCA	18	12 (2)	5 (1)
GO	33	34 (16)	14 (6)
HT	228	254 (159)	77 (38)
PRCA	19	21(7)	8 (1)
RA	11	20 (14)	6 (3)

Table S2. 2 Summary results of TWAS fine-mapping in UK Biobank (five methods).

Trait	Risk regions with GWAS or TWAS signals	FABIO	FOCUS	FOGS	GIFT	cTWAS
AS	18	20 (13)	48 (9)	65 (8)	15 (10)	16 (9)
BRCA	14	8 (1)	10 (1)	52 (3)	7 (1)	10 (1)
GO	30	28 (14)	36 (12)	69 (8)	24 (10)	22 (10)
HT	205	210 (141)	202 (118)	192 (95)	195 (125)	185 (122)
PRCA	16	18(5)	15 (1)	57 (0)	12 (3)	10 (2)
RA	10	18 (12)	20 (10)	75 (13)	16 (10)	18 (10)

Table S2. 3 Summary results of TWAS fine-mapping in UK Biobank (down-sampled).

Trait	GWAS risk regions	Significant TWAS genes	TWAS risk regions	Risk regions with GWAS or TWAS signals	FABIO	FOCUS	FOGS
AS	8	1	1	9	9 (5)	21 (2)	62 (3)
BRCA	5	0	0	5	7 (1)	5 (0)	45 (1)
GO	10	11	8	16	21 (8)	22 (4)	79 (6)
HT	59	20	17	66	141 (44)	76 (22)	143 (27)
PRCA	4	0	0	4	6 (3)	6 (1)	42 (0)
RA	4	5	3	6	12 (5)	9 (3)	38 (3)

2.6 Supplementary figures

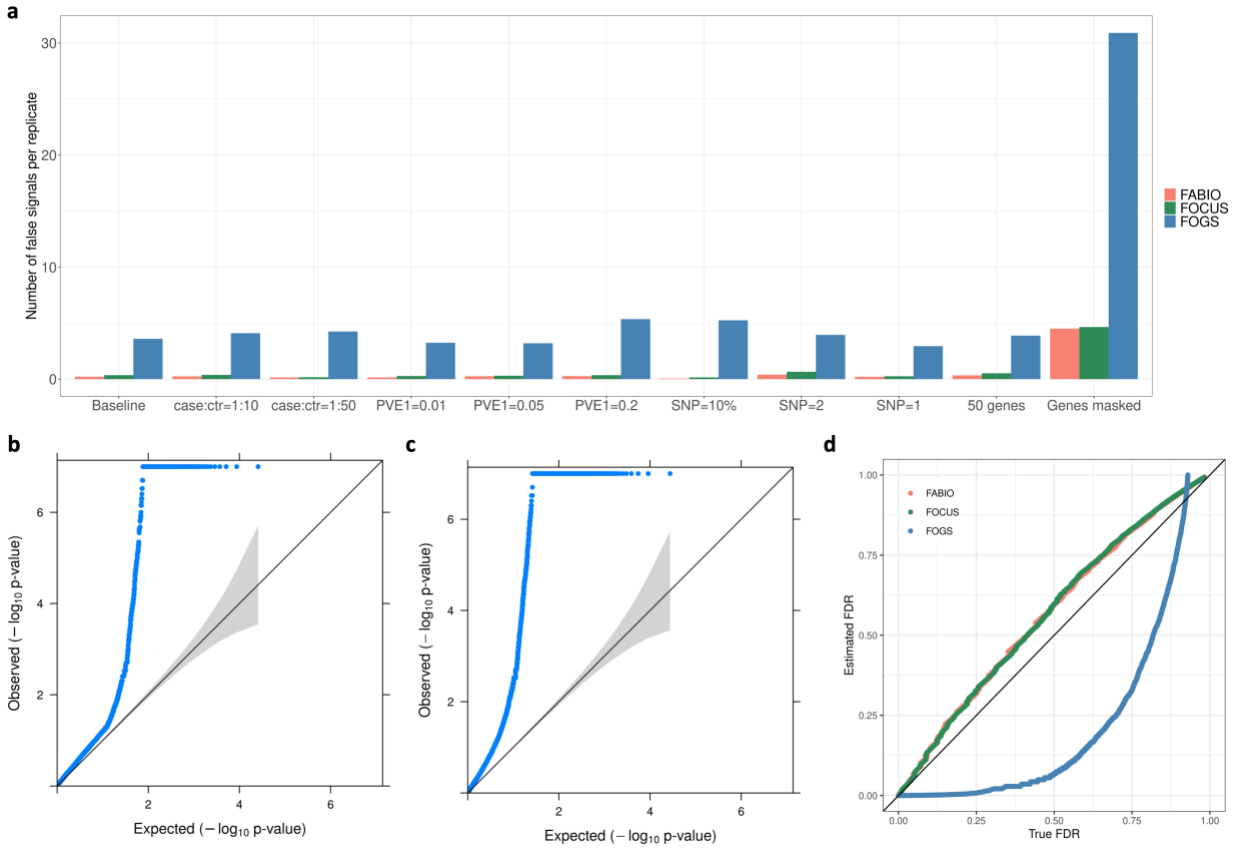

Figure S2. 1 Performance of different methods in controlling false signals.

(a) Number of false signal genes per simulation replicate under different complete null settings (the first 10 settings) and the masked gene null setting (the last setting). (b) Quantile-quantile plot of -log10 p-values for testing the non-causal genes using FOGS under the complete null simulation setting. (c) Quantile-quantile plot of -log10 p-values for testing the non-causal genes using FOGS under the masked null simulation setting. (d) Comparisons among different methods for the estimate of false discovery rate (FDR) under the alternative baseline simulation setting. Compared methods include FABIO (salmon pink), FOCUS (green) and FOGS (blue). The estimated FDR (y-axis) is plotted against the true FDR (x-axis).

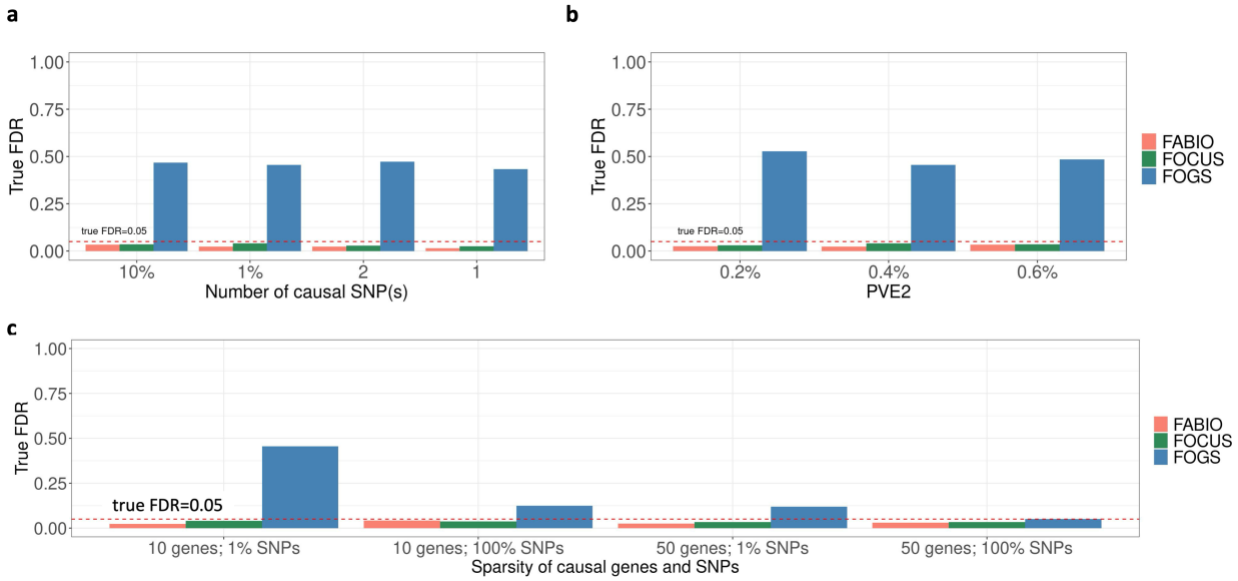

Figure S2. 2 True FDR at an estimated FDR threshold of 0.05 under different conditions.

The red dashed line indicates a true FDR of 0.05. (a) Under different numbers of causal SNP(s). (b) Under different percentages of PVE_2. (c) Under different sparsity of causal genes and SNPs.

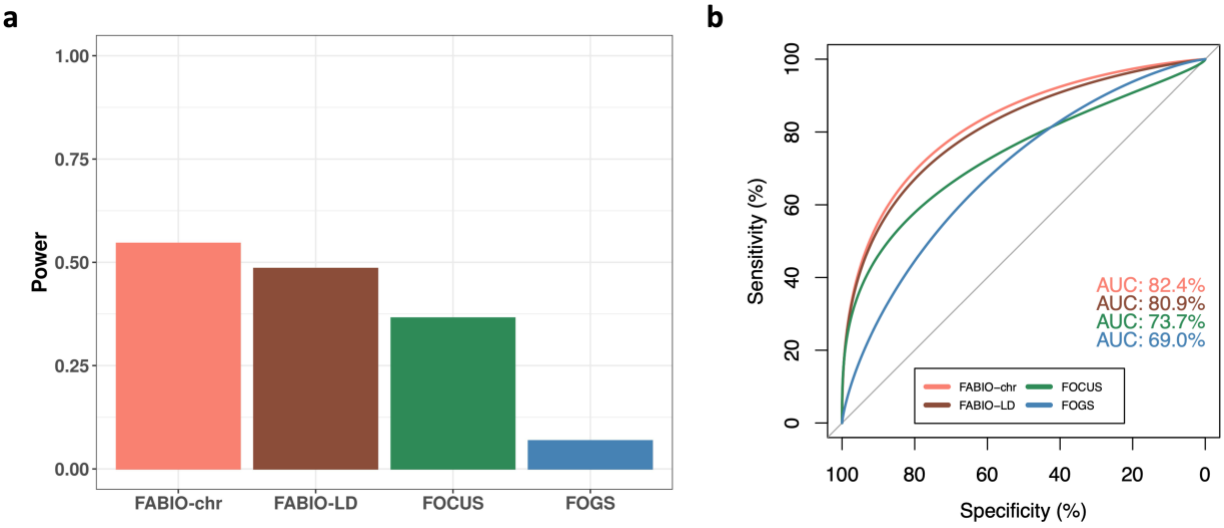

Figure S2. 3 Performance of different methods under the baseline simulation setting.

(a) Power comparison for different methods based on a true false discovery rate (FDR) of 0.05. FABIO was applied to either analyze all LD blocks jointly (FABIO-chr) or one LD block at a time (FABIO-LD). FOCUS and FOGS were only applied to analyze one LD block at a time. (b) ROC curves of different methods with AUCs recorded in the plot.

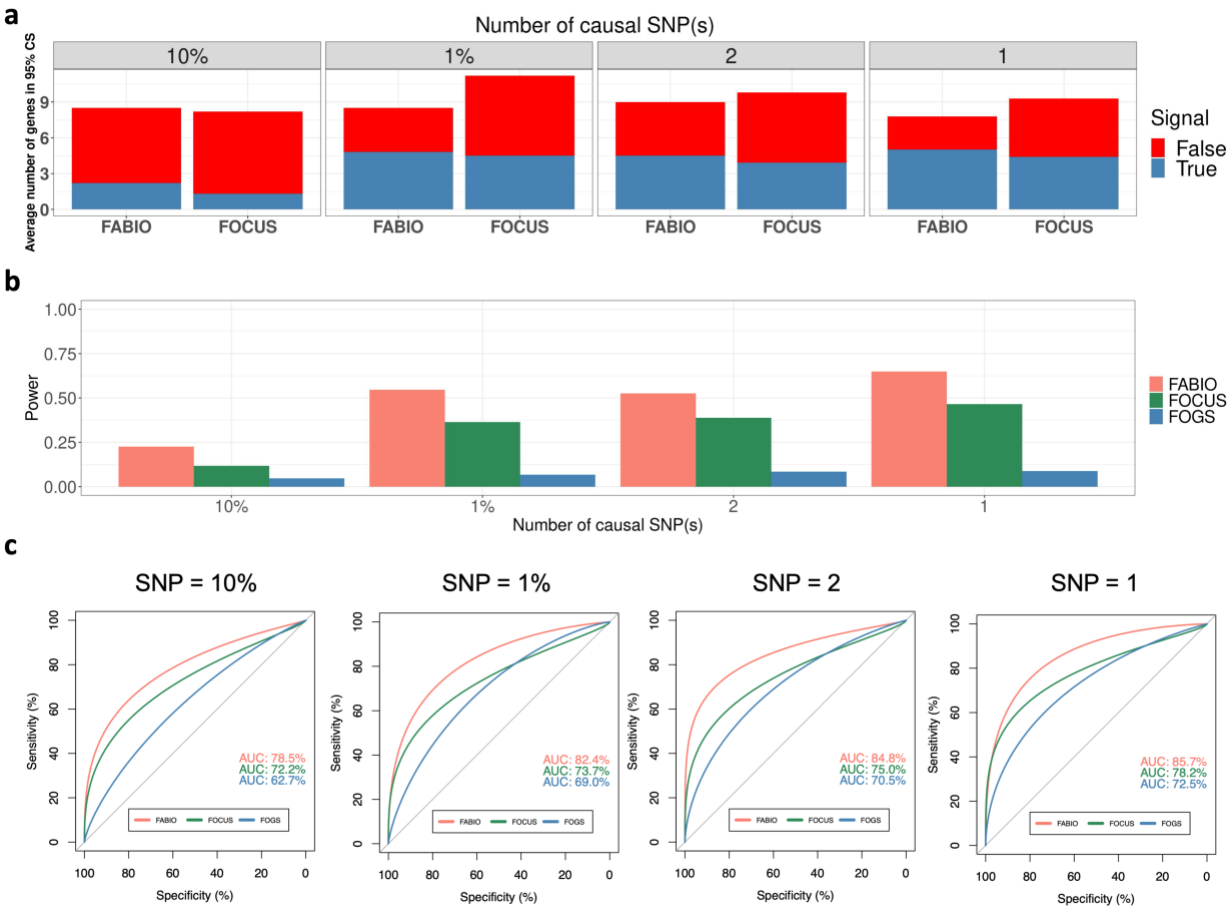

Figure S2. 4 Performance of different methods in simulations with the change of the number of causal SNP(s).

(a) Average number of genes in 95% credible set (CS) defined by FABIO or FOCUS and the number of true signal genes in 95% CS. (b) Power comparison for different methods based on a true false discovery rate (FDR) of 0.05. (c) ROC curves of different methods with AUCs recorded.

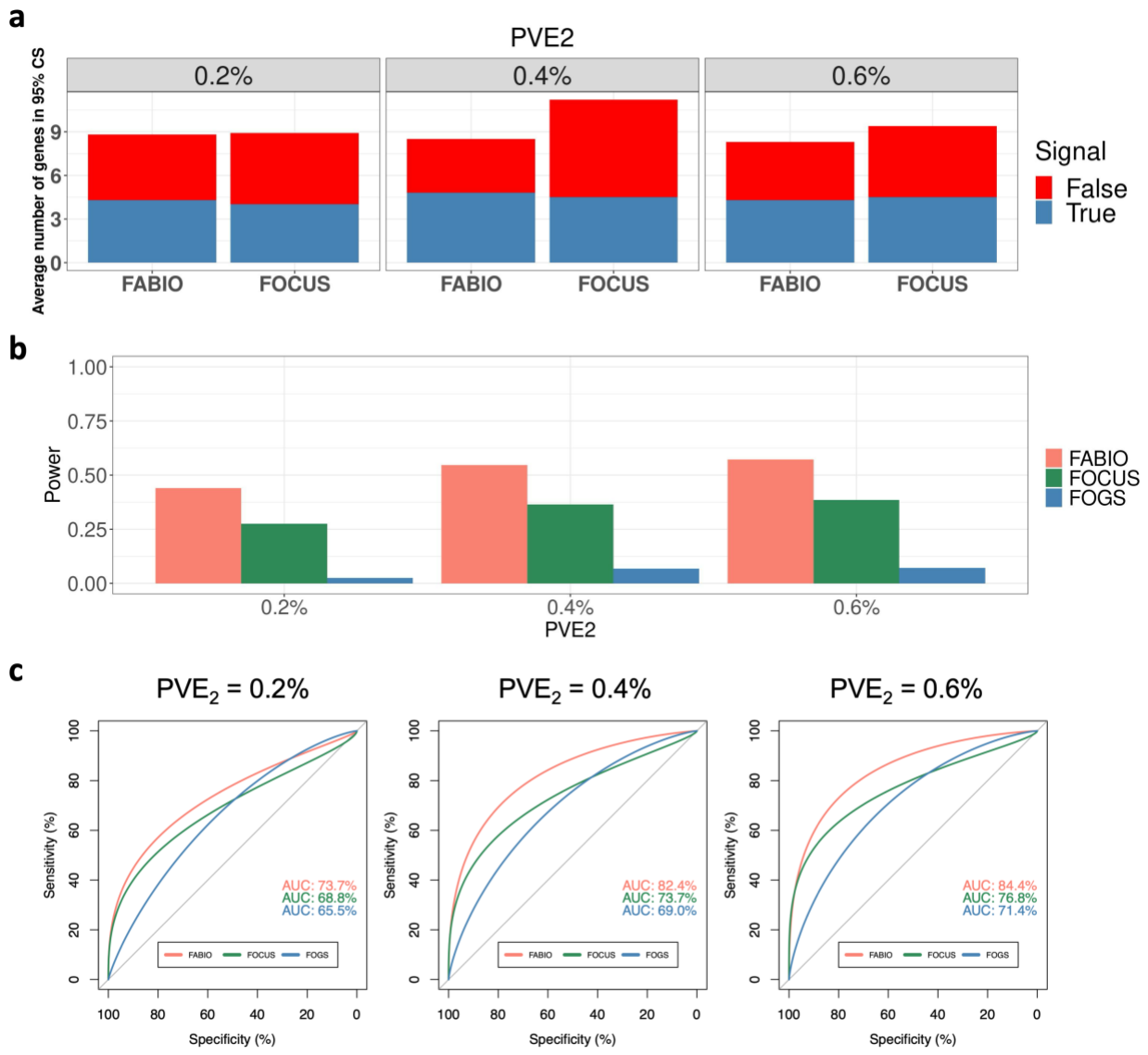

Figure S2. 5 Performance of different methods in simulations with the change of PVE_2.

(a) Average number of genes in 95% credible set (CS) defined by FABIO or FOCUS and the number of true signal genes in 95% CS. (b) Power comparison for different methods based on a true false discovery rate (FDR) of 0.05. (c) ROC curves of different methods with AUCs recorded.

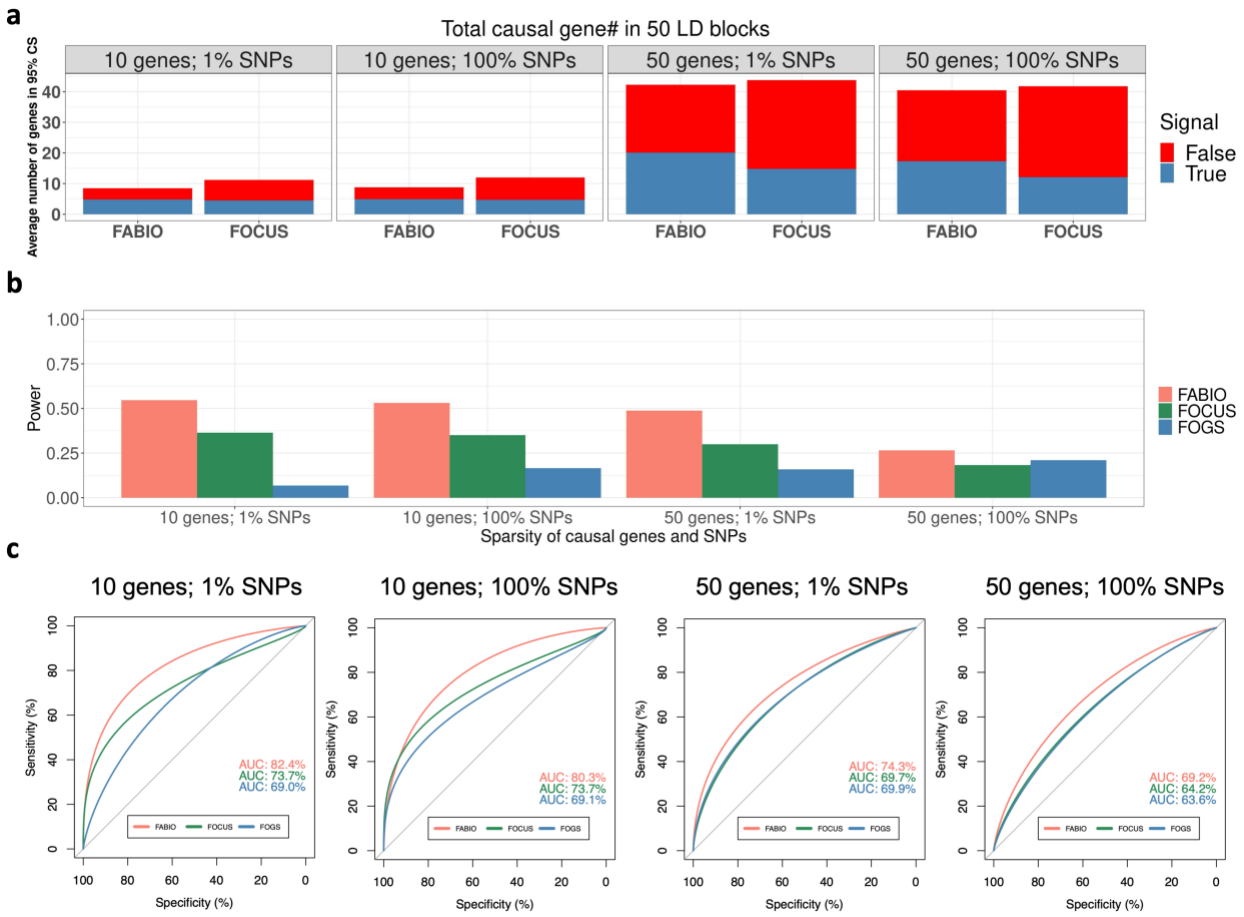

Figure S2. 6 Performance of different methods in simulations with the change of sparsity of causal genes and SNPs.

(a) Average number of genes in 95% credible set (CS) defined by FABIO or FOCUS and the number of true signal genes in 95% CS. (b) Power comparison for different methods based on a true false discovery rate (FDR) of 0.05. (c) ROC curves of different methods with AUCs recorded.

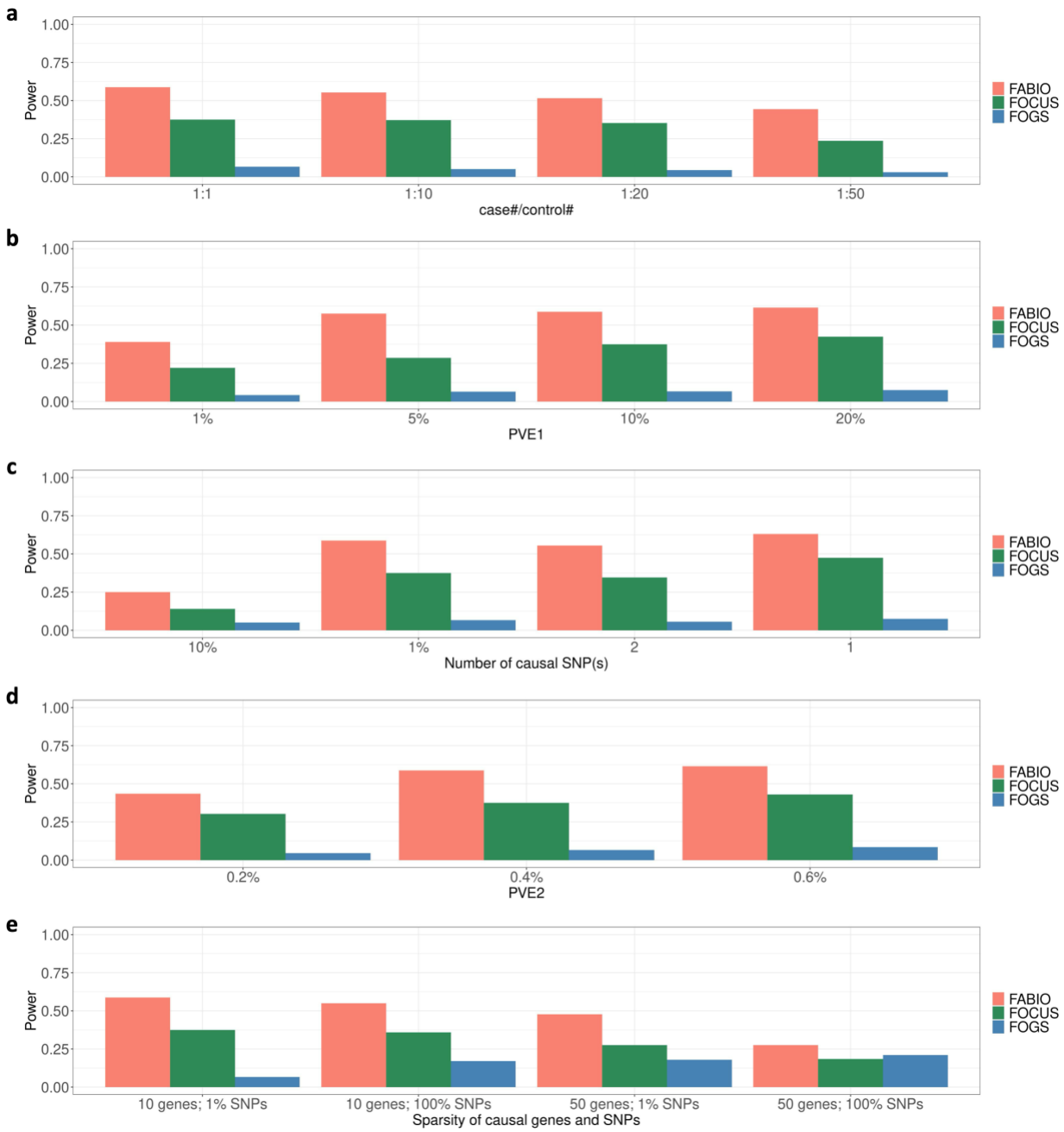

Figure S2. 7 Power comparison for different methods based on an estimated false discovery rate (FDR) threshold of 0.05.

(a) under different case/control ratios. (b) under different PVE1. (c) under different numbers of causal SNP(s). (d) under different PVE2. (e) under different sparsity of causal genes and SNPs.

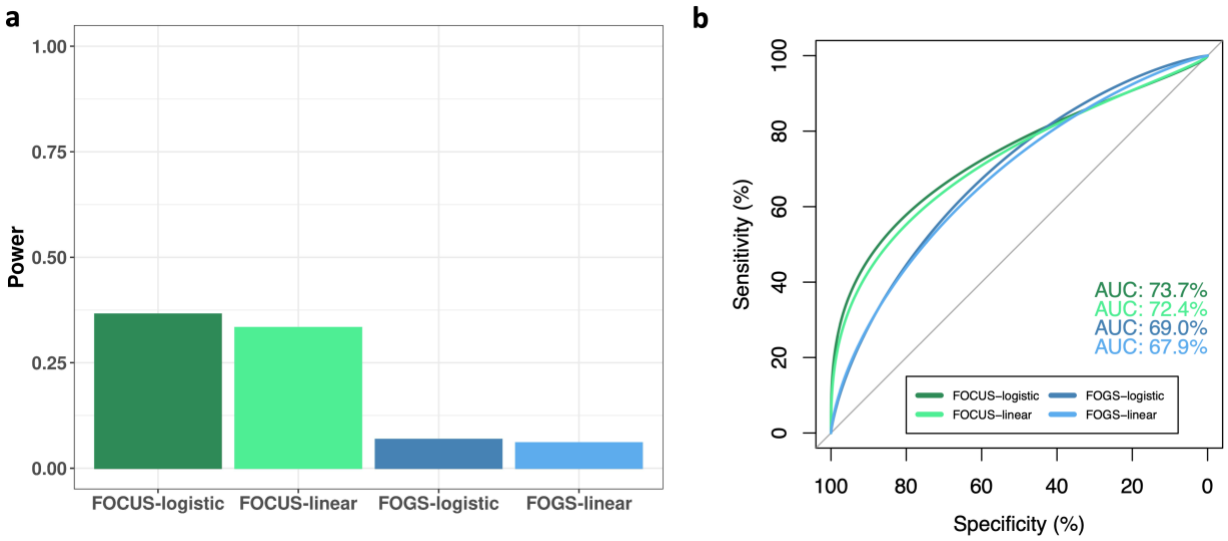

Figure S2. 8 Performance of FOCUS and FOGS with GWAS summary statistics generated by either logistic regression or linear regression.

(a) Power comparison based on a true false discovery rate (FDR) threshold of 0.05. (b) ROC curves with AUCs recorded.

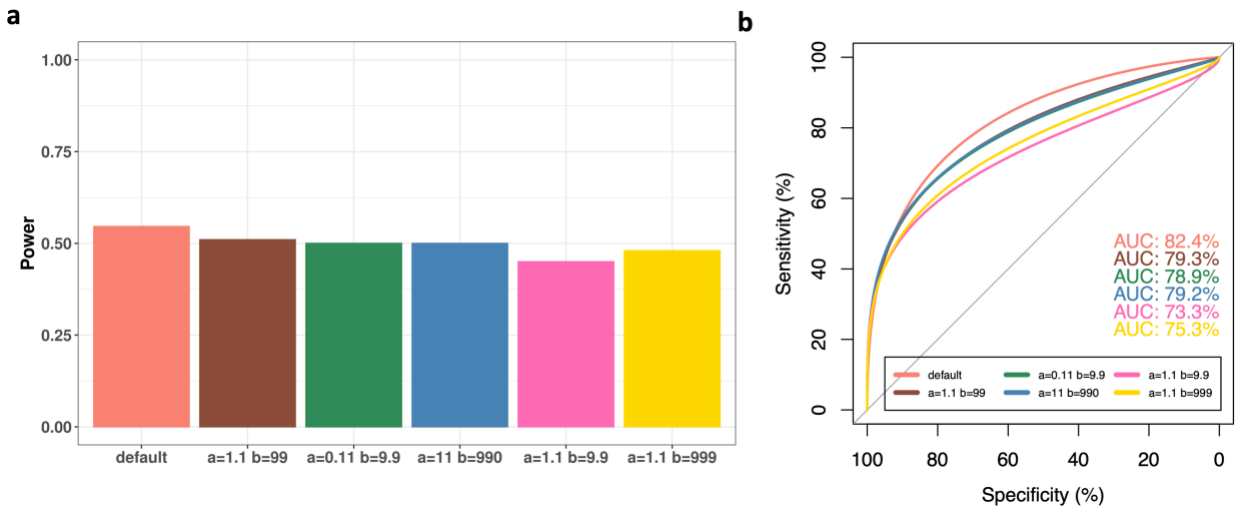

Figure S2. 9 Performance of FABIO with default or beta prior on π.

(a) Power comparison based on a true false discovery rate (FDR) threshold of 0.05. (b) ROC curves with AUCs recorded.

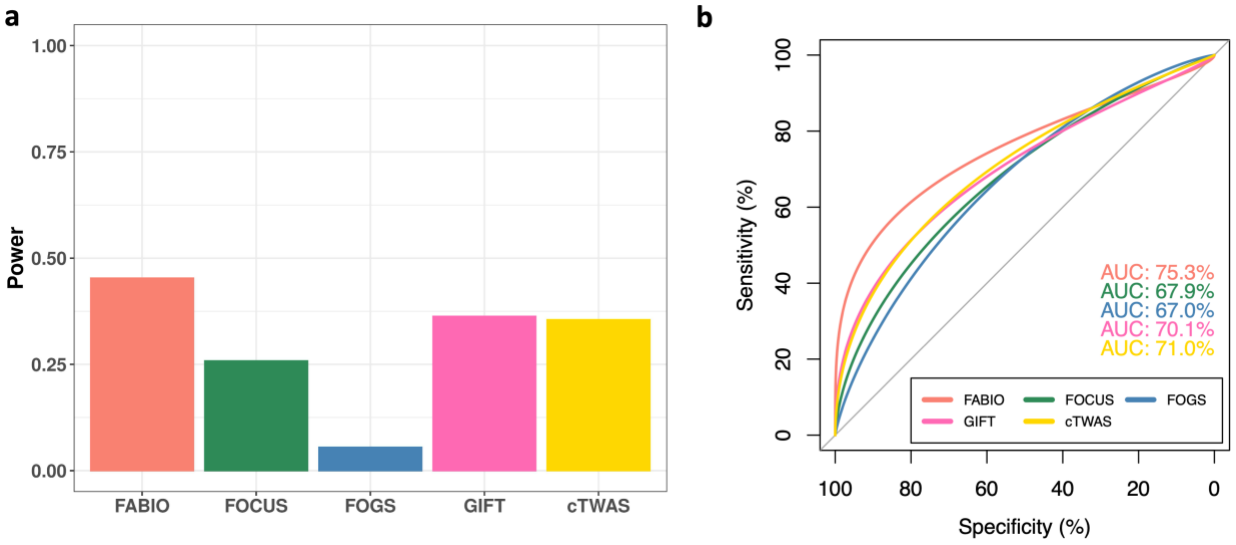

Figure S2. 10 Performance of all five methods.

(a) Power comparison based on a true false discovery rate (FDR) threshold of 0.05. (b) ROC curves with AUCs recorded.

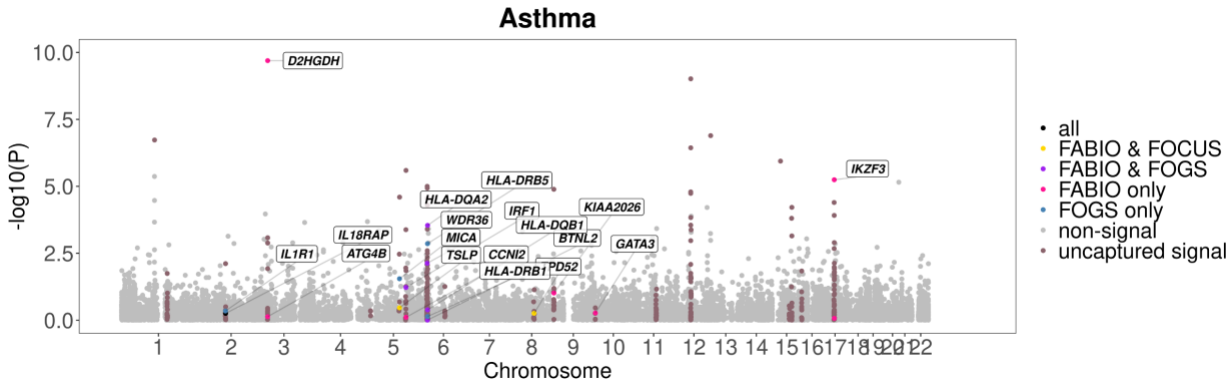

Figure S2. 11 TWAS Manhattan plot of asthma.

Genes identified in TWAS fine-mapping analysis are labeled in black boxes. Each gene is represented as a dot with x-axis indicating its genomic location and y-axis indicating its -log10 of p-value in the marginal TWAS association test. The dot of each gene is then colored based on different categories: (1) only identified by FABIO; (2) only identified by FOCUS; (3) only identified by FOGS; (4) identified by both FABIO and FOCUS; (5) identified by both FABIO and FOGS; (6) identified by both FOCUS and FOGS; (7) identified by all three methods; (8) located in a known risk region but missed by all three methods (uncaptured signal); (9) a gene shown no significance in both TWAS and GWAS analyses (non-signal).

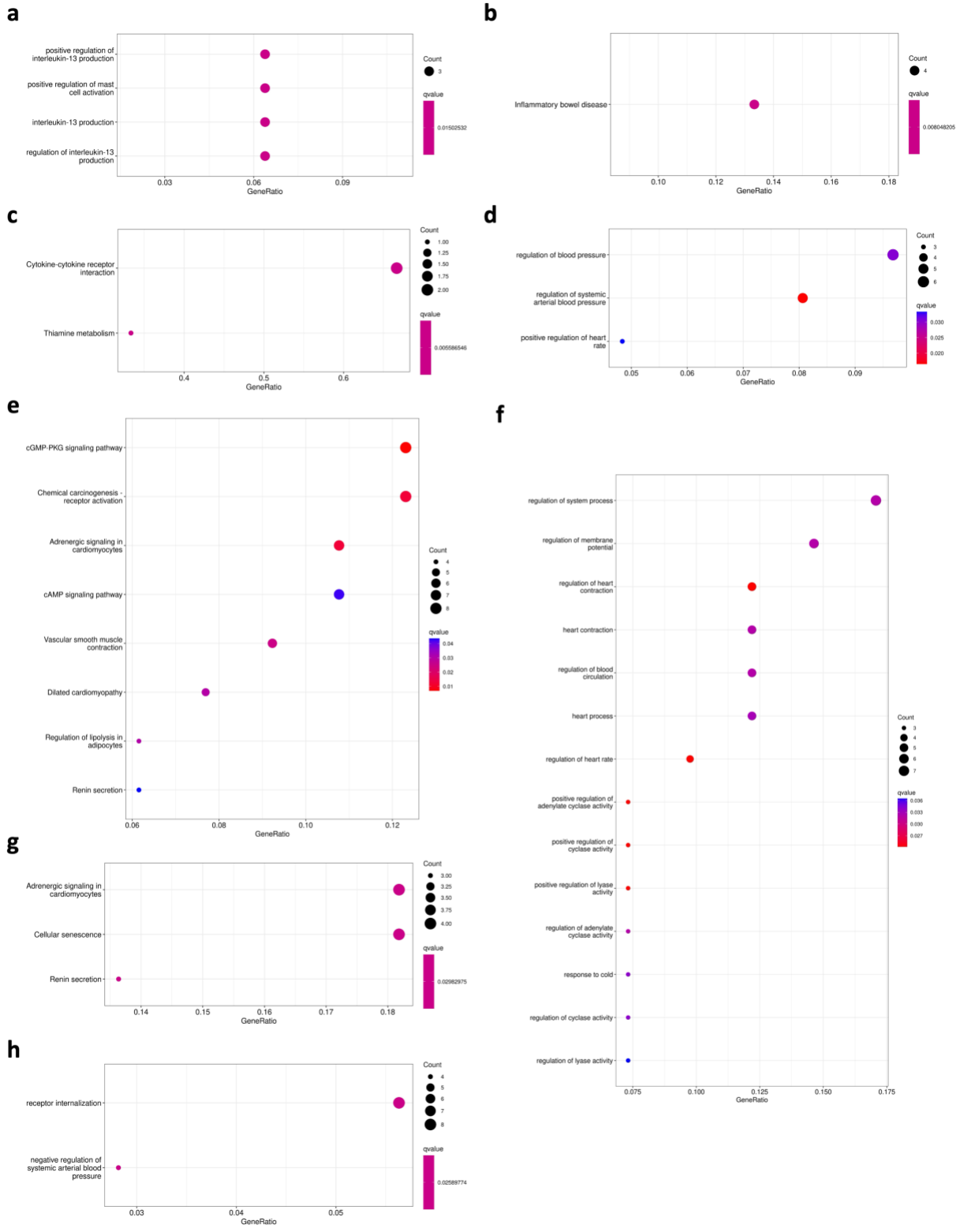

Figure S2. 12 Results of the Gene Ontology and KEGG pathway enrichment analysis.

Dot plots show the significant terms of identified genes identified by different fine-mapping methods, with color gradients representing statistical significance based on q-value. (a) Gene Ontology enriched pathways for asthma using

significant genes identified by FABIO. (b) KEGG enriched pathways for asthma using significant genes identified by FABIO. (c) KEGG enriched pathways for asthma using significant genes identified by FOCUS. (d) Gene Ontology enriched pathways for hypertension using significant genes identified by FABIO. (e) KEGG enriched pathways for hypertension using significant genes identified by FABIO. (f) Gene Ontology enriched pathways for hypertension using significant genes identified by FOCUS. (g) KEGG enriched pathways for hypertension using significant genes identified by FOCUS. (h) Gene Ontology enriched pathways for hypertension using significant genes identified by FOGS.

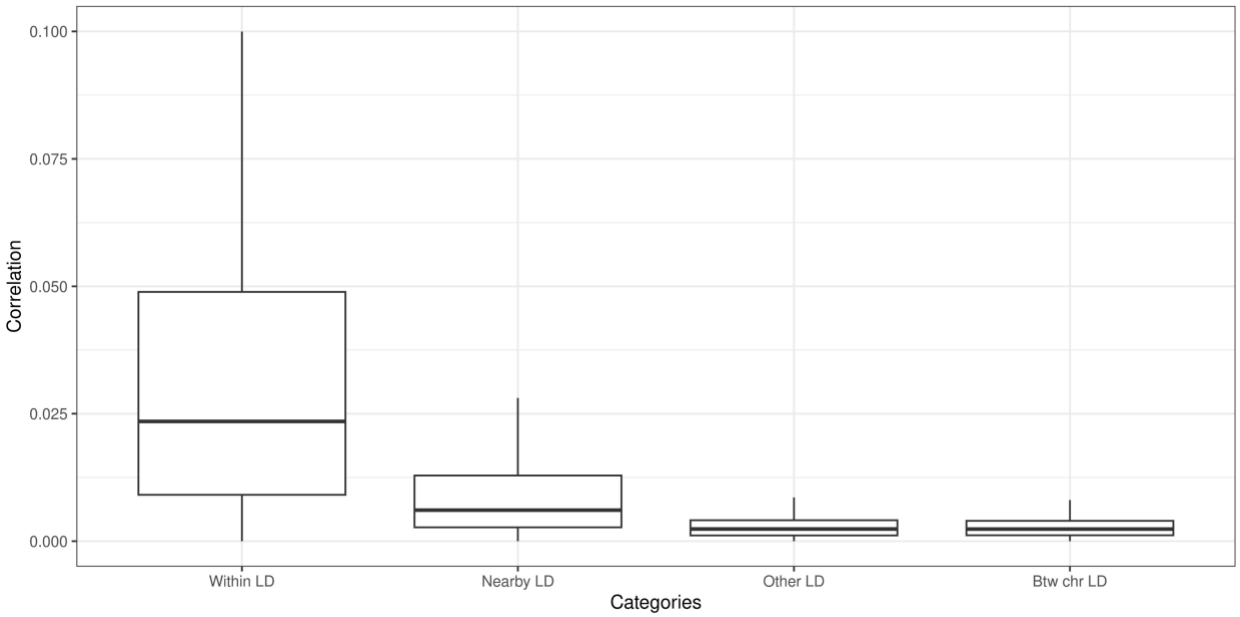

Figure S2. 13 Box plot of pairwise correlation of GReX for four distinct groups of gene pairs.

(1) Within LD: within each LD block, (2) Nearby LD: between two adjacent LD blocks, (3) Other LD: between two distant LD blocks, and (4) Btw chr LD: between one LD block on chromosome 1 and one on chromosome 2.

Chapter 3 mFABIO: An Integrative Multi-tissue TWAS Fine-mapping Approach to Prioritize Causal Genes and Tissues underlying Binary Traits

3.1 Introduction

Transcriptome-wide association studies (TWAS) integrate gene expression data with genome-wide association studies (GWAS) to identify potentially causal gene-trait associations (Gamazon et al., 2015; Gusev et al., 2016; Zhu et al., 2016). Conventional TWAS methods test the marginal association between the genetically regulated expression (GReX) of a single gene and a trait through a two-step process (Gamazon et al., 2015; Gusev et al., 2016; Wainberg et al., 2019). First, expression prediction models are built using local (cis-) single nucleotide polymorphisms (SNPs) in a reference dataset. Second, these models are used to impute GReX in the GWAS cohort and test its association with the trait of interest. By aggregating SNP-level signals into a single gene-based test, TWAS can increase statistical power and provide mechanistic insights into disease biology. However, a key limitation of these marginal approaches is confounding by nearby genes, as neighboring genes often exhibit correlated expression and are regulated by shared cis-SNPs in high linkage disequilibrium (LD) (Mancuso et al., 2019; Wainberg et al., 2019). Consequently, a gene identified by marginal TWAS may simply be a non-causal "tag gene" rather than the true causal one.

To address the confounding inherent in marginal analyses, TWAS fine-mapping methods were developed, which can prioritize causal genes by jointly modeling all genes within a locus of interest (Liu et al., 2024; Mancuso et al., 2019; Zhao et al., 2024). Similar to marginal TWAS, these methods follow a two-step process. First, they build GReX prediction models for all genes within a genomic locus using a reference expression panel (Gamazon et al., 2015; Zou & Hastie, 2005). Second, after imputing expression into a GWAS dataset, they apply a joint statistical model to identify the most likely causal gene(s) at that locus. For example, FOCUS (fine-mapping of causal gene sets) extends the Bayesian sparse regression model from GWAS fine-mapping toward

TWAS applications (Mancuso et al., 2019). GIFT (gene-based integrative fine-mapping through conditional TWAS) performs conditional TWAS analysis by explicitly adjusting for the GReX of other local genes to fine-map putatively causal genes (Liu et al., 2024). cTWAS (causal-TWAS) assesses the focal genes' association with a trait through a Bayesian regression model, while controlling for the effects of nearby genes and variants (Zhao et al., 2024). By analyzing multiple genes simultaneously, these fine-mapping methods provide a more comprehensive view of genetic associations. This joint modeling enables them to better dissect complex association signals and distinguish causal genes from those that are merely correlated.

Despite their advantages, current TWAS fine-mapping methods have a primary issue of their reliance on single-tissue gene expression reference data. They risk overlooking the true causal gene if the disease-relevant activity occurs in an unanalyzed tissue. Many complex diseases involve multiple tissues, where disease-associated genetic variants alter gene expression in a tissue-specific manner (Consortium, 2020). A gene's regulatory network and protein interactions can also vary significantly across different tissues (Barabási et al., 2011). To address this, a recently developed method directly models gene-tissue pairs by integrating GWAS with expression reference data from multiple tissues (Strober et al., 2025). However, there is a second, more fundamental limitation: most fine-mapping methods are designed for quantitative traits, that can be measured on a continuous scale, such as height or blood pressure. When applied to case-control studies, they still apply the same underlying linear regression framework, often relying on GWAS summary statistics generated under a similar linear assumption (Hautakangas et al., 2022; Mullins et al., 2021). While linear models can effectively approximate logistic regression for the small effect sizes typical of single SNPs in GWAS (Zhou et al., 2013), the aggregated effect sizes of GReX in TWAS can be substantially larger. Consequently, applying a linear model to a binary trait in a TWAS fine-mapping context may not accurately capture the binary nature of the outcome. Such modeling leads to a loss of statistical power and reduced performance for disease traits, which has been shown in a previous study (Zhang et al., 2024).

Here, we introduce mFABIO (**m**ulti-tissue **F**ine-m**A**pping of causal genes for **BI**nary **O**utcomes), a novel TWAS fine-mapping method that addresses the aforementioned limitations. Unlike existing methods, mFABIO utilizes a probit model to directly analyze binary outcomes and jointly models gene-tissue pairs across multiple tissues within a given locus. Similar to other approaches, it also accounts for potential horizontal pleiotropic effects. This comprehensive model

considers correlations in genetically regulated expression (GReX) among genes and tissues, thereby enhancing fine-mapping precision. We first validated mFABIO through comprehensive simulations, and then analyzed six binary disease traits from the UK Biobank, comparing it to four single-tissue TWAS fine-mapping methods (FABIO (Zhang et al., 2024), FOCUS (Mancuso et al., 2019), GIFT (Liu et al., 2024), cTWAS (Zhao et al., 2024)) and one multi-tissue method (TGFM (Strober et al., 2025)). In the UK Biobank analysis, mFABIO identified at least 14.8% more evidence-supported gene-tissue pairs compared to all competing methods. While increasing discovery, it also improved specificity by reducing the number of candidate gene-tissue pairs by at least 50.8% compared to single-tissue approaches. Notably, mFABIO prioritized multiple biologically plausible gene-tissue pairs, such as *D2HGDH* in lung tissue for asthma, *CYBRD1* in breast mammary tissue for breast cancer, and *CCR6* in spleen tissue for rheumatoid arthritis. These results establish mFABIO as a powerful and precise tool for multi-tissue TWAS fine-mapping of binary disease traits.

3.2 Materials and methods

3.2.1 Model for multi-tissue TWAS fine-mapping of binary traits

We consider a genome-wide association study (GWAS) of a binary trait conducted on n individuals, where $\boldsymbol{y} \in \{0,1\}^n$ represents the n-vector of binary phenotypes. Our objective is to use transcriptome-wide association study (TWAS) fine-mapping to identify genes whose genetically regulated expression (GReX) is associated with the trait and to prioritize the tissues most relevant to these associations. Different from existing single-tissue TWAS fine-mapping methods (Liu et al., 2024; Mancuso et al., 2019; Zhang et al., 2024; Zhao et al., 2024), we leverage tissue-specific GReX constructed from multiple tissues and jointly model all gene-tissue pairs within a defined genomic region, while explicitly accounting for the binary nature of the outcome trait. To do so, we consider K different tissues and examine genomic regions one at a time. For each genomic region of focus, we assume there are M genes in total and denote p_k as the number of genes with GReX constructed in tissue k ($p_k \leq M$). These GReX values are assumed to have been pre-computed using standard software such as SuSiE (Wang et al., 2020). Let $\widehat{\boldsymbol{G}}_{n\times p_k}$ denote the corresponding GReX matrix for these genes across n individuals in tissue k. We then define the overall GReX matrix $\widehat{\boldsymbol{G}}$ as an $n \times p$ matrix for all p gene-tissue pairs in that region as the

concatenation of the K tissue-specific GReX matrices: $\widehat{\boldsymbol{G}} \coloneqq \left[\widehat{\boldsymbol{G}}_{n\times p_1}, \widehat{\boldsymbol{G}}_{n\times p_2}, \cdots, \widehat{\boldsymbol{G}}_{n\times p_K}\right]_{n\times p}$, where $p = \sum_{k=1}^{K} p_k$. We denote $\boldsymbol{X}$ as an $n \times s$ genotype matrix for the same n individuals, capturing s SNPs within the region of focus. To facilitate computation, we center and standardize each column of $\widehat{\boldsymbol{G}}$ as well as each column of $\boldsymbol{X}$ to have a mean of zero and a standard deviation of one following (LaPierre et al., 2021). To accommodate the binary nature of the trait, we consider the following probit model linking the GReX to the binary outcome:

$$P(y_i = 1 \big| \widehat{\boldsymbol{G}}_i, \boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}) = 1 - P(y_i = 0 \big| \widehat{\boldsymbol{G}}_i, \boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \Phi\left(\mu + \widehat{\boldsymbol{G}}_i{}^T \boldsymbol{\beta} + \boldsymbol{X}_i{}^T \boldsymbol{\alpha}\right) (i = 1, \cdots, n) \quad (3.1)$$

Here $y_i \in \{0,1\}$ is the binary trait for the i'th individual; $\widehat{\boldsymbol{G}}_i$ is a p-vector of GReX for p gene-tissue pairs for the i'th individual (i.e. the i'th row of $\widehat{\boldsymbol{G}}$); $\boldsymbol{\beta}$ is a p-vector of corresponding effect sizes for the gene-tissue pairs; $\boldsymbol{X}_i$ is an s-vector of genotypes for the i'th individual (i.e. the i'th row of $\boldsymbol{X}$); $\boldsymbol{\alpha}$ is an s-vector of corresponding effect sizes for SNPs, also known as the horizontal pleiotropic effects, capturing the SNPs' effects on the trait not mediated through gene expression; μ is a scalar representing the intercept; and Φ is the cumulative distribution function (CDF) of the standard normal distribution.

The model above can be equivalently re-parameterized using a liability threshold framework. Specifically, following (Albert & Chib, 1993), we introduce a latent variable z_i for each individual $i \in \{1, \dots, n\}$, representing their underlying liability. This yields the following equivalent latent variable representation of Equation (1):

$$y_i = \begin{cases} 1 & if\ z_i > 0 \\ 0 & if\ z_i \leq 0 \end{cases}, \quad (3.2)$$

$$z_i = \mu + \widehat{\boldsymbol{G}}_i{}^T \boldsymbol{\beta} + \boldsymbol{X}_i{}^T \boldsymbol{\alpha} + \varepsilon_i \qquad \varepsilon_i \sim N(0,1). \quad (3.3)$$

We place independent Gaussian priors on each element of the SNP-effect vector $\boldsymbol{\alpha}$ across the s SNPs:

$$\boldsymbol{\alpha} \sim N(\boldsymbol{0}, \sigma_\alpha^2 \boldsymbol{I}_s), \quad (3.4)$$

where we further assume that σ_α^2 follows a conjugate and weakly informative inverse-gamma prior $\sigma_\alpha^2 \sim IG(1,1)$.

3.2.2 Decomposition of gene-tissue pair effects

Our goal is to identify the potential causal gene-tissue pair underlying the trait, by inferring the tissue-specific posterior probability that the GReX of each gene is associated with the trait. To do so, we posit that only a small subset of the p possible gene-tissue pairs displays non-zero effects. To capture this sparse structure, we adopt the Sum of Single Effects (SuSiE) prior (Wang et al., 2020) for the gene-tissue effect sizes $\boldsymbol{\beta}$ in Equation (3.3), which is decomposed as a sum of L single-effect components:

$$\boldsymbol{\beta} \coloneqq \sum_{l=1}^{L} \beta_l \boldsymbol{\gamma}_l. \tag{3.5}$$

Here we assume that there are at most L gene-tissue pairs in the region with non-zero effects (L is set to be 10 in the present study following (Wang et al., 2020)). Each $\beta_l \boldsymbol{\gamma}_l$ term is a single-effect component, where β_l is a scalar effect-size parameter following a normal prior $N\left(0, \sigma_\beta^2\right)$, for $l = 1,2,\dots,L$. For the l'th single-effect component ($l \in \{1,2,\dots,L\}$), we use a p-vector of binary indicators $\boldsymbol{\gamma}_l$ to indicate which gene-tissue pair is responsible for the l'th component: the element in $\boldsymbol{\gamma}_l$ that corresponds to the l'th non-zero effects is 1 while all the other elements in $\boldsymbol{\gamma}_l$ are 0's. We further assume a prior distribution $\boldsymbol{\gamma}_l \sim Multinomial(1, \boldsymbol{\pi})$, allowing each single effect to correspond to a specific column of $\widehat{\boldsymbol{G}}$. Here, the prior non-zero probability for each element of $\boldsymbol{\gamma}_l$ is determined by each element of the p-vector $\boldsymbol{\pi} \coloneqq (\pi_1, \dots, \pi_p)$. The structure of $\boldsymbol{\pi}$ will be introduced in the following section.

3.2.3 Hierarchical prior for variable selection

To infer causal probabilities at both the gene and gene-tissue pair levels, we first define π_j as the j'th element of the p-vector $\boldsymbol{\pi}$ as defined in the preceding section. Here, the j'th index corresponds to the specific pair of the k'th tissue and the m'th gene, thus $j \in \{1, \cdots, KM\}$ for K tissues and M genes. The prior probability $\pi_j \coloneqq P(\gamma_{lj} = 1)$, represents the probability that the j'th gene-tissue pair is responsible for the l'th single-effect component. We define this prior in a hierarchical fashion. Specifically, following (Ročková & George, 2014), we first denote the vector of prior probabilities for selecting each of the M unique genes as an M-vector $\boldsymbol{\pi}_{\boldsymbol{G}} = (\pi_{g_1}, \dots, \pi_{g_M})$. For the m'th unique gene g_m, we then denote the corresponding vector of prior probabilities for selecting each tissue given that the gene g_m has been selected as a K-vector $\boldsymbol{\pi}_{\boldsymbol{T}|g_m} = (\pi_{t_1|g_m}, \dots, \pi_{t_K|g_m})$. Afterwards, we assume a *priori* on π_j as:

$$\begin{aligned}\pi_j &= P(select\ pair\ \{g_m, t_k\}) \\ &= P(select\ gene\ g_m) \times P(select\ tissue\ t_k |\ gene\ g_m\ is\ selected)\end{aligned}$$

$$= \pi_{g_m} \times \pi_{t_k|g_m} \tag{3.6}$$

where $\sum_{j=1}^{p} \pi_j$ is ensured to be 1 under this structure.

We further specify Dirichlet priors on the gene-level probability vector $\boldsymbol{\pi_G}$ and on each of the tissue probability vectors $\boldsymbol{\pi_{T|g_m}}$: $\boldsymbol{\pi_G} \sim Dirichlet(1, \dots, 1)$ and $\boldsymbol{\pi_{T|g_m}} \sim Dirichlet(1, \dots, 1)$, to represent an uninformative starting belief that all genes, and all tissues within a gene, are equally likely to be selected. This choice is statistically motivated for two primary reasons. First, the Dirichlet distribution's support is the probability simplex, which is the valid domain for probability vectors that must sum to one. Second, and critically for computational feasibility, it is the conjugate prior for the categorical selection process. This property of conjugacy ensures that the posterior distributions of the probability vectors remain in the Dirichlet family, making their estimation tractable within our following variational Bayes inference, and allowing the model to learn data-driven, gene- and tissue-specific probabilities empirically.

3.2.4 Scalable variational inference

With the above model settings, our goal is to infer the posterior probability that each element of $\boldsymbol{\gamma}_l$ equals one. This is achieved through the posterior estimates of the nested vectors $\boldsymbol{\pi_G}$ and $\boldsymbol{\pi_{T|*}}$, which together enable the identification of causal genes, and subsequently, the causal tissues underlying these genes. To ensure scalable inference, we developed a variational Bayes algorithm. Specifically, we denote $\boldsymbol{\theta}$ as the set of hyperparameters defined in the preceding section, where $\boldsymbol{\theta} = (\sigma_\mu^2, \sigma_\alpha^2, \sigma_\beta^2)$, and $\boldsymbol{z}$ as the latent vector containing all z_i's defined in Equations (3.2) and (3.3). We then employ an empirical Bayes approach to obtain an estimate of $\widehat{\boldsymbol{\theta}}$ (details in the Appendix B). With the estimated $\widehat{\boldsymbol{\theta}}$, we approximate the complex conditional posterior distribution $p\left(\boldsymbol{z}, \mu, \boldsymbol{\beta}, \boldsymbol{\alpha} \middle| \widehat{\boldsymbol{G}}, \boldsymbol{X}, \boldsymbol{y}, \widehat{\boldsymbol{\theta}}\right)$ using an approximate conditional posterior distribution of $q\left(\boldsymbol{z}, \mu, \boldsymbol{\beta}, \boldsymbol{\alpha} | \widehat{\boldsymbol{G}}, \boldsymbol{X}, \boldsymbol{y}, \widehat{\boldsymbol{\theta}}\right)$. For the simplicity of the notations, we will omit the conditioning on $(\widehat{\boldsymbol{G}}, \boldsymbol{X}, \boldsymbol{y}, \widehat{\boldsymbol{\theta}})$ in the following text. We first assume that the approximate conditional posterior distribution $q(\boldsymbol{z}, \mu, \boldsymbol{\beta}, \boldsymbol{\alpha})$ is in the form of the following factorization:

$$q(\boldsymbol{z}, \mu, \boldsymbol{\beta}, \boldsymbol{\alpha}) = q(\boldsymbol{z}) \times q(\mu) \times q(\boldsymbol{\alpha}) \times \prod_{l=1}^{L} q(\beta_l) q(\boldsymbol{\gamma}_l)$$
$$= \left(\prod_{i=1}^{n} q(z_i)\right) \times q(\mu) \times q(\boldsymbol{\alpha}) \times \left(\prod_{l=1}^{L} q(\beta_l) q(\boldsymbol{\gamma}_l | \boldsymbol{\pi_G}, \boldsymbol{\pi_{T|*}}) q(\boldsymbol{\pi_G}) \left(\prod_{*=g_1}^{g_M} q(\boldsymbol{\pi_{T|*}})\right)\right) \tag{3.7}$$

which contains seven components: $q(z_i)$, $q(\mu)$, $q(\boldsymbol{\alpha})$, $q(\beta_l)$, $q(\boldsymbol{\gamma}_l|\boldsymbol{\pi}_{\boldsymbol{G}}, \boldsymbol{\pi}_{\boldsymbol{T}|*})$, $q(\boldsymbol{\pi}_{\boldsymbol{G}})$, and $q(\boldsymbol{\pi}_{\boldsymbol{T}|*})$. For the approximate posterior $q(z_i)$, we assume it follows a truncated normal distribution, whose parameters are updated based on the observed binary trait y_i. For the approximate posterior $q(\mu)$, we assume it follows a normal distribution, given μ has a normal prior as well as the normal-normal conjugation. Similarly, for the approximate posterior $q(\boldsymbol{\alpha})$, we assume it follows a multivariate normal distribution, given its multivariate normal prior defined earlier. For the approximate posterior $q(\beta_l)$, we assume it also follows a normal distribution for $l = 1,2, \dots, L$. For the approximate posterior $q(\boldsymbol{\gamma}_l|\boldsymbol{\pi}_{\boldsymbol{G}}, \boldsymbol{\pi}_{\boldsymbol{T}|*})$, we assume it is in the form of a multinomial distribution, with

$$q\left(\boldsymbol{\gamma}_l|\boldsymbol{\pi}_{\boldsymbol{G}}, \boldsymbol{\pi}_{\boldsymbol{T}|*}\right) \sim Multinomial(1, \tilde{\boldsymbol{\pi}}_l), \tag{3.8}$$

where the p-vector $\tilde{\boldsymbol{\pi}}_l = \left(\tilde{\pi}_{l1}, \cdots, \tilde{\pi}_{lp}\right)$ represents the posterior probability that each gene-tissue pair represents the l'th non-zero effect, and the corresponding estimate is known as the posterior inclusion probability (PIP). For the approximate posteriors $q(\boldsymbol{\pi}_{\boldsymbol{G}})$ and $q(\boldsymbol{\pi}_{\boldsymbol{T}|*})$, both are assumed to follow the Dirichlet distributions, and the posterior Dirichlet distribution $q(\boldsymbol{\pi}_{\boldsymbol{G}})$ over the M unique genes is used to obtain the gene-level PIP. With the specified form of the approximate posterior $q(\boldsymbol{z}, \mu, \boldsymbol{\beta}, \boldsymbol{\alpha})$, we aim to minimize its Kullback-Leibler (KL) divergence with the actual posterior $p(\boldsymbol{z}, \mu, \boldsymbol{\beta}, \boldsymbol{\alpha})$. Minimizing the KL divergence is equivalent to maximizing the evidence lower bound (ELBO), which can be achieved by an iterative Bayesian stepwise selection algorithm following the strategy of SuSiE (Wang et al., 2020). The detailed inference steps for each parameter are provided in the Appendix B.

3.2.5 Calculation of posterior inclusion probability (PIP)

To obtain the association evidence of each gene-tissue pair, we calculate the posterior probability that the j'th gene-tissue pair has a non-zero effect, or PIP_j, which equals to that the j'th gene-tissue pair is selected by at least one of the L single-effect components. Within our variational approximation, we assume the selections for each component are independent in the posterior. Thus, we first calculated the posterior probability that none of the L single-effect component selects the j'th gene-tissue pair:

$$q(\text{the } j\text{'th pair is not selected}) = \prod_{l=1}^{L}\left(1 - \tilde{\pi}_{lj}\right), \tag{3.9}$$

and the PIP for the j'th gene-tissue pair can be calculated as:

$$PIP_j = 1 - \prod_{l=1}^{L}\left(1 - \tilde{\pi}_{lj}\right), \;\; j = 1, \dots, p \tag{3.10}$$

To obtain the association evidence of each gene, we calculate the overall posterior probability that the gene g_m is involved in the phenotype through at least one of its tissues, or PIP_{g_m}. Within our variational approximation, the parameter vector $\boldsymbol{\pi_G}$ represents the probabilities of selecting each gene, so we can naturally estimate the gene-level PIP using the posterior mean of the corresponding gene selection probability. Specifically, assuming the final approximate posterior distribution for $\boldsymbol{\pi_G}$ is: $q(\boldsymbol{\pi_G}) = Dirichlet(\tilde{\delta}_{g_1}, \dots, \tilde{\delta}_{g_M})$, and $\boldsymbol{\pi_G} \coloneqq (\pi_{g_1}, \dots, \pi_{g_M})$, we have:

$$PIP_{g_m} = E_q[\pi_{g_m}] = \frac{\tilde{\delta}_{g_m}}{\sum_{m'=1}^{M} \tilde{\delta}_{g_{m'}}}, \quad m = 1, \dots, M \tag{3.11}$$

The PIP in our model serves as an important measure of evidence for the gene or gene-tissue pair associated with the outcome trait. We refer to our method and inference algorithm as the **m**ulti-tissue **F**ine-m**A**pping of causal genes for **BI**nary **O**utcomes (mFABIO). mFABIO is implemented as an R package, freely available at https://xiangzhou.github.io/software/.

3.2.6 Simulations

We conducted simulations to assess the performance of mFABIO and compare it with four single-tissue TWAS fine-mapping methods (FABIO, FOCUS, GIFT, cTWAS) and one multi-tissue method (TGFM). We used real genotype data from the UK Biobank, simulating both gene expression and a binary outcome trait. Specifically, we randomly sampled 50,000 individuals of European ancestry in UK Biobank to serve as the GWAS data and another 500 individuals of European ancestry (disjoint from GWAS samples) to serve as the eQTL mapping data. Our simulations focused on 426,593 SNPs on chromosome 1, which were partitioned into 133 independent LD blocks using LDetect (Berisa & Pickrell, 2015), following the same procedure of (Liu et al., 2024; Mancuso et al., 2019; Zhang et al., 2024; Zhao et al., 2024). From these, we randomly selected 50 LD blocks, each containing at least five genes. Within these selected blocks (containing 5-43 genes per block; mean = 16, median = 13), we extracted cis-SNPs located within 100 kb upstream of the transcription start site (TSS) and 100 kb downstream of the transcription end site (TES) of each gene, following (Liu et al., 2024; Zhang et al., 2024). We retained 789 genes that had at least 10 cis-SNPs, and these were used for the following simulations.

We then simulated the genetic architecture of gene expression across multiple tissues in both the GWAS and eQTL mapping panels. Following (Strober et al., 2025), we simulated all 789 selected genes to be expressed in three tissues and designated 50% of these gene-tissue pairs as

cis-genetically heritable. For each heritable pair, we randomly assigned five causal cis-SNPs: three shared across tissues and two specific to each tissue, and assumed each causal cis-SNP to explain a fixed proportion of gene expression variance. Following (Yuan et al., 2020; Zhang et al., 2024), we denoted the proportion of gene expression variance explained by genetic effects for each heritable gene-tissue pair as PVE_1. While all expressed gene-tissue pairs were considered during fine-mapping, only these heritable pairs were assigned simulated causal effects on the final trait. In the eQTL mapping panel, we applied SuSiE to the corresponding real genotypes and simulated gene expression data to generate prediction models for each gene-tissue pair.

We simulated the binary trait in the GWAS panel using a logistic regression model based on the simulated gene expression data. The probability of the binary trait $(y_i = 1)$ for the *i*'th individual was determined by a latent variable η_i as:

$$P(y_i = 1|\eta_i) = \frac{1}{1+\exp(-\eta_i)}. \quad (3.12)$$

The latent variable η_i links the trait to the predicted gene expression and is defined as a weighted sum of causal gene effects across three tissues, following the structure from (Hu et al., 2019):

$$\eta_i = u + w_1\widehat{\boldsymbol{g}}_{1i}^T\boldsymbol{v_1} + w_2\widehat{\boldsymbol{g}}_{2i}^T\boldsymbol{v_2} + w_3\widehat{\boldsymbol{g}}_{3i}^T\boldsymbol{v_3} + \epsilon_i, \quad \epsilon_i \sim N\left(0, 1 - \left[n_g \cdot PVE_2\right]\right), \quad (3.13)$$

where u is an intercept that controls the case:control ratio as it equals to the expectation of log(case#/control#); w_1, w_2, w_3 are the contributing weights for each tissue, with $w_1 + w_2 + w_3 = 1$; $\widehat{\boldsymbol{g}}_{*i}$ is an n_g-vector of simulated gene expression of the selected causal gene-tissue pairs for the *i*'th individual in the corresponding tissue. For each tissue, we randomly selected n_g causal gene-tissue pairs from the heritable pairs, and denote the proportion of the phenotype's variance explained by each causal gene-tissue pair as PVE_2, following (Yuan et al., 2020; Zhang et al., 2024). And $\boldsymbol{v}_*$ is an n_g-vector of genes' causal effect sizes in the corresponding tissue, with each element $\boldsymbol{v}_{*j}$ set to be: $\boldsymbol{v}_{*j} = \sqrt{PVE_2/PVE_1}$, following (Yuan et al., 2020; Zhang et al., 2024). Notably, we intentionally used a logistic regression model instead of a probit model to assess our method's performance under model misspecifications.

In the simulations, we first examined the calibration of test statistics from different methods under the null simulation settings, where we set $PVE_2 = 0\%$ so that none of the gene-tissue pairs were associated with the trait. Besides the null simulations, we also examined the power of different methods on detecting the causal genes among tissues under the alternative simulation settings with non-zero PVE_2. For both the null and alternative simulations, we started from a

baseline simulation setting. We then varied one parameter at a time on top of the baseline setting to examine the influence of different parameters. Specifically, for PVE_1, we set it to be 5%, 7.5%, 10% (baseline setting), following the same settings of (Strober et al., 2025). For the number of causal cis-SNPs for each heritable gene-tissue pair, we set it to be either 5 (baseline setting), or a random number between 3 and 7 (inclusive) following (Strober et al., 2025). For PVE_2, we varied its value to be either 0% (null setting), 0.2%, 0.4% (baseline setting), or 0.6% following (Yuan et al., 2020; Zhang et al., 2024). For the number of causal genes per tissue (n_g), we fixed it to be 10 following (Zhang et al., 2024). In terms of the contributing weights for each tissue, we considered three settings following (Hu et al., 2019): (1) $w_1 = 1, w_2 = w_3 = 0$; (2) $w_1 = w_2 = 1/2, w_3 = 0$ (baseline setting); and (3) $w_1 = w_2 = w_3 = 1/3$. For the binary trait, we examined different case:control ratios by varying the intercept u, using 1:1 as the baseline setting ($u = 0$) while exploring other ratios of 1:10 ($u = \log(0.1)$), and 1:50 ($u = \log(0.02)$).

We performed 100 simulation replicates for each simulation setting. In each replicate, we applied all the methods except TGFM to analyze one LD block at a time following their design. In contrast, TGFM was applied to each chromosome, as its software is designed to internally partition chromosome-level data into overlapping 3 Mb loci for analysis. For multi-tissue methods (mFABIO and TGFM), we directly applied them on gene-tissue pairs, while for single-tissue methods (FABIO, FOCUS, GIFT, and cTWAS), we applied them in two ways: (1) applied them on each tissue with corresponding tissue-specific predicted gene expression levels following their original designs, then combined the results across tissues; (2) naturally extended their designs to analyze gene-tissue pairs jointly, by preparing the gene-tissue pair level input files of cis-predicted expression models, and then applied them on gene-tissue pairs. We refer to the resulting methods in (2) as FABIO-gt, FOCUS-gt, GIFT-gt, and cTWAS-gt, respectively. In either approach, we summarized the results across simulation replicates. In the analysis, we first examined the calibration of test statistics from different methods under the null simulation settings, by examining the number of false discoveries at an estimated false discovery rate (FDR) of 0.05. The calculation of the estimated FDR will be introduced in the following paragraph. We then evaluated the power of different methods at the FDR of 0.05 under the alternative simulation settings. Since we knew the ground truth in the simulations, we first compared the power of different methods under a true FDR threshold of 0.05. Specifically, to compute the true FDR in the simulations, we first divided 100 simulation replicates into five groups, with 20 replicates per group. In each group, we obtained

the test statistics from each method for all gene-tissue pairs across the 20 replicates, either in the form of p-values or PIPs. For each method in turn, we sorted the test statistics from the most significant (i.e. the smallest p-value or the largest PIP) to the least significant (i.e. the largest p-value or the smallest PIP). For each ordered test statistic in the sorted list, we counted the number of false positives detected above the threshold defined by this specific test statistic. The number of false positives, divided by the total number of gene-tissue pairs called significant at that threshold, is the true FDR. We then determined threshold that corresponds to a true FDR of 0.05, based on which we further calculated power. Power is calculated as the number of true positives above the threshold in the sorted list divided by the total number of causal gene-tissue pairs. Afterwards, we obtained the average power across five groups to minimize stochasticity (Liu et al., 2024; Yuan et al., 2020; Zhang et al., 2024).

We also compared the methods at gene level. For multi-tissue methods mFABIO and TGFM that output gene-level PIPs, we directly used the same way to calculate the power at a true FDR of 0.05, based on the gene-level PIPs. For single-tissue methods FABIO, FOCUS, and cTWAS that only output PIPs at gene-tissue pair level, we first estimated the gene-level PIP by calculating the probability that the gene is causal in at least one tissue with the assumption of independence: $PIP_{gene} = 1 - \prod_m (1 - PIP_m)$. Then we calculate the power at a true FDR of 0.05, based on the estimated gene-level PIPs. For single-tissue method GIFT that only outputs p-values at gene-tissue pair level, we first estimated the gene-level p-values using Fisher's method (Fisher, 1970) to combine the p-values with the assumption of independence. Then we calculated the power at a true FDR of 0.05, based on the estimated gene-level p-values.

Because we can only compute the true FDR in the simulations but not in the real data, we also compared the power of different methods under an estimated FDR threshold of 0.05. We estimated FDR using different approaches: the Benjamini-Hochberg approach (Benjamini & Hochberg, 1995) to estimate FDR using p-values from the frequentist method (GIFT), and the local FDR approach (Efron, 2007) to estimate FDR using PIPs from the Bayesian methods (mFABIO, TGFM, FABIO, FOCUS, and cTWAS). Specifically for the local FDR approach, we first calculated the local FDR for all the gene-tissue pairs, which is effectively 1-PIP (Liu et al., 2024; Zhang et al., 2024). We then sorted gene-tissue pairs based on these local FDRs from small to large. For each gene-tissue pair in turn, we estimated FDR for that pair by aggregating the local FDRs from the first pair up to that pair in the sorted list. Afterwards, we identified the cutoff value

of PIP/local FDR that corresponds to a targeted FDR threshold of 0.05. This is the significant threshold we use that corresponds to an estimated FDR of 0.05. We also compared the methods at gene level, by first estimating the gene-level PIPs or p-values as illustrated previously, and then we calculated the power at an estimated FDR of 0.05.

3.2.7 Real datasets

3.2.7.1 Genotype-Tissue Expression (GTEx) data

We constructed the genetically regulated expression (GReX) using publicly available cis-predicted expression models (https://doi.org/10.7910/DVN/8IPOPK) derived from the Genotype-Tissue Expression (GTEx) project (Strober et al., 2025). These models were built using data from individuals of European ancestry. To mitigate heterogeneity in expression quantitative trait loci (eQTL) sample sizes, the original 47 GTEx tissues were aggregated into 38 meta-tissues with more homogeneous sample sizes. For each meta-tissue, models included only protein-coding genes, with cis-eQTLs defined as variants within 500 Kb of a gene's transcription start site (TSS). A series of quality control filters were applied: we removed variants with a minor allele frequency (MAF) < 0.05, strand-ambiguous variants, and those not present in the UK Biobank dataset. This filtering resulted in a final set of 10,545,304 variants for model building. Prediction models for each gene-tissue pair were then generated by applying SuSiE (Wang et al., 2020) to the eQTL summary statistics and in-sample linkage disequilibrium (LD) matrices. More details on the model construction can be found in (Strober et al., 2025). Ultimately, this process yielded models for 799 to 5,496 genes per meta-tissue, comprising a total of 13,700 unique genes and 119,270 unique gene-tissue pairs.

3.2.7.2 UK Biobank (UKBB) data

The UK Biobank data consists of 487,409 individuals and 92,693,895 imputed autosome SNP variants (Bycroft et al., 2018). We followed the same sample QC procedure as Neale Lab (https://github.com/Nealelab/UK_Biobank_GWAS/tree/master/imputed-v2-gwas) to retain a total of 337,198 individuals of European ancestry. To ensure compatibility with the cis-predicted expression models described previously, the UKBB genotype data was then filtered to retain only the 10,545,304 variants used to build those models.

3.2.8 Real data application

We performed fine-mapping on six binary disease traits through integrating the GTEx cis-predicted expression models on different meta-tissues with the UKBB genotype data. The examined disease traits include asthma, breast cancer, gout, hypertension, prostate cancer, and rheumatoid arthritis. These disease traits were selected following (Yang & Zhou, 2022; Zhang et al., 2024) and have a wide range of prevalence (0.02 to 0.38). For the two sex-related traits (breast cancer and prostate cancer), we limited our analysis to female individuals (for breast cancer) or male individuals (for prostate cancer). For a given trait in UKBB, we kept all individuals of European ancestry with self-reported case status of that trait in the GWAS data. Same as the simulations, SuSiE-based cis-predicted expression models were used for all the comparing methods (mFABIO, TGFM, FABIO, FOCUS, GIFT, and cTWAS). For the multi-tissue methods mFABIO and TGFM, we directly used the cis-predicted expression models from all 38 meta-tissues. In contrast, for the single-tissue methods (FABIO, FOCUS, GIFT, and cTWAS), we naturally extended their designs to analyze gene-tissue pairs jointly, a strategy that showed superior performance in simulations over single-tissue analyses. Same as the simulations, we prepared the gene-tissue pair level input files of cis-predicted expression models, and then applied them on gene-tissue pairs. The genotype input data was then prepared based on each method's design. For mFABIO and FABIO, which are built for binary outcomes, we used individual-level GWAS genotypes. For the other four methods (TGFM, FOCUS, GIFT, and cTWAS), which require summary statistics, we generated these by performing a logistic regression on the GWAS data using PLINK 2.0 (Chang et al., 2015), adjusting for the top 10 genotype PCs and sex. We then paired either the individual-level GWAS or the GWAS summary statistics with the variant weights from cis-predicted expression models as input. We analyzed 1,433 independent LD blocks (Berisa & Pickrell, 2015) across all 22 autosomal chromosomes one at a time, except for TGFM. We applied TGFM to 2,682 overlapping 3 Mb loci spanning the entire genome following its own design. We also performed marginal TWAS analysis for each tissue-disease pair using FUSION (Gusev et al., 2016) with the same GWAS summary statistics and the variant weights. For a fair comparison across methods, we used an estimated FDR threshold of 0.05 for all methods to declare significance, as described in details in the simulations.

To validate our findings, we further evaluated whether significant gene-tissue pairs (estimated FDR < 0.05) were enriched in tissues with known disease relevance. We established this relevance by querying PubMed for publications over the last 10 years for each disease-tissue combination.

The three GTEx tissue categories with the highest publication counts for each disease were designated potentially "disease-relevant". This literature review yielded the following classifications:

- asthma: lung, skin (GTEx tissues: Skin_Sun_Exposed_Lower_leg and Skin_Not_Sun_Exposed_Suprapubic), and blood (GTEx tissue: Whole_Blood)
- breast cancer: breast mammary tissue, blood (GTEx tissue: Whole_Blood), and liver
- gout: blood (GTEx tissue: Whole_Blood), liver, and skin (GTEx tissues: Skin_Sun_Exposed_Lower_leg and Skin_Not_Sun_Exposed_Suprapubic)
- hypertension: blood (GTEx tissue: Whole_Blood), heart (GTEx tissues: Heart_Atrial_Appendage and Heart_Left_Ventricle), and artery (GTEx tissues: Artery_Aorta, Artery_Coronary, and Artery_Tibial)
- prostate cancer: prostate, blood (GTEx tissue: Whole_Blood), and liver
- rheumatoid arthritis: blood (GTEx tissue: Whole_Blood), fibroblast (GTEx tissue: Cells_Cultured_fibroblasts), and spleen

Based on this classification, we generated the absolute counts and calculated the proportion of each method's significant gene-tissue pairs that fell within the relevant tissues for a given disease.

3.3 Results

3.3.1 Method overview

mFABIO is described in the Materials and Methods section, with technical details provided in the Appendix B and a method schematic shown in Figure 3.1.

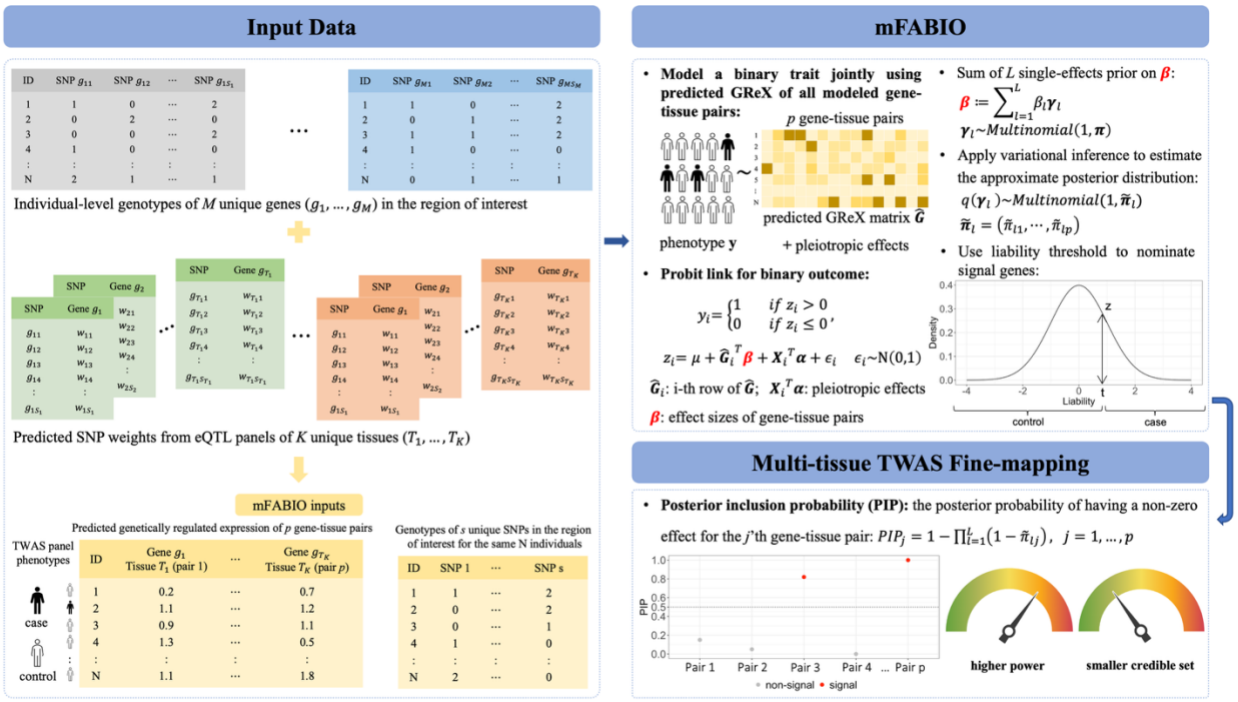

Figure 3. 1 Schematic overview of mFABIO for multi-tissue TWAS fine-mapping of binary outcomes.

Similar to other existing single-tissue and multi-tissue methods, mFABIO is a two-step TWAS fine-mapping approach that requires the predicted genetically regulated expression (GReX) in the study cohort as input. Shown under the "Input Data" section, mFABIO analyzes one genomic region at a time. For each genomic region containing s SNPs and M genes, mFABIO begins with individual-level genotypes of s SNPs within the region of interest. It also takes as input the SNP prediction weights for each of the M genes, which can be obtained from eQTL mapping cohorts across different tissues using standard methods such as SuSiE. These genotypes and SNP weights are then combined to generate predicted GReX for the p gene-tissue pairs in the study cohort. Additionally, mFABIO requires the genotypes of all s SNPs used in the GReX modeling to account for potential horizontal pleiotropic effects. Shown under the "mFABIO" section, mFABIO explicitly models the binary nature of the outcome trait through a latent variable z with a sum of L single-effects prior on the gene-tissue pair effect sizes β. It also simultaneously models all gene-tissue pairs on a region of interest to account for the GReX correlation, through the input of individual-level GReX matrix. We apply variational inference to estimate the approximate posterior distribution of the single-effect selection vector γ_l, and use the posterior inclusion probability (PIP) as the evidence for the gene-tissue pair associated with the binary outcome trait ("Multi-tissue TWAS Fine-mapping" section). * Icons used in this figure are from BioRender.com.

Briefly, mFABIO is a multi-tissue TWAS fine-mapping method designed to identify gene-tissue pairs whose GReX is associated with a binary trait of interest. Different from existing TWAS fine-mapping approaches, mFABIO explicitly models the binary nature of the outcome trait through a probit model, and jointly analyzes gene-tissue pairs across multiple tissues within a given region. This joint modeling accounts for GReX correlations among genes and tissues to enhance the precision of TWAS fine-mapping. We adapt the variational inference within an iterative Bayesian stepwise selection algorithm of SuSiE, making mFABIO scalable to large biobank scale datasets.

3.3.2 mFABIO controls false signals in null simulations

We first assessed the false positive rates of all methods across eight null simulation settings ($PVE_2 = 0\%$, and other parameters were varied one at a time on top of the baseline setting) using an estimated FDR threshold of 0.05. All methods demonstrated well-controlled false positives, identifying fewer than 0.50 false signals on average per replicate at both the gene-tissue pair (Figure S3.1a) and gene levels (Figure S3.1b). The multi-tissue methods, mFABIO and TGFM, were particularly stringent, showing no false positives under most settings. The only exceptions were an average of 0.20 false-positive gene-tissue pairs and 0.10 false-positive genes when the case:control ratio was 1:50 for both mFABIO and TGFM, and an average of 0.10 false-positive gene-tissue pairs and 0.10 false-positive genes when there were 3 causal tissues for TGFM. The single-tissue methods also exhibited low error rates. On average, their false positive identifications ranged from 0.10 to 0.23 at the gene-tissue pair level and from 0.08 to 0.18 at the gene level across all null scenarios.

3.3.3 mFABIO produces calibrated PIPs and improves power for fine-mapping causal signals in alternative simulations

We also evaluated how calibrated the PIPs/p-values are from different methods in terms of controlling the false discovery rate (FDR) at an estimated FDR threshold of 0.05 in our alternative simulations. Under the baseline setting at the gene-tissue pair level, the true FDR for multi-tissue methods was well-calibrated to this threshold, with mFABIO and TGFM both at 0.04. In contrast, the true FDR for all single-tissue methods (FABIO-gt, FABIO, FOCUS-gt, FOCUS, GIFT-gt, GIFT, cTWAS-gt, and cTWAS) was inflated, ranging from 0.13 to 0.16 (Figure S3.2a). Similar results were observed under different case:control ratios (Figure S3.2a) and numbers of causal tissue(s) (Figure S3.2b), as well as other simulation settings (Figures S3.2c-d). Across all settings, the true FDR that corresponds to the estimated FDR of 0.05 was on average 0.04 for both mFABIO and TGFM, while ranged from 0.12 to 0.15 for single-tissue methods. The superiority of the multi-tissue approach was also evident at the gene level, where the average true FDR under an estimated FDR threshold of 0.05 was 0.04 for mFABIO and TGFM, versus 0.11 to 0.14 for single-tissue methods (Figures S3.3a-d).

Next, we evaluated the power of different methods in detecting causal gene-tissue pairs and genes under alternative settings. To ensure a fair comparison that accounts for the previously observed FDR inflation, we first computed the power based on a true FDR threshold of 0.05. We

found that mFABIO is more powerful than other methods under all alternative simulation settings. For example, in the baseline setting at a true FDR of 0.05, mFABIO substantially outperformed the other approaches. For identifying causal gene-tissue pairs, mFABIO achieved a power of 34.5%, compared to 27.5% for TGFM and a range of 15.0%-19.0% for the single-tissue methods (Figure 3.2a). This trend held for identifying causal genes, where mFABIO's power was 36.3%, versus 28.9% for TGFM and 15.8%-19.5% for the single-tissue methods (Figure 3.2b).

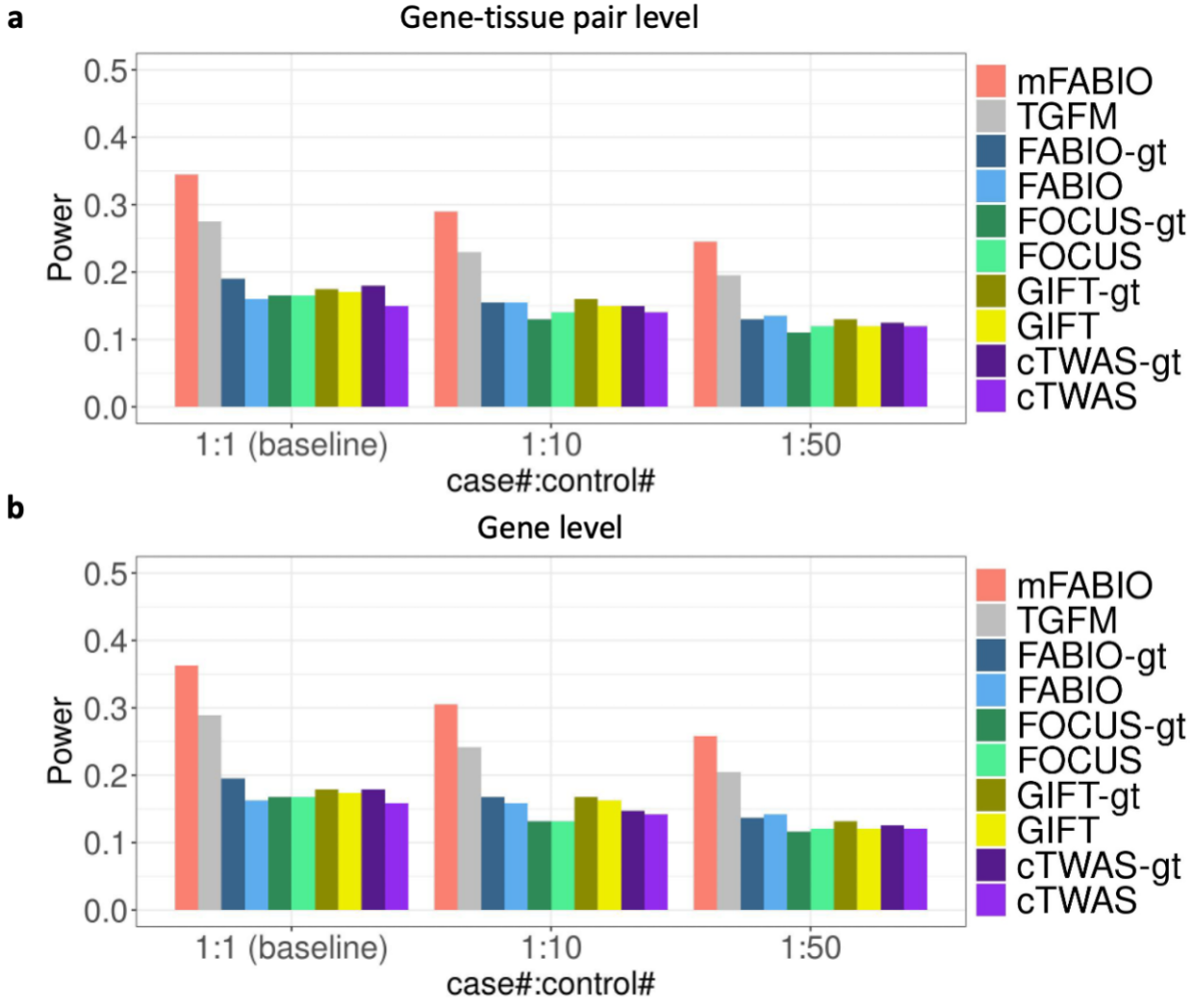

Figure 3. 2 Power of methods under the true FDR threshold of 0.05 across different case:control ratios.

(a) Results at gene-tissue pair level. **(b)** Results at gene level.

We carefully examined the influence of different simulation parameters on the power of different methods. In terms of case:control ratio, we found that the power decreased for all methods as the case:control ratio became more imbalanced. However, the relative performance ranking

remained consistent, with the two multi-tissue methods outperforming the single-tissue methods and mFABIO ranking as the most powerful in all conditions. For instance, at the gene-tissue pair level, as the ratio decreased from 1:1 to 1:10 and 1:50, mFABIO's power went from 34.5% to 29.0% and 24.5%. Following the same trend, TGFM's power decreased from 27.5% to 23.0% and 19.5%, while the power for single-tissue methods dropped from a range of 15.0%-19.0% down to 11.0%-13.0% (Figure 3.2a). A similar trend was observed at the gene level (Figure 3.2b). Next, we investigated how the number of causal tissues affects the power, and we found that the power of all methods decreased as the number of causal tissues increased from one to three. This was an expected result of our simulation design, as the total effect size across causal gene-tissue pairs was fixed, and scenarios with more causal tissues had a smaller effect size for each pair. Despite this overall power reduction, the performance ranking of different methods remained consistent. At the gene-tissue pair level, mFABIO was the most powerful, with its power decreasing from 41.0% (with one causal tissue) to 29.3% (with three). For comparison, TGFM's power dropped from 35.0% to 24.3%, while the power range for single-tissue methods fell from 23.0%-28.0% to 14.3%-17.7% under the same conditions (Figure 3.3a). A similar trend was also observed at the gene level (Figure 3.3b).

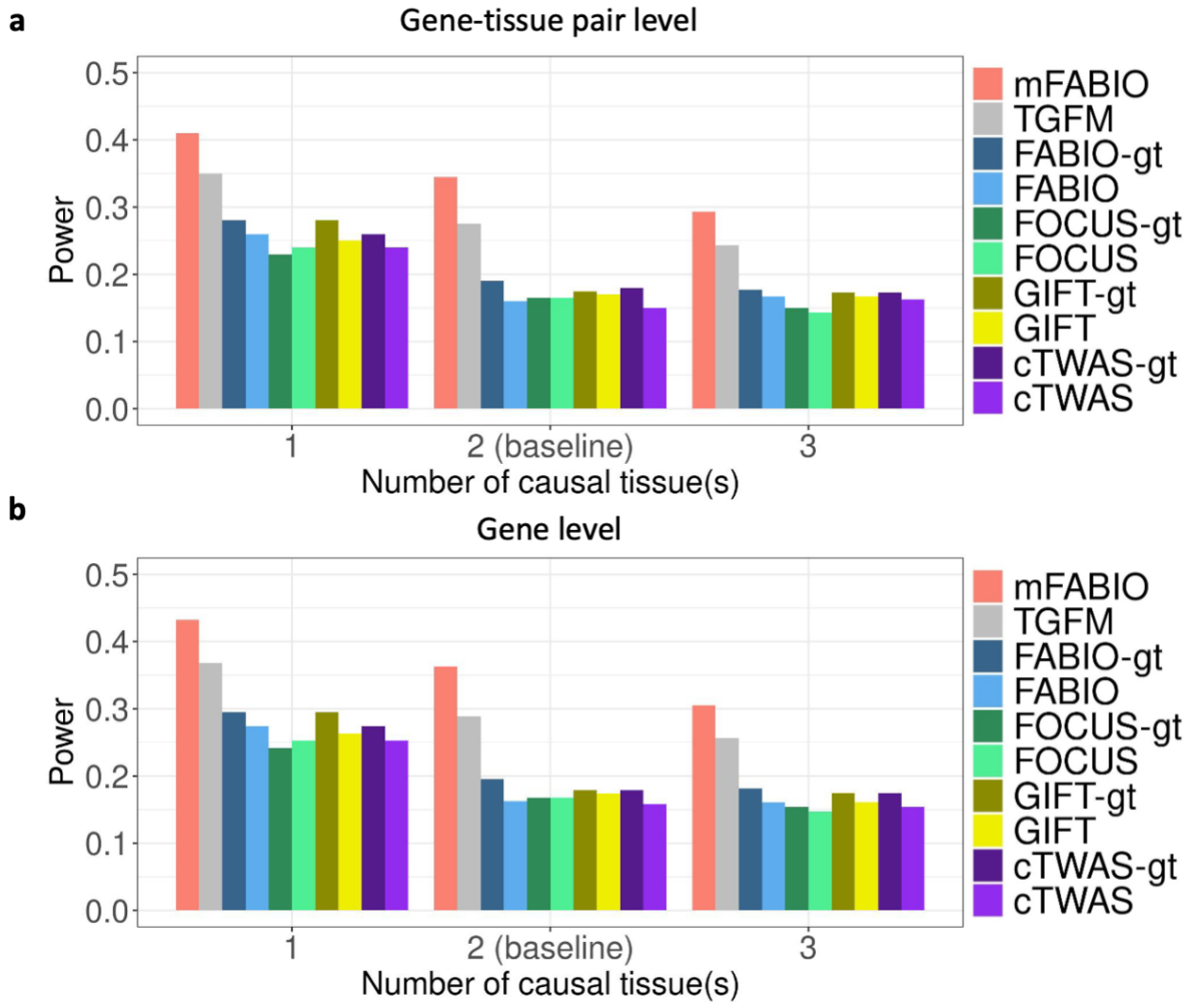

Figure 3. 3 Power of methods under the true FDR threshold of 0.05 across different numbers of causal tissue(s). **(a)** Results at gene-tissue pair level. **(b)** Results at gene level.

In terms of SNP effects on gene expression (PVE_1), we observed that the performance of all methods improved with increasing PVE_1. Their performance ranking remained stable, with mFABIO consistently outperforming TGFM and both multi-tissue methods surpassing the single-tissue methods, at both gene-tissue pair level (Figure S3.4a) and gene level (Figure S3.4b). Specifically, for the most challenging setting ($PVE_1 = 5\%$), mFABIO still achieved a power of 19.0% at the gene-tissue pair level and 19.5% at the gene level. This was considerably higher than TGFM (15.0% and 15.4%, respectively) and the single-tissue methods (which ranged from 7.5%-10.5% and 7.7%-10.8%, respectively). Additionally, we had a simulation setting fixing PVE_1 at 10% and randomizing the number of causal cis-SNPs, which yielded power trends similar to the

baseline setting at both gene-tissue pair level (Figure S3.4a) and gene level (Figure S3.4b). Finally, we investigated how power was affected by the gene's effects on the outcome trait (PVE_2). We found that the power of all methods grew as PVE_2 increased, with mFABIO consistently remaining the most powerful method. At gene-tissue pair level, as PVE_2 was raised from 0.2% to 0.4% and 0.6%, mFABIO's power rose from 25.5% to 34.5% and 46.5%, respectively. Under the same conditions, TGFM's power increased from 20.5% to 27.5% and 37.5%. The power range for single-tissue methods also improved, increasing from 11.0%-14.0% to 15.0%-19.0% and finally to 21.0%-26.0% (Figure S3.5a). A similar trend was also observed at the gene level (Figure S3.5b).

The power comparisons described above were based on a true FDR threshold of 0.05, a metric that can only be calculated in simulations. To better mimic a real-data scenario, we also compared power using an estimated FDR threshold of 0.05. We acknowledge this comparison is inherently biased: as shown above, this threshold is well-calibrated for multi-tissue methods but leads to significant FDR inflation for single-tissue methods (Figures S3.2 and S3.3), potentially boosting their apparent power. Despite this advantage, the single-tissue methods did not outperform mFABIO. At the gene-tissue pair level, mFABIO retained the highest power, achieving 34.5% in the baseline setting and averaging 30.6% across all 10 alternative simulations. In comparison, TGFM's power was 28.5% (baseline) and 25.0% (average), while the single-tissue methods' power had a range of 22.0%-26.0% (baseline) and 21.0%-23.1% (average). Though the power of single-tissue methods appeared more comparable under this threshold, it came at the cost of excessive false discoveries. The conclusion holds regardless of the case:control ratio (Figure S3.6a), the number of causal tissue(s) (Figure S3.6b), the SNP effect on gene expression (Figure S3.6c), the number of causal cis-SNPs (Figure S3.6c), and the gene's effects on the outcome trait (Figure S3.6d), as well as at the gene level (Figures S3.7a-d). For the single-tissue methods, we also observed the advantage of the joint analysis of all gene-tissue pairs by naturally extending their designs. This strategy was marginally more powerful, than applying those methods on each tissue and combining the results across tissues. The joint analysis yielded higher average power at both the gene-tissue pair level (21.2%-23.1% vs. 21.0%-21.5%) and the gene level (21.4%-22.8% vs. 20.8%-21.3%). Therefore, we exclusively used this more powerful joint analysis strategy for all single-tissue methods in our subsequent real-data applications.

3.3.4 Application to analyzing binary traits in UKBB

We performed multi-tissue TWAS fine-mapping analysis on six binary disease traits through integrating the GTEx cis-predicted expression models of 38 meta-tissues and the GWAS of UK Biobank (details in 3.2 Materials and methods). Specifically, we focused on 13,700 genes and 119,270 unique gene-tissue pairs and examined their associations with each of the six binary traits. The six binary traits include asthma (AS), breast cancer (BRCA), gout (GO), hypertension (HT), prostate cancer (PRCA), and rheumatoid arthritis (RA).

We first evaluated the results at gene-level among these traits. We identified 1,427 (1,355 unique) GWAS risk genes that contain at least one genome-wide significant SNP ($p < 5\times10^{-8}$) across the six disease traits. We performed the TWAS marginal analysis using FUSION for each pair of meta-tissue and disease trait, and considered those genes that are significant in at least one pair of meta-tissue and trait as TWAS risk genes. We identified a total of 479 (470 unique) TWAS risk genes across the six traits. Overlapping the results from the two analyses leads to a total of 1,766 (1,661 unique) risk genes that contain at least one genome-wide significant SNP or shows marginal TWAS significance (Table 3.1). We consider these risk genes more likely to be the potentially "true" causal genes compared to the rest.

Table 3. 1 Gene-level results of multi-tissue TWAS fine-mapping in UK Biobank.

Trait	mFABIO	TGFM	FABIO	FOCUS	GIFT	cTWAS
AS	58 (31)	37 (16)	108 (35)	155 (30)	98 (28)	92 (18)
BRCA	43 (19)	29 (10)	65 (8)	75 (10)	54 (9)	61 (7)
GO	38 (29)	30 (21)	72 (26)	85 (22)	87 (25)	75 (20)
HT	78 (50)	55 (30)	209 (52)	230 (41)	203 (49)	212 (40)
PRCA	21 (17)	15 (11)	55 (10)	85 (10)	57 (10)	62 (11)
RA	15 (8)	12 (5)	22 (3)	35 (4)	20 (3)	25 (4)

This table summarizes the number of discoveries for each of the six disease traits (rows) in the multi-tissue TWAS fine-mapping analysis of UK Biobank. A GWAS risk gene is defined as a gene that harbors at least one genome-wide significant SNP (p-value $< 5 \times 10^{-8}$) within 500 Kb of its transcription start site (TSS). A TWAS risk gene is defined as a gene with a marginal significant TWAS p-value after the Bonferroni correction in at least one of the meta-tissue-trait pairs. A risk gene with GWAS/TWAS evidence is defined as a gene that harbors at least one genome-wide significant SNP or shows marginal TWAS significance. The last six columns list the number of genes discovered by each of the six methods. The number in the bracket is the number of identified genes that are risk genes with GWAS/TWAS evidence. We used an estimated FDR threshold of 0.05 to declare significance for all methods in the fine-mapping analysis.

Same as what we did in the simulations, we estimated FDR using the test statistics generated by different methods and declared significant gene associations based on an estimated FDR threshold of 0.05 for all methods. mFABIO identified an average of 42 causal genes per disease, with 60.9% of those genes supported by GWAS or TWAS evidence (Table 3.1). As a comparison, the other multi-tissue method TGFM, identified an average of 29 causal genes per disease, with 52.2% of those genes supported by GWAS or TWAS evidence. Among the four single-tissue methods - FABIO, FOCUS, GIFT, and cTWAS - the average numbers of identified causal genes were 88, 110, 86, and 87, respectively, with corresponding proportions of GWAS/TWAS-supported genes at 25.2%, 17.6%, 23.9%, and 19.0% (Table 3.1). Although mFABIO and TGFM identified fewer causal gene candidates at an FDR of 0.05 compared to single-tissue methods, the proportions of GWAS/TWAS-supported genes were substantially higher. When considering the total number of GWAS/TWAS-supported genes identified across the six disease traits, mFABIO identified at least 14.9% more than any of the other methods (Table 3.1). Overall, mFABIO not only outperformed other methods in identifying a greater number of GWAS/TWAS-supported gene candidates, but also achieved the smallest candidate gene set with the highest proportion of supported genes compared to those single-tissue methods. These findings highlighted the superior performance of mFABIO at gene-level.

We then evaluated the identified gene-tissue pairs across different methods. In total, we identified 11,613 (11,076 unique) GWAS risk gene-tissue pairs, where the gene of each pair contains at least one genome-wide significant SNP ($p < 5\times10^{-8}$) associated with the corresponding trait. Additionally, we identified 2,152 (2,116 unique) TWAS risk gene-tissue pairs that showed significance in our marginal TWAS analysis for each meta-tissue and disease trait pair. Combining these two sets resulted in 13,148 (12,466 unique) risk gene-tissue pairs that either contain a genome-wide significant SNP or show marginal TWAS significance (Table 3.2). We consider these risk gene-tissue pairs to be more likely representative of the “true” associations with the corresponding disease traits.

Table 3. 2 Gene-tissue pair-level results of multi-tissue TWAS fine-mapping in UK Biobank.

Trait	mFABIO	TGFM	FABIO	FOCUS	GIFT	cTWAS
AS	41 (29)	33 (19)	86 (25)	111 (30)	88 (26)	78 (25)
BRCA	40 (33)	28 (20)	52 (15)	58 (15)	45 (18)	50 (21)
GO	95 (74)	45 (32)	181 (77)	215 (68)	155 (62)	175 (70)
HT	152 (120)	84 (58)	423 (126)	470 (117)	385 (116)	391 (106)
PRCA	37 (33)	20 (15)	95 (14)	101 (12)	87 (12)	89 (14)
RA	26 (13)	18 (8)	36 (6)	51 (8)	33 (6)	42 (6)

This table summarizes the number of identified gene-tissue pairs for each of the six disease traits (rows) in the multi-tissue TWAS fine-mapping analysis of UK Biobank. A GWAS risk gene-tissue pair is defined as its gene that harbors at least one genome-wide significant SNP (p-value $< 5 \times 10^{-8}$) within 500 Kb of its transcription start site (TSS). A TWAS risk gene-tissue pair is defined as a pair with a significant marginal TWAS p-value after the Bonferroni correction in the corresponding meta-tissue-trait analysis. A risk gene-tissue pair with GWAS/TWAS evidence is defined as its gene that harbors at least one genome-wide significant SNP or this pair shows marginal TWAS significance. The last six columns list the number of gene-tissue pairs discovered by each of the six methods. The number in the bracket is the number of identified gene-tissue pairs that are risk pairs with GWAS/TWAS evidence. We used an estimated FDR threshold of 0.05 to declare significance for all methods in the fine-mapping analysis.

At the gene-tissue pair-level, mFABIO identified an average of 65 causal gene-tissue pairs per disease, with 77.2% of those pairs supported by GWAS or TWAS evidence (Table 3.2). TGFM identified an average of 38 causal pairs per disease, with 66.7% of those pairs supported by GWAS or TWAS evidence. Among the four single-tissue methods - FABIO, FOCUS, GIFT, and cTWAS - the average numbers of identified causal pairs were 145, 167, 132, and 137, respectively, with corresponding proportions of GWAS/TWAS-supported pairs at 30.1%, 24.9%, 30.3%, and 29.3% (Table 3.2). Similar to our observations at gene-level, mFABIO and TGFM identified fewer causal gene-tissue pairs at an FDR of 0.05 compared to single-tissue methods, however, the proportions of GWAS/TWAS-supported pairs were substantially higher. Again, if we consider the total number of GWAS/TWAS-supported pairs identified across the six disease traits, mFABIO identified at least 14.8% more than any of the other methods (Table 3.2). Therefore, mFABIO had superior performance at the gene-tissue pair-level as well.

We further evaluated whether significant gene-tissue pairs (estimated FDR < 0.05) were enriched in the likely disease-relevant tissues based on PubMed literature (details in 3.2 Materials and methods). We first compared the absolute number of significant pairs found within these

tissues. By this metric, mFABIO identified the most relevant pairs for all diseases except hypertension (Figure 3.4a). The exception for hypertension is likely due to that, single-tissue methods are confounded by a much larger total number of discoveries, which may include a high rate of false positives as shown in Table 2. To account for this, we calculated the proportion of a method's significant pairs located in relevant tissues as a more robust metric. This analysis showed that mFABIO achieved the highest proportion of relevant discoveries for every disease (Figure 3.4b), with an overall average of 29.7%. This average was higher than that of TGFM (27.2%) and substantially higher than the range for single-tissue methods (12.5%–15.5%). These results demonstrate mFABIO's superior ability to prioritize biologically relevant gene-tissue associations in real-data applications.

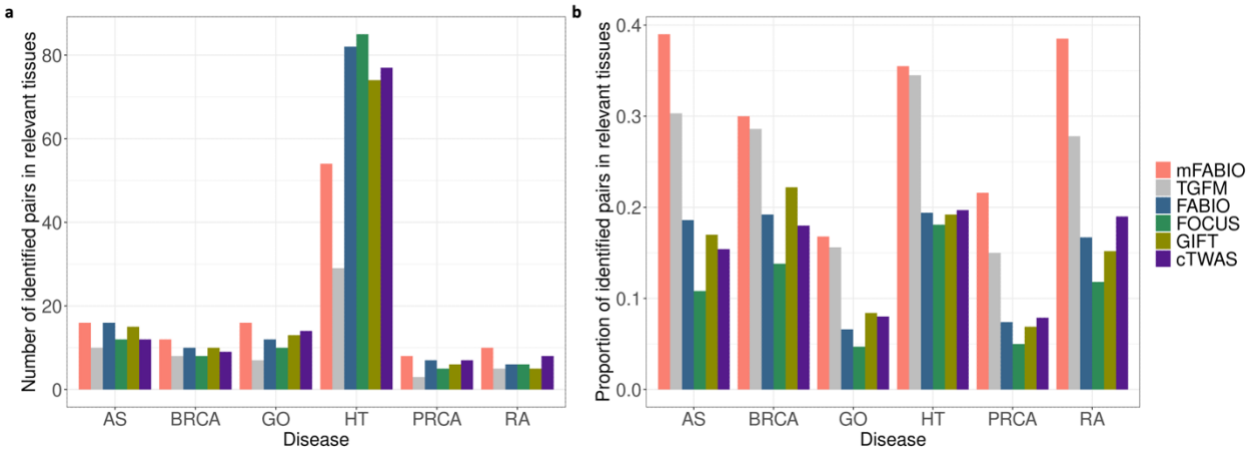

Figure 3. 4 Comparison of significant gene-tissue pairs located in relevant tissues.

(a) Absolute number of significant gene-tissue pairs located in relevant tissues determined by PubMed literature for each method across diseases. **(b)** Proportion of a method's significant gene-tissue pairs located in relevant tissues determined by PubMed literature across diseases.

We list a few gene-tissue examples that are uniquely identified by mFABIO. The first example is *D2HGDH* (D-2-hydroxyglutarate dehydrogenase) in lung tissue associated with asthma (gene-tissue PIP = 1, marginal TWAS p-value = 2.94×10^{-6}). *D2HGDH* has been associated with asthma in a previous GWAS (Zhu et al., 2018) through its cis-SNP rs34290285, which is consistent with our GWAS results on asthma that highlighted the same marker within the gene locus (Figure 3.5a). And the previous lung eQTL analysis of the top SNPs revealed that the G allele of rs34290285 is associated with reduced expression of *D2HGDH* mRNA in the lung (Zhu et al.,

2018). This downregulation leads to the accumulation of the metabolite D-2-hydroxyglutarate (D-2HG). Critically, in murine models, accumulating D-2HG was shown to potently suppress key hallmarks of allergic asthma, including airway hyperreactivity, Th2 cell responses, and IgE-mediated inflammation (Tharakan et al., 2023). This provides a direct mechanistic pathway connecting lower *D2HGDH* expression in the lung to a protective effect against asthma.

The second example is *CYBRD1* (cytochrome b reductase 1) in breast mammary tissue associated with breast cancer (gene-tissue PIP = 1). Notably, while the top cis-SNP in its coding region (rs13020413) did not reach genome-wide significance in previous studies and our GWAS (Figure 3.5b), *CYBRD1* was flagged as a significant gene in our TWAS pairing the breast mammary tissue with breast cancer (marginal TWAS p-value = 8.90×10^{-6}). This finding is biologically relevant, as *CYBRD1* (also known as *DCYTB*) plays a crucial role in iron metabolism (McKie et al., 2001), a process linked to tumor-infiltrating lymphocytes (TILs) and regulated by hypoxia-inducible factor (HIF) in malignant breast cells (Boult et al., 2008). Furthermore, *CYBRD1* serves as a prognostic marker for breast cancer, as its increased expression is associated with a favorable prognosis and is implicated in regulating cancer cell proliferation and apoptosis (Lemler et al., 2017). The successful identification of *CYBRD1* demonstrates that mFABIO can nominate credible, tissue-specific candidate genes that lack strong statistical support from traditional GWAS.

The third example is *CCR6* (C-C chemokine receptor type 6) in spleen associated with rheumatoid arthritis (gene-tissue PIP = 0.98). While the top cis-SNP for *CCR6* in our GWAS (rs1571878) was not genome-wide significant (Figure 3.5c), it has been previously associated with rheumatoid arthritis (RA) in another GWAS study (Ge et al., 2021). Also, CCR6 is a significant gene candidate in our TWAS pairing spleen with RA (marginal TWAS p-value = 3.92×10^{-6}). The involvement of the spleen in RA is well-established, as spleen is a major hub for immune activity, and its enlargement is a characteristic of Felty's syndrome, a known complication of severe RA (Owlia et al., 2014). This connection is further strengthened by mechanistic evidence from animal models. One study found that in mice with collagen-induced arthritis, regulatory T-cells (Tregs) in the spleen significantly increased their expression of *CCR6* compared to the healthy control (Li et al., 2015). This suggests that during arthritic inflammation, splenic Tregs upregulate *CCR6*, a key receptor that mediates the migration of immune cells to inflammatory sites.

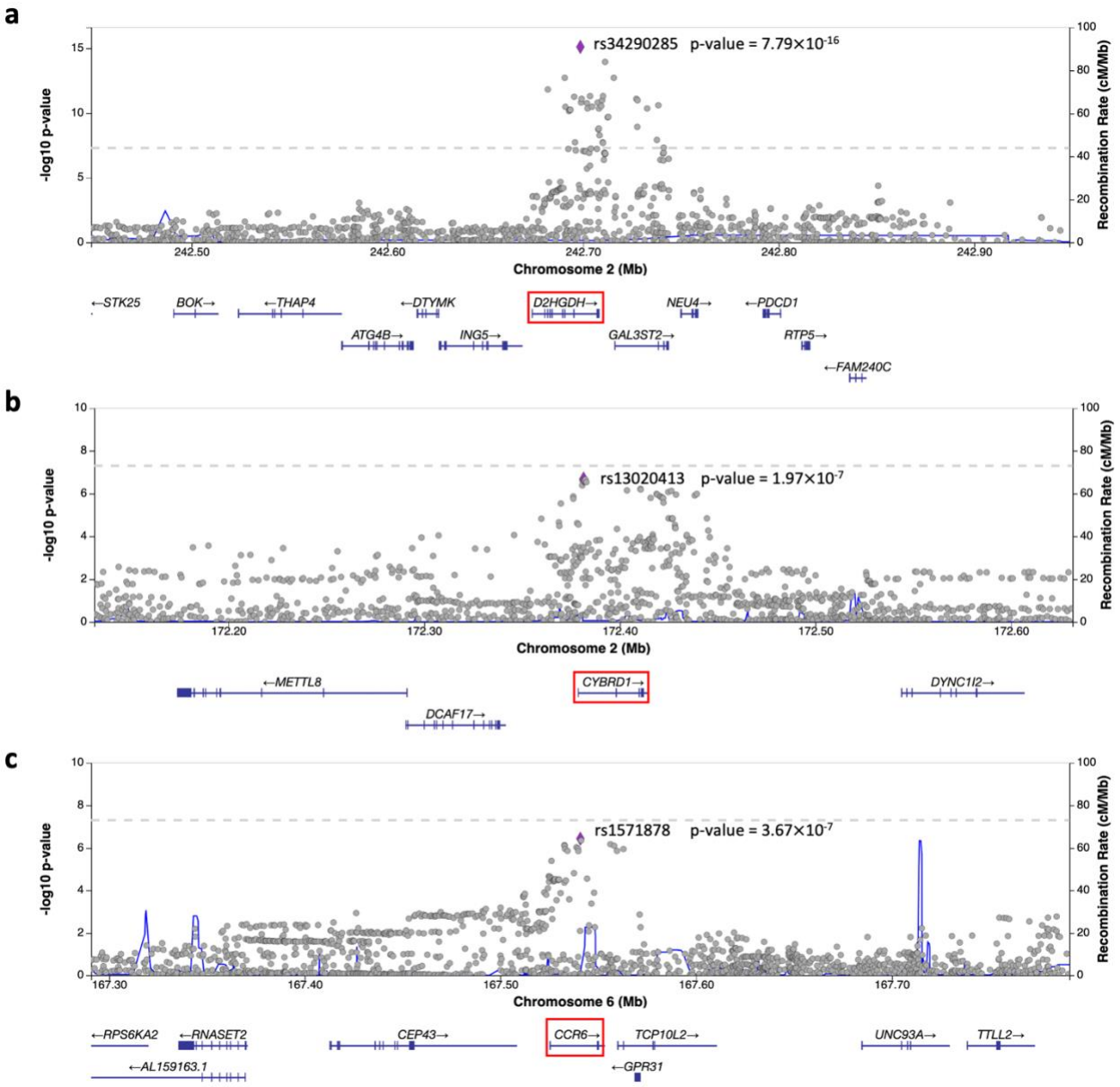

Figure 3. 5 LocusZoom plots of selected gene candidates uniquely identified by mFABIO.

(a) Plot of the corresponding locus with the GWAS p-value of the top cis-SNP rs34290285 in *D2HGDH* for asthma. **(b)** Plot of the corresponding locus with the GWAS p-value of the top cis-SNP rs13020413 in *CYBRD1* for breast cancer. **(c)** Plot of the corresponding locus with the GWAS p-value of the top cis-SNP rs1571878 in *CCR6* for rheumatoid arthritis. All these LocusZoom plots were generated using (Boughton et al., 2021).

3.4 Discussion

In this study, we introduce mFABIO, a novel multi-tissue TWAS fine-mapping method specifically designed for binary outcomes. Unlike existing approaches, mFABIO integrates a probit model to appropriately reflect the binary nature of disease traits and jointly models gene

expression across multiple tissues to account for GReX correlations among gene-tissue pairs. This joint modeling enables mFABIO to yield well-calibrated test statistics and achieve greater statistical power than current single-tissue and multi-tissue methods developed for continuous traits. Through extensive simulations and real-world applications to six disease phenotypes in the UK Biobank, we demonstrate the practical advantages of mFABIO.

mFABIO models the relationship between genetically regulated expression (GReX) and binary outcomes using a probit link function. This introduces a latent liability score, where an individual is classified as having a disease if their score surpasses a certain threshold. Although a common alternative is the logistic link function that models disease probability directly, both approaches often yield similar results in practice (Czado & Munk, 2000). We confirmed this robustness in our own simulations. Even when we generated binary outcomes using a logistic model, which is a direct misspecification of mFABIO's probit assumption, the method still retained reasonable power. This result shows that mFABIO can effectively capture GReX-outcome relationships even when its underlying modeling assumptions are not perfectly met.

We acknowledge that mFABIO is not without limitations. Firstly, mFABIO's probit link function requires individual-level GWAS data to infer its latent liability scores. This precludes the direct use of more widely available GWAS summary statistics. While requiring individual-level data may restrict the application of mFABIO, it helps mitigate biases commonly observed in methods that make use of only GWAS summary statistics, especially when there is a LD matrix mismatch between the reference panel and the analyzed GWAS data (Benner et al., 2017; Yang et al., 2012). Secondly, like other methods in this study, mFABIO uses a two-stage TWAS fine-mapping strategy. It relies on a reference eQTL mapping study to construct the gene expression prediction model in the first stage, and applies the SNP weights inferred in the eQTL mapping study to construct GReX in the GWAS data for TWAS fine-mapping in the second stage. This convenient strategy ignores uncertainty from the initial expression modeling stage, which can lead to a loss of statistical power (Liu et al., 2021; Yuan et al., 2020). A key future direction is to develop a joint-likelihood framework for mFABIO that simultaneously models both stages, thus accounting for this uncertainty. Finally, the performance of the existing methods including mFABIO depends on the ancestry of the reference eQTL panel. In our application, the European-ancestry GTEx panel was a good match for the European-ancestry UKBB cohort we analyzed. However, applying such TWAS fine-mapping approaches to multi-ancestry GWAS data would

fail to capture diverse genetic architectures. Developing TWAS fine-mapping methods that can leverage or account for ancestry diversity is a critical next step for the field. Even with these limitations, mFABIO stands as a robust and powerful method for fine-mapping causal tissues and genes for binary outcomes.

3.5 Supplementary figures

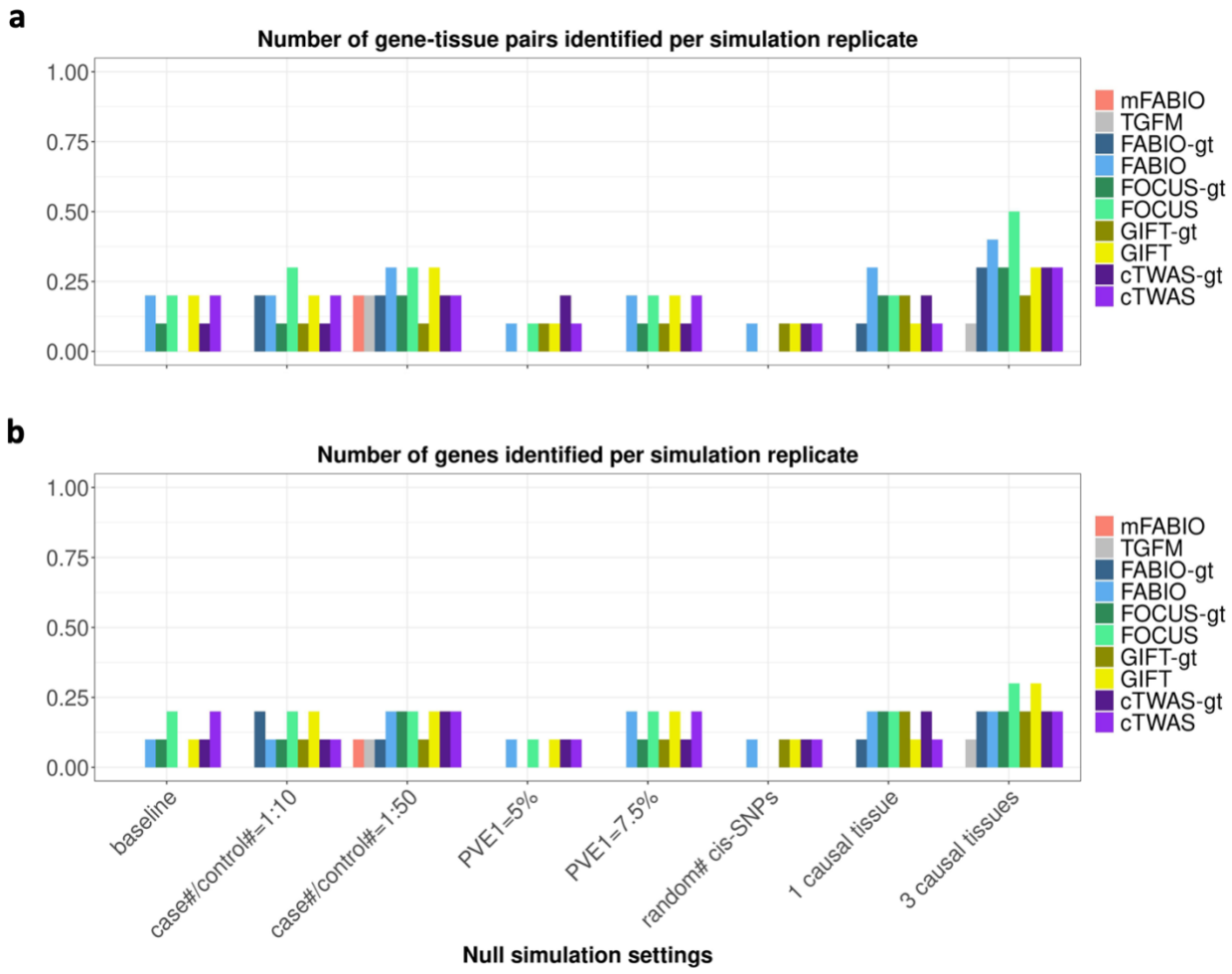

Figure S3. 1 Number of false signals in null simulations.

(a) The average number of false positive gene-tissue pairs per simulation replicate across null settings using an estimated FDR threshold of 0.05. (b) The average number of false positive genes per simulation replicate across null settings using an estimated FDR threshold of 0.05.

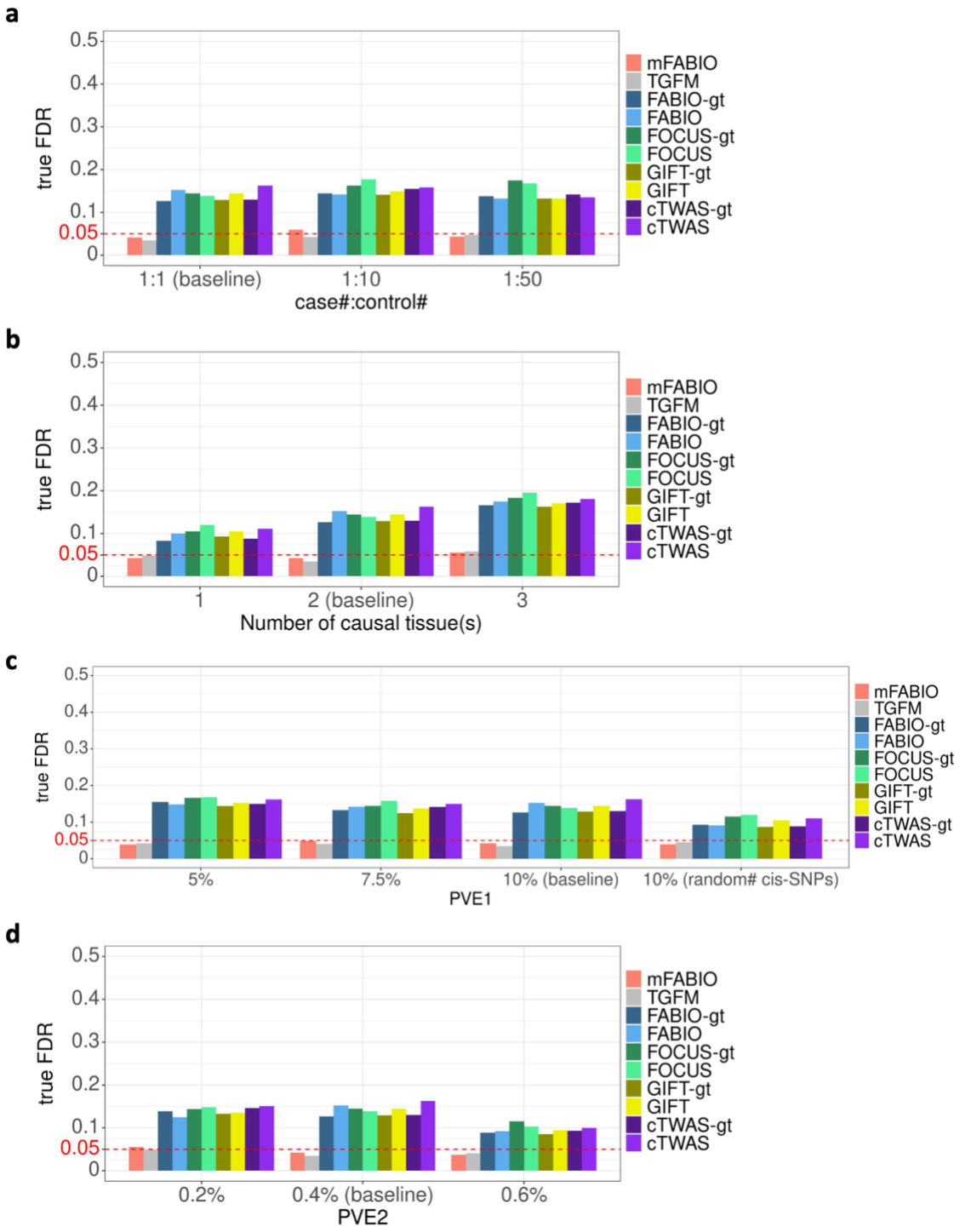

Figure S3. 2 The true false discovery rate (FDR) in alternative simulations at the gene-tissue pair level.

We calculated the true FDR under an estimated FDR threshold of 0.05 to evaluate the calibration of the methods in different simulation settings: (a) different case:control ratios; (b) different numbers of causal tissue(s); (c) different proportions of gene expression variance explained by genetic effects (PVE1) and different numbers of causal cis-SNPs; (d) different proportions of the phenotype's variance explained by causal gene-tissue pairs (PVE2).

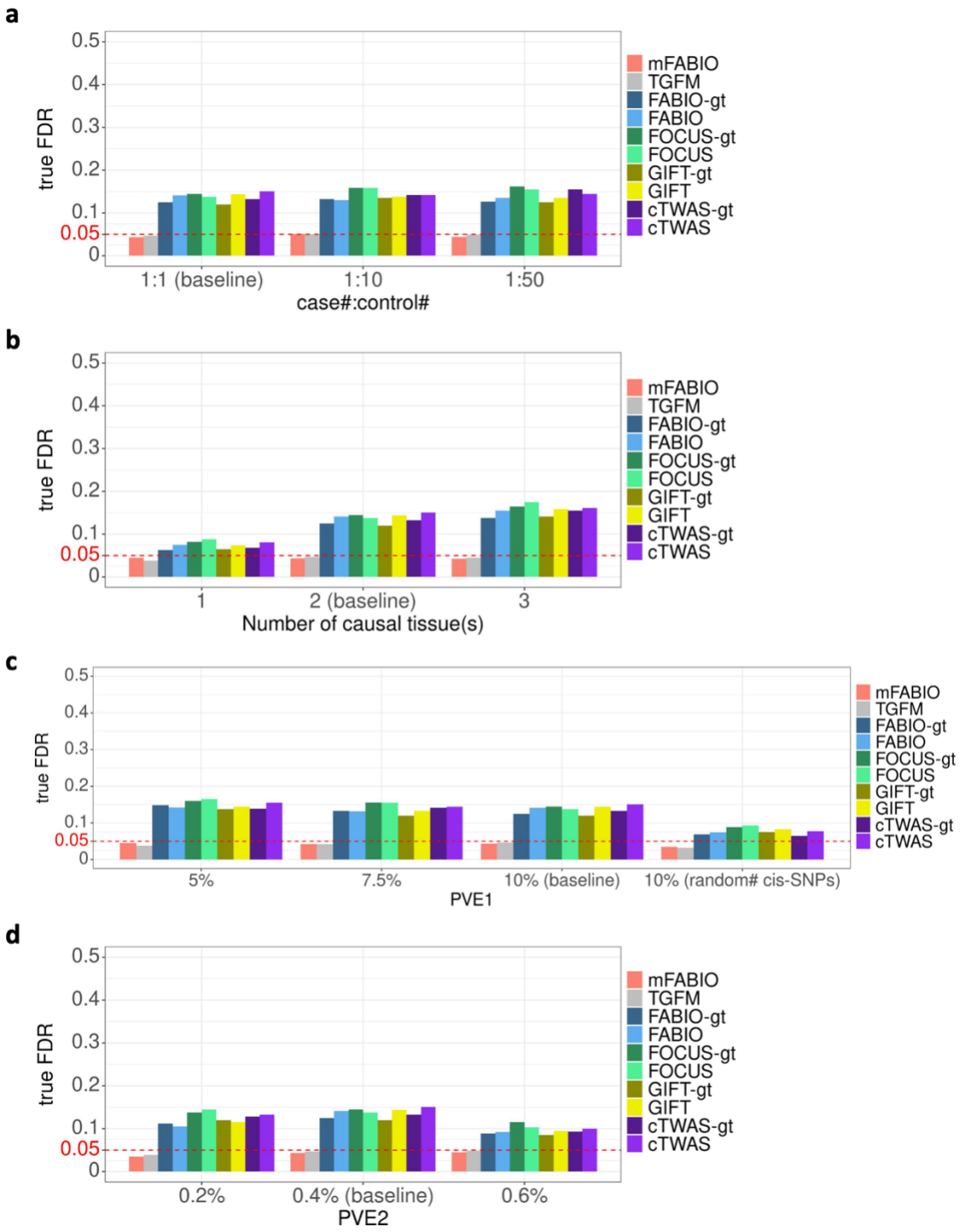

Figure S3. 3 The true false discovery rate (FDR) in alternative simulations at the gene level.

We calculated the true FDR under an estimated FDR threshold of 0.05 to evaluate the calibration of the methods in different simulation settings: (a) different case:control ratios; (b) different numbers of causal tissue(s); (c) different proportions of gene expression variance explained by genetic effects (PVE1) and different numbers of causal cis-SNPs; (d) different proportions of the phenotype's variance explained by causal gene-tissue pairs (PVE2).

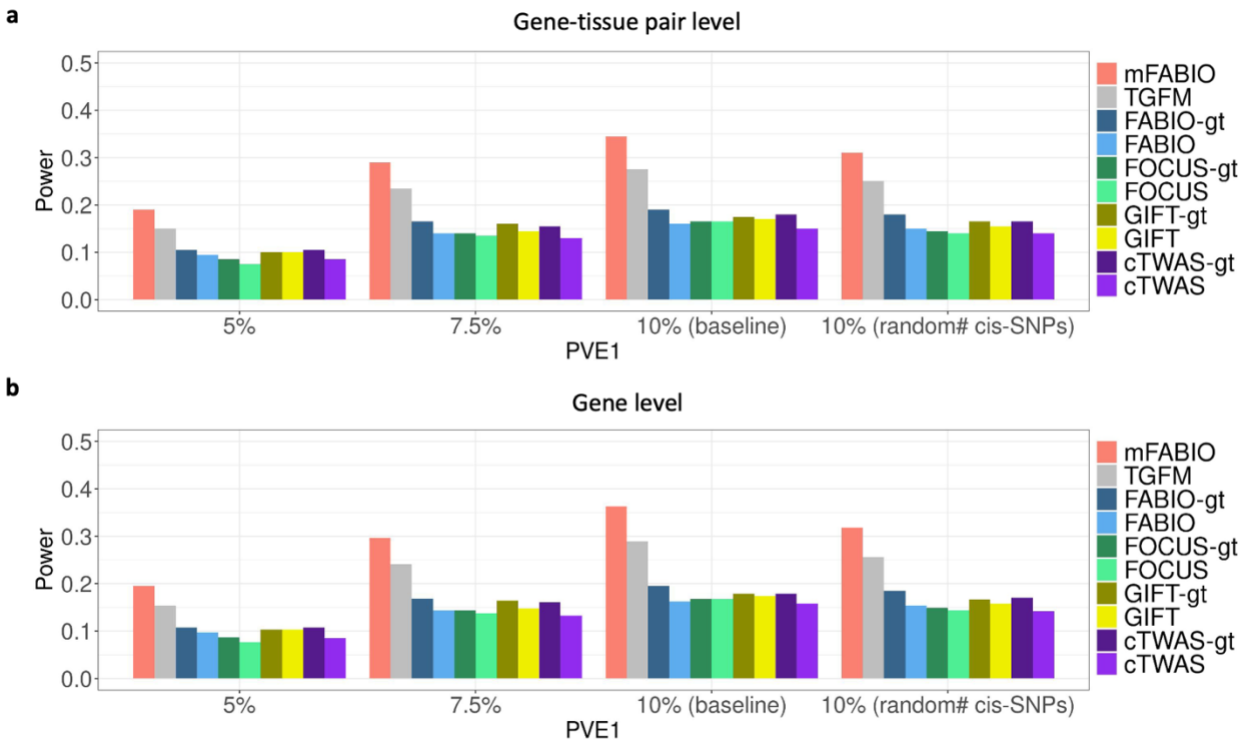

Figure S3. 4 Power of methods under the true FDR threshold of 0.05 across different PVE1 and numbers of causal cis-SNPs.

(a) Results at gene-tissue pair level. (b) Results at gene level.

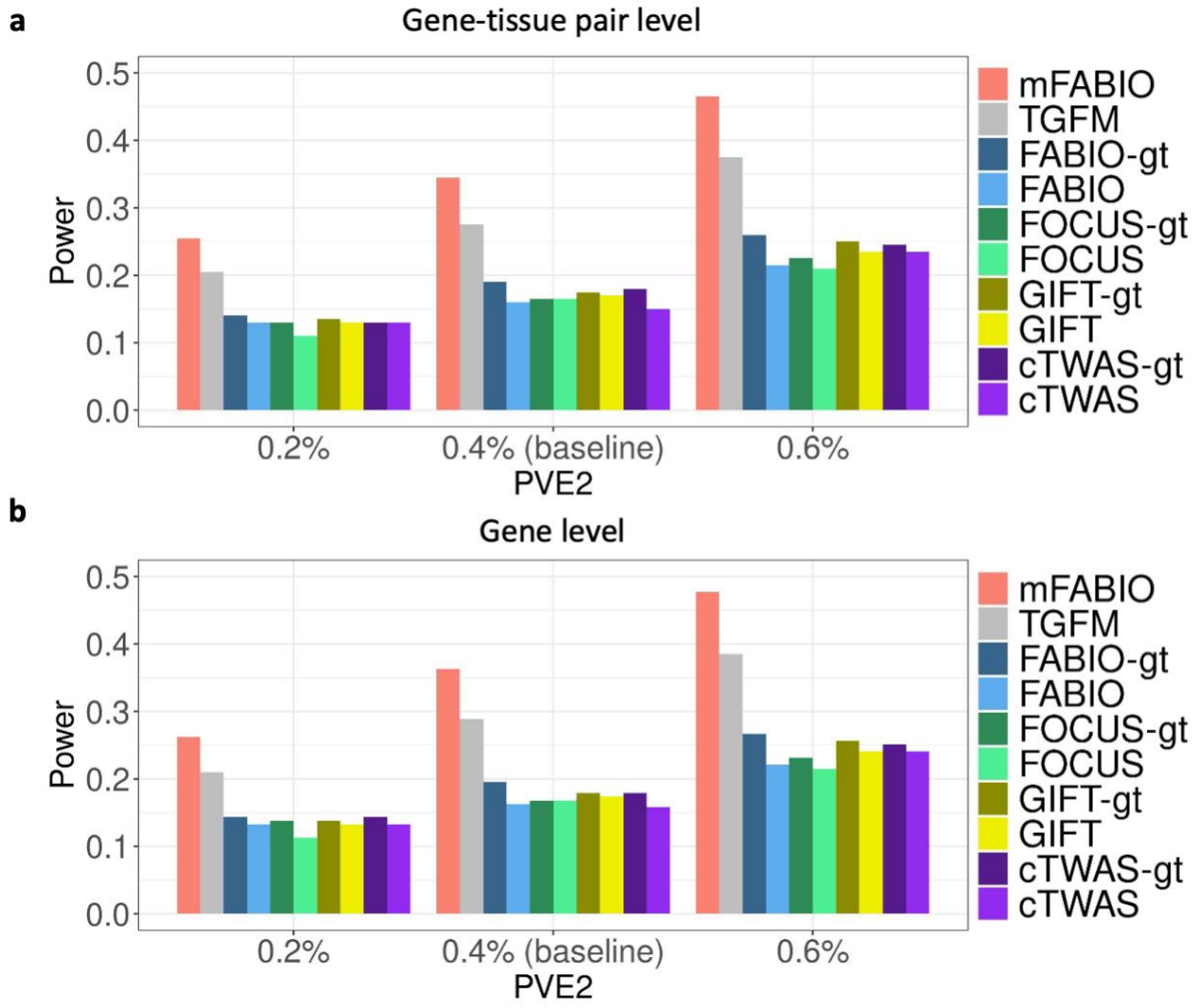

Figure S3. 5 Power of methods under the true FDR threshold of 0.05 across different PVE2.

(a) Results at gene-tissue pair level. (b) Results at gene level.

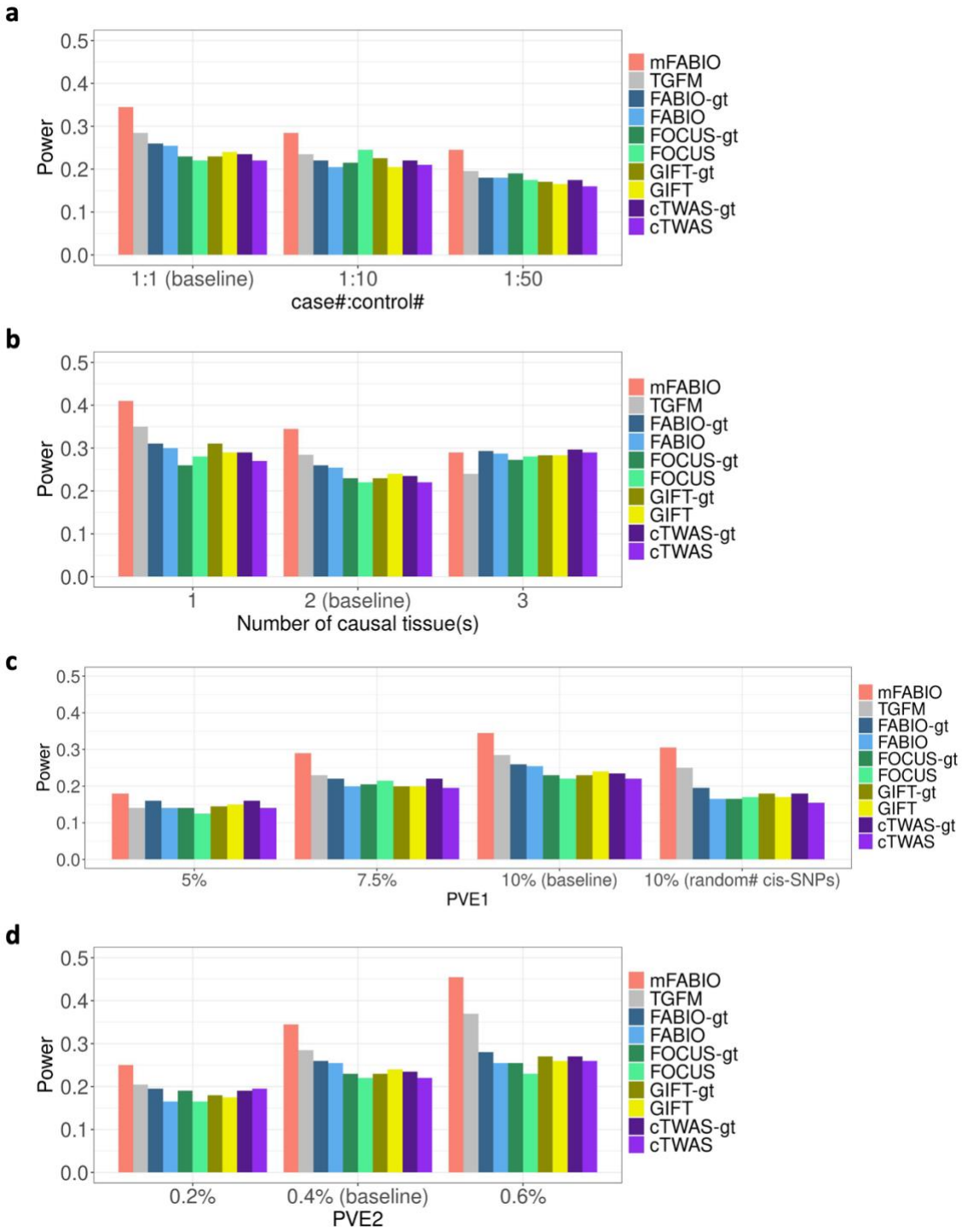

Figure S3. 6 Statistical power in alternative simulations at the gene-tissue pair level.

We calculated the power under an estimated FDR threshold of 0.05 to evaluate the performance of the methods in different simulation settings: (a) different case:control ratios; (b) different numbers of causal tissue(s); (c) different proportions of gene expression variance explained by genetic effects (PVE1) and different numbers of causal cis-SNPs; (d) different proportions of the phenotype's variance explained by causal gene-tissue pairs (PVE2).

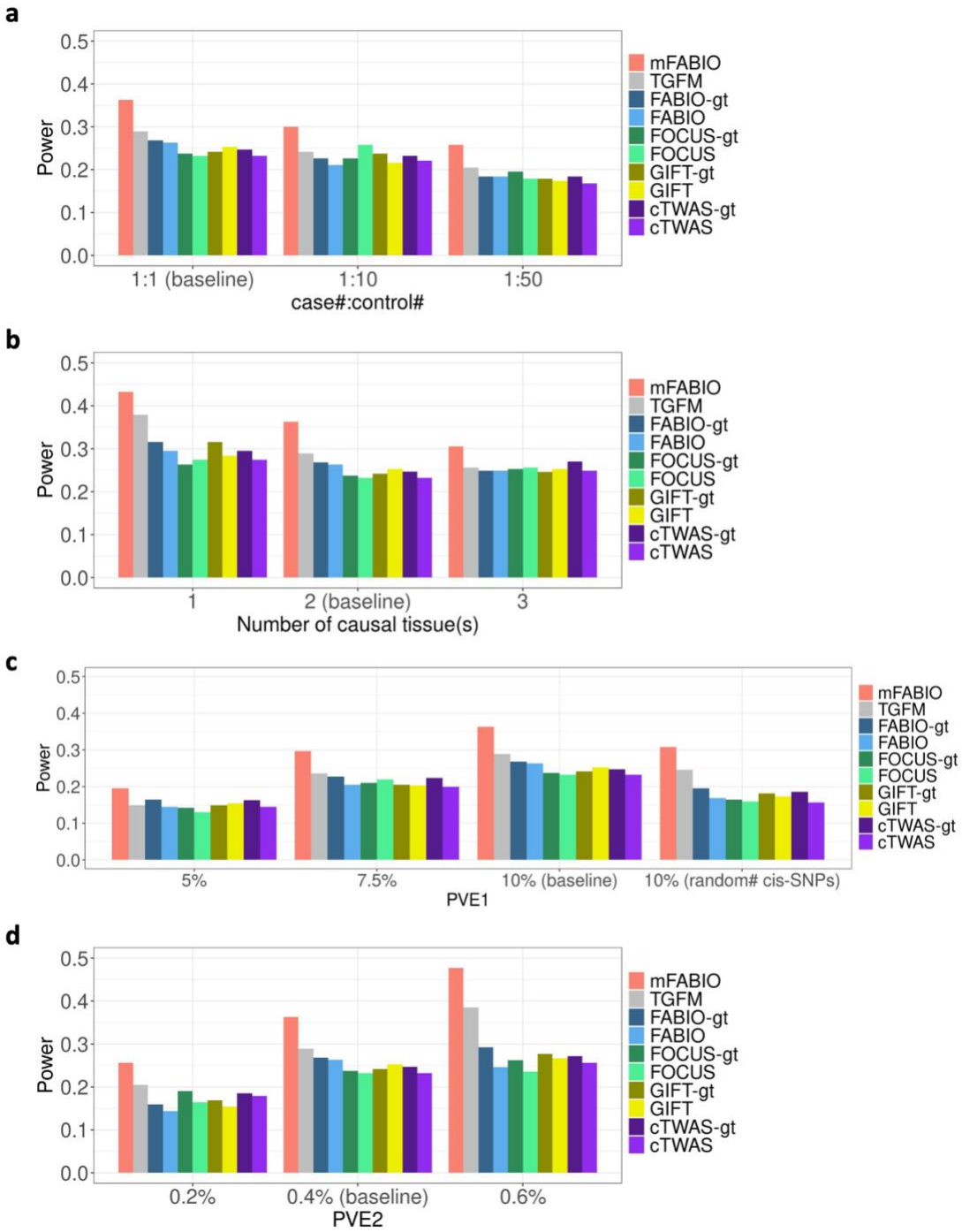

Figure S3. 7 Statistical power in alternative simulations at the gene level.

We calculated the power under an estimated FDR threshold of 0.05 to evaluate the performance of the methods in different simulation settings: (a) different case:control ratios; (b) different numbers of causal tissue(s); (c) different proportions of gene expression variance explained by genetic effects (PVE1) and different numbers of causal cis-SNPs; (d) different proportions of the phenotype's variance explained by causal gene-tissue pairs (PVE2).

Chapter 4 Multi-cytokine based TWAS for Seven Inflammatory Skin Disorders Identify Candidate Causal Genes in Keratinocytes

4.1 Introduction

Inflammatory skin diseases such as psoriasis, atopic dermatitis, and alopecia areata involve complex genetics and gene-gene/gene-environment interactions (Bieber, 2023). To reveal disease-associated genetic loci and provide a greater understanding of their immunopathogenesis, genome-wide association studies (GWAS) have been successfully undertaken in these diseases in the past two decades (Common et al., 2019; Elder, 2009; Tamari & Hirota, 2014). Although many genetic signals identified through GWAS in these diseases have been shown to play regulatory roles in immunocytes, especially T-cells (Ho & Kupper, 2019), their role(s) in other cell types, such as keratinocytes, the most common cell type in the skin, are not well-studied.

Transcriptome-wide association studies (TWAS) can facilitate the prioritization of disease-associated risk genes in GWAS, as they can identify potential causal genes associated with each locus (Gusev et al., 2016). TWAS examine one gene at a time, model its genetically regulated expression (GReX) in a given tissue, and assess the association between GReX and the phenotype of interest. Ideally, GReX is modeled through reference expression quantitative trait loci (eQTL) data in a disease-corresponding tissue. For example, recent studies have modeled GReX in skin, dermal fibroblast, lymphocyte, and whole blood to nominate disease-specific candidate genes in atopic dermatitis (Wu et al., 2023) and psoriasis (Zhu et al., 2022). However, eQTLs can be context-specific, dependent on intrinsic and extrinsic factors (Zhernakova et al., 2017), such as microbial stimulation (Kim-Hellmuth et al., 2017). In addition, it has been demonstrated that capturing context-specific and context-shared components of expression can improve the prediction of gene expression (Thompson et al., 2022). In the case of skin diseases, many of which are driven, or influenced, by cytokine interactions between keratinocytes and immune-cells (e.g., IL-17 from Th17; interferon and IFN-y from Th1; IL-13 from Th2), the regulation of gene expression can be cytokine-dependent. For instance, in atopic dermatitis, expression of the gene

encoding filaggrin (FLG), a major genetic risk locus for atopic dermatitis (Palmer et al., 2006), was demonstrated to be reduced in the keratinocytes exposed to the Th2 cytokines IL4 and IL-13 (Howell et al., 2009).

In this study, we improve the power of TWAS by *i)* generating eight sets of condition-specific GReX prediction models in keratinocytes for different cytokine stimulations for a total of 400 RNA-seq datasets from 50 individuals, and *ii)* using independent component analysis (ICA) to integrate the GReX models from the different cytokine stimulations to conduct multi-condition TWAS on seven inflammatory skin diseases with publicly available GWAS summary statistics. We further validated the results through pathogenesis-relevant data using scRNA-seq and spatial transcriptomic data. TWAS modeled on immune-activated epithelial cells are very limited, and our study highlights that by leveraging multi-condition we can provide complementary information for enhancing power.

4.2 Materials and methods

4.2.1 GWAS and risk loci of seven common skin disorders

We obtained GWAS summary statistics for seven inflammatory skin diseases: acne (Mitchell et al., 2022) (20,165 cases and 595,231 controls), alopecia areata (Betz et al., 2015) (3,253 cases and 7,543 controls), atopic dermatitis (Ried et al., 2015) (21,000 cases and 95,000 controls), psoriasis (Patrick et al., 2018) (11,024 cases and 16,336 controls), systemic lupus erythematosus (Bentham et al., 2015) (7,219 cases and 15,991 controls), systemic sclerosis (López-Isac et al., 2019) (9,846 cases and 18,333 controls), and vitiligo (Jin et al., 2016) (4,680 cases and 39,586 controls). For each disease locus (i.e., region with top association signal p-value $\leq 5\times10^{-8}$), we computed the Bayesian credible intervals (BCIs) within ±500kb of a GWAS signal using the corresponding GWAS summary statistics. For psoriasis, we computed the BCIs based on our more recent meta-analysis (Dand et al., 2023) to include potential novel disease loci. Details on the calculation of the BCIs can be found in our previous publication (Stuart et al., 2022).

4.2.2 GReX modeling

A cohort of 50 individuals was genotyped, and 400 RNA-seq samples were collected in keratinocytes for eight cytokine stimulations with different concentrations of stimuli: unstimulated, IFNα (5ng/ml), IFNγ (5ng/ml), IL-4 (10ng/ml), IL-13 (10ng/ml), IL-17A (10ng/ml),

IL-17A(10ng/ml)+TNFα(10ng/ml), and TNFα (10ng/ml) stimulated. Details on cell culture are provided in the Appendix C. The expression quantification of the RNA-seq data has been described previously (Tsoi et al., 2019; Tsoi et al., 2020). After removing lowly expressed genes with zero counts in at least half of the individuals, we obtained 18,599 genes, including 13,279 protein-coding genes annotated from GENCODE (Harrow et al., 2012). We performed PEER (Stegle et al., 2012) to remove confounding effects. The samples were genotyped using the Illumina Infinium CoreExome array, and after haplotyping Minimac (Das et al., 2016) was used for imputation using 1000 Genomes Project (GRCh37/hg19) as a reference panel (Consortium, 2012). Only genetic variants with imputation quality greater than 0.7 were used. In each cytokine condition, we used the genotypes of 50 individuals and their corresponding expression profile of genes to model the genetically regulated component of each gene's expression (GReX) by applying the Bayesian Sparse Linear Mixed Model (BSLMM) (Zhou et al., 2013). Statistical details on BSLMM are provided in the Appendix C.

4.2.3 TWAS analysis

We conducted TWAS in each condition for each disorder using S-PrediXcan (Barbeira et al., 2018), which leverages information from GWAS summary statistics and the BSLMM GReX models. Then for each disorder, we integrated single-condition TWAS results across all eight cytokine conditions for each gene and conducted a multi-condition TWAS using S-MultiXcan (Barbeira et al., 2019), which jointly models the phenotype through a linear regression on predicted GReX in all eight conditions. Both single-condition and multi-condition TWAS provide p-value as the evidence of association between the gene and the phenotype. Statistical details on the TWAS methods are provided in the Appendix C. We further applied the TWAS fine-mapping method mFABIO (introduced in Chapter 3) on psoriasis, to jointly model all gene-condition pairs simultaneously, and nominated candidate gene-condition pairs.

4.2.4 Colocalization analysis

We conducted colocalization analysis using FastENLOC (Hukku et al., 2021; Hukku et al., 2022; Pividori et al., 2020; Wen et al., 2017) to integrate genetic associations from QTL and GWAS data. FastENLOC was applied separately to each disease under each cytokine condition, focusing on protein-coding genes within LD blocks that contained at least one significant gene identified in the multi-condition TWAS analysis. For each tested gene, FastENLOC calculates a

gene-level colocalization probability (GCP), representing the likelihood that the gene contains at least one colocalized SNP. Genes with a GCP > 0.5 were considered having evidence of colocalization in the analysis. We further applied the Wilcoxon rank-sum test to compare the GCP of genes among the TWAS significant genes against that among the TWAS non-significant genes.

4.2.5 scRNA-seq and spatial transcriptomics data analysis

We used the scRNA of psoriatic skin as described before (Ma et al., 2023). The collected scRNA-seq data were processed and analyzed using R package Seurat (Hao et al., 2021) following the standard practice. The spatial transcriptomics data collected from CosMx platform were processed and analyzed using R package Giotto (Dries et al., 2021) following the standard practice. We combined the scaled expression level of multiple genes across cells as the module score to represent the expression level of targeted genes in different cell types.

4.3 Results

We leveraged the gene expression profiles from keratinocytes under different cytokine conditions to conduct multi-condition TWAS in seven inflammatory skin diseases: acne, alopecia areata (AA), atopic dermatitis (AD), psoriasis, systemic sclerosis (SSc), systemic lupus erythematosus (SLE), and vitiligo (Figure 4.1). We applied the Bayesian Sparse Linear Mix Model (BSLMM) to generate a prediction model for genetically regulated expression (GReX) using RNA-seq data from cytokine-stimulated keratinocytes from 50 genotyped individuals. With eight sets of GReX prediction models, we applied S-PrediXcan to conduct TWAS using each disorder's GWAS summary statistics and performing TWAS in each condition independently. We then integrated across all eight single-condition TWAS for testing of gene-trait associations as multi-condition TWAS using S-MultiXcan. This analytical workflow is summarized and available at: https://github.com/superggbond/Multi-context-TWAS.

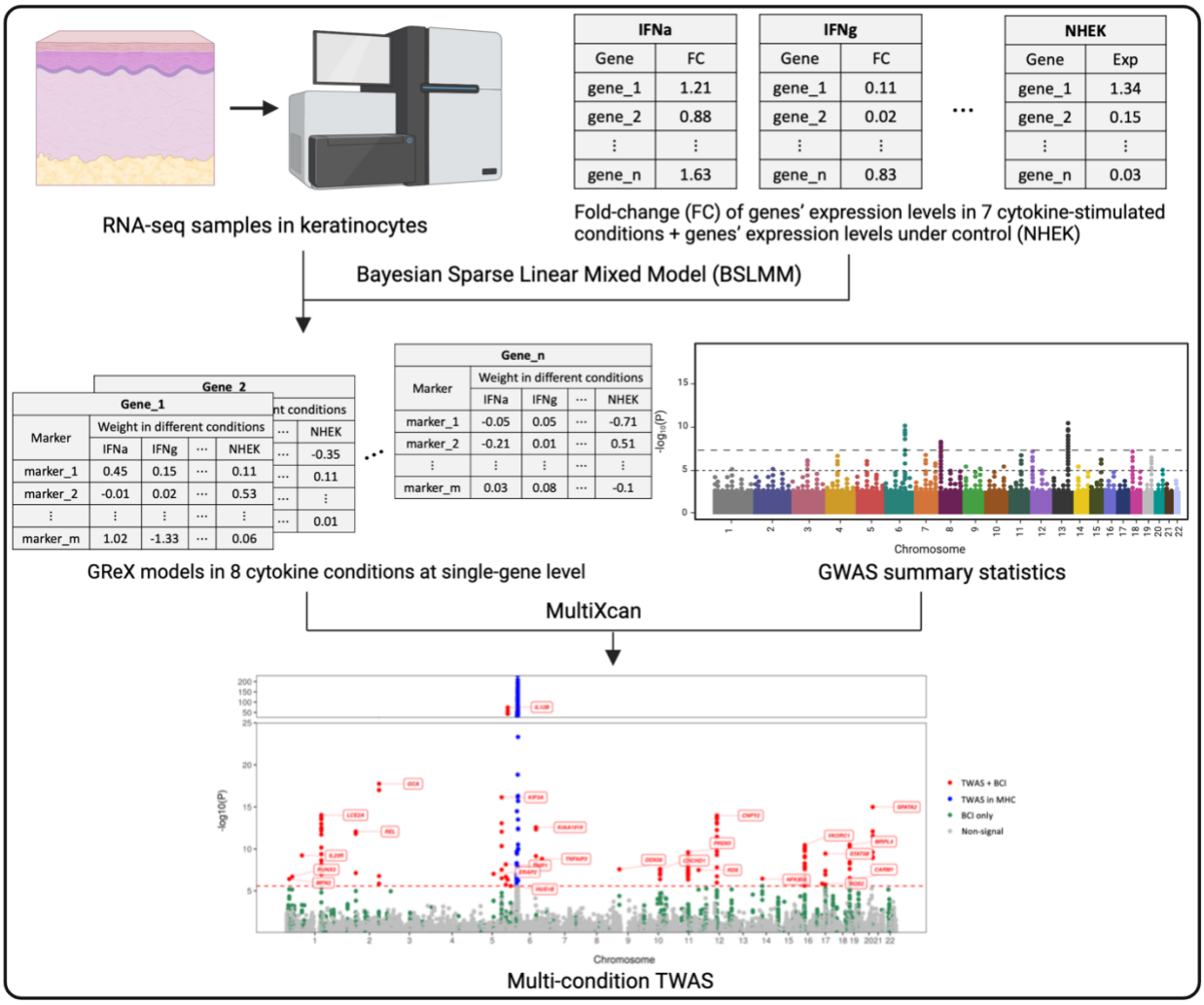

Figure 4. 1 Workflow of the multi-condition TWAS on a skin disease.

BSLMM is a hybrid approach between linear mix models (LMMs) and sparse regression models, outperforming other polygenic modeling methods in estimating the proportion of variance in phenotypes explained (PVE) by available genotypes and predicting phenotypes across a wide range of genetic architectures (Zhou et al., 2013). We applied BSLMM to model changes in the gene expression in unstimulated cells and for each cytokine stimulation. Single nucleotide polymorphism (SNP) weights were generated from each gene's GReX model and used in the TWAS analysis. We retained 2,751,472 SNPs shared across the genotyped cohort and the GWAS of all diseases except SSc. For SSc, where fewer markers are covered in the GWAS, 1,900,427 SNPs were shared among the genotyped cohort and across all the diseases. On average, 273±148 cis-SNPs were included in the GReX model per gene for SSc, and 405±185 cis-SNPs were

included per gene for acne, AA, AD, psoriasis, SLE, and vitiligo. Details on the model evaluation are provided in the Appendix C.

Significant genes from the TWAS were identified after adjusting for multiple testing using Bonferroni correction for each cytokine stimulation under each inflammatory skin disease. Details of the number of TWAS significant genes or protein-coding genes identified in each stimulation can be found in Table S4.1 and S4.2, respectively. For each disease, we counted the number of genes significant in at least one of the eight cytokine conditions (Table 4.1, Single-condition analysis). For instance, in VIT, there were 488 significant gene-condition pairs (Table S4.2) for 223 unique TWAS-significant protein-coding genes (Table 4.1, Single-condition analysis). The number of unique TWAS genes across different cytokine stimulations ranges from 124 to 355 among the seven skin diseases, with SLE having the highest number of significant TWAS genes and SSc the fewest. The number of Bayesian credible intervals (BCIs) for susceptibility regions of each disease is listed in Table 4.1, as well as the number of protein-coding genes that overlap with the BCIs. We extended the exon regions (exon region ± 500 kb) when overlapping a gene with a BCI. The number of identified TWAS signals overlapping with BCIs, and the number of BCIs containing such signals are summarized in Table 4.1, Single-condition analysis. Among the seven diseases, psoriasis had the highest number of TWAS protein-coding genes (119) overlapping with BCIs and most BCIs (25) containing such TWAS signals.

Table 4. 1 Multi-condition TWAS have higher power in revealing disease-associated genes.

			Acne	AA	AD	PSO	SSc	SLE	VIT
Total # of genes						18,599			
# of PC[1] genes						13,279			
# of BCIs[2]			41	10	12	119	11	39	40
# of PC genes that overlap BCIs			363	139	269	1,277	161	406	354
Single-condition analysis*	**# of TWAS sig.**	**genes[3]**	279	244	142	283	124	355	306
		PC genes[3]	205	179	114	210	84	260	223
		PC genes that overlap BCIs[3]	63	28	43	119	13	55	76
	# of BCIs that overlap TWAS sig. PC genes		17	4	7	25	4	15	20
Multi-condition analysis*	**# of TWAS sig.**	**genes**	404	322	221	293	182	458	430
		PC genes	288	241	167	215	138	328	324
		PC genes that overlap BCIs	72	40	57	122	13	57	90
	# of BCIs that overlap TWAS sig. PC genes		17	5	7	25	4	15	21

[1] PC: protein-coding

[2] Bayesian credible intervals (BCIs) of genomic regions identified from GWAS for each disorder

[3] The number of genes here represents the union set of TWAS targets across all seven different cytokine stimulations.

AA: alopecia areata; AD: atopic dermatitis; PSO: psoriasis: SSc: scleroderma; SLE: lupus; VIT: vitiligo

* Comparable counts of genes/BCIs share the same gradient grey shading.

To prioritize causal genes for different inflammatory skin diseases, we applied S-MultiXcan, leveraging information from seven cytokine stimulations in primary keratinocytes. Significant TWAS genes (after Bonferroni correction) range from 182 to 458, and 138 to 328 are significant TWAS protein-coding genes (Table 4.1, Multi-condition analysis), with the highest number of TWAS genes for SLE and the lowest for SSc. The rank of the number of TWAS protein-coding genes follows the same order as the number of significant genes across the seven diseases (Table 4.1, Multi-condition analysis). We partitioned the genome into 1,522 independent LD loci defined by LDetect (Berisa & Pickrell, 2016), a commonly used approach in GWAS and TWAS to identify regions of interest. This resulted in up to 94 genes (mean = 12, SD = 13) and up to 81 protein-coding genes (mean = 9, SD = 10) per LD locus. The density of TWAS signals per LD locus and the number of LD loci containing only one TWAS-significant protein-coding gene

varied. For example, we identified 215 TWAS significant protein-coding genes for psoriasis from 34 independent LD loci, suggesting that there are multiple causal gene candidates for many of these LD loci; in fact, only 11 LD loci contain a single TWAS significant protein-coding gene (Table S4.3). On the contrary, in vitiligo, the TWAS protein-coding genes came from 113 independent LD loci, and 58 LD loci encompass a single TWAS significant protein-coding gene, the most among the seven diseases. This is consistent with the total number of identified TWAS signals, given that vitiligo has the second-greatest number of TWAS signals. We further observed that, among LD loci containing a single TWAS-significant protein-coding gene, the majority did not overlap with BCIs identified by traditional GWAS (Table S4.3). Across the seven diseases, we identified an average of 42 such loci per disease, with an average of 39 (92.9%) harboring a single TWAS signal not overlapping with BCIs. These findings highlight the ability of our multi-condition TWAS analysis to nominate risk loci with a single, well-supported gene signal beyond what is captured by traditional GWAS. Details of TWAS signals and LD loci numbers can be found in Table 4.1 and Table S4.3, respectively.

Overall, we observed a greater number of genes in multi-condition TWAS than the single-condition TWAS (Table 4.1), with most of the genes identified by single-condition TWAS also covered by multi-condition TWAS (Figure S4.1). We also revealed more protein-coding genes that overlap with known BCIs in multi-condition TWAS, although the numbers of overlapping BCIs are similar to the single-condition TWAS, ranging from 4 to 25 among the seven inflammatory skin diseases (Table 4.1). Multi-condition significant TWAS protein-coding genes that overlap with BCIs are summarized in Manhattan plots for each disease (Figure 4.2 and Figure S4.2). For each locus, we compared the significant genes identified by multi-condition TWAS or by single-condition TWAS against those non-significant genes. First, we calculated the distance between the transcript start site (TSS) of a gene and the most significant GWAS signal in the locus of that gene. In all seven diseases, we observed shorter distances for genes identified by TWAS compared to non-significant genes (Figure S4.3a). Notably, the distance in genes identified by both the multi- and single- condition TWAS are generally shorter, or at least comparable, to that in genes identified exclusively by single only- or multi only- condition TWAS. Across all seven diseases, the average distance of identified genes shared by both multi- and single-condition TWAS (148 Kb) was significantly shorter than that of genes identified exclusively by multi-condition TWAS (216 Kb) or single-condition TWAS (219 Kb), with p-values of 5.42×10^{-5} and

1.83×10^{-11}, respectively. Next, for the genes identified by the TWAS analysis, we compared changes in expression levels in lesional versus non-lesional skin for four of the diseases: AA, acne, AD, and psoriasis. Significantly, we found that the magnitude of dysregulation (measured by the average of absolute log2FC) was higher in genes identified exclusively by multi-condition TWAS than in non-significant genes for AA, acne, and psoriasis, by 58.5% ($p = 1.81\times10^{-3}$), 17.6% ($p = 2.05\times10^{-2}$), and 28.4% ($p = 2.38\times10^{-3}$), respectively. (Figure S4.3b). Additionally, the magnitude was also higher than in genes identified exclusively by single-condition TWAS for all four diseases (Figure S4.3b). These results concord with the previous S-MultiXcan study, highlighting that we can obtain enhanced power by leveraging different conditions from same individuals.

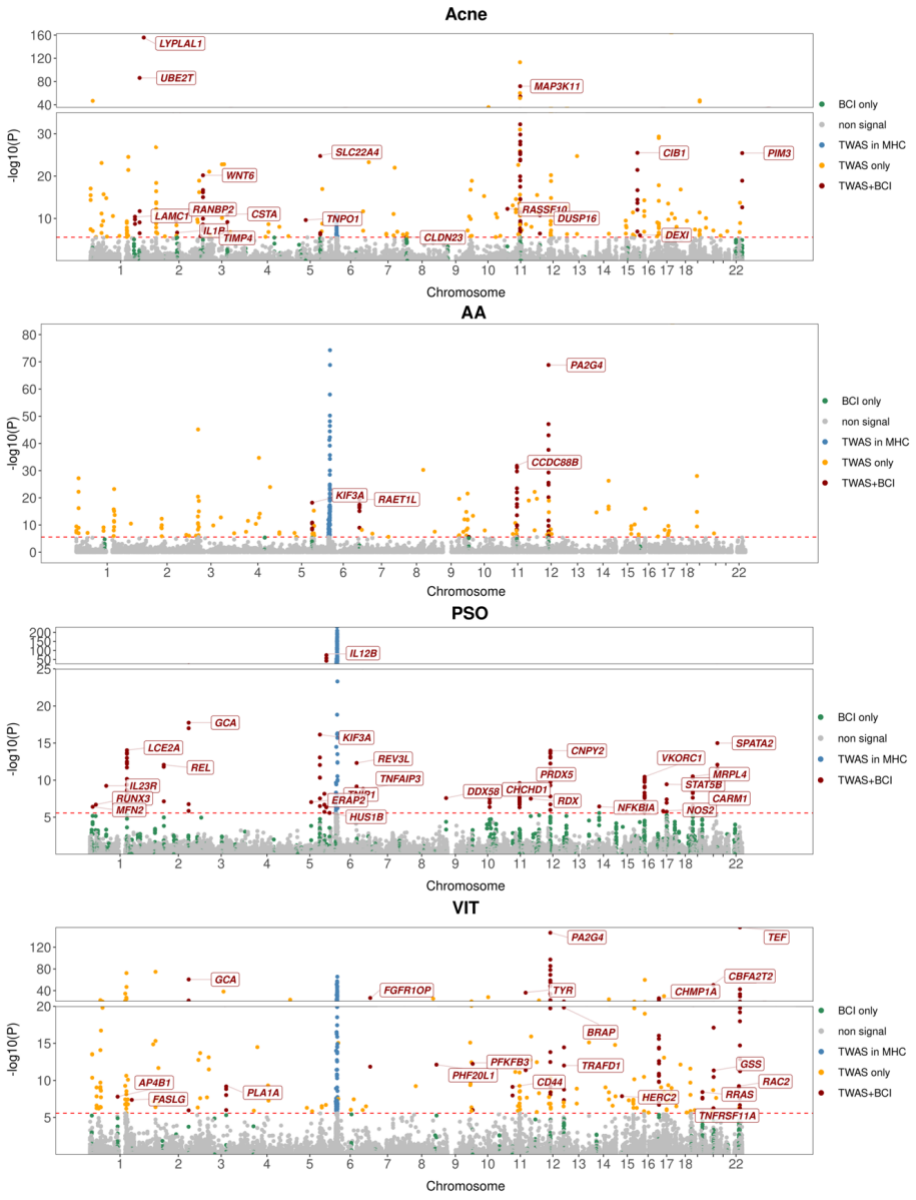

Figure 4. 2 Multi-condition TWAS Manhattan plots of signal genes for different diseases.

Genes overlapping BCIs without TWAS significance are labeled in green; TWAS significant genes in MHC region are labeled in blue; TWAS significant genes overlapping with BCIs are labeled in red; All other non-signals are grey. The top TWAS significant gene for each BCI is indicated in red box.

We then compared significance level in multi-condition TWAS against single-condition TWAS for some of the notable causal gene candidates, and studied their dysregulation in cytokine-stimulated keratinocytes (Tsoi et al., 2019; Tsoi et al., 2020) (Table 4.2). This revealed that leveraging shared eQTLs across multiple conditions improves our ability to identify potentially causal genes, as the TWAS signals are usually more significant in multi-condition TWAS than in the most significant single-condition TWAS (Table 4.2).

We highlight five candidate causal genes here. *ERAP2* encodes an aminopeptidase involved in antigen presentation (Hattori & Tsujimoto, 2013) and is a candidate causal gene for psoriasis (Figure 4.2), with eQTLs for *ERAP2* also shown to be associated with psoriasis in different populations (Popa et al., 2016; Wiśniewski et al., 2018). While *ERAP2* did not pass the Bonferroni-corrected p-value threshold in any of the single-condition TWAS (lowest $p = 7.05\times10^{-3}$ in IL-17A+TNF-stimulation, $p = 0.70$ in IL-17A-stimulation, and $p = 0.17$ in TNF-stimulation), the multi-condition TWAS boosted the p-value to 9.09×10^{-8} (Table 4.2). Notably, the human endoplasmic reticulum aminopeptidases (ERAPs) are encoded by two genes in the same locus: *ERAP1* and *ERAP2*. *ERAP2* is the only significant gene in LD locus chr5:95963519-97315601 in our multi-condition TWAS ($p = 1.44\times10^{-3}$ for *ERAP1*), suggesting this to be a strong candidate in this disease locus. When investigating the effect sizes (i.e., weights) of genetic markers involved in our GReX models of *ERAP2*, we noticed that markers in BCIs had higher median absolute weights than those outside of BCIs in IL-17A+TNFα stimulation, which is the best condition where we achieved the lowest p-value in single-condition TWAS (Figure 4.3a-b). Furthermore, when correlating the observed gene expression level versus the predicted gene expression level using all genetic markers versus only markers in BCIs, we observed better performance when all markers in GReX models were included (Figure 4.3c; $r = 0.51$ and $p = 1.5\times10^{-4}$ vs. $r = 0.035$ and $p = 0.81$ when using only BCI markers), highlighting the advantage of BSLMM that takes into consideration all markers in the region for gene expression modeling.

Table 4. 2 Identification of causal gene candidates for different skin immune-mediated disorders.

Disorder	Gene	Single-condition TWAS		Multi-condition TWAS	DEG analysis	
		Best condition[1]	p-value	p-value	2^MLE[2]	p-value[3]
Acne	*WNT6*	IFNγ	3.34×10^{-14}	6.64×10^{-21}	0.52	6.97×10^{-6}
	WNT10A	IFNγ	1.17×10^{-3}	1.10×10^{-10}	0.41	1.45×10^{-73}
	LGR6	IL17A+TNFα	1.03×10^{-6}	8.30×10^{-10}	0.55	2.74×10^{-8}
	CCDC138	TNFα	8.12×10^{-9}	3.81×10^{-10}	0.66	7.14×10^{-24}
	MAP3K11	TNFα	9.93×10^{-34}	1.20×10^{-72}	1.57	5.11×10^{-37}
	SEMA4B	TNFα	3.46×10^{-10}	3.56×10^{-22}	1.65	1.02×10^{-24}
AA	*RASGRP2*	IL17A+TNFα	5.91×10^{-6}	2.19×10^{-6}	0.59	3.45×10^{-4}
	RAET1L	TNFα	6.43×10^{-13}	3.74×10^{-20}	2.00	1.93×10^{-55}
AD	*LCE5A*	IFNγ	1.80×10^{-20}	1.88×10^{-17}	0.61	5.47×10^{-6}
	SPRR3	IFNγ	6.47×10^{-13}	5.45×10^{-14}	0.50	3.04×10^{-16}
	CRCT1	IL13	1.84×10^{-23}	6.47×10^{-22}	0.66	6.01×10^{-6}
	TCHHL1	IL4	1.90×10^{-5}	3.45×10^{-15}	0.59	2.24×10^{-4}
PSO	*LCE1F*	IL17A	2.18×10^{-5}	1.03×10^{-7}	0.57	3.49×10^{-3}
	LCE3D	TNFα	1.64×10^{-9}	1.98×10^{-12}	1.58	8.52×10^{-3}
	DDX58	TNFα	4.13×10^{-8}	2.60×10^{-8}	1.61	4.02×10^{-12}
	ERAP2	IL17A+TNFα	7.05×10^{-3}	9.09×10^{-8}	1.68	2.49×10^{-14}
	PRSS53	IL17A+TNFα	3.65×10^{-12}	8.28×10^{-11}	1.52	2.36×10^{-5}
SLE	*KCP*	IL17A+TNFα	2.41×10^{-28}	9.34×10^{-33}	0.59	1.21×10^{-5}
	ITGAM	TNFα	9.62×10^{-32}	6.25×10^{-48}	3.18	5.81×10^{-26}
VIT	*IFIH1*	IL17A+TNFα	2.41×10^{-10}	2.82×10^{-22}	2.27	2.87×10^{-38}
	SLC37A1	TNFα	1.42×10^{-7}	2.59×10^{-6}	1.87	1.16×10^{-32}

[1] The gene achieves the most significant p-value in this cytokine condition in the single-condition TWAS analysis.

[2] MLE: fold change based on maximum likelihood estimation

[3] The adjusted p-value of the gene in the DEG analysis in the best condition.

WNT10A is a member of the large *WNT* family that plays an important role in multiple signaling pathways starting from embryonic development (Miller, 2001), and a recent GWAS meta-analysis supported *WNT10A* as a causal gene for acne predisposition (Mitchell et al., 2022). Our study provided complementary information and foundation through the integrative analysis to suggest *WNT10A* as a causal gene for acne in this locus (Figure 4.2). Similar to *ERAP2*, *WNT10A* did not achieve significance after Bonferroni correction in any of the single conditions (lowest $p = 1.17\times10^{-3}$ in IFNγ-stimulated condition), while the multi-condition TWAS boosted the p-value to 1.10×10^{-10} (Table 4.2). Markers in BCIs also had a higher median absolute weight than those outside of BCIs in the best cytokine stimulation (IFNγ), where we achieved the most significant

p-value in the single-condition TWAS (Figure 4.3d-e). Similar to what we observed for *ERAP2*, only using markers in BCIs led to poor prediction in gene expression level (Figure 4.3f).

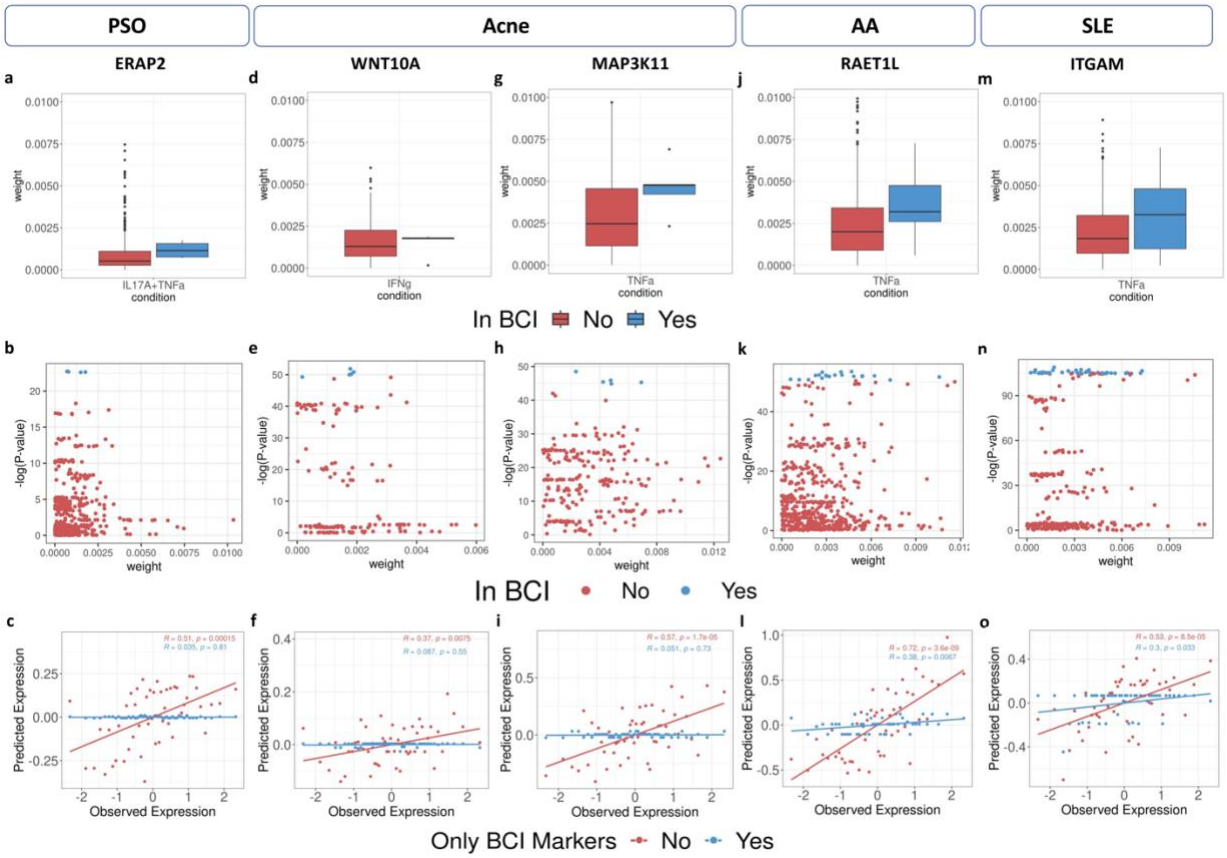

Figure 4. 3 Plots for GReX modeling.

(a) Box plot of absolute weights of genetic markers involved in the GReX model of *ERAP2* in IL17A+TNFα-stimulated condition. (b) Dot plot shows the modeled genetic markers' p-values from GWAS against absolute weights in the GReX model of *ERAP2* in IL17A+TNFα-stimulated condition. (c) Scatter plot with regression line shows the correlation between the predicted and observed gene expression level of *ERAP2* in IL17A+TNF-stimulated condition. (d) Box plot of absolute weights of genetic markers involved in the GReX model of *WNT10A* in IFNγ-stimulated condition. (e) Dot plot shows the modeled genetic markers' p-values from GWAS against absolute weights in the GReX model of *WNT10A* in IFNγ-stimulated condition. (f) Scatter plot with regression line shows the correlation between the predicted and observed gene expression level of *WNT10A* in IFNγ-stimulation. (g), (j), (m) Box plot of absolute weights of genetic markers involved in the GReX model of *MAP3K11*, *RAET1L*, and *ITGAM*, respectively, in TNFα-stimulated keratinocytes. (h), (k), (n) Dot plot shows the modeled genetic markers' p-values from GWAS against absolute weights in the GReX model of MAP3K11, *RAET1L*, and *ITGAM*, respectively, in TNF-stimulated keratinocytes. (i), (l), (o) Scatter plot with regression line shows the correlation between the predicted and observed gene expression level of *MAP3K11*, *RAET1L*, and *ITGAM*, respectively, in TNF-stimulated keratinocytes.

We further identified three other notable causal gene candidates in keratinocytes (Figure 4.2): *MAP3K11* for acne, *RAET1L* (*ULBP6*) for AA, and *ITGAM* for SLE. Both *MAP3K11* and *RAET1L* have been implicated in these diseases through GWAS (Mitchell et al., 2022; Petukhova

et al., 2010), and interestingly, previous work has highlighted that *ITGAM* is a risk factor for discoid cutaneous lupus than SLE (Jarvinen et al., 2010). Our integrative approach provides independent and supporting data for these genes as the strongest causal candidates in these loci. For these three genes, multi-condition TWAS dramatically increased the significance of the identified association (Table 4.2). The markers in BCIs also had higher median absolute weights for these three genes (Figure 4.3g-h, 4.3j-k, 4.3m-n, respectively). Significantly, only using markers in BCIs led to inferior prediction in the gene expression level of these three genes when comparing with our BSLMM approach (Figure 4.3i, 4.3l, 4.3o, respectively).

We next applied mFABIO (introduced in Chapter 3) to fine-map the psoriasis signals, integrating our individual-level GWAS data with the multi-condition GReX from eight cytokine stimulations. The results demonstrate a dramatic improvement in resolution. As detailed in Table 4.1, while a conventional multi-condition TWAS identified 293 significant genes, mFABIO narrowed this down to just 34 significant gene-condition pairs, representing 31 unique genes. This refined set showed strong concordance with the initial findings: 25 of the 31 genes were also identified by the conventional TWAS. Importantly, all 31 mFABIO-prioritized genes were contained within the BCIs indicated in Table 4.1, and all five key gene-condition pairs highlighted for psoriasis in Table 4.2 were validated. These findings show that mFABIO not only provides robust, supplementary evidence for conventional TWAS associations but also substantially refines them by pinpointing a much smaller, higher-confidence set of candidates.

We used genetic colocalization (FastENLOC) (Hukku et al., 2021; Hukku et al., 2022; Pividori et al., 2020; Wen et al., 2017) to evaluate if the disease association shares same signal with the genes' eQTL results under each cytokine condition, focusing on protein-coding genes within LD blocks that contained at least one significant gene identified in the multi-condition TWAS analysis. Specifically, we overlapped the colocalization results with TWAS-significant genes under the top two cytokine conditions highlighted in Table S4.4, which had the largest numbers and proportions of identified TWAS-significant protein-coding genes. For each disease-cytokine condition pair, we computed the percentage of colocalized genes (i.e. gene-level colocalization probability GCP > 0.5) among *(i)* TWAS-significant genes and *(ii)* genes from the same locus but were not significant in TWAS. We observed that, for all 14 disease-cytokine pairs listed in Table S4.4, the percentage of colocalized genes was consistently higher among the TWAS-significant group than among the non-significant group (Figure S4.4). We applied the

Wilcoxon rank-sum test to compare the GCP of genes among those two groups and observed that all 14 pairs individually had significantly ($p < 0.05$) higher GCP of genes among the TWAS-significant group (Table S4.5), as well as combining them all together ($p = 1.03\times10^{-15}$). Notably, in 10 out of 14 pairs, only the TWAS significant genes showed colocalization, and 89% of the colocalized genes (63 out of 71) are TWAS significant, indicating that the colocalization analysis effectively controlled false positives. Details on the number of genes and p-values are summarized in Table S4.5. We further overlapped colocalization results with top candidates listed in Table 4.2, and we have demonstrated strong evidence of colocalization for: *WNT6* (IFNγ, GCP = 0.78), *WNT10A* (IFNγ, GCP = 0.95), and *SEMA4B* (TNFα, GCP = 0.80) for acne; *RAET1L* (TNFα, GCP = 0.86) for AA; *CRCT1* (IL-13, GCP = 0.95) for AD; *LCE3D* (TNFα, GCP = 0.77) for PSO; *ITGAM* (TNFα, GCP = 0.90) for SLE; and *IFIH1* (IL-17A+TNFα, GCP = 0.96) for VIT.

We also evaluated our TWAS results using different single cell expression platforms. First, by using psoriasis scRNA-seq, we highlighted that most of the TWAS targets have higher expressions in keratinocytes compared with other cell types (Figure S4.5). We then investigated how cell-cell interactions would modulate the TWAS targets in keratinocytes by studying the TWAS results in single-cell resolution spatial expression data using the CosMX platform (Figure 4.4a). We first identified individual TWAS target genes with significant spatial expression patterns, and their expressions were elevated among keratinocytes interacting with other cell types (compared to keratinocytes not interacting). We identified four TWAS signal genes (Figure 4.4b), including *DDX58* (Figure 4.4c), which have higher expression in keratinocytes interacting with type 2B conventional dendritic cells (cDC2B). Then we calculated a module score (defined as the scaled expression level of genes) combining all TWAS signal genes covered in the psoriasis CosMx dataset. It revealed that the module score is higher in the keratinocytes interacting with the CD16+ dendritic cells ($CD16^+$ DC; $p=4.87x10^{-16}$), Langerhans cells (LC; $p=1.77x10^{-3}$), and Tc17 ($p= 3.23x10^{-3}$) (Figure 4.4d-f, respectively). Figure 4.4g illustrates how these immunocytes interact with keratinocytes in the *in-situ* space. Significantly, these keratinocytes interacting cells are the ones that show the highest expressions of TNF or IL-17A (for Tc17) (Figure 4.4h). These results highlight how TWAS target gene expression in keratinocytes is altered by the proximity of TNF/IL17 expressing immunocytes, and further show that $CD16^+$ DC, LC, and Tc17 are major sources of these cytokines in psoriatic skin.

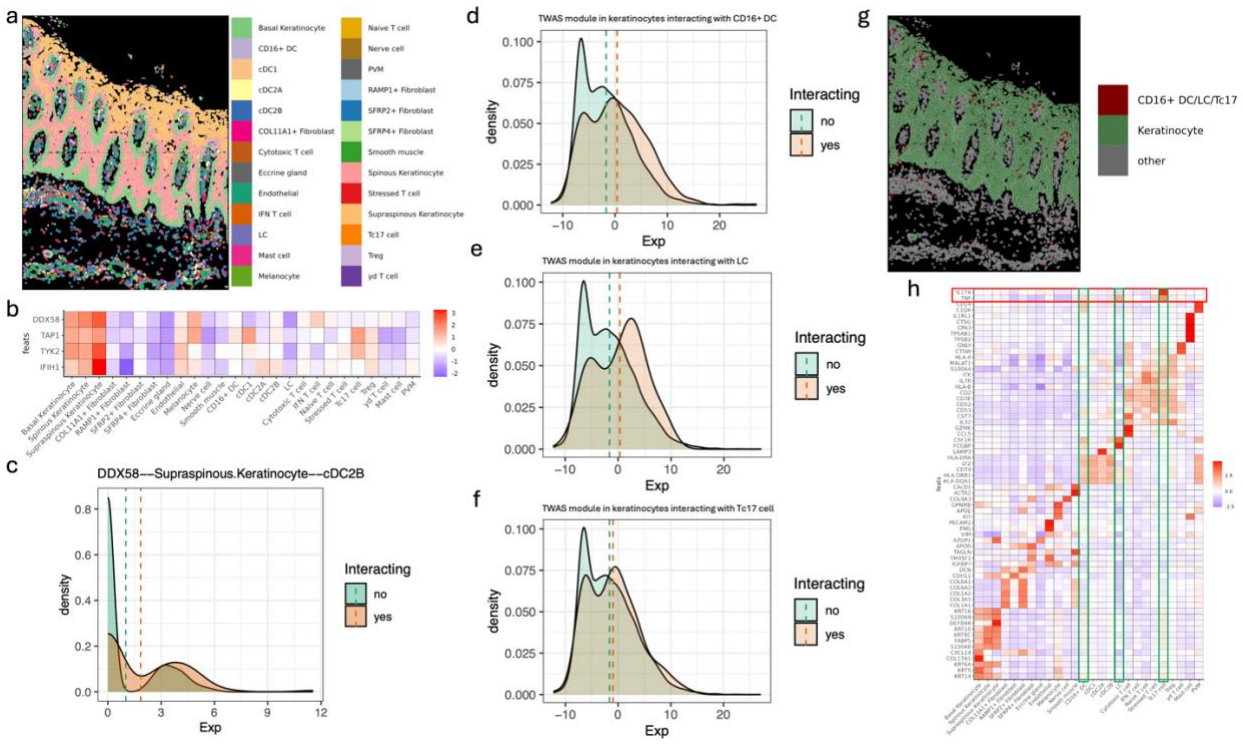

Figure 4. 4 Plots for the analysis of spatial transcriptomics.

(a) In-situ plot of cells from CosMx dataset with cell type annotated. (b) Heatmap of selected TWAS signals across different cell types. (c) Density plot of the expression of DDX5B in two interacting types of cells. (d-f) Density plot of the TWAS module score in two interacting types of cells. (g) In-situ plot of cells from CosMx dataset indicating the interaction between CD16+ DC/LC/Tc17 and keratinocytes. (h) Heatmap of representative marker genes across different cell types. The highlighted cytokine stimulus is indicated by the red box, and the highlighted cell types are indicated by green boxes.

4.4 Discussion

The success of GWAS has facilitated functional genomics studies to identify gene targets for disease-associated loci. For complex skin diseases, most of these functional studies have focused on the regulatory roles of these susceptibility regions in immune cells (Bangert et al., 2011; Cao et al., 2019; Das et al., 2019; Ilkovitch, 2011; Sawada & Gallo, 2021; Tamagawa-Mineoka, 2015). Thus, the regulatory roles of skin disease signals identified from GWAS are much less understood in cell types other than immunocytes, such as keratinocytes and fibroblasts. In this study, we conducted TWAS for seven common inflammatory skin diseases to explore the regulatory roles of identified GWAS loci in keratinocytes. In addition to a single-condition eQTL reference to model the gene expression, we leverage eQTL studies of keratinocytes under seven different inflammatory stimuli to perform a multi-condition TWAS, establishing a more powerful

approach for the identification of causal genes and their associated inflammatory activating stimuli.

Although the sample size of the eQTL studies we utilized was relatively small (sample size = 50), our study revealed unique significant TWAS genes. We also emphasized the increased power of multi-condition TWAS for each of the seven inflammatory skin diseases analyzed, with the multi-condition TWAS covering 85.4% of the single-condition TWAS significant protein-coding genes. The small percentage of single-condition TWAS-only significant signals can be attributed to unique and highly cytokine-specific signaling pathways. For example, *NR4A2* did not reach significance in our multi-condition TWAS for acne, while it is significant in our single-condition TWAS in IFNα-stimulation. We also evaluated the performance by calculating the distance between the TSS and the most significant GWAS signal within the LD locus of that gene, finding that genes identified exclusively by multi-condition TWAS had a shorter distance than those identified exclusively by single-condition TWAS. In addition, genes identified exclusively by multi-condition TWAS have consistent higher magnitude in disease dysregulation. Through multi-condition TWAS, we also investigated the number of independent components of variation kept among our seven cytokine conditions, which can be treated as the number of synthetic independent conditions (SICs). Over half of the genes had 4 or 5 SICs, which indicated that the gene expression in keratinocytes is correlated among different cytokine conditions. These observations above indicated that, given relatively small eQTL reference panels, multi-condition TWAS can still be a more powerful framework to integrate multiple conditions while controlling for false positives as it considers the correlation between conditions. To complement our TWAS analysis, we also performed colocalization analysis. It revealed that, in all examined disease-cytokine pairs, TWAS-significant genes consistently displayed higher rates of colocalization than their non-significant counterparts. This finding supports the reliability of the gene candidates identified through our TWAS approach.

Although multi-condition TWAS is generally more robust, the numbers of BCIs containing TWAS signals did not differ significantly from the single-condition TWAS when overlapping the signals with disease-specific BCIs calculated from the corresponding GWAS risk loci. This implies that, within a disease-specific BCI, multi-condition TWAS identified more disease-associated genes than single-condition TWAS. This was confirmed when we narrowed down our list of candidate genes. We highlighted five examples of gene candidates by searching for TWAS-

significant genes that are also differentially expressed in keratinocytes in specific cytokine stimulation. Among them, *ERAP2* and *WNT10A* are two genes that did not achieve significance in any of the single-condition TWAS but reached a level of significance in the multi-condition TWAS. We also observed in all five examples that multi-condition TWAS dramatically increased the significance of their association.

In addition to the single-gene candidates, our data show that by integrating the findings from the TWAS studies with single-cell and spatial-seq data, we are able to infer deeper biological insights into the disease mechanisms involved. Our study provides a framework that integrates the BSLMM-predicted GReX in multiple conditions to highlight the regulatory roles of GWAS signals in keratinocytes for common skin disorders. The genomic data we used in this study also suggests that the response QTL is essential to studying these complex immune-mediated conditions.

4.5 Supplementary tables

Table S4. 1 Number of single-condition TWAS significant genes identified in each cytokine condition.

	Acne	AA	AD	PSO	SSc	SLE	VIT
IFNα	70	73	40	114	26	133	97
IFNγ	57	72	30	125	25	111	74
IL13	73	70	28	125	22	126	76
IL17A	62	65	29	122	29	138	85
IL17A+TNFα	68	51	30	128	23	131	81
IL4	62	62	31	111	26	134	98
TNFα	58	71	28	126	20	118	78
NHEK	67	74	27	126	21	120	89

Table S4. 2 Number of single-condition TWAS significant protein-coding genes identified in each cytokine condition.

	Acne	**AA**	**AD**	**PSO**	**SSc**	**SLE**	**VIT**
IFNα	45	56	33	79	20	96	67
IFNγ	47	51	26	94	16	80	54
IL13	53	53	25	98	20	100	57
IL17A	45	43	27	91	20	96	63
IL17A+TNFα	55	39	25	98	15	102	59
IL4	43	45	24	82	18	95	70
TNFα	41	52	24	95	13	85	53
NHEK	52	56	24	93	13	91	65

Table S4. 3 Number of LD loci that contain TWAS signals.

		Acne	AA	AD	PSO	SSc	SLE	VIT
# of LD loci containing	TWAS sig. genes	136	98	91	36	74	127	142
	TWAS sig. PC* genes	109	81	66	34	55	100	113
	only one TWAS sig. gene	65	53	57	9	48	60	75
	only one TWAS sig. PC gene	58	45	41	11	34	46	58
	only one TWAS sig. PC gene which is overlapping BCIs	6	2	1	1	1	5	3
	only one TWAS sig. PC gene which is not overlapping BCIs	52	43	40	10	33	41	55

* PC: protein-coding

Table S4. 4 Multi-condition TWAS highlight cytokine-specific associations.

Among the PC genes that are significant in the multi-condition TWAS, the table highlights the number (and percentage) of genes that exhibit the most significant single-condition TWAS in each cytokine condition.

	IFNα	**IFNγ**	**IL13**	**IL17A**	**IL17A+TNFα**	**IL4**	**TNFα**	**NHEK**	**SUM**
Acne	39	32	**42**	32	**41**	30	30	42	288
(%)	(13.54)	(11.11)	**(14.58)**	(11.11)	**(14.24)**	(10.42)	(10.42)	(14.58)	(100.00)
AA	**37**	27	**35**	20	29	20	32	41	241
(%)	**(15.35)**	(11.20)	**(14.52)**	(8.30)	(12.03)	(8.30)	(13.28)	(17.01)	(100.00)
AD	20	20	**25**	17	20	15	**23**	27	167
(%)	(11.98)	(11.98)	**(14.97)**	(10.18)	(11.98)	(8.98)	**(13.77)**	(16.17)	(100.00)
PSO	24	**32**	24	21	**31**	20	24	39	215
(%)	(11.16)	**(14.88)**	(11.16)	(9.77)	**(14.42)**	(9.30)	(11.16)	(18.14)	(100.00)
SSc	**23**	13	**20**	19	12	16	18	17	138
(%)	**(16.67)**	(9.42)	**(14.49)**	(13.77)	(8.70)	(11.59)	(13.04)	(12.32)	(100.00)
SLE	**52**	37	36	**43**	38	35	36	51	328
(%)	**(15.85)**	(11.28)	(10.98)	**(13.11)**	(11.59)	(10.67)	(10.98)	(15.55)	(100.00)
VIT	**55**	40	34	39	38	34	**42**	42	324
(%)	**(16.98)**	(12.35)	(10.49)	(12.04)	(11.73)	(10.49)	**(12.96)**	(12.96)	(100.00)

* Top two largest numbers and proportions except for the control (NHEK) are bolded for each skin disorder.

Table S4. 5 Summary of colocalization analysis for different disease-cytokine pairs.

		coloc+TWAS+[1]	TWAS+[2]	coloc+TWAS-[3]	TWAS-[4]	p-value[5]
Acne	**IL13**	2	42	0	75	2.19×10^{-2}
	IL17A+TNFα	5	41	0	57	4.58×10^{-8}
AA	**IFNα**	3	37	0	50	1.61×10^{-3}
	IL13	6	35	3	43	1.70×10^{-2}
AD	**IL13**	5	25	0	56	4.77×10^{-3}
	TNFα	1	23	1	42	1.74×10^{-3}
PSO	**IFNγ**	4	32	0	35	4.01×10^{-2}
	IL17A+TNFα	6	31	0	28	3.31×10^{-2}
SSc	**IFNα**	2	23	0	49	9.47×10^{-3}
	IL13	3	20	1	70	3.91×10^{-2}
SLE	**IFNα**	4	52	0	74	2.22×10^{-5}
	IL17A	6	43	0	81	1.22×10^{-2}
VIT	**IFNα**	9	55	0	115	5.29×10^{-3}
	TNFα	7	42	3	70	7.28×10^{-3}
Total		63	501	8	845	1.03×10^{-15}

[1] coloc+TWAS+: the number of colocalized genes that are also significant in TWAS

[2] TWAS+: the number of TWAS significant genes

[3] coloc+TWAS-: the number of colocalized genes that are not significant in TWAS

[4] TWAS-: the number of TWAS non-significant genes

[5] p-value: p-value of the Wilcoxon rank-sum test, where we test whether the colocalization probability of genes among the TWAS significant genes is higher than that among the TWAS non-significant genes

4.6 Supplementary figures

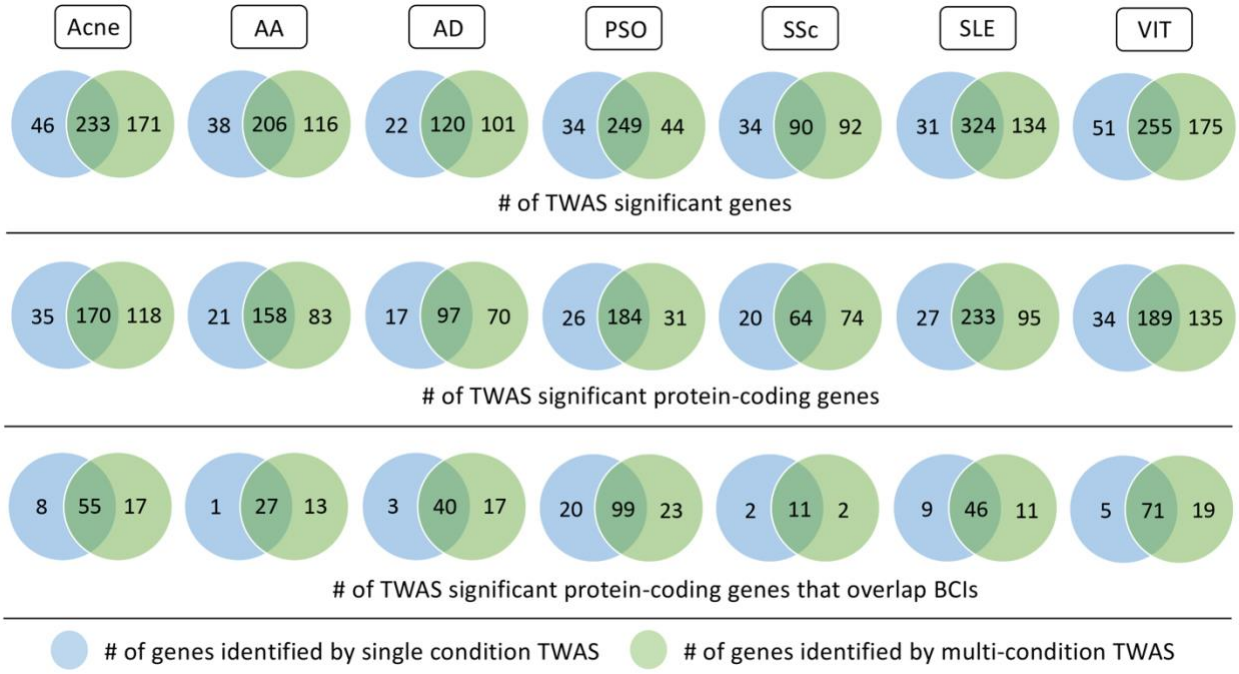

Figure S4. 1 Venn diagrams showing numbers of genes specifically identified by single condition TWAS, multi-condition TWAS, or shared by both methods.

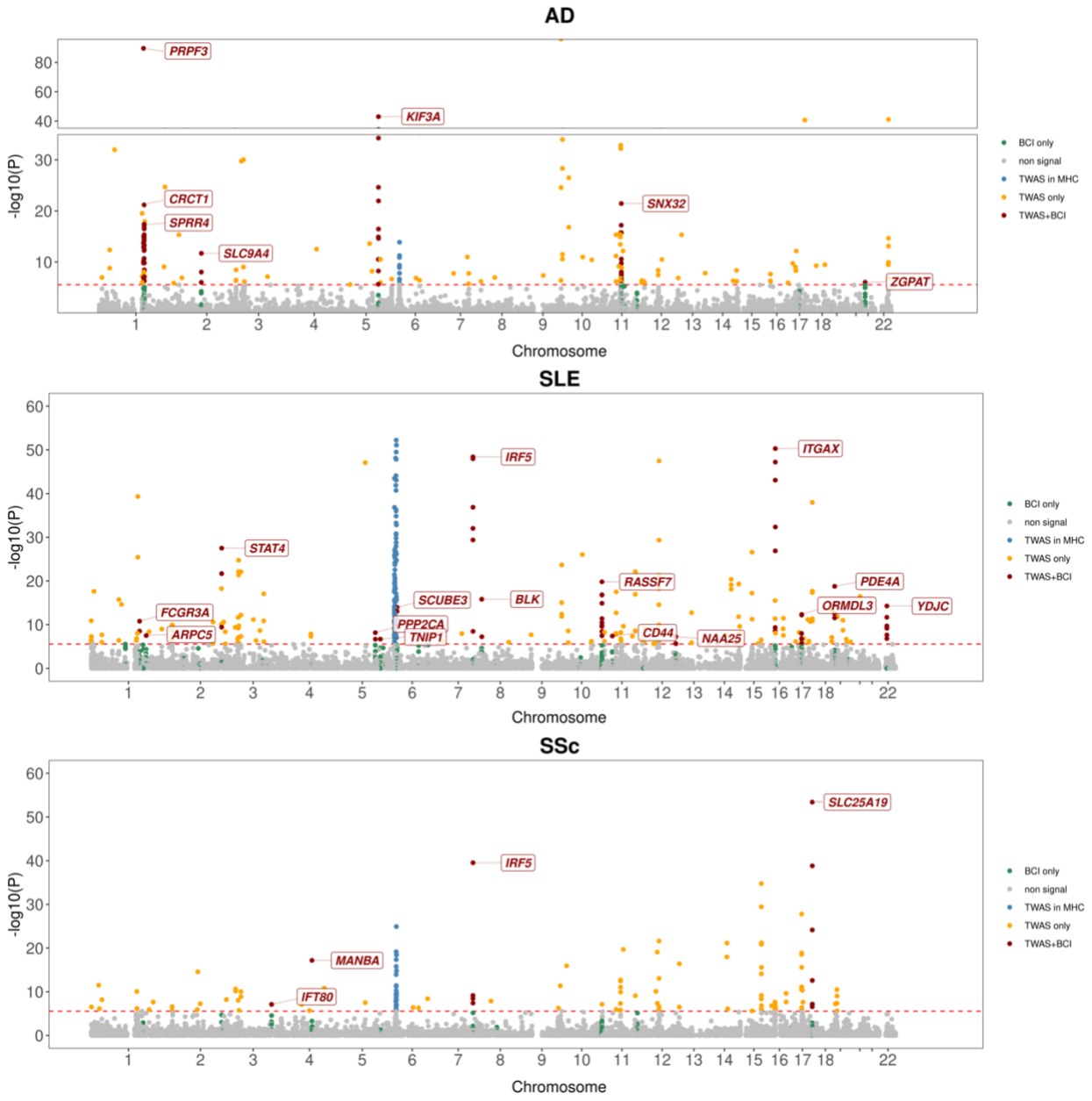

Figure S4. 2 TWAS Manhattan plots of signal genes for different diseases.

Genes overlapping BCIs without TWAS-significance are labeled in green; TWAS significant genes in MHC region are labeled in blue; TWAS significant genes overlapping BCIs are labeled in red; All other non-signals are grey. The top TWAS significant gene for each BCI is indicated with boxed red name.

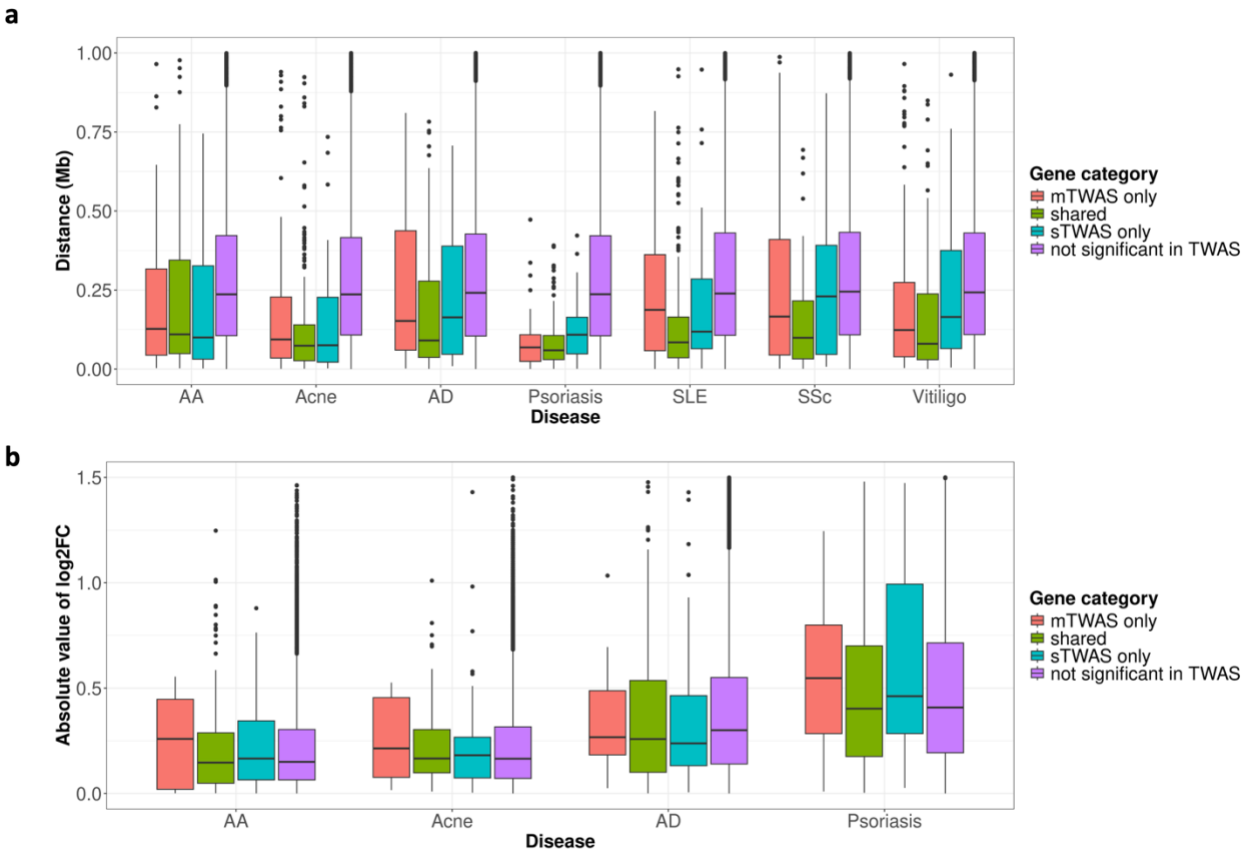

Figure S4. 3 Boxplot of measurements for different gene categories across diseases.

mTWAS only: significant genes identified exclusively by multi-condition TWAS; sTWAS only: significant genes identified exclusively by single-condition TWAS; shared: significant genes identified by both TWAS; not significant in TWAS: genes shown no significancy in either of the TWAS. (a) The distance between the transcript start site of each gene and the most significant GWAS signal within the LD locus of that gene. (b) The absolute value of log2 fold change (log2FC) comparing the gene expression levels in lesional versus non-lesional skin.

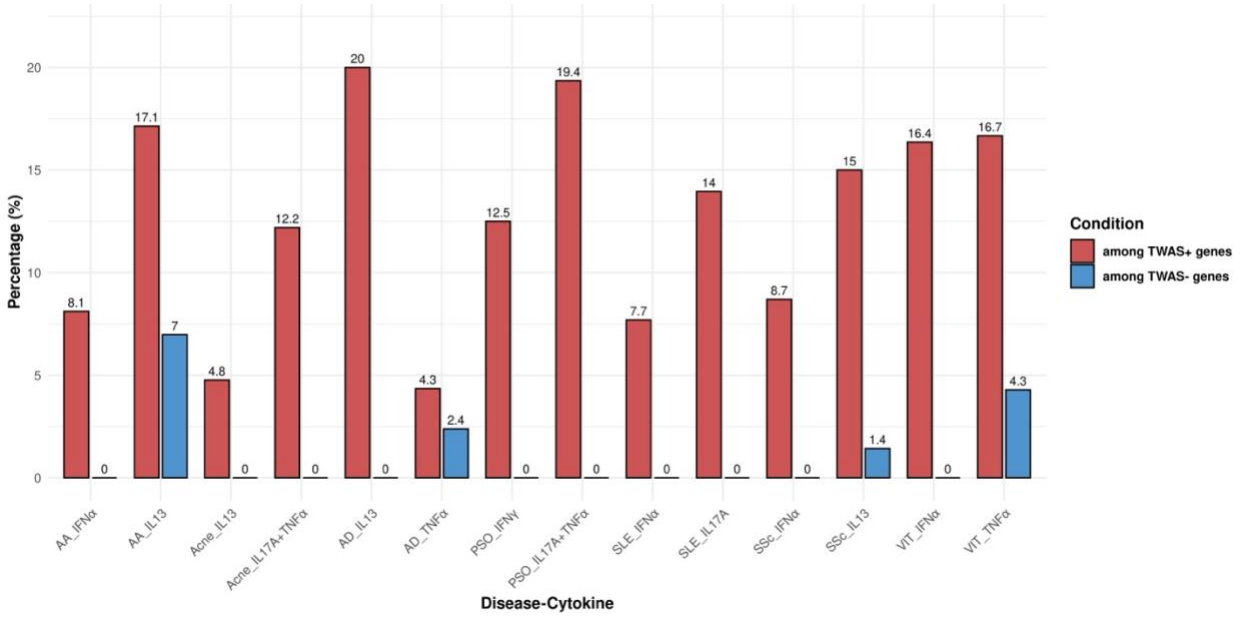

Figure S4. 4 Percentages of colocalized genes among TWAS significant (TWAS+) genes and among TWAS non-significant (TWAS-) genes across disease-cytokine pairs.

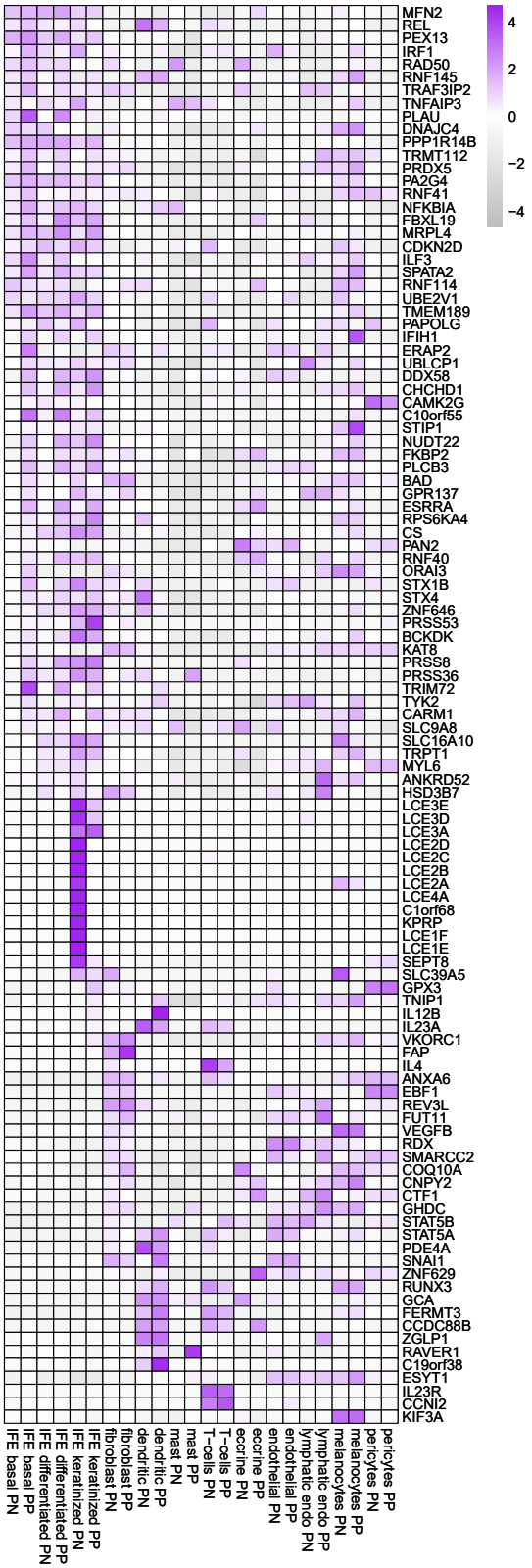

Figure S4. 5 Heatmap of TWAS signal expression across different cell types in scRNA-seq data.

PP: lesional skin of psoriasis; PN: non-lesional skin of psoriasis.

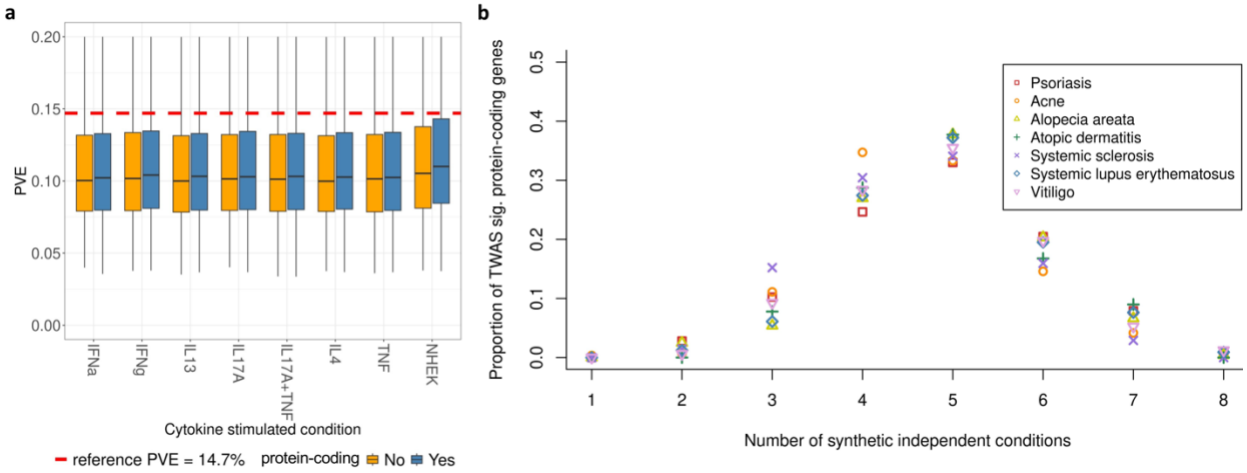

Figure S4. 6 Evaluation plots.

(a) Boxplot of PVEs in BSLMM GReX models grouped by protein-coding and non-protein-coding genes in different conditions. (b) Number of synthetic independent conditions (SICs) across all seven cytokine stimulations for different skin diseases in multi-condition TWAS analysis.

Chapter 5 Conclusion

The increasing availability of GWAS datasets (including those from the biobanks) and multi-tissue Expression Quantitative Trait Loci (eQTL) datasets has fueled the rapid development of Transcriptome-wide Association Studies (TWAS). TWAS enhance our ability to identify gene-trait associations by moving beyond the single-variant-level discovery offered by GWAS alone. This dissertation addresses key challenges in this evolving field by advancing TWAS methodologies, with a specific focus on statistical fine-mapping. Furthermore, it demonstrates the translational power of these techniques by applying both conventional and fine-mapping frameworks to investigate the genetic architecture of inflammatory skin disorders through the integration of multi-omics data. In this chapter, I will review these projects, discuss their limitations, and propose promising directions for future research.

5.1 Review of projects

In Chapter 2, we addressed two critical limitations in existing TWAS fine-mapping methods. First, current methods are primarily designed for quantitative traits and rely on linear models. While this approach is a reasonable approximation in GWAS where single-variant effect sizes are small, TWAS often reveals much larger genetically regulated expression (GReX) effects. Consequently, using linear models for binary disease traits in TWAS may not accurately capture the dichotomous nature of the outcome, leading to a significant loss of statistical power. Second, existing frameworks analyze GReX correlations only within local genomic regions defined by linkage disequilibrium (LD) blocks. This narrow scope overlooks correlations that can span across regions due to long-range LD or co-expression. Furthermore, by restricting the analysis to these blocks, these methods risk missing causal genes located elsewhere, which inflates residual error variance and further diminishes the power to detect true signals. To overcome these challenges, we developed FABIO. Our framework introduces two key innovations:

1. It explicitly models the binary nature of the outcome using a probit model, improving accuracy and power.

2. It simultaneously analyzes all genes across an entire chromosome, effectively accounting for GReX correlations both within and between LD blocks. This chromosome-wide approach enhances fine-mapping precision by reducing residual error.

FABIO also implements an efficient Markov Chain Monte Carlo (MCMC) algorithm for model inference, making it scalable to large, biobank-scale datasets, while avoiding the high computational cost associated with directly modeling binary outcomes using a probit link. Compared with other existing linear model-based methods, FABIO outperforms approaches such as FOGS, GIFT, and cTWAS. Even compared to the fastest existing method, FOCUS, FABIO can still achieve comparable running times. Through comprehensive simulations and an application to six binary disease traits in the UK Biobank, we demonstrate FABIO's superior power and accuracy in prioritizing putatively causal genes for binary outcomes.

In Chapter 3, we introduced mFABIO, a novel framework that extends our fine-mapping approach to a multi-tissue context. We developed mFABIO to address two persistent challenges in existing TWAS fine-mapping methods. First, most methods are limited by their reliance on single-tissue reference data. This is a significant drawback, as many complex diseases involve pathogenic mechanisms across multiple tissues, with genetic variants often altering gene expression in a tissue-specific manner. Consequently, single-tissue analyses risk overlooking the true causal gene if its relevant activity occurs in an unstudied tissue. Second, as demonstrated in Chapter 2, a more fundamental limitation is the inappropriate use of linear models for binary disease traits. This practice fails to accurately capture the dichotomous nature of case-control outcomes, leading to a loss of statistical power and reduced fine-mapping performance. mFABIO overcomes both limitations simultaneously. It explicitly models the binary nature of disease traits using a probit model while jointly analyzing gene-tissue pairs across multiple tissues within a given genomic region. This joint modeling strategy effectively accounts for complex correlation structures among genes and tissues, thereby enhancing fine-mapping precision. To ensure scalability to biobank-sized datasets, mFABIO employs an efficient variational inference algorithm and, like other modern approaches, accounts for potential horizontal pleiotropic effects. We validated mFABIO through comprehensive simulations and an analysis of six binary diseases from the UK Biobank. By comparing it against both single-tissue and a state-of-the-art multi-tissue method, we show that mFABIO is a powerful and precise tool for the multi-tissue fine-mapping of complex binary traits.

In Chapter 4, we investigated the complex genetic architecture of seven inflammatory skin disorders, including psoriasis, atopic dermatitis, and vitiligo. Recognizing that TWAS for skin diseases are limited by a lack of immune-activated cellular models, we created a more biologically relevant context by leveraging transcriptomic data from keratinocytes under eight distinct pro-inflammatory cytokine stimulations. We first conducted a large-scale, multi-condition marginal TWAS. The statistical power of this analysis was enhanced through two key steps: (1) we generated eight sets of condition-specific gene expression (GReX) prediction models from 400 RNA-seq datasets, and (2) we integrated these models using independent component analysis (ICA) to capture shared and unique regulatory signals across conditions. We then validated the identified gene-trait associations by contextualizing them with single-cell and spatial transcriptomic data. Finally, to move from broad associations to causal gene prioritization, we applied our mFABIO framework to the multi-condition psoriasis data. This analysis successfully resolved complex association signals from the marginal TWAS, demonstrating mFABIO's ability to pinpoint putative causal genes within a realistic, multi-condition biological system. Overall, this study establishes a powerful workflow that leverages multi-condition experimental data to first discover and then refine genetic signals, providing deeper insights into the mechanisms of inflammatory skin diseases.

5.2 Limitations and future directions

Through these three interconnected chapters, this dissertation introduces novel TWAS fine-mapping methods for binary traits, centered on a probit model that directly links genetically regulated expression (GReX) to disease status. Chapter 2 establishes this framework, Chapter 3 extends it to a multi-tissue context, and Chapter 4 demonstrates its practical power in dissecting the genetics of complex skin diseases. While this work addresses key gaps, important limitations and future directions remain.

First, our new methods require individual-level genotype and phenotype data. This presents a practical barrier to its widespread application, as access to such individual-level data is often restricted due to privacy concerns and institutional data sharing policies, precluding its use with the vast public repositories of GWAS summary statistics. However, this requirement is a deliberate trade-off made to enhance analytical accuracy. By working directly with individual-level data, our model accurately computes linkage disequilibrium (LD) from the target sample, thereby mitigating

the pervasive issue of LD mismatch that affects methods reliant on external reference panels. Other alternative methods must approximate LD using external reference panels (e.g., 1000 Genomes), which can introduce substantial bias if the genetic ancestry of the reference panel differs from that of the GWAS cohort. A critical future direction is to adapt our framework to be compatible with summary-level data, but this transition is non-trivial. The core challenge stems from our method's use of a probit model, creating a statistical mismatch with the logistic regression-based summary statistics that are the standard for public GWAS of binary traits. This issue reflects a broader challenge in the field, as fitting any generalized linear model like a logistic or probit regression requires an individual-level likelihood function that cannot be reconstructed from summary data. Consequently, the field has developed several effective strategies, such as using linear models as a robust approximation (Gusev et al., 2016) or transforming summary statistics to an underlying continuous liability scale (Bulik-Sullivan et al., 2015; Lee et al., 2011). Informed by these approaches, a primary goal for our future work is to develop a novel and principled transformation methodology specifically tailored to our framework - one that converts publicly available log-odds ratios into the probit-equivalent statistics our model requires. Successfully developing such a method would be a major contribution, combining the unique advantages of our probit-based approach with the immense practical power of existing summary datasets and dramatically broadening the impact of our method.

Second, like most other existing TWAS fine-mapping methods (Mancuso et al., 2019; Wu & Pan, 2020), our two-stage fine-mapping approach ignores uncertainty from the expression modeling stage. In the first stage, we build predictive models of gene expression using a reference panel, yielding a set of SNP weights for each gene. In the second stage, these weights are treated as fixed, error-free parameters to test for association with the GWAS trait. This sequential process conveniently modularizes the analysis, but it critically fails to propagate the substantial uncertainty from the initial expression modeling stage. The SNP weights are merely estimates derived from a finite reference sample and are subject to both sampling variation and model misspecification. By ignoring this uncertainty, our final fine-mapping results, such as the posterior inclusion probabilities (PIPs) for genes, are likely to be overly confident, potentially leading to an inflated rate of false discoveries (Hu et al., 2019; Mancuso et al., 2019). A more principled and robust future direction would be to develop a joint-likelihood framework that models the entire generative process in a single, unified step. Such a framework would simultaneously analyze the molecular

QTL data (e.g., eQTLs) and the GWAS summary statistics, allowing the uncertainty in the expression prediction to be naturally propagated through to the final gene-trait association analysis. While computationally demanding, this approach would yield more accurately calibrated and reliable evidence for causal genes, representing a significant advance in the statistical rigor of TWAS fine-mapping.

Third, our methods, like the vast majority of current TWAS approaches, depend on well-matched ancestry between the eQTL reference panel and the GWAS cohort. This requirement stems from the fact that both LD patterns and the genetic architecture of gene expression can differ substantially across ancestral populations (Consortium, 2015; Stranger et al., 2012). Applying an expression model trained in one population (e.g., European) to a GWAS cohort from another (e.g., African or East Asian) can lead to inaccurate genetic prediction of expression, which in turn severely diminishes statistical power and can produce misleading results (Martin et al., 2019). This dependency creates a major bottleneck, restricting the application of these powerful tools primarily to European-ancestry populations for whom large-scale reference datasets are most available. As a consequence, this limitation contributes to the widening gap in health equity by preventing the full benefits of genomic research from reaching diverse and underrepresented groups. Therefore, a critical next step for the field is the development of novel statistical frameworks that can effectively model and leverage multi-ancestry data. Future methods should move beyond the single-ancestry paradigm by jointly modeling data from diverse populations, which could borrow information across groups to build more robust and powerful predictive models. Such trans-ancestry tools would not only improve the accuracy of TWAS fine-mapping but are also essential for ensuring that the insights gained from genomic medicine are equitable and broadly applicable to all of humanity.

Finally, as demonstrated in the analyses within Chapter 4, we applied TWAS fine-mapping in concert with colocalization methods to interpret association signals. While often used together, it is crucial to distinguish between their distinct and complementary roles, and to understand why TWAS or TWAS fine-mapping alone can be insufficient. The primary goal of TWAS fine-mapping is to answer the "which gene" question by prioritizing causal genes within a locus. However, a significant association from this analysis can arise from two very different scenarios: either a true causal effect or a spurious one driven by confounding by LD. For instance, a gene may be strongly implicated simply because its top eQTL is in high LD with a separate, true causal

variant for the disease. In this case, TWAS fine-mapping would correctly identify the gene with the strongest association signal, yet this association would be non-causal. This is precisely the ambiguity that colocalization is designed to resolve. Colocalization addresses the "how" question by formally testing the hypothesis of vertical pleiotropy - that a single causal variant is responsible for both the molecular (eQTL) and the complex trait (GWAS) signals. By distinguishing between a shared causal variant and two distinct, linked variants, it provides the mechanistic evidence needed to rule out confounding by LD. Therefore, these methods form a powerful partnership: TWAS fine-mapping acts as a wide lens to nominate the most likely causal gene(s), while colocalization serves as a microscope to validate the specific mechanistic link, together building a more robust and compelling causal story from a GWAS locus.

5.3 Summary

In conclusion, this dissertation bridges a critical gap between statistical methodology and biological application in the field of transcriptome-wide association studies. By developing a novel suite of fine-mapping tools - FABIO and mFABIO - specifically engineered for binary disease traits, we have established a more robust and accurate framework for linking genetically regulated gene expression to disease risk. The progression from a single-tissue model to a multi-tissue, multi-condition application demonstrates a complete workflow, moving from foundational theory to practical implementation. The successful application of these methods to the complex landscape of inflammatory skin disorders showcases their power not just to identify statistical associations, but to resolve them into biologically plausible causal hypotheses. This body of work provides the research community with powerful new tools for immediate use and establishes a clear path forward for developing the next generation of more accurate, equitable, and integrative models of genetic risk for complex human diseases.

Appendices

Appendix A: Chapter 2 Supplementary Text

A.1 Prior specification of π and σ^2

Our method involves two hyper parameters, π and σ^2: π controls the proportion of non-zero $\boldsymbol{\alpha}$ values in equation (2.3) of the main text; and σ^2 controls the expected magnitude of the non-zero $\boldsymbol{\alpha}$ values. Following equation (2.1), we first place a uniform prior on log π:

$$\log(\pi) \sim \mathrm{U}(\log(1/p), \log(1)) . \tag{A.1}$$

A common alternative is a uniform distribution directly on π (Guan & Stephens, 2011), while it would place large weights on large numbers of variables having non-zero effect in the model than a uniform prior on $\log(\pi)$. For example, it would assume a 50% prior probability to the event that more than half of the genes have non-zero effects, and a 90% prior probability to the event that more than 10% of the genes have non-zero effects. However, only a small proportion of genes are likely to contribute to the phenotype as we expected in practice, and the uniform prior on log π is thus more suitable in this case. We also provide an alternative option of a beta prior distribution on π: $\pi \sim beta(a, b)$, where users can choose the values of a and b if they have prior knowledge to determine the prior expectation and the variance of π.

Following equation (2.1), we then re-parameterize the above model in terms of more interpretable quantities. Specifically, we define PVE as the total proportion of variance in latent variables $\boldsymbol{z}$ in equation (2.3) explained by the sparse effects of causal genes, and PVE can be represented as a function of $\widehat{\mathbf{G}}$ and $\boldsymbol{\alpha}$:

$$\mathrm{PVE}\left(\widehat{\mathbf{G}}, \boldsymbol{\alpha}\right) := \frac{\mathrm{V}(\widehat{\mathbf{G}}\boldsymbol{\alpha})}{\mathrm{V}(\widehat{\mathbf{G}}\boldsymbol{\alpha})+1}, \tag{A.2}$$

where the function $\mathrm{V}(x)$ is defined as:

$$\mathrm{V}(x) := \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 . \tag{A.3}$$

These definitions ensure that PVE will lie in the interval $[0,1]$. PVE reflects how well we could predict the latent variable $\boldsymbol{z}$ in our model equation (2.3) from the GReX if we knew the optimal $\boldsymbol{\alpha}$.

Since PVE is a function of $\boldsymbol{\alpha}$, whose prior distribution depends on parameters π and σ^2, the prior distribution for PVE will then depend on the priors assigned to these parameters. To simplify our model, we further aim to choose a prior for $\boldsymbol{\alpha}$ to induce the prior for PVE to be roughly uniform on 0 and 1. Following (Zhou et al., 2013), we introduce h as an approximation (a ratio of expectations rather than an expectation of ratios) to the expectation of PVE:

$$h := \frac{p\pi s\sigma^2}{p\pi s\sigma^2+1}, \tag{A.4}$$

where p, π and σ^2 are defined the same as before, s is the average variance of expression level across genes: $s = \frac{1}{np}\sum_{i=1}^{p}\sum_{j=1}^{n} g_{ij}^2$, g_{ij} is the ijth element of the matrix $\widehat{\mathbf{G}}$. Here in equation (A.4), $p\pi$ is the expected number of causal genes in the model, and σ^2 is the expected effect size variance of each causal gene, and s is the average variance of genes' predicted expression level. Therefore, the whole term $p\pi s\sigma^2$ represents the expected variance contributed by the sparse gene effects. Through re-parameterizing the model in terms of (h, π) instead of (σ^2, π), now we can specify the uniform prior distribution for h, which is independent of π:

$$h \sim \mathrm{U}(0,1), \tag{A.5}$$

Since h approximates PVE, this prior distribution will lead to a reasonable uniform prior distribution for PVE. However, we only treat h as the approximation to PVE for prior specification, we use PVE's definition in equation (A.2) when estimating it directly from the real data.

A.2 Posterior sampling

Our goal is to obtain the posterior inclusion probability (PIP) of $\boldsymbol{\alpha}$, which represents the evidence for gene-trait association. To do so, we introduce a vector of binary indicators $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_p) \in \{0,1\}^p$ to facilitate computation (George & McCulloch, 1993), where each γ_i indicates whether the corresponding α_i is non-zero. Our point-normal priors for $\boldsymbol{\alpha}$ can then be written as:

$$\gamma_i \sim \mathrm{Bernoulli}(\pi), \tag{A.6}$$

$$\boldsymbol{\alpha}_{\boldsymbol{\gamma}} \sim \mathrm{MVN}_{|\boldsymbol{\gamma}|}\left(0, \sigma^2\mathbf{I}_{|\boldsymbol{\gamma}|}\right), \tag{A.7}$$

$$\boldsymbol{\alpha}_{-\boldsymbol{\gamma}} \sim \delta_0, \tag{A.8}$$

where $\boldsymbol{\alpha}_{\boldsymbol{\gamma}}$ denotes the sub-vector of $\boldsymbol{\alpha}$ corresponding to the entries $\{i: \gamma_i = 1\}$; $\boldsymbol{\alpha}_{-\boldsymbol{\gamma}}$ denotes the sub-vector of $\boldsymbol{\alpha}$ corresponding to the other entries $\{i: \gamma_i = 0\}$; and $|\boldsymbol{\gamma}|$ denotes the number of non-zero entries in $\boldsymbol{\gamma}$. We use MCMC to obtain posterior samples of $(h, \pi, \boldsymbol{\gamma})$ on the product space $(0,1) \times (0,1) \times \{0,1\}^p$, which is given by:

$$P(h, \pi, \boldsymbol{\gamma}|\boldsymbol{y}) \propto P(\boldsymbol{y}|h, \pi, \boldsymbol{\gamma}) P(h) P(\boldsymbol{\gamma}|\pi) P(\pi) , \tag{A.9}$$

We also sample the posteriors of the latent variables $\boldsymbol{z}$ using the conditional distribution $\mathrm{P}(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\alpha})$:

$$z_i|y_i = 1, \boldsymbol{\alpha} \sim \mathrm{N}\left(\mu + \widehat{\boldsymbol{g}}_{\gamma_i}{}^T \boldsymbol{\alpha}, 1\right) \text{ left truncated at 0 ,} \tag{A.10}$$

$$z_i|y_i = 0, \boldsymbol{\alpha} \sim \mathrm{N}\left(\mu + \widehat{\boldsymbol{g}}_{\gamma_i}{}^T \boldsymbol{\alpha}, 1\right) \text{ right truncated at 0 .} \tag{A.11}$$

Conditional on the latent variables $\boldsymbol{z}$, the posterior sampling for the parameters $(h, \pi, \boldsymbol{\gamma})$ will be based on the marginal likelihood $\mathrm{P}(h, \pi, \boldsymbol{\gamma}|\boldsymbol{z})$. After integrating $\boldsymbol{\alpha}$ out, the marginal likelihood can be represented as:

$$P(\boldsymbol{z}|h, \pi, \boldsymbol{\gamma}) \propto |\sigma^{-2}\boldsymbol{\Omega}|^{\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{z}-\mathbf{1}_n\bar{z})^T \mathbf{P} (\boldsymbol{z}-\mathbf{1}_n\bar{z})} . \tag{A.12}$$

where $\boldsymbol{\Omega} = (\widehat{\mathbf{G}}_{\boldsymbol{\gamma}}^T \widehat{\mathbf{G}}_{\boldsymbol{\gamma}} + \sigma^{-2}\mathbf{I}_{|\boldsymbol{\gamma}|})^{-1}$, $\widehat{\mathbf{G}}_{\boldsymbol{\gamma}}$ is an $n \times |\boldsymbol{\gamma}|$ sub-matrix of $\widehat{\mathbf{G}}$ that only contains predicted expression level of genes with non-zero effects, $\bar{z} = \frac{1}{n}\sum_{i=1}^n z_i$ is the sample mean of $\boldsymbol{z}$, and $\mathbf{P} = \mathbf{I}_n - \widehat{\mathbf{G}}_{\boldsymbol{\gamma}} \boldsymbol{\Omega} \widehat{\mathbf{G}}_{\boldsymbol{\gamma}}^T$.

We use a standard Metropolis-Hastings algorithm to draw posterior samples of $(h, \pi, \boldsymbol{\gamma})$ from the above marginal distribution, starting with a rank-based proposal distribution for $\boldsymbol{\gamma}$ (Guan & Stephens, 2011). The advantage of this rank-based distribution is to focus more on examining genes with stronger marginal associations. To be more detailed, we first calculate the p-value for each gene using the standard LMM algorithm (Verbeke et al., 1997), and then rank genes based on their p values from small to large. To put more weights on genes that have higher rank in the single-gene tests, we propose with a mixture distribution $Q_p = 0.3U_p + 0.7G_p$ following (Zhou et al., 2013). Here, U_p is a uniform distribution on 1, 2, …, p and G_p is a geometric distribution truncated to 1, 2, …, p with its parameter chosen to give a mean of 2000. We denote $\boldsymbol{\gamma}^+ = \{i: \gamma_i = 1\}$ and propose the new $\boldsymbol{\gamma}$ by randomly choosing from one of the following options:

Option 1 (with probability 0.4): generate a number r from Q_p until the covariate with rank r is not in $\boldsymbol{\gamma}^+$, then add this covariate to $\boldsymbol{\gamma}^+$;

Option 2 (with probability 0.4): randomly (uniformly) select a covariate in $\boldsymbol{\gamma}^+$ and remove it from $\boldsymbol{\gamma}^+$;

Option 3 (with probability 0.2): select two covariates by the above two options and then switch their indicator values.

We then apply the random walk proposals based on uniform distributions for h and $\log(\pi)$: we update h by adding a random variable from $\mathrm{U}(-0.1,0.1)$ to the current value, and update $\log(\pi)$ by adding a random variable from $\mathrm{U}(-0.05,0.05)$ to the current value. New values that beyond the range from 0 to 1 will be reflected back. To improve the theoretical MCMC convergence, we also apply a "small world proposal" (Guan & Krone, 2007) in addition to the local proposals illustrated above. In the "small world proposal", we compound multiple local moves instead of a single move to induce longer-range proposals on the proposal distributions. The number of compounded local moves is uniformly drawn from a uniform distribution ranging from 1 to 20.

After obtaining the posterior samples of $\boldsymbol{z}$ and the parameters $(h, \pi, \boldsymbol{\gamma})$, we sample the posterior of $\boldsymbol{\alpha}$ using its conditional distribution:

$$\boldsymbol{\alpha}_{\boldsymbol{\gamma}} | \boldsymbol{z}, h, \pi, \boldsymbol{\gamma} \sim \mathrm{MVN}_{|\boldsymbol{\gamma}|}\left(\boldsymbol{\Omega}\widehat{\mathbf{G}}_{\boldsymbol{\gamma}}^{T}\boldsymbol{z}, \boldsymbol{\Omega}\right), \tag{A.13}$$

$$\boldsymbol{\alpha}_{-\boldsymbol{\gamma}} | \boldsymbol{z}, h, \pi, \boldsymbol{\gamma} \sim \delta_0 . \tag{A.14}$$

At the end, we sample μ in equation (2.3) based on the conditional distribution:

$$\mu | \boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\gamma} \sim \mathrm{N}\left(\frac{1}{n}\mathbf{1}_{\boldsymbol{n}}^{T}\left(\boldsymbol{z} - \widehat{\mathbf{G}}_{\boldsymbol{\gamma}}\boldsymbol{\alpha}_{\boldsymbol{\gamma}}\right), \frac{1}{n}\right). \tag{A.15}$$

A.3 Model specifications for FOCUS

FOCUS examines genomic loci one at a time. For a particular locus that contains p SNPs for m genes, FOCUS relies on a multiple linear regression model to model a quantitative trait $\mathbf{y}$ as a linear combination of gene expression levels in the form of $\mathbf{y} = \widehat{\mathbf{G}}\boldsymbol{\alpha} + \lambda\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{y}$ is a centered phenotype vector for n individuals, $\widehat{\mathbf{G}}$ is an n × m predicted genetically regulated expression (GReX) matrix, $\boldsymbol{\alpha}$ is an m × 1 vector of corresponding gene effects, $\boldsymbol{\beta}$ is an n × 1 vector of covariant, λ is a scalar of the corresponding covariant effect, and $\boldsymbol{\varepsilon}$ is a random noise term with $\mathrm{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$ and $\mathrm{V}[\boldsymbol{\varepsilon}] = \mathbf{I}_{\mathrm{n}}\sigma_{\mathrm{e}}^2$. Same as FABIO, FOCUS relies on the predicted GReX in the form of $\widehat{\mathbf{G}} = \mathbf{X}\boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is a p × m weight matrix representing the SNP effects on gene expression estimated from a separate expression mapping study. Because of the linear modeling framework, the above model can also be translated into a model based on marginal GWAS summary statistics,

in the form of the estimated SNP effects on gene expression, SNP LD matrix, and SNPs Z-scores from marginal GWAS analysis.

A.4 Model specifications for FOGS

FOGS examines genomic loci one at a time. For a particular locus with m genes, FOGS further examines one gene at a time. For a gene of focus, FOGS first divides the cis-SNPs in the locus into two sets: a set of p SNPs with non-zero effects on the expression of the gene of focus and another set of q SNPs with non-zero effects on the expression of other genes in the locus. Afterwards, FOGS relies on a multiple linear regression model in the form of $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{V}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$, where $\mathbf{y}$ is a centered phenotype vector for n individuals, $\mathbf{X}$ is a centered n × p genotype matrix at p SNPs with non-zero effects on expression of the target gene, $\mathbf{V}$ is a centered n × q genotype matrix at q SNPs with non-zero effects on expression of the remaining m-1 genes, and $\boldsymbol{\beta}$ is a p × 1 vector of the SNP effects on the target gene expression while $\boldsymbol{\alpha}$ is a q × 1 vector of the SNP effects on the remaining genes. FOGS tests the causal effect between the phenotype and the target gene under the null hypothesis H_0: $\boldsymbol{\beta} = \mathbf{0}$. For the target gene with p SNPs, FOGS further uses ridge regression to estimate the p conditional Z scores for those SNPs by converting marginal effects obtained from GWAS summary statistics to joint effects. Afterwards, to aggregate Z scores of multiple SNPs from the same gene, FOGS applies a weighted adaptive sum of powered score (SPU) test with eQTL-derived weights, and generates the p-value for each gene as the evidence of causality. Since the underlying truth for the effect sizes and directions of SNPs is unknown, FOGS constructs a class of SPU tests using a positive integer γ as the tuning parameter. As γ increases, $\mathrm{SPU}(\gamma)$ becomes more powered to detect signals where a sparse set of SNPs display effects on phenotype. FOGS further selects the optimal value of γ, where $\mathrm{SPU}(\gamma)$ achieves the smallest p-value. FOGS uses $\gamma \in \{1,2,3,4,5,6\}$ as the default setting; thus FOGS is expected to be more powerful when the true SNP effects on the trait are not sparse. Overall, FOGS uses marginal GWAS summary statistics, in the form of the estimated SNP effects on gene expression, SNP LD matrix, and SNPs Z-scores from marginal GWAS analysis.

Appendix B: Chapter 3 Supplementary Text

B.1 Variational inference overview

As illustrated in the main text, we aim to minimize the Kullback-Leibler (KL) divergence between the approximate posterior and the actual posterior. Here we denote all the parameters in our model as $\boldsymbol{\Theta} = (\mu, \beta_l, \boldsymbol{\gamma}_l, \boldsymbol{\pi_G}, \boldsymbol{\pi}_{\boldsymbol{T}|*}, \boldsymbol{\alpha}, \sigma_\alpha^2, \sigma_\beta^2)$. Based on our model definitions in main text Equations (3.2)-(3.5), minimizing the KL divergence is equivalent to maximizing the evidence of lower bound (ELBO) defined as:

$$ELBO = E_q[\log p(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\Theta})] - E_q[\log q(\boldsymbol{z}, \boldsymbol{\Theta})]$$

where $p(\cdot)$ denotes the prior probability and $q(\cdot)$ denotes the approximation posterior.

For the intercept μ, we assume a normal distribution: $\mu \sim N(0, \sigma_\mu^2)$, where $\sigma_\mu^2 \to \infty$ as a non-informative prior following the common practice in Bayesian approaches, which is equivalent to treating it as a fixed effect. For the prior of the scalar effect-size parameter β_l, we assume a normal distribution: $N(0, \sigma_\beta^2)$, for $l = 1,2,\dots,L$, and σ_β^2 has a weakly informative inverse-gamma prior: $\sigma_\beta^2 \sim InverseGamma(1,1)$. For the prior of $\boldsymbol{\gamma}_l$, which is a p-vector, we assume it to follow a multinomial distribution: $\boldsymbol{\gamma}_l \sim Mult(1, \boldsymbol{\pi})$, where $\boldsymbol{\pi}$ is a p-vector of prior probabilities that remain the same for all L effects. We further denote $\boldsymbol{\gamma}_{lj}$ as the causal status of j'th gene-tissue pair taking value 0 or 1, and its corresponding prior probability π_j is determined by hyperparameters $\boldsymbol{\pi_G}$ and $\boldsymbol{\pi}_{\boldsymbol{T}|*}$ following the Equation (5) in the main text. We assume symmetric Dirichlet priors for both of the hyperparameters: $\boldsymbol{\pi_G} \sim Dirichlet(1,\dots,1)$ and $\boldsymbol{\pi}_{\boldsymbol{T}|*} \sim Dirichlet(1,\dots,1)$, to represent an uninformative starting belief that all genes, and all tissues within a gene, are equally likely to be selected. Therefore, the prior probability that each gene-tissue pair displays the l'th non-zero effect is equally to be $1/p$. For the prior of the SNP effect sizes $\boldsymbol{\alpha}$, we assume it to follow a multivariate normal distribution: $\boldsymbol{\alpha} \sim N(\mathbf{0}, \sigma_\alpha^2 \boldsymbol{I})$, and σ_α^2 also has a weakly informative inverse-gamma prior: $\sigma_\alpha^2 \sim InverseGamma(1,1)$.

With these prior specifications, we then apply an iterative algorithm to update the parameters, by identifying the optimal approximate posterior distribution $q^*(\cdot)$ to maximize the ELBO. The update for any single parameter (or variable) is proportional to the exponentiated expectation of the log joint probability, taken with respect to all other parameters and variables. We will share the derivation of the approximate posterior distributions and how we update the parameters in the following sections.

B.2 Update of latent variable *z*

We first update the latent variable vector $\boldsymbol{z}$ by finding the optimal approximate posterior distribution $q^*(z_i)$ for each individual's latent variable, given the current state of all other model parameters. Assuming $\log p(\boldsymbol{z}) = \prod_{i=1}^{n} p(z_i)$ following (Held & Holmes, 2006), we first isolate the terms in ELBO that are related to each z_i, and get:

$$\log q^*(z_i) \propto E_q[\log p(y_i|z_i) + \log p(z_i|\eta_i)]$$

Based on the Equation (3.2) in the main text, $p(y_i|z_i)$ is an indicator function:

$$p(y_i|z_i) = \begin{cases} 1, & if\ (y_i = 1\ and\ z_i \geq 0)\ or\ (y_i = 0\ and\ z_i < 0) \\ 0, & otherwise \end{cases}.$$

Therefore, it enforces the optimal approximate posterior distribution $q^*(z_i)$ to be truncated. Based on the Equation (3.3) in the main text, z_i follows a prior normal distribution: $z_i \sim N(\eta_i, 1)$, where $\eta_i = \mu + \widehat{\boldsymbol{G}}_i^{\,T}\boldsymbol{\beta} + \boldsymbol{X}_i^{\,T}\boldsymbol{\alpha}$, and $\log p(z_i|\eta_i) = -\frac{1}{2}\log 2\pi - \frac{1}{2}(z_i - \eta_i)^2$. Combining these results above, only when $(y_i = 1\ and\ z_i \geq 0)\ or\ (y_i = 0\ and\ z_i < 0)$, we have:

$$\log q^*(z_i) \propto E_q\left[\log 1 + (-\frac{1}{2}\log 2\pi - \frac{1}{2}(z_i - \eta_i)^2)\right]$$

$$\propto E_q\left[-\frac{1}{2}(z_i - \eta_i)^2\right] \propto -\frac{1}{2}({z_i}^2 - 2z_i E_q[\eta_i]) \propto -\frac{1}{2}\big(z_i - E_q[\eta_i]\big)^2$$

Therefore, the optimal posterior distribution of z_i conditional on $\boldsymbol{y}$ and other given parameter estimates is a truncated normal with mean $E_q[\eta_i] = E_q[\mu + \widehat{\boldsymbol{G}}_i^{\,T}\boldsymbol{\beta} + \boldsymbol{X}_i^{\,T}\boldsymbol{\alpha}]$:

$$z_i|y_i = 1, \mu, \boldsymbol{\beta}, \boldsymbol{\alpha} \sim N\big(E_q[\eta_i], 1\big), \qquad truncated\ to\ [0, +\infty)$$

$$z_i|y_i = 0, \mu, \boldsymbol{\beta}, \boldsymbol{\alpha} \sim N\big(E_q[\eta_i], 1\big), \qquad truncated\ to\ (-\infty, 0)$$

If $y_i = 1$, then $z_i \in [0, +\infty)$ and we can update $E_q[z_i]$ using the posterior mean as:

$$E_q[z_i] = E_q[\eta_i] + \frac{\phi(E_q[\eta_i])}{\Phi(E_q[\eta_i])}$$

if $y_i = 0$, then $z_i \in (-\infty, 0)$ and we can update $E_q[z_i]$ using the posterior mean as:

$$E_q[z_i] = E_q[\eta_i] - \frac{\phi(E_q[\eta_i])}{\Phi(-E_q[\eta_i])}$$

where $\boldsymbol{\phi}$ is the probability density function (PDF) of the standard normal distribution, and $\boldsymbol{\Phi}$ is the cumulative distribution function (CDF) of the standard normal distribution. These updates are performed for all $\boldsymbol{i = 1, \ldots, n}$, resulting in a new vector of expectations $\boldsymbol{E_q[z]}$, which is then used to update the other parameters in the model.

B.3 Update of μ

Next, we update those regression parameters that determine $\boldsymbol{z}$. Starting with the intercept μ, which is treated as a fixed effect in the model, we update the point estimate $\tilde{\mu}$ by finding the value that maximizes the ELBO, holding all other parameters fixed. We first isolate relevant terms from the ELBO, and get:

$$\tilde{\mu} = \arg\max_{\mu} E_q[\log p(\boldsymbol{z}|\mu, \boldsymbol{\beta}, \boldsymbol{\alpha})]$$

By minimizing the sum of squared errors between the observations and their expected values, then we have:

$$\tilde{\mu} = \arg\min_{\mu} E_q\left[\left\|\boldsymbol{z} - (\mu\mathbf{1} + \widehat{\boldsymbol{G}}\boldsymbol{\beta} + \boldsymbol{X}\boldsymbol{\alpha})\right\|^2\right]$$

We take the derivative with respect to μ and set it to zero, and get the final update rule as:

$$\tilde{\mu} = \frac{1}{n}\sum_{i=1}^{n} E_q\left[z_i - \widehat{\boldsymbol{G}}_i{}^T\boldsymbol{\beta} - \boldsymbol{X}_i{}^T\boldsymbol{\alpha}\right]$$

B.4 Update of α and σ_α^2

Given the SNP effect sizes $\boldsymbol{\alpha}$ has a multivariate normal prior: $\boldsymbol{\alpha} \sim N(\mathbf{0}, \sigma_\alpha^2\boldsymbol{I})$, we assume $q^*(\boldsymbol{\alpha}) = N(\boldsymbol{\mu_\alpha}, \boldsymbol{\Sigma_\alpha})$ due to the normal-normal conjugation. We first isolate those terms in ELBO that involve the vector $\boldsymbol{\alpha}$:

$$\begin{aligned}
\log q^*(\boldsymbol{\alpha}) &\propto E_q[\log p(\boldsymbol{\alpha}|\sigma_\alpha^2) + \log p(\boldsymbol{z}|\mu, \boldsymbol{\beta}, \boldsymbol{\alpha})] \\
&\propto -\frac{1}{2}E_q\left[\frac{1}{\sigma_\alpha^2}\right]\boldsymbol{\alpha}^T\boldsymbol{\alpha} + E_q\left[-\frac{1}{2}\left\|\left(\boldsymbol{z} - \mu\mathbf{1} - \widehat{\boldsymbol{G}}\boldsymbol{\beta}\right) - \boldsymbol{X}\boldsymbol{\alpha}\right\|^2\right] \\
&\propto E_q\left[\boldsymbol{z} - \mu\mathbf{1_n} - \widehat{\boldsymbol{G}}\boldsymbol{\beta}\right]^T\boldsymbol{X}\boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^T(\boldsymbol{X}^T\boldsymbol{X} + E_q\left[\frac{1}{\sigma_\alpha^2}\right]\boldsymbol{I_s})\boldsymbol{\alpha}
\end{aligned}$$

Then based on the log-kernel of a multivariate normal distribution, we have:

$$\boldsymbol{\Sigma_\alpha} = (\boldsymbol{X}^T\boldsymbol{X} + E_q\left[\frac{1}{\sigma_\alpha^2}\right]\boldsymbol{I_s})^{-1}$$

$$\boldsymbol{\mu_\alpha} = \boldsymbol{\Sigma_\alpha}(\boldsymbol{X}^T E_q[\boldsymbol{z} - \mu\boldsymbol{1_n} - \widehat{\boldsymbol{G}}\boldsymbol{\beta}])$$

Then we focus on the update of σ_α^2, given that we place an inverse-gamma prior on σ_α^2: $\sigma_\alpha^2 \sim InverseGamma(a_\alpha, b_\alpha)$, and initially $a_\alpha = b_\alpha = 1$. Again, we isolate those terms in the log joint probability that involve σ_α^2:

$$\log q^*(\sigma_\alpha^2) \propto \log p(\sigma_\alpha^2) + E_q[\log p(\boldsymbol{\alpha}|\sigma_\alpha^2)]$$

$$\propto -(a_\alpha + 1)\log\sigma_\alpha^2 - \frac{b_\alpha}{\sigma_\alpha^2} + E_q\left[-\frac{s}{2}\log 2\pi - \frac{s}{2}\log\sigma_\alpha^2 - \frac{1}{2\sigma_\alpha^2}\boldsymbol{\alpha}^T\boldsymbol{\alpha}\right] \propto$$

$$\propto -(a_\alpha + 1)\log\sigma_\alpha^2 - \frac{b_\alpha}{\sigma_\alpha^2} - \frac{s}{2}\log\sigma_\alpha^2 - \frac{1}{2\sigma_\alpha^2}E_q[\boldsymbol{\alpha}^T\boldsymbol{\alpha}]$$

$$= -\left(a_\alpha + \frac{s}{2} + 1\right)\log\sigma_\alpha^2 - \frac{1}{\sigma_\alpha^2}\left(b_\alpha + \frac{E_q[\boldsymbol{\alpha}^T\boldsymbol{\alpha}]}{2}\right)$$

Here, s is the total number of SNPs in the region of interest defined in the main text. Based on the log-kernel of an inverse-gamma distribution, the optimal posterior for σ_α^2 is also an inverse-gamma distribution: $q^*(\sigma_\alpha^2) = InverseGamma(\tilde{a}_\alpha, \tilde{b}_\alpha)$, with

$$\tilde{a}_\alpha = a_\alpha + \frac{s}{2}$$

$$\tilde{b}_\alpha = b_\alpha + \frac{E_q[\boldsymbol{\alpha}^T\boldsymbol{\alpha}]}{2}$$

where $E_q[\boldsymbol{\alpha}^T\boldsymbol{\alpha}] = trace(\boldsymbol{\Sigma_\alpha}) + \boldsymbol{\mu_\alpha}^T\boldsymbol{\mu_\alpha}$, based on the updated $\boldsymbol{\alpha}$ defined above.

B.5 Update of $\boldsymbol{\beta_l}$, $\boldsymbol{\sigma_\beta^2}$, and $\boldsymbol{\gamma_l}$

Now we move to the parameters for the Sum of Single Effects (SuSiE) prior placed on the gene-tissue effect sizes $\boldsymbol{\beta}$. As defined in the main text Equation (3.4), $\boldsymbol{\beta}$ is decomposed as a sum of L single-effect components. Each single-effect component is denoted as $\beta_l\boldsymbol{\gamma}_l$, where β_l is a shared scalar effect-size parameter following a normal prior $N(0, \sigma_\beta^2)$, for $l = 1,2,\ldots,L$; and $\boldsymbol{\gamma}_l$ is a p-vector of binary indicators to indicate which gene-tissue pair is selected by the l'th component. We assume a shared prior distribution $\boldsymbol{\gamma}_l \sim Multinomial(1, \boldsymbol{\pi})$, for $l = 1,2,\ldots,L$, allowing each single effect to select a specific column of $\widehat{\boldsymbol{G}}$. For each of the L components, we update its selection

probabilities with $q^*(\boldsymbol{\gamma}_l)$ and its effect size distribution with $q^*(\beta_l)$. After all L components have been updated, we update the shared effect size variance with $q^*(\sigma_\beta^2)$.

For each component l, we first consider a residual vector $\boldsymbol{r}_l$, which serves as the target signal for the l'th component to explain and is denoted as:

$$\boldsymbol{r}_l \coloneqq \boldsymbol{z} - \mu\mathbf{1}_{\boldsymbol{n}} - \boldsymbol{X}\boldsymbol{\alpha} - \sum_{l' \neq l} \beta_{l'}\boldsymbol{\gamma}_{l'}$$

Given the conjugate prior distribution $\boldsymbol{\gamma}_l \sim Multinomial(1, \boldsymbol{\pi})$, $q^*(\boldsymbol{\gamma}_l)$ is also a multinomial distribution with one trial, denoted as $q^*(\boldsymbol{\gamma}_l) = Multinomial(1, \tilde{\boldsymbol{\pi}}_l)$, and $\tilde{\boldsymbol{\pi}}_l = (\tilde{\pi}_{l1}, \ldots, \tilde{\pi}_{lp})$ for the total of p gene-tissue pairs. Then we have:

$$\log \tilde{\pi}_{lj} \propto E_q\left[\log p\left(\boldsymbol{z}|\mu, \boldsymbol{\alpha}, \beta_l, \boldsymbol{\gamma}_{lj} = 1\right)\right] + E_q\left[\log p\left(\gamma_{lj} = 1\right)\right]$$

Here, $E_q\left[\log p\left(\boldsymbol{z}|\mu, \boldsymbol{\alpha}, \beta_l, \gamma_{lj} = 1\right)\right]$ represents the log Bayes factor, for including the j'th gene-tissue pair in a single-variable regression model for the residual $\boldsymbol{r}_l$. It quantifies the evidence from the data in favor of a model where the j'th gene-tissue pair has an effect, versus a null model with only noise, and we denote it as $\log BF_j$. Here we denote the null model as $model_0: \boldsymbol{r}_l = \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{I}_n)$, and the alternative model as $model_j: \boldsymbol{r}_l = \widehat{\boldsymbol{G}}_{,j}\beta_l + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{I}_n)$, where $\widehat{\boldsymbol{G}}_{,j}$ is the j'th column of $\widehat{\boldsymbol{G}}$. We further denote that when $\boldsymbol{\gamma}_{lj} = 1$, the optimal posterior distribution for β_l as $q^*\left(\beta_l|\, \gamma_{lj} = 1\right) = N(\mu_{\beta_{l|j}}, \sigma^2_{\beta_{l|j}})$, given the normal-conjugate prior on β_l. We have:

$$\log q^*\left(\beta_l|\, \gamma_{lj} = 1\right) \propto \log p(\boldsymbol{r}_l|\beta_l) + \log p(\beta_l) \propto \left(-\frac{n}{2}\log 2\pi - \frac{1}{2}\left\|\boldsymbol{r}_l - \widehat{\boldsymbol{G}}_{,j}\beta_l\right\|^2\right) + \left(-\frac{\beta_l^2}{2\sigma_\beta^2}\right)$$

$$\propto \left(\widehat{\boldsymbol{G}}_{,j}^{\ T}\boldsymbol{r}_l\right)\beta_l - \frac{1}{2}\left(\widehat{\boldsymbol{G}}_{,j}^{\ T}\widehat{\boldsymbol{G}}_{,j} + \frac{1}{\sigma_\beta^2}\right)\beta_l^2$$

Based on the log-kernel of the normal distribution, we solve $\mu_{\beta_{l|j}}$ and $\sigma^2_{\beta_{l|j}}$ as:

$$\sigma^2_{\beta_{l|j}} = \left(\widehat{\boldsymbol{G}}_{,j}^{\ T}\widehat{\boldsymbol{G}}_{,j} + \frac{1}{\sigma_\beta^2}\right)^{-1}$$

$$\mu_{\beta_{l|j}} = \sigma^2_{\beta_{l|j}}\left(\widehat{\boldsymbol{G}}_{,j}^{\ T}\boldsymbol{r}_l\right)$$

Then we have the log Bayes factor as:

$$
\begin{aligned}
\log BF_j &= \log p\left(\boldsymbol{r}_l|model_j\right) - \log p(\boldsymbol{r}_l|model_0) \\
&= \log p\left(\boldsymbol{r}_l|\beta_l, model_j\right) + \log p\left(\beta_l|model_j\right) - \log p\left(\beta_l|\boldsymbol{r}_l, model_j\right) \\
&\quad - \log p(\boldsymbol{r}_l|model_0) \\
&= \left(-\frac{n}{2}\log 2\pi - \frac{1}{2}\left\|\boldsymbol{r}_l - \widehat{\boldsymbol{G}}_{,j}\mu_{\beta_{l|j}}\right\|^2\right) + \left(-\frac{1}{2}\log 2\pi\sigma_\beta^2 - \frac{\mu_{\beta_{l|j}}^2}{2\sigma_\beta^2}\right) \\
&\quad - \left(-\frac{1}{2}\log 2\pi\sigma_{\beta_{l|j}}^2\right) - \left(-\frac{n}{2}\log 2\pi - \frac{1}{2}\boldsymbol{r}_l{}^T\boldsymbol{r}_l\right) \\
&= \mu_{\beta_{l|j}}\widehat{\boldsymbol{G}}_{,j}{}^T\boldsymbol{r}_l - \frac{1}{2}\mu_{\beta_{l|j}}^2\widehat{\boldsymbol{G}}_{,j}{}^T\widehat{\boldsymbol{G}}_{,j} - \frac{1}{2}\log\sigma_\beta^2 - \frac{\mu_{\beta_{l|j}}^2}{2\sigma_\beta^2} + \frac{1}{2}\log\sigma_{\beta_{l|j}}^2 \\
&= \frac{\mu_{\beta_{l|j}}^2}{2\sigma_{\beta_{l|j}}^2} + \frac{1}{2}\log\sigma_{\beta_{l|j}}^2 - \frac{1}{2}\log\sigma_\beta^2
\end{aligned}
$$

For $E_q\left[\log p\left(\boldsymbol{\gamma}_{lj} = 1\right)\right]$, it depends on the hierarchical priors we introduced in the main text Equation (3.5). Following the same annotations in the main text, we assume that the *j*'th gene-tissue pair correspond to the specific pair of the *k*'th unique tissue and the *m*'th unique gene. Then we have:

$$E_q\left[\log p\left(\boldsymbol{\gamma}_{lj} = 1\right)\right] = E_q\left[\log p\left(\pi_{g_m}\right)\right] + E_q\left[\log p\left(\pi_{t_k|g_m}\right)\right]$$

We have Dirichlet priors denoted as:

$$\boldsymbol{\pi}_{\boldsymbol{G}} \sim Dirichlet(\delta_{g_1}, \dots, \delta_{g_M})$$

$$\boldsymbol{\pi}_{\boldsymbol{T}|g_m} \sim Dirichlet(\delta_{t_1|g_m}, \dots, \delta_{t_K|g_m})$$

And Dirichlet distribution is the conjugate prior distribution of the multinomial distribution, so both $q^*(\boldsymbol{\pi}_{\boldsymbol{G}})$ and $q^*(\boldsymbol{\pi}_{\boldsymbol{T}|*})$ are also Dirichlet distributions, and we denote them as:

$$q^*(\boldsymbol{\pi}_{\boldsymbol{G}}) = Dirichlet(\tilde{\delta}_{g_1}, \dots, \tilde{\delta}_{g_M})$$

$$q^*\left(\boldsymbol{\pi}_{\boldsymbol{T}|g_m}\right) = Dirichlet(\tilde{\delta}_{t_1|g_m}, \dots, \tilde{\delta}_{t_K|g_m})$$

It is also known that if a probability vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$ follows a Dirichlet distribution: $\boldsymbol{\theta} \sim Dirichlet(\delta_1, \dots, \delta_N)$, then the expectation of the logarithm of its component θ_i ($i \in \{1,2, \dots, N\}$) can be calculated as:

$$E[\log\theta_i] = \psi(\delta_i) - \psi\left(\sum_{i=1}^{N}\delta_i\right)$$

where ψ is the digamma function. Then we have:

$$E_q[\log p(\gamma_{lj}=1)] = E_q[\log p(\pi_{g_m})] + E_q[\log p(\pi_{t_k|g_m})]$$
$$= \psi(\tilde{\delta}_{g_m}) - \psi\left(\sum_{i=1}^{M}\tilde{\delta}_{g_i}\right) + \psi(\tilde{\delta}_{t_k|g_m}) - \psi\left(\sum_{j=1}^{K}\tilde{\delta}_{t_j|g_m}\right)$$

Combining the calculation of $E_q[\log p(\boldsymbol{z}|\mu, \boldsymbol{\alpha}, \beta_l, \gamma_{lj}=1)]$ and $E_q[\log p(\gamma_{lj}=1)]$, now we have:

$$\log \tilde{\boldsymbol{\pi}}_{lj} \propto \log BF_j + E_q[\log p(\gamma_{lj}=1)]$$
$$= \frac{\mu^2_{\beta_{l|j}}}{2\sigma^2_{\beta_{l|j}}} + \frac{1}{2}\log\sigma^2_{\beta_{l|j}} - \frac{1}{2}\log\sigma^2_{\beta} + \psi(\tilde{\delta}_{g_m}) - \psi\left(\sum_{i=1}^{M}\tilde{\delta}_{g_i}\right) + \psi(\tilde{\delta}_{t_k|g_m})$$
$$- \psi\left(\sum_{j=1}^{K}\tilde{\delta}_{t_j|g_m}\right)$$

After calculating $\log \tilde{\boldsymbol{\pi}}_{lj}$ for all $j = 1, \dots, p$, we normalize them to get the final probabilities $\tilde{\boldsymbol{\pi}}_l = (\tilde{\pi}_{l1}, \dots, \tilde{\pi}_{lp})$, as the update of $q^*(\boldsymbol{\gamma}_l)$.

Then we update the approximate posterior distribution for the effect size β_l with $q^*(\beta_l)$. Given the prior: $\beta_l \sim N(0, \sigma^2_\beta)$ and the normal-normal conjugation, we assume $q^*(\beta_l) = N(\tilde{\mu}_{\beta_l}, \tilde{\sigma}^2_{\beta_l})$. As illustrated earlier, when the *j*'th gene-tissue pair is selected as the *l*'th effect component, we have $q^*(\beta_l|\,\boldsymbol{\gamma}_{lj}=1) = N(\mu_{\beta_{l|j}}, \sigma^2_{\beta_{l|j}})$, so the posterior $q^*(\beta_l)$ is the average of all these conditional posteriors, weighted by the probability of each selection $\tilde{\pi}_{lj}$ for $j = 1, \dots, p$. Therefore,

$$\tilde{\mu}_{\beta_l} = \sum_{j=1}^{p}\tilde{\pi}_{lj}\mu_{\beta_{l|j}}$$

$$\tilde{\sigma}^2_{\beta_l} = \sum_{j=1}^{p}\tilde{\pi}_{lj}\sigma^2_{\beta_{l|j}} + \left(\sum_{j=1}^{p}\tilde{\pi}_{lj}\mu^2_{\beta_{l|j}} - \left(\sum_{j=1}^{p}\tilde{\pi}_{lj}\mu_{\beta_{l|j}}\right)^2\right) = \sum_{j=1}^{p}\tilde{\pi}_{lj}\left(\mu^2_{\beta_{l|j}} + \sigma^2_{\beta_{l|j}}\right) - \tilde{\mu}^2_{\beta_l}$$

After we updated all the parameters for the *L* individual components, we update the effect size variance σ^2_β that is shared among all of them. We assume an inverse-gamma prior: $\sigma^2_\beta \sim InverseGamma(a_\beta, b_\beta)$, and initially $a_\beta = b_\beta = 1$. We isolate those terms in the log joint probability that involve σ^2_β:

$$\log q^*(\sigma_\beta^2) \propto \log p(\sigma_\beta^2) + E_q\left[\sum_{l=1}^{L} \log p(\beta_l|\sigma_\alpha^2)\right]$$

$$\propto -(a_\beta + 1)\log\sigma_\beta^2 - \frac{b_\beta}{\sigma_\beta^2} + \sum_{l=1}^{L} E_q\left[-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma_\beta^2 - \frac{\beta_l^2}{2\sigma_\beta^2}\right] \propto$$

$$\propto -(a_\beta + 1)\log\sigma_\beta^2 - \frac{b_\beta}{\sigma_\beta^2} - \frac{L}{2}\log\sigma_\beta^2 - \frac{1}{2\sigma_\beta^2}\sum_{l=1}^{L} E_q[\beta_l^2]$$

$$= -\left(a_\beta + \frac{L}{2} + 1\right)\log\sigma_\beta^2 - \frac{1}{\sigma_\beta^2}\left(b_\beta + \frac{\sum_{l=1}^{L} E_q[\beta_l^2]}{2}\right)$$

Based on the log-kernel of an inverse-gamma distribution, the optimal posterior for σ_β^2 is also an inverse-gamma distribution: $q^*(\sigma_\beta^2) = InverseGamma(\tilde{a}_\beta, \tilde{b}_\beta)$, with

$$\tilde{a}_\beta = a_\beta + \frac{L}{2}$$

$$\tilde{b}_\beta = b_\beta + \frac{\sum_{l=1}^{L} E_q[\beta_l^2]}{2}$$

B.6 Update of $\boldsymbol{\pi_G}$ and $\boldsymbol{\pi_{T|*}}$

Using the same annotations as the preceding section, we have the Dirichlet priors denoted as:

$$\boldsymbol{\pi_G} \coloneqq (\pi_{g_1}, \dots, \pi_{g_M}) \sim Dirichlet(\delta_{g_1}, \dots, \delta_{g_M})$$

$$\boldsymbol{\pi_{T|g_m}} \coloneqq (\pi_{t_1|g_m}, \dots, \pi_{t_K|g_m}) \sim Dirichlet\left(\delta_{t_1|g_m}, \dots, \delta_{t_K|g_m}\right),\ \ m = 1, \dots, M$$

We first isolate those terms in ELBO that involve $\boldsymbol{\pi_G}$:

$$\log q^*(\boldsymbol{\pi_G}) \propto \log p(\boldsymbol{\pi_G}) + E_q\left[\sum_{l=1}^{L} \log p\left(\boldsymbol{\gamma}_l|\boldsymbol{\pi_G}, \boldsymbol{\pi_{T|*}}\right)\right]$$

$$\propto \left(\sum_{m=1}^{M}\left(\delta_{g_m} - 1\right)\log\pi_{g_m}\right) + \left(\sum_{l=1}^{L}\sum_{j=1}^{p} q\left(\gamma_{lj} = 1\right)\log(\boldsymbol{\pi_G}|\gamma_{lj} = 1)\right)$$

For the term $\sum_{\boldsymbol{l=1}}^{\boldsymbol{L}}\sum_{\boldsymbol{j=1}}^{\boldsymbol{p}} \boldsymbol{q}\left(\boldsymbol{\gamma_{lj}} = \boldsymbol{1}\right)\boldsymbol{log}(\boldsymbol{\pi_G}|\boldsymbol{\gamma_{lj}} = \boldsymbol{1})$, we re-group the summation by gene since each gene only corresponds to partial of the total p gene-tissue pairs, then we have:

$$\log q^*(\boldsymbol{\pi_G}) \propto \left(\sum_{m=1}^{M}\left(\delta_{g_m}-1\right)\log \pi_{g_m}\right)+\left(\sum_{m=1}^{M}\left(\left(\sum_{l=1}^{L}\sum_{j\in g_m\, pairs}\tilde{\pi}_{lj}\right)\log \pi_{g_m}\right)\right)$$

$$=\sum_{m=1}^{M}\left(\delta_{g_m}+\left(\sum_{l=1}^{L}\sum_{j\in g_m\, pairs}\tilde{\pi}_{lj}\right)-1\right)\log \pi_{g_m}$$

where $g_m\, pairs$ represents the indices of all the gene-tissue pairs that contains gene g_m. And based on the log-kernel of the Dirichlet distribution, we update $\boldsymbol{\pi_G}$ with:

$$q^*(\boldsymbol{\pi_G}) = Dirichlet(\tilde{\delta}_{g_1},\dots,\tilde{\delta}_{g_M})$$

$$\tilde{\delta}_{g_m}=\delta_{g_m}+\left(\sum_{l=1}^{L}\sum_{j\in g_m\, pairs}\tilde{\pi}_{lj}\right),\ \ m=1,\dots,M$$

Then we isolate those terms in ELBO that involve $\boldsymbol{\pi_{T|*}}$, and take gene g_m as an example:

$$\log q^*\left(\boldsymbol{\pi_{T|g_m}}\right) \propto \log p\left(\boldsymbol{\pi_{T|g_m}}\right)+E_q\left[\sum_{l=1}^{L}\log p\left(\boldsymbol{\gamma}_l|\boldsymbol{\pi_G},\boldsymbol{\pi_{T|*}}\right)\right]$$

$$\propto \left(\sum_{k=1}^{K}\left(\delta_{t_k|g_m}-1\right)\log \pi_{t_k|g_m}\right)$$

$$+\left(\sum_{l=1}^{L}\sum_{j=1}^{p} q\left(\gamma_{lj}=1\right)\log(\boldsymbol{\pi_{T|g_m}}|\gamma_{lj}=1)\right)$$

For the term $\sum_{l=1}^{L}\sum_{j=1}^{p} q\left(\gamma_{lj}=1\right)\log(\boldsymbol{\pi_{T|g_m}}|\gamma_{lj}=1)$ here, similarly we re-group the summation by tissue, then we have:

$$\log q^*\left(\boldsymbol{\pi_{T|g_m}}\right) \propto \left(\sum_{k=1}^{K}\left(\delta_{t_k|g_m}-1\right)\log \pi_{t_k|g_m}\right)$$

$$+\left(\sum_{k=1}^{K}\left(\sum_{l=1}^{L}\sum_{j=pair\,\{g_m,t_k\}}\tilde{\pi}_{lj}\right)\log \pi_{t_k|g_m}\right)$$

$$=\sum_{k=1}^{K}\left(\delta_{t_k|g_m}+\left(\sum_{l=1}^{L}\sum_{j=index\,\{g_m,t_k\}}\tilde{\pi}_{lj}\right)-1\right)\log \pi_{t_k|g_m}$$

where $index\ \{g_m,t_k\}$ represents the index of the specific gene-tissue pair $\{g_m,t_k\}$. And based on the log-kernel of the Dirichlet distribution, we update $\boldsymbol{\pi_{T|g_m}}$ with:

$$q^*\left(\boldsymbol{\pi_{T|g_m}}\right) = Dirichlet(\tilde{\delta}_{t_1|g_m},\dots,\tilde{\delta}_{t_K|g_m})$$

$$\tilde{\delta}_{t_k|g_m} = \delta_{t_k|g_m} + \left(\sum_{l=1}^{L} \sum_{j=index\{g_m,t_k\}} \tilde{\pi}_{lj} \right), \quad k = 1, \ldots, K$$

B.7 Summary of variational inference

In summary, for each effect from 1 to L, we update the optimal posterior distributions for the parameters by maximizing the corresponding ELBO. The update steps above can be summarized in the form of Iterative Bayesian stepwise selection (IBSS) algorithm:

Iterative Bayesian stepwise selection (IBSS) Algorithm

Required input #1: $\widehat{\boldsymbol{G}}, \boldsymbol{X}, \boldsymbol{y}$

Required input #2: Number of the non-zero effects L, initial estimates of hyperparameters $\sigma_\alpha^2, \sigma_\beta^2, \delta_{g_1}, \ldots, \delta_{g_M}$, and $\delta_{t_1|g_m}, \ldots, \delta_{t_K|g_m}$ for $m = 1, \ldots, M$

1: Repeat

2: update latent variable vector $\boldsymbol{z}$

3: update regression parameters that determine $\boldsymbol{z}$: μ, $\boldsymbol{\alpha}$, and σ_α^2

4: for l in 1: L do

5: update $\boldsymbol{\gamma}_l$ based on the current $\boldsymbol{\pi}_{\boldsymbol{G}}$ and $\boldsymbol{\pi}_{\boldsymbol{T}|*}$

6: update β_l

7: end for

8: update σ_β^2 that is shared among all L components

9: update $\boldsymbol{\pi}_{\boldsymbol{G}}$ and $\boldsymbol{\pi}_{\boldsymbol{T}|*}$ based on the posterior distribution of the current $\boldsymbol{\gamma}_l$

10: Loop over previous steps until the convergence, where the change of ELBO between consecutive iterations is below 0.001

Appendix C: Chapter 4 Supplementary Text

C.1 Cell culture

Keratinocytes were isolated under the IRB-approved protocol #STU0000944, and grown in 12 well plate in 154 CF medium (Thermo Fisher #M154CF500) with human keratinocyte growth supplement (Thermo Fisher #S0015). Keratinocytes were grown to confluency at which time the complete medium (with supplements) was replaced by basal 154 CF medium (without supplements). After 8 hours of stimulation cells were harvested and RNA was isolated using RNeasy Plus Mini kit (Qiagen # 74136). RNA was analyzed by RNA Nano Chips (Agilent Technologies) and sequenced.

C.2 Bayesian Sparse Linear Mixed Model (BSLMM)

BSLMM predicts the GReX using a linear mixed model with one random effect term and sparsity-inducing priors on the regression coefficients:

$$\boldsymbol{y} = 1_n\mu + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{u} + \boldsymbol{\varepsilon}, \tag{C.1}$$

$$\beta_i \sim \pi \mathrm{N}(0, \sigma^2) + (1-\pi)\delta_0 \tag{C.2}$$

Here $\boldsymbol{y}$ is an n-vector of the expression level of a gene (after the PEER analysis) measured on n individuals in the eQTL reference study, $\mathbf{X}$ is an $n \times p$ matrix of genotypes measured on the same individuals at p genetic markers, $\boldsymbol{\beta}$ is a p-vector of genetic marker effects, 1_n is an n-vector of 1's, μ is a scalar representing the mean of gene expression, $\boldsymbol{u}$ is an n-vector of random effects, and $\boldsymbol{\varepsilon}$ is an n-vector of error terms. β_i is the i-th element of $\boldsymbol{\beta}$, which follows a point-normal distribution with π representing the proportion of non-zero $\boldsymbol{\beta}$ and δ_0, denoting a point mass at zero. By placing a uniform prior: $\log(\pi) \sim U(\log\left(\frac{1}{p}\right), \log(1))$, BSLMM reflects the uncertainty in π. In this study, we extracted cis-SNPs that resided within 100 kb of each gene's transcription start site (TSS) and that are present in the skin GWAS datasets. The extracted cis-SNPs are the p genetic markers in the model.

C.3 Evaluation of the BSLMM GReX models

The proportion of variance in phenotypes explained (PVE) is used to evaluate the performance of the GreX models generated using BSLMM. We observed that the PVEs were consistent in all cytokine stimulations, with the median ranging from 10.7% to 11.0%, while in the unstimulated condition (NHEK), the median PVE was slightly larger (12.3%). The PVE estimates in our models are comparable with those estimated from a previous study on the genetic architecture of gene expression in lymphoblastoid cell lines, which showed the median PVE to be 14.7% across eQTL target genes (Shang et al., 2020). Notably, while there were no distinct differences in PVE between protein-coding versus non-protein-coding genes across cytokine stimulations, we observed slightly higher PVE (15.3% in protein-coding vs. 13.8% non-protein-coding, $p = 0.02$) in unstimulated controls (Figure S4.6a).

C.4 TWAS analysis

TWAS was first conducted in each condition for each disorder using S-PrediXcan, which leverages information from GWAS summary statistics and the SNP weights from the BSLMM GReX models:

$$Z_g \approx \sum_{i \in \text{Model}_g} \beta_{ig} \frac{\hat{\sigma}_i}{\hat{\sigma}_g} \frac{\hat{\alpha}_i}{se(\hat{\alpha}_i)} \tag{C.3}$$

Here Z_g is the Z-score (Wald statistic) of the association between predicted GReX of gene g and a phenotype, β_{ig} is the weight of the *i*-th SNP in the GReX model of gene g predicted by BSLMM (the *i*-th element of $\boldsymbol{\beta}$ in equation (C.1)), $\hat{\alpha}_i$ is the GWAS regression coefficients for the *i*-th SNP, $\text{se}(\hat{\alpha}_i)$ is the standard error of $\hat{\alpha}_i$, $\hat{\sigma}_i$ is the estimated allele frequency variance of the *i*-th SNP, and $\hat{\sigma}_g$ is the estimated variance of the predicted GReX of gene g.

Then for each disorder, we integrated single-condition TWAS results across all eight cytokine conditions for each gene and conducted a multi-condition TWAS using S-MultiXcan, which jointly models the phenotype through a linear regression on predicted GReX in *k* conditions (*k*=8 in our study):

$$\boldsymbol{h} = \mathbf{T}\boldsymbol{g} + \boldsymbol{e} \tag{C.4}$$

Here $\boldsymbol{h}$ is a centered *m*-vector of phenotypes for the *m* individuals in GWAS, $\mathbf{T}$ is an $m \times k$ matrix where the *j*-th column contains the values of standardized predicted gene expression in the *j*-th condition, $\boldsymbol{g}$ is a *k*-vector with the *j*-th element representing the effect size of the *j*-th condition,

and $\boldsymbol{e}$ is the error term with variance σ_e^2. To quantify the significance of the inferred multi-condition gene-level association using GWAS summary statistics, S-MultiXcan uses the fact that the estimated $\boldsymbol{g}$ in equation (C.4) approximately follows a multivariate normal distribution: $\hat{\boldsymbol{g}} \sim \text{MVN}(\boldsymbol{g}, \sigma_e^2(\mathbf{T}^t\mathbf{T})^{-1})$. Then, under the null hypothesis of no association, it follows that $\hat{\boldsymbol{g}}^t \frac{\mathbf{T}^t\mathbf{T}}{\sigma_e^2}\hat{\boldsymbol{g}} \sim \chi_k^2$. Further replacing $\hat{\boldsymbol{g}}$ with estimates from the marginal regression, it shows that $\hat{\boldsymbol{g}}^t \frac{\mathbf{T}^t\mathbf{T}}{\sigma_e^2}\hat{\boldsymbol{g}} \approx \hat{\boldsymbol{Z}}^t Cor(\mathbf{T})^{-1}\hat{\boldsymbol{Z}}$, where $Cor(\mathbf{T})$ is the autocorrelation of $\mathbf{T}$, and $\hat{\boldsymbol{Z}}$ is a k-vector with the j-th element being the marginal Z-score calculated from equation (C.3) in the j-th condition. Both single-condition and multi-condition TWAS provide p-value as the evidence of association between the gene and the phenotype.

C.5 Cytokine-specific associations

In the multi-condition TWAS, the number of independent components of variation kept among the analyses are recorded for each gene and treated as the number of synthetic independent conditions (SICs) among the seven cytokine stimulations and unstimulated control. The SICs provide surrogation for the maximum number of independent genetic regulatory circuits among the different cytokine stimulations (Figure S4.6b). The results are consistent among the seven inflammatory skin diseases and suggest that most of the TWAS significant genes are shared in up to five independent signaling pathways among the seven cytokine stimulations we have examined.

The number and proportion of the multi-condition TWAS genes that have the most significant association with each disease for each cytokine condition are shown in Table S4.4. The highest proportion of significant TWAS genes was observed in IL-13-stimulation for acne (15%) and AD (15%), and IFNγ (15%) and IL-17A+TNFα (14%) for psoriasis. In the associations with AA, SSc, SLE, and vitiligo, the highest proportions (15%, 17%, 16%, and 17%, respectively) were observed in the IFNα-stimulated condition.

Bibliography

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, *88*(422), 669-679.

Banafea, G. H., Bakhashab, S., Alshaibi, H. F., Natesan Pushparaj, P., & Rasool, M. (2022). The role of human mast cells in allergy and asthma. *Bioengineered*, *13*(3), 7049-7064.

Bangert, C., Brunner, P. M., & Stingl, G. (2011). Immune functions of the skin. *Clinics in dermatology*, *29*(4), 360-376.

Barabási, A.-L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, *12*(1), 56-68.

Barbeira, A. N., Dickinson, S. P., Bonazzola, R., Zheng, J., Wheeler, H. E., Torres, J. M., Torstenson, E. S., Shah, K. P., Garcia, T., & Edwards, T. L. (2018). Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nature communications*, *9*(1), 1825.

Barbeira, A. N., Pividori, M., Zheng, J., Wheeler, H. E., Nicolae, D. L., & Im, H. K. (2019). Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS genetics*, *15*(1), e1007889.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289-300.

Benner, C., Havulinna, A. S., Järvelin, M.-R., Salomaa, V., Ripatti, S., & Pirinen, M. (2017). Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. *The American Journal of Human Genetics*, *101*(4), 539-551.

Bentham, J., Morris, D. L., Cunninghame Graham, D. S., Pinder, C. L., Tombleson, P., Behrens, T. W., Martín, J., Fairfax, B. P., Knight, J. C., & Chen, L. (2015). Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nature genetics*, *47*(12), 1457-1464.

Berisa, T., & Pickrell, J. K. (2015). Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics*, *32*(2), 283.

Berisa, T., & Pickrell, J. K. (2016). Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics*, *32*(2), 283.

Betz, R. C., Petukhova, L., Ripke, S., Huang, H., Menelaou, A., Redler, S., Becker, T., Heilmann, S., Yamany, T., & Duvic, M. (2015). Genome-wide meta-analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci. *Nature communications*, *6*(1), 5966.

Bieber, T. (2023). Disease modification in inflammatory skin disorders: opportunities and challenges. *Nature Reviews Drug Discovery*, *22*(8), 662-680.

Boughton, A. P., Welch, R. P., Flickinger, M., VandeHaar, P., Taliun, D., Abecasis, G. R., & Boehnke, M. (2021). LocusZoom. js: interactive and embeddable visualization of genetic association study results. *Bioinformatics*, *37*(18), 3017-3018.

Boult, J., Roberts, K., Brookes, M. J., Hughes, S., Bury, J. P., Cross, S. S., Anderson, G. J., Spychal, R., Iqbal, T., & Tselepis, C. (2008). Overexpression of cellular iron import proteins is associated with malignant progression of esophageal adenocarcinoma. *Clinical Cancer Research*, *14*(2), 379-387.

Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., Daly, M. J., Price, A. L., & Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics*, *47*(3), 291-295.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., & O'Connell, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203-209.

Cao, T., Shao, S., Fang, H., Li, B., & Wang, G. (2019). Role of regulatory immune cells and molecules in autoimmune bullous dermatoses. *Frontiers in Immunology*, *10*, 1746.

Carithers, L. J., & Moore, H. M. (2015). The genotype-tissue expression (GTEx) project. *Biopreservation and biobanking*, *13*(5), 307.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, *4*(1), s13742-13015-10047-13748.

Chen, C.-J., Tseng, C.-C., Yen, J.-H., Chang, J.-G., Chou, W.-C., Chu, H.-W., Chang, S.-J., & Liao, W.-T. (2018). ABCG2 contributes to the development of gout and hyperuricemia in a genome-wide association study. *Scientific reports*, *8*(1), 3137.

Cleophas, M., Joosten, L., Stamp, L. K., Dalbeth, N., Woodward, O. M., & Merriman, T. R. (2017). ABCG2 polymorphisms in gout: insights into disease susceptibility and treatment approaches. *Pharmacogenomics and personalized medicine*, 129-142.

Common, J., Barker, J., & van Steensel, M. A. (2019). What does acne genetics teach us about disease pathogenesis? *British Journal of Dermatology*, *181*(4), 665-676.

Consortium, G. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, *369*(6509), 1318-1330.

Consortium, G. P. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*(7422), 56.

Consortium, G. P. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68.

Corren, J. (2013). Role of interleukin-13 in asthma. *Current allergy and asthma reports*, *13*(5), 415-420.

Czado, C., & Munk, A. (2000). Noncanonical links in generalized linear models–when is the effort justified? *Journal of statistical planning and inference*, *87*(2), 317-345.

Dale, B. L., & Madhur, M. S. (2016). Linking inflammation and hypertension via LNK/SH2B3. *Current opinion in nephrology and hypertension*, *25*(2), 87-93.

Dand, N., Stuart, P. E., Bowes, J., Ellinghaus, D., Nititham, J., Saklatvala, J. R., Teder-Laving, M., Thomas, L. F., Traks, T., & Uebe, S. (2023). GWAS meta-analysis of psoriasis identifies new susceptibility alleles impacting disease mechanisms and therapeutic targets. *medRxiv*.

Das, D., Akhtar, S., Kurra, S., Gupta, S., & Sharma, A. (2019). Emerging role of immune cell network in autoimmune skin disorders: an update on pemphigus, vitiligo and psoriasis. *Cytokine & Growth Factor Reviews*, *45*, 35-44.

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., Vrieze, S. I., Chew, E. Y., Levy, S., & McGue, M. (2016). Next-generation genotype imputation service and methods. *Nature genetics*, *48*(10), 1284-1287.

Devallière, J., & Charreau, B. (2011). The adaptor Lnk (SH2B3): an emerging regulator in vascular cells and a link between immune and inflammatory signaling. *Biochemical pharmacology*, *82*(10), 1391-1402.

Dries, R., Zhu, Q., Dong, R., Eng, C.-H. L., Li, H., Liu, K., Fu, Y., Zhao, T., Sarkar, A., & Bao, F. (2021). Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome biology*, *22*, 1-31.

Eckenstaler, R., & Benndorf, R. A. (2021). The role of ABCG2 in the pathogenesis of primary hyperuricemia and gout—An update. *International journal of molecular sciences*, *22*(13), 6678.

Efron, B. (2007). Size, power and false discovery rates.

Elder, J. T. (2009). Genome-wide association scan yields new insights into the immunopathogenesis of psoriasis. *Genes & Immunity*, *10*(3), 201-209.

Fisher, R. A. (1970). Statistical methods for research workers. In *Breakthroughs in statistics: Methodology and distribution* (pp. 66-70). Springer.

Gamazon, E. R., Wheeler, H. E., Shah, K. P., Mozaffari, S. V., Aquino-Michaels, K., Carroll, R. J., Eyler, A. E., Denny, J. C., Consortium, G., & Nicolae, D. L. (2015). A gene-based association method for mapping traits using reference transcriptome data. *Nature genetics*, *47*(9), 1091-1098.

Gao, B., & Zhou, X. (2024). MESuSiE enables scalable and powerful multi-ancestry fine-mapping of causal variants in genome-wide association studies. *Nature genetics*, *56*(1), 170-179.

Ge, X., Frank-Bertoncelj, M., Klein, K., McGovern, A., Kuret, T., Houtman, M., Burja, B., Micheroli, R., Shi, C., & Marks, M. (2021). Functional genomics atlas of synovial fibroblasts defining rheumatoid arthritis heritability. *Genome biology*, *22*, 1-39.

George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American statistical Association*, *88*(423), 881-889.

Guan, Y., & Krone, S. M. (2007). Small-world MCMC and convergence to multi-modal distributions: From slow mixing to fast mixing.

Guan, Y., & Stephens, M. (2011). Bayesian variable selection regression for genome-wide association studies and other large-scale problems.

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W., Jansen, R., De Geus, E. J., Boomsma, D. I., & Wright, F. A. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nature genetics*, *48*(3), 245-252.

Gusev, A., Mancuso, N., Won, H., Kousi, M., Finucane, H. K., Reshef, Y., Song, L., Safi, A., Consortium, S. W. G. o. t. P. G., & McCarroll, S. (2018). Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nature genetics*, *50*(4), 538-548.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., & Zager, M. (2021). Integrated analysis of multimodal single-cell data. *Cell*, *184*(13), 3573-3587. e3529.

Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B. L., Barrell, D., Zadissa, A., & Searle, S. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome research*, *22*(9), 1760-1774.

Hattori, A., & Tsujimoto, M. (2013). Endoplasmic reticulum aminopeptidases: biochemistry, physiology and pathology. *The journal of biochemistry*, *154*(3), 219-228.

Hautakangas, H., Winsvold, B. S., Ruotsalainen, S. E., Bjornsdottir, G., Harder, A. V., Kogelman, L. J., Thomas, L. F., Noordam, R., Benner, C., & Gormley, P. (2022).

Genome-wide analysis of 102,084 migraine cases identifies 123 risk loci and subtype-specific risk alleles. *Nature genetics*, *54*(2), 152-160.

Held, L., & Holmes, C. C. (2006). Bayesian auxiliary variable models for binary and multinomial regression.

Ho, A. W., & Kupper, T. S. (2019). T cells and the skin: from protective immunity to inflammatory skin disorders. *Nature Reviews Immunology*, *19*(8), 490-502.

Howell, M. D., Kim, B. E., Gao, P., Grant, A. V., Boguniewicz, M., DeBenedetto, A., Schneider, L., Beck, L. A., Barnes, K. C., & Leung, D. Y. (2009). Cytokine modulation of atopic dermatitis filaggrin skin expression. *Journal of Allergy and Clinical Immunology*, *124*(3), R7-R12.

Hu, Y., Li, M., Lu, Q., Weng, H., Wang, J., Zekavat, S. M., Yu, Z., Li, B., Gu, J., & Muchnik, S. (2019). A statistical framework for cross-tissue transcriptome-wide association analysis. *Nature genetics*, *51*(3), 568-576.

Hukku, A., Pividori, M., Luca, F., Pique-Regi, R., Im, H. K., & Wen, X. (2021). Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations. *The American Journal of Human Genetics*, *108*(1), 25-35.

Hukku, A., Sampson, M. G., Luca, F., Pique-Regi, R., & Wen, X. (2022). Analyzing and reconciling colocalization and transcriptome-wide association studies from the perspective of inferential reproducibility. *The American Journal of Human Genetics*, *109*(5), 825-837.

Ilkovitch, D. (2011). Role of immune-regulatory cells in skin pathology. *Journal of Leukocyte Biology*, *89*(1), 41-49.

Jarvinen, T. M., Hellquist, A., Koskenmies, S., Einarsdottir, E., Panelius, J., Hasan, T., Julkunen, H., Padyukov, L., Kvarnstrom, M., Wahren-Herlenius, M., Nyberg, F., D'Amato, M., Kere, J., & Saarialho-Kere, U. (2010). Polymorphisms of the ITGAM gene confer higher risk of discoid cutaneous than of systemic lupus erythematosus. *PLoS One*, *5*(12), e14212. https://doi.org/10.1371/journal.pone.0014212

Jin, Y., Andersen, G., Yorgov, D., Ferrara, T. M., Ben, S., Brownson, K. M., Holland, P. J., Birlea, S. A., Siebert, J., & Hartmann, A. (2016). Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. *Nature genetics*, *48*(11), 1418-1424.

Jones, L., Gordon-Smith, K., & Perry, A. (2021). Genome-wide association study of over 40,000 bipolar disorder cases provides novel biological insights. *Nature genetics*.

Kim-Hellmuth, S., Bechheim, M., Pütz, B., Mohammadi, P., Nédélec, Y., Giangreco, N., Becker, J., Kaiser, V., Fricker, N., & Beier, E. (2017). Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nature communications*, *8*(1), 266.

Klebanov, L., & Yakovlev, A. (2007). Diverse correlation structures in gene expression data and their utility in improving statistical inference.

Kuenzig, M. E., Bishay, K., Leigh, R., Kaplan, G. G., Benchimol, E. I., & Team, C. S. R. (2018). Co-occurrence of asthma and the inflammatory bowel diseases: a systematic review and meta-analysis. *Clinical and translational gastroenterology*, *9*(9), e188.

LaPierre, N., Taraszka, K., Huang, H., He, R., Hormozdiari, F., & Eskin, E. (2021). Identifying causal variants by fine mapping across multiple studies. *PLoS genetics*, *17*(9), e1009733.

Lappalainen, T., Sammeth, M., Friedländer, M. R., 't Hoen, P. A., Monlong, J., Rivas, M. A., Gonzalez-Porta, M., Kurbatova, N., Griebel, T., & Ferreira, P. G. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, *501*(7468), 506-511.

Lee, S. H., Wray, N. R., Goddard, M. E., & Visscher, P. M. (2011). Estimating missing heritability for disease from genome-wide association studies. *The American Journal of Human Genetics*, *88*(3), 294-305.

Lemler, D. J., Lynch, M. L., Tesfay, L., Deng, Z., Paul, B. T., Wang, X., Hegde, P., Manz, D. H., Torti, S. V., & Torti, F. M. (2017). DCYTB is a predictor of outcome in breast cancer that functions via iron-independent mechanisms. *Breast Cancer Research*, *19*, 1-18.

Li, L., Chen, Z., von Scheidt, M., Li, S., Steiner, A., Güldener, U., Koplev, S., Ma, A., Hao, K., & Pan, C. (2022). Transcriptome-wide association study of coronary artery disease identifies novel susceptibility genes. *Basic research in cardiology*, *117*(1), 6.

Li, N., Wei, W., Yin, F., Chen, M., Ma, T. R., Wu, Q., Zhou, J. R., Zheng, S.-G., & Han, J. (2015). The abnormal expression of CCR4 and CCR6 on Tregs in rheumatoid arthritis. *International Journal of Clinical and Experimental Medicine*, *8*(9), 15043.

Liu, L., Yan, R., Guo, P., Ji, J., Gong, W., Xue, F., Yuan, Z., & Zhou, X. (2024). Conditional transcriptome-wide association study for fine-mapping candidate causal genes. *Nature genetics*, *56*(2), 348-356.

Liu, L., Zeng, P., Xue, F., Yuan, Z., & Zhou, X. (2021). Multi-trait transcriptome-wide association studies with probabilistic Mendelian randomization. *The American Journal of Human Genetics*, *108*(2), 240-256.

Liu, X., Miao, Y., Liu, C., Lu, W., Feng, Q., & Zhang, Q. (2023). Identification of multiple novel susceptibility genes associated with autoimmune thyroid disease. *Frontiers in Immunology*, *14*, 1161311.

López-Isac, E., Acosta-Herrera, M., Kerick, M., Assassi, S., Satpathy, A. T., Granja, J., Mumbach, M. R., Beretta, L., Simeón, C. P., & Carreira, P. (2019). GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. *Nature communications*, *10*(1), 4955.

Ma, F., Plazyo, O., Billi, A. C., Tsoi, L. C., Xing, X., Wasikowski, R., Gharaee-Kermani, M., Hile, G., Jiang, Y., Harms, P. W., Xing, E., Kirma, J., Xi, J., Hsu, J. E., Sarkar, M. K., Chung, Y., Di Domizio, J., Gilliet, M., Ward, N. L., . . . Gudjonsson, J. E. (2023). Single cell and spatial sequencing define processes by which keratinocytes and fibroblasts amplify inflammatory responses in psoriasis. *Nat Commun*, *14*(1), 3455. https://doi.org/10.1038/s41467-023-39020-4

Mancuso, N., Freund, M. K., Johnson, R., Shi, H., Kichaev, G., Gusev, A., & Pasaniuc, B. (2019). Probabilistic fine-mapping of transcriptome-wide association studies. *Nature genetics*, *51*(4), 675-682.

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature genetics*, *51*(4), 584-591.

Matsuo, H., Ichida, K., Takada, T., Nakayama, A., Nakashima, H., Nakamura, T., Kawamura, Y., Takada, Y., Yamamoto, K., & Inoue, H. (2013). Common dysfunctional variants in ABCG2 are a major cause of early-onset gout. *Scientific reports*, *3*(1), 2014.

Matsuo, H., Takada, T., Ichida, K., Nakamura, T., Nakayama, A., Ikebuchi, Y., Ito, K., Kusanagi, Y., Chiba, T., & Tadokoro, S. (2009). Common defects of ABCG2, a high-capacity urate exporter, cause gout: a function-based genetic analysis in a Japanese population. *Science translational medicine*, *1*(5), 5ra11-15ra11.

Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., Reynolds, A. P., Sandstrom, R., Qu, H., & Brody, J. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science*, *337*(6099), 1190-1195.

McKie, A. T., Barrow, D., Latunde-Dada, G. O., Rolfs, A., Sager, G., Mudaly, E., Mudaly, M., Richardson, C., Barlow, D., & Bomford, A. (2001). An iron-regulated ferric reductase associated with the absorption of dietary iron. *Science*, *291*(5509), 1755-1759.

Méndez-Enríquez, E., & Hallgren, J. (2019). Mast cells and their progenitors in allergic asthma. *Frontiers in Immunology*, *10*, 821.

Miller, J. R. (2001). The wnts. *Genome biology*, *3*, 1-15.

Mitchell, B. L., Saklatvala, J. R., Dand, N., Hagenbeek, F. A., Li, X., Min, J. L., Thomas, L., Bartels, M., Jan Hottenga, J., & Lupton, M. K. (2022). Genome-wide association meta-analysis identifies 29 new acne susceptibility loci. *Nature communications*, *13*(1), 702.

Morley, M., Molony, C. M., Weber, T. M., Devlin, J. L., Ewens, K. G., Spielman, R. S., & Cheung, V. G. (2004). Genetic analysis of genome-wide variation in human gene expression. *Nature*, *430*(7001), 743-747.

Mullins, N., Forstner, A. J., O'Connell, K. S., Coombes, B., Coleman, J. R., Qiao, Z., Als, T. D., Bigdeli, T. B., Børte, S., & Bryois, J. (2021). Genome-wide association study of more

than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature genetics*, *53*(6), 817-829.

Nakagawa, P., Gomez, J., Grobe, J. L., & Sigmund, C. D. (2020). The renin-angiotensin system in the central nervous system and its role in blood pressure regulation. *Current hypertension reports*, *22*(1), 7.

Owlia, M. B., Newman, K., & Akhtari, M. (2014). Felty's syndrome, insights and updates. *The open rheumatology journal*, *8*, 129.

Palmer, C. N., Irvine, A. D., Terron-Kwiatkowski, A., Zhao, Y., Liao, H., Lee, S. P., Goudie, D. R., Sandilands, A., Campbell, L. E., & Smith, F. J. (2006). Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nature genetics*, *38*(4), 441-446.

Patrick, M. T., Stuart, P. E., Raja, K., Gudjonsson, J. E., Tejasvi, T., Yang, J., Chandran, V., Das, S., Callis-Duffin, K., & Ellinghaus, E. (2018). Genetic signature to provide robust risk assessment of psoriatic arthritis development in psoriasis patients. *Nature communications*, *9*(1), 4178.

Petukhova, L., Duvic, M., Hordinsky, M., Norris, D., Price, V., Shimomura, Y., Kim, H., Singh, P., Lee, A., & Chen, W. V. (2010). Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. *Nature*, *466*(7302), 113-117.

Pividori, M., Rajagopal, P. S., Barbeira, A., Liang, Y., Melia, O., Bastarache, L., Park, Y., Consortium, G., Wen, X., & Im, H. K. (2020). PhenomeXcan: Mapping the genome to the phenome through the transcriptome. *Science advances*, *6*(37), eaba2083.

Popa, O. M., Cherciu, M., Cherciu, L. I., Dutescu, M. I., Bojinca, M., Bojinca, V., Bara, C., & Popa, L. O. (2016). ERAP1 and ERAP2 gene variations influence the risk of psoriatic arthritis in Romanian population. *Archivum immunologiae et therapiae experimentalis*, *64*, 123-129.

Rael, E. L., & Lockey, R. F. (2011). Interleukin-13 signaling and its role in asthma. *World Allergy Organization Journal*, *4*(3), 54-64.

Ray, A., & Cohn, L. (1999). Th2 cells and GATA-3 in asthma: new insights into the regulation of airway inflammation. *The Journal of clinical investigation*, *104*(8), 985-993.

Ried, J. S., Li, J., Zuo, X. B., Zheng, X. D., Yin, X. Y., Sun, L. D., McAleer, M. A., O'Regan, G. M., & Fahy, C. M. (2015). Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nature genetics*, *47*(12), 1449-1456.

Ročková, V., & George, E. I. (2014). EMVS: The EM approach to Bayesian variable selection. *Journal of the American statistical Association*, *109*(506), 828-846.

Rudemiller, N. P., Lund, H., Priestley, J. R., Endres, B. T., Prokop, J. W., Jacob, H. J., Geurts, A. M., Cohen, E. P., & Mattson, D. L. (2015). Mutation of SH2B3 (LNK), a genome-wide association study candidate for hypertension, attenuates Dahl salt-sensitive hypertension via inflammatory modulation. *Hypertension*, *65*(5), 1111-1117.

Saleh, M. A., McMaster, W. G., Wu, J., Norlander, A. E., Funt, S. A., Thabet, S. R., Kirabo, A., Xiao, L., Chen, W., & Itani, H. A. (2015). Lymphocyte adaptor protein LNK deficiency exacerbates hypertension and end-organ inflammation. *The Journal of clinical investigation*, *125*(3), 1189-1202.

Sawada, Y., & Gallo, R. L. (2021). Role of Epigenetics in the Regulation of Immune Functions of the Skin. *Journal of Investigative Dermatology*, *141*(5), 1157-1166.

Schaid, D. J., Chen, W., & Larson, N. B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nature Reviews Genetics*, *19*(8), 491-504.

Schoettler, N., Rodríguez, E., Weidinger, S., & Ober, C. (2019). Advances in asthma and allergic disease genetics: Is bigger always better? *Journal of Allergy and Clinical Immunology*, *144*(6), 1495-1506.

Shang, L., Smith, J. A., Zhao, W., Kho, M., Turner, S. T., Mosley, T. H., Kardia, S. L., & Zhou, X. (2020). Genetic architecture of gene expression in European and African Americans: an eQTL mapping study in GENOA. *The American Journal of Human Genetics*, *106*(4), 496-512.

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., & Smoller, J. W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics*, *14*(7), 483-495.

Song, Y., Zhou, X., Kang, J., Aung, M. T., Zhang, M., Zhao, W., Needham, B. L., Kardia, S. L., Liu, Y., & Meeker, J. D. (2021). Bayesian hierarchical models for high-dimensional mediation analysis with coordinated selection of correlated mediators. *Statistics in medicine*, *40*(27), 6038-6056.

Song, Y., Zhou, X., Zhang, M., Zhao, W., Liu, Y., Kardia, S. L., Diez Roux, A. V., Needham, B. L., Smith, J. A., & Mukherjee, B. (2020). Bayesian shrinkage estimation of high dimensional causal mediation effects in omics studies. *Biometrics*, *76*(3), 700-710.

Spain, S. L., & Barrett, J. C. (2015). Strategies for fine-mapping complex traits. *Human molecular genetics*, *24*(R1), R111-R119.

Stegle, O., Parts, L., Piipari, M., Winn, J., & Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nature protocols*, *7*(3), 500-507.

Stranger, B. E., Montgomery, S. B., Dimas, A. S., Parts, L., Stegle, O., Ingle, C. E., Sekowska, M., Smith, G. D., Evans, D., & Gutierrez-Arcelus, M. (2012). Patterns of cis regulatory variation in diverse human populations. *PLoS genetics*, *8*(4), e1002639.

Strober, B. J., Zhang, M. J., Amariuta, T., Rossen, J., & Price, A. L. (2025). Fine-mapping causal tissues and genes at disease-associated loci. *Nature genetics*, *57*(1), 42-52.

Stuart, P. E., Tsoi, L. C., Nair, R. P., Ghosh, M., Kabra, M., Shaiq, P. A., Raja, G. K., Qamar, R., Thelma, B., & Patrick, M. T. (2022). Transethnic analysis of psoriasis susceptibility in South Asians and Europeans enhances fine mapping in the MHC and genome wide. *Human Genetics and Genomics Advances*, *3*(1).

Takada, T., Ichida, K., Matsuo, H., Nakayama, A., Murakami, K., Yamanashi, Y., Kasuga, H., Shinomiya, N., & Suzuki, H. (2014). ABCG2 dysfunction increases serum uric acid by decreased intestinal urate excretion. *Nucleosides, Nucleotides and Nucleic Acids*, *33*(4-6), 275-281.

Tamagawa-Mineoka, R. (2015). Important roles of platelets as immune cells in the skin. *Journal of dermatological science*, *77*(2), 93-101.

Tamari, M., & Hirota, T. (2014). Genome-wide association studies of atopic dermatitis. *The Journal of dermatology*, *41*(3), 213-220.

Tharakan, A., Kumar, A., Allegood, J., Cowart, L. A., Conrad, D. H., & Martin, R. K. (2023). D-2-hydroxyglutarate suppresses allergic sensitization in a murine model of experimental asthma. *Allergy*, *78*(11), 3014.

Thompson, M., Gordon, M. G., Lu, A., Tandon, A., Halperin, E., Gusev, A., Ye, C. J., Balliu, B., & Zaitlen, N. (2022). Multi-context genetic modeling of transcriptional regulation resolves novel disease loci. *Nature communications*, *13*(1), 5704.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267-288.

Tsoi, L. C., Rodriguez, E., Degenhardt, F., Baurecht, H., Wehkamp, U., Volks, N., Szymczak, S., Swindell, W. R., Sarkar, M. K., & Raja, K. (2019). Atopic dermatitis is an IL-13–dominant disease with greater molecular heterogeneity compared to psoriasis. *Journal of Investigative Dermatology*, *139*(7), 1480-1489.

Tsoi, L. C., Rodriguez, E., Stölzl, D., Wehkamp, U., Sun, J., Gerdes, S., Sarkar, M. K., Hübenthal, M., Zeng, C., & Uppala, R. (2020). Progression of acute-to-chronic atopic dermatitis is associated with quantitative rather than qualitative changes in cytokine responses. *Journal of Allergy and Clinical Immunology*, *145*(5), 1406-1415.

Verbeke, G., Molenberghs, G., & Verbeke, G. (1997). *Linear mixed models for longitudinal data*. Springer.

Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A. N., Knowles, D. A., Golan, D., Ermel, R., Ruusalepp, A., Quertermous, T., & Hao, K. (2019). Opportunities and challenges for transcriptome-wide association studies. *Nature genetics*, *51*(4), 592-599.

Wang, G., Sarkar, A., Carbonetto, P., & Stephens, M. (2020). A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *82*(5), 1273-1300.

Wen, J., Zhang, J., Zhang, H., Zhang, N., Lei, R., Deng, Y., Cheng, Q., Li, H., & Luo, P. (2024). Large-scale genome-wide association studies reveal the genetic causal etiology between air pollutants and autoimmune diseases. *Journal of Translational Medicine*, *22*(1), 392.

Wen, X., Pique-Regi, R., & Luca, F. (2017). Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLoS genetics*, *13*(3), e1006646.

Wills-Karp, M., & Chiaramonte, M. (2003). Interleukin-13 in asthma. *Current opinion in pulmonary medicine*, *9*(1), 21-27.

Wills-Karp, M. (2004). Interleukin-13 in asthma pathogenesis. *Immunological reviews*, *202*(1), 175-190.

Wiśniewski, A., Matusiak, Ł., Szczerkowska-Dobosz, A., Nowak, I., Łuszczek, W., & Kuśnierczyk, P. (2018). The association of ERAP1 and ERAP2 single nucleotide polymorphisms and their haplotypes with psoriasis vulgaris is dependent on the presence or absence of the HLA-C* 06: 02 allele and age at disease onset. *Human immunology*, *79*(2), 109-116.

Wrigley, R., Phipps-Green, A. J., Topless, R. K., Major, T. J., Cadzow, M., Riches, P., Tausche, A.-K., Janssen, M., Joosten, L. A., & Jansen, T. L. (2020). Pleiotropic effect of the ABCG2 gene in gout: involvement in serum urate levels and progression from hyperuricemia to gout. *Arthritis research & therapy*, *22*(1), 45.

Wu, C., & Pan, W. (2020). A powerful fine-mapping method for transcriptome-wide association studies. *Human genetics*, *139*(2), 199-213.

Wu, H., Ke, X., Huang, W., Shi, W., Yao, S., Duan, Y.-Y., Tian, W., Dong, S.-S., Xue, H.-Z., & Guo, Y. (2023). Multitissue Integrative Analysis Identifies Susceptibility Genes for Atopic Dermatitis. *Journal of Investigative Dermatology*, *143*(4), 602-611. e614.

Yang, J., Ferreira, T., Morris, A. P., Medland, S. E., Consortium, G. I. o. A. T., Replication, D. G., Consortium, M.-a., Madden, P. A., Heath, A. C., Martin, N. G., & Montgomery, G. W. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature genetics*, *44*(4), 369-375.

Yang, S., & Zhou, X. (2022). PGS-server: accuracy, robustness and transferability of polygenic score methods for biobank scale studies. *Briefings in Bioinformatics*, *23*(2).

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology*, *16*(5), 284-287.

Yuan, Z., Zhu, H., Zeng, P., Yang, S., Sun, S., Yang, C., Liu, J., & Zhou, X. (2020). Testing and controlling for horizontal pleiotropy with probabilistic Mendelian randomization in transcriptome-wide association studies. *Nature communications*, *11*(1), 3861.

Zeng, P., & Zhou, X. (2017). Non-parametric genetic prediction of complex traits with latent Dirichlet process regression models. *Nature communications*, *8*(1), 456.

Zhang, D.-H., Yang, L., Cohn, L., Parkyn, L., Homer, R., Ray, P., & Ray, A. (1999). Inhibition of allergic inflammation in a murine model of asthma by expression of a dominant-negative mutant of GATA-3. *Immunity*, *11*(4), 473-482.

Zhang, H., He, K., Li, Z., Tsoi, L. C., & Zhou, X. (2024). FABIO: TWAS fine-mapping to prioritize causal genes for binary traits. *PLoS genetics*, *20*(12), e1011503.

Zhao, S., Crouse, W., Qian, S., Luo, K., Stephens, M., & He, X. (2024). Adjusting for genetic confounders in transcriptome-wide association studies improves discovery of risk genes of complex traits. *Nature genetics*, *56*(2), 336-347.

Zhernakova, D. V., Deelen, P., Vermaat, M., Van Iterson, M., Van Galen, M., Arindrarto, W., Van't Hof, P., Mei, H., Van Dijk, F., & Westra, H.-J. (2017). Identification of context-dependent expression quantitative trait loci in whole blood. *Nature genetics*, *49*(1), 139-145.

Zhou, X., Carbonetto, P., & Stephens, M. (2013). Polygenic modeling with Bayesian sparse linear mixed models. *PLoS genetics*, *9*(2), e1003264.

Zhou, Y., Liang, Z.-S., Jin, Y., Ding, J., Huang, T., Moore, J. H., Zheng, Z.-J., & Huang, J. (2022). Shared genetic architecture and causal relationship between asthma and cardiovascular diseases: a large-scale cross-trait analysis. *Frontiers in genetics*, *12*, 775591.

Zhu, D., Yao, S., Wu, H., Ke, X., Zhou, X., Geng, S., Dong, S., Chen, H., Yang, T., & Cheng, Y. (2022). A transcriptome-wide association study identifies novel susceptibility genes for psoriasis. *Human molecular genetics*, *31*(2), 300-308.

Zhu, Z., Lee, P. H., Chaffin, M. D., Chung, W., Loh, P.-R., Lu, Q., Christiani, D. C., & Liang, L. (2018). A genome-wide cross-trait analysis from UK Biobank highlights the shared genetic architecture of asthma and allergic diseases. *Nature genetics*, *50*(6), 857-864.

Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M. R., Powell, J. E., Montgomery, G. W., Goddard, M. E., Wray, N. R., & Visscher, P. M. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature genetics*, *48*(5), 481-487.

Zhuang, Z., Yao, M., Wong, J. Y., Liu, Z., & Huang, T. (2021). Shared genetic etiology and causality between body fat percentage and cardiovascular diseases: a large-scale genome-wide cross-trait analysis. *BMC medicine*, *19*(1), 100.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301-320.