Title
Integrative statistical methods to understand the genetic basis of complex trait

Author
Kichaev, Gleb

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Integrative statistical methods to understand the genetic basis of complex traits

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Bioinformatics

by

Gleb Kichaev

2018

ABSTRACT OF THE DISSERTATION

Integrative statistical methods
to understand the genetic basis
of complex traits

by

Gleb Kichaev
Doctor of Philosophy in Bioinformatics
University of California, Los Angeles, 2018
Professor Bogdan Pasaniuc, Chair

The Genome-wide Association study (GWAS) is one of the primary tools for understanding the genetic basis of complex traits. In this dissertation I introduce enhanced statistical methods to do integrative GWAS analysis with functional genomic data. First, I describe an integrative fine-mapping framework to prioritize causal variants at known GWAS risk loci. Next, I expand upon this framework to exploit genetic heterogeniety across human populations to improve statistical efficiency. I then consider a new inference strategy to reduce the computational burden of the methodology. Finally, I propose a new approach for GWAS discovery that leverages functional genomic data through polygenic modeling.

The dissertation of Gleb Kichaev is approved.

Jason Ernst

Paivi Elizabeth Pajukanta

Eleazar Eskin

Bogdan Pasaniuc, Committee Chair

University of California, Los Angeles

2018

To my parents . . .

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I would like to thank my brilliant advisor, Bogdan Pasaniuc, whose patience and endless support during the last five years have been invaluable to my development as a scientist.

I would also like to express my gratitude to Alkes Price, who graciously hosted me at the Harvard School of Public Health in the summer of 2016, and who has continued to provide mentorship and guidance since. I would also like to acknowledge my other friends and collaborators at Harvard: Peter Kraft, Sasha Gusev, Po-Ru Loh, Hillary Finucane, Steven Gazal, Armin Schoech, Yakir Reshef, Gaurav Bhatia, and Carla Marquez Luna.

I owe a great deal of my success to my lab mates, who provided the friendly, intellectual atmosphere that led to so many of the ideas found in these pages: Robert Brown, Huwenbo Shi, Wen-Yun Yang, Nick Mancuso, Kathy Burch, Malika Freund, Ruth Johnson, Megan Roytman, Valerie Arboleda, Claudia Giambartolomei, Megan Major, and Arunabha Majumdar. Thank you all for your camaraderie.

I would also like to thank my committee members Paivi Pajukanta, Jason Ernst, and Eleazar Eskin for their support. Finally, I would like to thank the Biomedical Big Data training fellowship which provided two years of funding and training during my PhD.

VITA

2006–2010	B.S., (Cell & Developmental Biology), UCSB, Santa Barbara, California.
2011-2012	Graduate Certificate, (Bioinformatics), Stanford University. Stanford, California.
2012–2014	M.S. (Biostatistics), UCLA, Los Angeles, California.
2013-curr	Graduate Student Researcher. UCLA. Bogdan Pasaniuc Lab
Sum 2016	Visiting Doctoral Scholar. Harvard Univiersty. Alkes Price Lab.
Fall 2013	Teaching Assistant, UCLA. Department of Biostatistics. Biostatistics 100A - Introduction to Biostatistics
Fall 2017	Teaching Assistant, UCLA. Department of Computer Science. Computer Science 226 - Machine Learning in Bioinformatics

PUBLICATIONS

Selected publications (of 22)

Kichaev, Gleb, et al. ”Integrating functional data to prioritize causal variants in statistical fine-mapping studies.” PLoS genetics 10.10 (2014): e1004722.

Kichaev, Gleb, and Pasaniuc, Bogdan. ”Leveraging functional-annotation data in trans-ethnic fine-mapping studies.” The American Journal of Human Genetics 97.2 (2015): 260-271.

Kichaev, Gleb, Roytman, Megan, et al. "Improved methods for multi-trait fine mapping of pleiotropic risk loci." Bioinformatics 33.2 (2017): 248-255.

Kichaev, Gleb, et al. "Leveraging polygenic functional enrichment to improve GWAS power." bioRxiv (2017): 222265. (*Under Review at The American Journal of Human Genetics*)

CHAPTER 1

Introduction

Most complex traits and common diseases are in part, genetically determined [5]. This motivates the study of these traits through the lens of human genetics, which provides a rational set of tools for mapping disease susceptibility to risk genes [6]. Most prominently, Genome-wide Association Studies (GWAS), an experimental design wherein genetic markers known as Single Nucleotide Polymorphisims (SNPs) are surveyed in large cohorts of individuals and then (marginally) tested for association, have now identified tens of thousands of loci spanning hundreds of human traits [7, 8, 9].

While GWAS have been successful in robustly identifying risk loci with high statistical confidence, they typically do not yield the true causal SNPs or genes, thus limiting the mechanistic insight into the underlying biology. This is because the majority of loci that are uncovered by GWAS reside outside of known coding regions [10, 11]. A corollary to this observation is that genetic effects on complex traits are most likely mediated through alterations in gene regulation rather than the gene product itself [9, 12]. Efforts by large consortia such as the ENCODE and Roadmap have been instrumental in illuminating the non-coding genome by mapping functional regulatory elements in many diverse cell types and tissues [13, 14]. These maps not only facilitate interpretation of GWAS findings, but can also serve as an orthogonal source of signal when performing GWAS analyses. To bridge the gap between statistical association and disease biology, this dissertation introduces new methods to perform integrative analysis of functional genomic and population-scale GWAS data.

Linkage disequilibrium (LD) is a phenomenon that arises when nearby SNPs are inherited on the same haplotype. This induces a rich correlation structure between genetic

markers across the genome [15]. In the context of GWAS, the correlation structure manifests most strikingly when clusters of neighboring variants all appear to have significant marginal associations to a trait under study. One of the central challenges in post-GWAS analysis is isolating the variants that are truly impacting a trait from ones that are simply correlated due to LD. The process of statistically resolving causal variants from their correlated neighbors is referred to as "fine-mapping" [16]. The primary objective of fine-mapping is to minimize the number of SNPs one would need to follow-up in biological assays while maximizing the number of true causal variants discovered. This prioritization can be accomplished probabilistically [17, 18, 1]. Prior to the methods developed in this dissertation, probabilistic fine-mapping would either require access to individual level data, be conducted under a simplifying assumptions that each GWAS locus harbors at most a single causal variant, or neglect functional data [17, 1, 19]. In the first part of this dissertation, I develop a statistical framework that overcomes these limitations [20], generalize the approach to operate with multiple continental populations [21], and propose a computationally efficient inference strategy [22].

In detail, Chapter 2 of this dissertation introduces a powerful statistical framework for fine-mapping with functional data called *Probabilistic Annotation INgraTOR* (PAINTOR). As the name suggests, this is a probabilistic fine-mapping method that incorporates functional annotation data using a hierarchical Bayesian model. The key features of this method is that it allows for multiple causal variants at any GWAS risk locus and facilitates unbiased selection of relevant, trait-specific functional data through an empirically estimated prior. Critically, inference is carried out directly on summary data (Z-scores), allowing us to tap in to the largest, most powerful GWAS while simultaneously obviating the need to access individual level genetic data. I demonstrate in real and simulated data that this approach achieves state of the art performance in fine-mapping resolution. This work was published in *PLoS Genetics* [20] and was subsequently featured as research highlight in *Nature Genetics* [23].

While early GWAS were primarily conducted in individuals of European descent, there has been an increasing effort to study complex trait genetics in non-European co-

horts [24]. Empirically, there is evidence that GWAS signal identified in Europeans generalizes to non-European populations despite having divergent histories [25, 26]. Motivated by this observation, in Chapter 3, I expand upon the integrative framework developed in Chapter 2 to leverage genetic diversity across continental populations to improve fine-mapping of causal variants. I demonstrate in real and simulated data that by explicitly modeling multi-ethnic LD patterns, we are able to gain substantial improvements in fine-mapping resolution. Leveraging functional priors further enhances performance. This work was published in *The American Journal of Human Genetics* [21] for which I was awarded the *C.W. Cotterman Award* in 2015.

The core inference procedure in the PAINTOR framework developed in Chapters 2 and 3 integrates over combinations of causal SNPs that can explain the observed GWAS signal. In the extreme case where each one is potentially allowed to be causal, this operation would require $O(2^M)$ likelihood evaluations for M SNPs. Clearly, this is intractable for even a modestly-sized locus. We evade these difficulties in Chapters 2 and 3 by limiting the number of total causal SNPs k to be a small number, yielding a more reasonable $O(M^k)$ computational complexity. However, this still remains computationally inefficient as the vast majority of causal combinations are unlikely to explain the data (e.g a set of SNPs with Z-scores close to zero). In Chapter 4, I propose an efficient Importance Sampling procedure that dramatically improves runtime without sacrificing fine-mapping accuracy. In addition, earlier iterations of PAINTOR required heuristic approximations of the causal effect sizes. In this chapter, I adopted a more principled Bayesian treatment of the causal effects [27, 28, 29]. This not only improved statistical robustness of the model, but also facilitated a reduction in the computational footprint of each likelihood evaluation from $O(M^3)$ to $O(K^3)$. This work as published in *Bioinformatics* [22] where I was a co-first author jointly with Megan Roytman. My contribution to this work was developing the Importance Sampling procedure, making the model fully Bayesian, and writing and maintaining the software implementation. This paper also introduced a way to do multi-trait fine-mapping which was lead by Megan Roytman.

Genetic mapping of diseases and complex traits involves a large number of statistical

tests that imposes a heavy multiple-testing correction burden, necessitating stringent p-value thresholds that limits power [30, 31]. In the view of emerging functional genomics data, a single p-value threshold applied uniformly across the genome will be sub-optimal as not all SNPs, aprior, have the same chance of tagging phenotypically relevant signal [32, 33]. For example, a SNP that is in strong linkage disequilibrium (LD) with multiple variants that fall within DNase Hypersensitivity Sites (DHS) is more likely to be associated with a trait than a low LD SNP in an intergenic region. Leveraging this intuition to re-calibrate p-value thresholds when performing association tests is the focus of the second part of this dissertation.

Most complex traits have overwhelmingly polygenic architectures and exhibit enrichment for heritability in certain functional categories of variants [9, 11, 34]. In the final chapter of this dissertation, I describe an integrative method that leverages polygenic functional enrichment to improve association power. The method, Functionally Informed Novel Discovery Of Risk loci (FINDOR), uses a comprehensive assortment of functional annotations that broadly capture coding, conserved, regulatory and LD-related bio-features of the genome. I demonstrate in simulations that the method is well-calibrated under the null and delivers substantial increases in power at true causal loci. An application of FINDOR to 27 independent traits from the UK Biobank spanning approximately 460K individuals, discovered an additional 583 new independent loci. At the writing of this dissertation, this work is still under review and is available as pre-print on the *bioRxiv*.

CHAPTER 2

Probabilistic Annotation Integrator

2.1 Introduction

Recent breakthroughs in high throughput genotyping technologies have ushered in the era of genome-wide association studies (GWAS) that have reproducibly identified thousands of genetic variants associated to many diseases and complex traits [10]. GWAS leverage the linkage disequilibrium (LD) patterns among genetic markers for probing genetic variation beyond the typed variants. Thus, it is often the case that the associated variant is not itself biologically causal, but rather, a proxy as a result of LD. Identification of causal variants underlying risk loci is performed within fine-mapping studies [35, 36, 37] through sequencing (or array typing and imputation) followed by variant prioritization using marginal association statistics or posterior probabilities [1, 38, 19]. Using these measures, a set of top candidate variants is selected for testing in functional experiments to validate biological causality.

Many statistical approaches have been introduced for fine-mapping ranging from a simple ranking of marginal association statistics to Bayesian approaches that integrate elaborate priors [1, 39, 40, 41, 42, 43, 44, 45, 17, 46, 47]. Due to the fact that fine-mapping can be casted as a variable selection problem, both LASSO-like procedures that estimate empirical probabilities of inclusion for SNPs based on sub-sampling [44], as well as Bayesian approaches that perform joint multipoint inference to compute posterior inclusion probabilities [45] have been proposed. The inclusion probabilities provided by these methods offer a natural way to prioritize variants in fine-mapping. However, although neither of the two variable selection approaches assume a fixed number of causal variants,

they both require individual level data which is often not readily available. Ranking of SNPs for follow-up analysis can also be performed based on correlation-adjusted t-scores that explicitly take into account the correlation structure among variants, thus requiring individual level data [43] as well. Recent works [1, 39, 40] have proposed to estimate posterior probabilities and credible sets for variants to be causal under the simplifying assumption of single causal per locus. A key advantage of such approaches is that they only require marginal association statistics which are readily available for large-scale data sets.

Large-scale initiatives such as The Encyclopedia of DNA Elements (ENCODE) [13] have ascribed functional importance to more than 80% of the human genome and have provided a genome-wide catalogue of regulatory regions. This functional annotation data can be used jointly with the standard association signal to gain insights into the genetic basis of common traits. Indeed, variants associated with certain ENCODE genomic functional annotations such as DNase I Hypersensitive Sites, transcription factor binding sites and expression quantitative loci are enriched among GWAS hits [48, 49, 50, 51, 2], with recent work demonstrating that it is possible to integrate such data with the GWAS association signal to identify novel risk loci [41]. However, existing integrative frameworks typically either assume a single causal variant per risk locus [41] that is likely to be incorrect at many risk loci [41, 52, 35, 19, 53, 54, 55, 56] or do not make use of functional data [57, 58]. Although ENCODE functional annotation data are clearly beneficial for fine-mapping [51], a rigorous statistical framework for integrating the different types of information for the purpose of prioritizing plausible causal variants is currently lacking.

In this work we introduce PAINTOR (Probabilistic Annotation INTegratOR), a framework to combine external functional annotations (sets of variants that localize within certain genomic features, e.g. enhancers, repressors) with genetic association data (the strength of association between genetic variants and the phenotype) to improve the prioritization of causal variants in fine-mapping studies. As compared to existing approaches that only rely on the strength of association between genotype and phenotype [59, 1, 38], our framework combines two orthogonal lines of evidence to estimate variant-specific

probabilities for causality: functional relevance and genotype-phenotype association. These probabilities can then be used for prioritization of variants for functional validation studies to determine biological causality. More specifically, we incorporate the external functional annotation data through an Empirical Bayes prior [60] with parameters inferred from targeted fine-mapping data, obviating the need to make assumptions on which tissue-specific annotation is relevant to the trait of interest. Finally, budgetary constraints will invariably restrict the number of potential variants that can be validated in functional studies. We address this issue by proposing a cost-to-benefit optimization framework to guide the design of experimental follow-up studies.

We use extensive simulations starting from the 1000 Genomes data to show that our approach improves resolution of statistical fine-mapping and is superior to existing frameworks. In our simulations of a trait with a heritability of $h_g^2 = 0.25$ across 100 risk loci, one needs to test in functional assays an average of 12.3 SNPs per locus to identify 90% of all causal variants if using our approach. In addition, if causal variants are preferentially enriched within certain genomic regions [48, 50, 41, 2], PAINTOR further reduces the average number of SNPs per locus needed to capture 90% of the causal variants to 10.4. We show in simulations that the enrichment estimates provided by PAINTOR are largely unbiased, a fact that we can subsequently use to search for the annotations most phenotypically relevant. We then demonstrate an application of our approach using data from a large-scale meta-analysis study of blood lipid phenotypes (triglycerides (TG), total cholesterol (TC), high density lipoprotein (HDL), low density lipoprotein (LDL)[61])and find that causal variants at risk loci are preferentially enriched within coding regions and significantly depleted from repressed regions. In real data PAINTOR is able to reduce the size of the 90% confidence set from an average 17.5 to 13.5 SNPs per locus, a reduction consistent to simulation results. We provide software implementing our framework freely available to the research community at
http://bogdan.bioinformatics.ucla.edu/software/paintor/.

2.2 Results

2.2.1 Overview of statistical fine-mapping with functional annotation

To illustrate PAINTOR, consider the case of two risk loci that are fine-mapped through sequencing to elucidate the causal variant(s) driving the phenotype (Figure 2.1). The observed association statistics at all SNPs at these loci are a function of the causal variants, their effect size and the locus-specific LD structure. We use a multivariate normal approximation to connect the LD structure of a fine-mapping locus to the association statistics (e.g. association z-scores) which allows for the possibility of modeling multiple causal variants – an important feature since the number of causals variants per locus is typically unknown a priori. We integrate functional annotation data through an Empirical Bayes prior [60] such that the prior probability of a variant to be causal is governed by its membership to functional classes (see Methods). We perform maximum likelihood estimation over all fine-mapping loci using a variant of the Expectation Maximization algorithm to infer the parameters of the model, followed by estimation of the probabilities for each variant to be causal (see Methods). Intuitively, PAINTOR up-weights variants residing in certain functional annotations (e.g. transcription start sites) while down-weighting variants within annotations less relevant to the trait (e.g. intergenic). The weight associated to each functional annotation is inferred from the data itself without making any ad-hoc assumptions on which tissue-specific annotations are relevant to the trait of interest. The main output of PAINTOR is a probability for each variant to be causal that can be used for selection of SNPs to be tested for biological causality in functional assays.

2.2.2 Functional annotation data improves statistical fine mapping performance

Various approaches for fine-mapping have been proposed, ranging from methods that require individual genotype data to methods that take as input summary association data and integrate functional annotations (see Table 2.9). We used simulations to compare PAINTOR to previously proposed methods. It is generally the case that in fine-mapping

studies several risk loci are simultaneously sequenced (or densely genotyped) and a set of plausible causal SNPs is selected for follow-up in functional assays. Therefore, we simulated fine-mapping data sets for a disease with phenotypic variance explained of $h_g^2 = 0.25$ across 100 risk loci, each 10Kb in size for N = 10,000 individuals (see Methods). We created three synthetic "functional annotations" that roughly correspond to coding exons (2.2% of all variants), transcription start sites (2.2% of all variants), and DNase Hypersensitivity Sites (30.7% of all variants) and enriched them with causal variants at 9.5, 5.7 and 3.7-fold to approximately match what we observed in real data (see below). Each simulation resulted in approximately 64 loci that harbor at least one causal variant with 34 harboring a single causal variant and the remaining harboring multiple causal variants (see Methods). We compared all approaches across only loci with at least one causal variant.

We find that prioritizing variants using PAINTOR posterior probabilities achieves superior accuracy over existing methodologies (Figure 2.2, Table 2.1). Our approach identifies more causal variants at all selection thresholds, and is a consequence of PAINTOR's ability to model multiple causal variants while incorporating functional priors. For example, in order to find (50%, 90%) of all causal variants one needs to select an average of (1.3, 10.4) SNPs per locus if using PAINTOR. In contrast, ranking SNPs using frameworks that assume a single causal variant, such as Maller et al.[1] and fgwas[41], require (2.7, 25.4) and (2.0, 21.5) SNPs per locus, respectively. In general, we observe an increase in performance for methods that incorporate functional data and allow for multiple causal variants at a risk locus (Tables 2.1 and 2.2). Despite having access to individual level data, variable selection strategies[44, 45] were less accurate than PAINTOR in our simulations (Figure 2.2, Table 2.1). Ranking SNPs based on correlation-adjusted t-scores [43] was superior to existing methodologies, however, still failed to achieve the same level of accuracy of PAINTOR, requiring an average of (2.0,13.3) SNPs per locus to find (50%, 90%) of all causal variants. Across all methodologies, the relative performance holds irrespective of whether SNPs are prioritized across all fine-mapping loci or within each locus independently (generally the latter strategy is sub-optimal (Table 2.1)). Finally, we note

that iterative conditioning, a method typically used to detect multiple independent signals, performs worse than the prioritization strategies described here (see Figure 2.5) [19]. Interestingly, as the number of SNPs selected for follow-up increases, the naive approach of selecting based on association p-value alone attains high accuracy, most likely due to the much smaller set of assumptions as compared to other methods.

2.2.3 Factors impacting fine-mapping performance

Having established that PAINTOR increases fine-mapping accuracy over existing methods in simulations, we next explored the gain in performance attributable to having access to functional annotation data. We find that prioritizing variants using PAINTOR with functional data increases accuracy at all significance thresholds. For example, in order to find (50%, 90%) of all causal variants one needs to select an average of (1.3, 10.4) SNPs per locus if integrating functional data as opposed to (1.7, 12.3) if excluding annotation data. We note that our approach that does not empirically estimate the prior, but uses the known prior information does not lead to superior performance over PAINTOR in these simulations (see Table 2.1) reflecting the fact that the prior probabilities for each SNP are accurately estimated. Furthermore, as the size of the fine-mapping locus is increased, PAINTOR continues to outperform simpler approaches. In particular, to resolve 90% of the causal variants for loci (10Kb, 25Kb, 50Kb) in size, one needs to select (27.4, 52.3, 110.7) SNPS per locus if ranking on posterior probabilities assuming a single causal variant as opposed to (11.4, 16.0, 24.1) SNPs per locus if ranking using PAINTOR (see Table 2.3).

We next sought to determine at what types of loci is functional prior data providing the biggest increase in accuracy. Loci where the association signal is strong (i.e. loci where the p-value at the causal variants are in the top quartile across all loci with at least 1 causal variant) do not gain much from integration of functional annotation data, with the number of SNPs required to find 90% of the causal variants decreasing by only 6.5%. On the other hand, at loci where the association signal is weak (i.e. loci where the p-value at the causal variants are in the bottom quartile) we observe a 21.4% decrease in the

total number of SNPs to be followed-up to find 90% of all causal variants (see Table 2.12). This suggests that as the causal status for a SNP becomes increasingly ambiguous on the basis of association data alone (e.g. small effect size), the importance of incorporating additional sources of information is magnified.

It is not guaranteed that the true causal variant will be present in the fine-mapping data set due to technical reasons (e.g. capture sequencing technology or imputation accuracy). To explore this scenario, we simulated fine-mapping data sets at a locus 100Kb in size after which we masked the true causal(s) from the data (see Methods). To measure fine-mapping performance when causal variant is absent from the data, we looked at the distance in base-pairs between variants in the top N SNPs to the true masked causal SNP. As expected, we observed a decrease in performance when causal variants are absent from the fine-mapping dataset (e.g. the average median distance to the true causal variant in the set of top 5 SNPs increases by 6% when the causal variant is masked, see Table 2.4). The rather small nominal decrease in localization distance suggests that accurate localization may be attained even in the absence of the causal variant.

Alternatively, we can recast the observed improvement in causal variant localization when incorporating functional annotations as a decrease in size of the set of SNPs to account for a fixed amount of posterior probability mass. We extend existing work for single-locus fine-mapping [1, 39, 40, 19] to define an ρ-level causal set as the set of top SNPs (rank-ordered based on probabilities) across all fine-mapping loci that consume an ρ fraction of the *total* posterior probability mass. We observe a reduction in the number of SNPs within the 90%, 95% and 99% confidence sets when using functional annotations as compared to no functional data (see Table 2.2). In addition, although PAINTOR with annotation yields fewer SNPs with high probability than the PAINTOR with no annotation (232.8 vs 265.2 at a threshold of > 0.1), having access to annotation yields more simulated causals with high posterior probability (78.6 vs 73.8 at a threshold of > 0.1)(see Table 2.10).

2.2.4 Estimation of relevant annotation data for fine mapping

A vast resource for functional annotations is the ENCODE project[13], which has ascribed regulatory biological function to a large fraction of the human genome and has shown that regulatory DNA regions are highly cell-specific. Coupling this insight with the fact that for most complex diseases the relevant tissues are unknown, stresses the importance of carefully selecting cell-specific annotations for any specific trait [51]. A byproduct of our framework is the estimation of enrichment of causal variants within functional annotations (i.e. the ratio of prior probability of causality for SNPs within annotation versus those outside the annotation). Therefore, we can use PAINTOR to infer which functional annotations show significant effect on the probability of causality and use only those annotations to estimate probability of causality. To assess how accurately PAINTOR can recapitulate functional enrichment, we simulated fine-mapping studies over 100 loci with a synthetic functional annotation (see Methods) and either enriched or depleted causal variants within this annotation. We also compared our approach to fgwas[41] as it too is capable of inferring enrichment from summary data. Figure 2.3 demonstrates that both PAINTOR and fgwas are able to provide unbiased estimates of enrichment. However, we find that PAINTOR is more efficient than fgwas, and has a smaller variance attached to those estimates. We note that as causal variants become increasingly depleted from functional categories, fgwas tends to fail to converge (e.g. fgwas fails in nearly 21% of cases for simulations with 8-fold depletions). Finally, we assessed PAINTOR and fgwas for more realistic annotation data (i.e. contiguous segments in the genome) and find that both methods attain very similar results (see Figure 2.7).

2.2.5 Selecting the optimal number of SNPs for functional testing

Although PAINTOR (and previous methods) provide a quantification of the probability of each variant to be causal that can be used to rank variants based on their plausible causality, it remains unclear how to choose the number of variants to test in functional assays. The optimum number is constrained by the budget of the study and by an im-

plicit cost to benefit ratio for selecting the optimal number of SNPs to be followed up. We propose a framework that assumes that every causal variant identified adds a benefit (B) while every selected variant is tested at a cost (C); therefore, the utility function we propose to maximize is $U = B * N_c - C * N_t$, where N_c is the total number of true causal variants from the total number of selected SNPs (N_t). We note that the ratio $r = B/C$ is the critical parameter of the utility function. Using the results from simulations with functional annotation enrichment described above, we assessed the capacity of the proposed utility function in selecting the number of SNPs for follow-up under various values for the ratio $r = B/C$ (Figure 2.4). For example, at a ratio $r = 10$ (the benefit of finding a causal outweighs 10 times the cost of testing 1 SNP), the utility is maximized by selecting approximately 3.5 SNPs per locus for validation resulting in 72.6% of causal variants successfully identified (see Figure 2.9).

Selection of a set of variants for follow-up is usually performed based on a threshold on posterior probability or based on credible sets that account for a given amount (e.g. $\rho = 90\%$) of the probability of capturing all causal variants[1, 39]. We assessed these two strategies for selecting variants for functional testing within the context of our benefit-to-cost framework. We find that a posterior probability threshold of (0.9, 0.5, 0.1) roughly corresponds to optimizing benefit-to-cost-ratios of $r = (1.25, 5, 10)$. These results suggest that a simple translation of the arbitrary thresholds on posterior probabilities into cost-to-benefit optimum is attainable. In a similar fashion, we can assess credible sets within our cost-to-benefit framework. For example, the 90% credible set yields an average of 393 SNPs which is approximately 88% of the optimum for a benefit-to-cost of $r = 10$.

2.2.6 Application to meta-analysis data of lipid phenotypes

To validate our approach, we applied PAINTOR to association summary data from a large meta-analysis of four lipid traits. Our goal was to build a model that incorporated all the independent sources of available information (i.e. association signals alongside carefully selected functional annotations) to produce a prioritization of plausible causal SNPs for

these phenotypes. We used the GWAS hits reported by Teslovich et al. [61] under the assumption that these regions contain causal variants and therefore well-suited to fine-map using PAINTOR. We first ran our method on 450 cell-type-specific annotations (see Methods) and fit the model to each annotation independently on both the original and densely imputed data sets for all four traits. Consistent to previous works, we observe that imputation consistently enhances the signal of enrichment [62, 2, 41]; for example, for HDL, the relative probability for causality for coding exons increases from 7.4 to 12.4 from using the original data to 1KG-imputed data (see Table 2.17). This effect is most likely due to the availability of more variants through imputation thus being able to localize the association signal to genomic annotation more accurately. Across the four traits in general, we see consistent signal of increased relative probability for causality within transcribed regions (e.g. exons and transcription start sites (TSS)) and a depletion of causal variants in repressed regions; for example, for TG, the coding exons show a log2 relative probability for causality of 3.4 while the repressed regions show an log2 relative probability of -1.6.

Having identified functional annotations that are relevant to the four traits of interest (see Table 2.5), we devised trait-specific PAINTOR models that included the top marginal annotations in conjunction with the association statistics to estimate the probability of causality for all SNPs from the risk loci on the densely imputed data sets (see Methods). Table 2.6 shows the HDL SNPs that attain a posterior PAINTOR probability greater than 0.9 (results for the other traits are displayed in the Tables 2.13,2.14,2.15). Unsurprisingly, the majority of these top SNPs localize in functional elements and attain a high marginal association statistic. We observe an abundance of liver associated cell types, DNase Hypersensitivity Sites, and genic elements annotated to these top SNPs. Notably, PAINTOR identifies four non-synonymous variants (rs7607980, rs1260326, rs5110, rs13107325), two of which were not reported in the initial Teslovich et al. findings. Overall by incorporating functional annotations we see a marked improvement in fine-mapping resolution across all four traits as indicated by a reduction in the 90% confidence sets relative to PAINTOR models with no annotations of 19.0%, 34.9%, 50.6%, and 24.2% for HDL, LDL, TC, and TG, respectively (Table 2.7). This corresponds to approximately an average reduction of

17.5 to 13.5 SNPs per locus across the four traits.

2.3 Discussion

Recent efforts by large consortia such the ENCODE have provided a genomic map of regulatory regions and have shown that GWAS associated variants are preferentially enriched within these regions. In this work, we propose a principled approach to unifying these genomic features with the standard association signal to improve the localization accuracy in fine-mapping studies. Our method relies on empirical data to select trait-specific genomic annotations, thus removing the need for ad-hoc selection of relevant functional annotations a priori. Through simulated and real data results, we have shown that our integrative framework is able to reduce the number of variants that need to be investigated to identify causal variants that alter risk of disease.

Our method shares similarities to recent integrative approaches proposed in the context of GWAS [41]. Although conceptually both approaches integrate functional and association signal, the two methodologies are fundamentally distinct in their aims. Whereas [41] seeks to identify novel risk loci by leveraging functional information, we instead propose our method as way to refine signal at known GWAS loci. This fundamental distinction leads to different statistical models and optimization procedures allowing for superior accuracy for refining association signal through fine-mapping. In addition our method addresses a limitation of [41] by allowing for the possibility of multiple causal variants at a risk locus.

Several hierarchical Bayesian methods have been developed that combine prior information with genomic association data to help prioritize variants in various contexts [57, 58]. The main contribution of our approach is that we explicitly account for LD between SNPs which we can learn from external reference panels such as the 1000 Genomes. Additionally, because we do not take a fully Bayesian approach [58] (i.e. integrate over the entire hyper-parameter space), we are able to devise computationally efficient algorithms that allow our method to search over the ever-increasing number of functional annota-

tions (e.g. ENCODE) to identify the most informative subset while retaining the ability to model multiple causal variants.

We have shown that PAINTOR can unbiasedly estimate enrichment of causal variants in different functional elements on the basis of summary association data alone. This may prove to be particularly important as access to individual genotype data is more cumbersome than summary-level statistics. The unbiased nature of the estimation procedure may provide clues to the genetic basis of common traits. For example our results suggest that although coding variants are more likely to be causal than regulatory variants, the majority of the genetic variation contributing to the trait at these risk loci may lie within regulatory as opposed to coding regions due to the larger number of variants residing in regulatory regions. This is consistent with recent work that concluded that variants in regulatory regions show a higher contribution to traits than coding variants, however, such an analysis required individual level data [2].

One interesting implication of our results is that while higher-order functional data is very useful for gleaning insight into to the genetic architecture of human diseases genome-wide [49, 2], the main component of accuracy in a fine-mapping study is the sample size (see Table 2.11). Consequently, the success of a fine-mapping experiment may hinge on first obtaining an adequate sample size and then augmenting that sample size with functional data. These findings are largely in-line with what was previously reported in the context of GWAS [41].

In this work we have applied our framework to known risk loci identified in GWAS in the search for plausible causal variants. As future work, our approach could be extended to risk loci that do not pass a genome-wide stringency, potentially leading to discovery of novel risk loci. Additionally, risk loci for related traits that are known to share a genetic basis could potentially be combined, leading to an increase in power to identify variants that contribute to both traits. Finally, we anticipate that the approximations of the non-centrality parameters could be handled in a more principled fashion using a Bayesian approach that integrates a prior distribution of effect sizes. We leave a thorough investigation of these directions as future work.

2.4 Methods

2.4.1 PAINTOR Probabilistic Model

A standard approach to model the strength of association of genotype to phenotype is through the Z-score. For a continuous phenotype, the trait values are marginally regressed on each SNP and the corresponding Z-score is taken to be the Wald statistic (i.e $\frac{\hat{\beta}}{SE(\hat{\beta})}$), which is distributed $N(0,1)$ under the null. For case-control designs, the Z-score can also be obtained through the standard test statistic for two proportions (assuming equal sample sizes of $\frac{N}{2}$): $\frac{\sqrt{N}(f_i^+ - f_i^-)}{\sqrt{2f_i(1-f_i)}}$, where $f_i^+(f_i^-)$ denotes the frequency of the SNP in the cases (controls) and $f = \frac{f_i^+ + f_i^-}{2}$. We define a fine-mapping locus as a contiguous region of the genome flanking a GWAS "hit" on both sides. Let Z_j be a vector of Z-scores from the j^{th} locus $(1 \leq j \leq L)$ of length N_j. In addition, let Σ_j be the corresponding LD matrix of pairwise correlation coefficients for locus j that can be derived directly from individual level data if available, or approximated using an appropriate reference panel such as the 1000 Genomes. We obtain K annotations $(1 \leq k \leq K)$ from external repositories (e.g. ENCODE[13]) and for each SNP i, create a $(K+1)$-length binary annotation vector $A_{i,j}$, where $A_{i,j,k} = 1$ if the i^{th} SNP at the j^{th} locus is part of annotation k. For example, one such annotation could be all coding sites and the annotation vector will contain a 1 only if the SNP is located within coding region. We note that $A_{i,j,0} = 1 \; \forall \; i, j$ and serves to represent the "baseline" annotation whose corresponding coefficient can be interpreted as the baseline prior odds for causality of any SNP within the set of fine-mapping loci. Let γ_k be the effect size of the k_{th} annotation on the probability of a SNP being causal and the non-centrality parameter, $\lambda_{i,j}$, be the standardized effect size of SNP i at locus j. Finally, let C_j be an indicator vector of causality where $C_{i,j} = 1$ if SNP i at locus j is causal and 0 otherwise. Now, we can define the likelihood of the data relative to these terms as:

$$
\begin{aligned}
L(Z;\gamma,\lambda,A) &= \sum_{C\in\mathcal{C}} P(Z,C;\gamma,\lambda,A) \\
&= \prod_j \sum_{C_j\in\mathcal{C}_j} P(Z_j|C_j;\lambda_j)P(C_j;\gamma,A_{*,j}) \quad \text{(By independence of each locus)}
\end{aligned} \tag{2.1}
$$

where the sum is taken across all causal indicator vector sets $\mathcal{C}$. We note that in order to keep the enumeration of the causal vector sets combinatorially tractable, we restrict the total number of potential causal variants at each locus to three or less in practice (see Figure 2.10 for assessment of run time versus number of causal variants considered). We define the annotation effect on the causal probability through a standard logistic model:

$$
\begin{aligned}
P(C_j;\gamma) &= \prod_i P(C_{ij};\gamma) \\
P(C_{ij};\gamma) &= \left(\frac{1}{1+\exp(\gamma^T A_{ij})}\right)^{C_{ij}} \left(\frac{1}{1+\exp(-\gamma^T A_{ij})}\right)^{1-C_{ij}}
\end{aligned} \tag{2.2}
$$

and relate the causal set of SNPs to the observed association Z-scores under a standard multivariate normal assumption[63, 64, 65] as:

$$
P(Z_j|C_j;\lambda_j) = \mathcal{N}(Z_j;\Sigma_j(\lambda_j\circ C_j),\Sigma_j) \quad \text{(p.d.f of multivariate normal)} \tag{2.3}
$$

where $\lambda_j \circ C_j$ denotes the elemental pairwise multiplication between two vectors. A summary of model parameters can be found in Table 2.8.

2.4.2 Model Fitting

In order to compute the probability of causality, we must first fit the data to our model. We accomplish this through a maximum likelihood estimation over γ. The formulation of our approach lends itself to the standard Expectation Maximization (EM) algorithm. The E-step of the EM involves computing at each locus independently, the posterior probability of each $C_j \in \mathcal{C}_j$ using an application of Bayes Theorem:

$$
P(C_j|Z_j,\gamma^{(t)},\lambda) = \frac{P(Z_j|C_j;\lambda_j)P(C_j;\gamma^{(t)}))}{\sum_{C_j\in\mathcal{C}_j} P(Z_j|C_j;\lambda_j)P(C_j;\gamma^{(t)})} \tag{2.4}
$$

To obtain the posterior probability, $P(C_{ij}|Z_j;\gamma^{(t)},\lambda)$, for each $\text{SNP}_{i,j}$ we marginalize across all $C_j = (C_{1j}, C_{2j}, ..., C_{N_jj})$ such that $C_{ij} = 1$.

$$P(C_{ij}|Z_j,\gamma^{(t)},\lambda) = \sum_{C_j \in \mathcal{C}_j : C_{ij}=1} P(C_j|Z_j,\gamma^{(t)},\lambda) \tag{2.5}$$

Despite the fact that posterior probabilities are calculated independently at each locus, we can set up the objective function to aggregate the results and borrow information across loci to compute estimates of $\gamma^{(t)}$. In doing so, we prevent over fitting of the data to any one locus, offering more robust estimates of the model parameters leading, in turn, to more accurate posterior probabilities. We define our Q function for the M step as follows

$$\begin{aligned}
Q(\gamma,\lambda|\gamma^{(t)},\lambda) &= \sum_j \sum_{C_j} P(C_j|Z_j,\gamma^{(t)},\lambda) \ln P(Z_j, C_j;\gamma^{(t)},\lambda_j) \\
&= \sum_j \sum_{C_j} P(C_j|Z_j,\gamma^{(t)},\lambda) \left(\ln P(C_j;\gamma^{(t)}) + \ln P(Z_j|C_j,\lambda_j)\right) \\
&= \sum_j \sum_{C_j} P(C_j|Z_j,\gamma^{(t)},\lambda) \ln P(C_j;\gamma^{(t)}) + \sum_j \sum_{C_j} P(C_j|Z_j,\gamma^{(t)},\lambda) \ln P(Z_j|C_j,\lambda_j) \\
&= Q(\gamma|\gamma^{(t)}) + Q(\lambda|\lambda)
\end{aligned}$$

thereby partitioning the likelihood, decoupling the estimation of $\gamma's$ from the $\lambda's$. We simplify $Q(\gamma|\gamma^{(t)})$ to obtain

$$\begin{aligned}
Q(\gamma|\gamma^{(t)},\lambda) &= \sum_j \sum_i \sum_{c_{ij}\in 0,1} P(c_{ij}|Z_j;\gamma^{(t)},\lambda) \ln P(c_{ij};\gamma^{(t)}) \\
&= -\sum_i \sum_j P(c_{ij}=1|Z_j;\gamma^{(t)},\lambda) \ln(1+\exp(\gamma^T A_{ij})) \\
&\quad -\sum_i \sum_j P(c_{ij}=0|Z_j;\gamma^{(t)},\lambda) \ln(1+\exp(-\gamma^T A_{ij}))
\end{aligned}$$

which is a concave function whose gradient is simply

$$\begin{aligned}\frac{\partial Q(\gamma|\gamma^{(t)},\lambda)}{\partial\gamma} &= -\sum_i\sum_j P(c_{ij}=1|Z_j;\gamma^{(t)},\lambda)\frac{1}{1+\exp(-\gamma^T A_{ij})}A_{ij} \\ &\quad +\sum_i\sum_j P(c_{ij}=0|Z_j;\gamma^{(t)},\lambda)\frac{1}{1+\exp(\gamma^T A_{ij})}A_{ij}\end{aligned}$$

We optimized this function using the NLopt C++ package's implementation of the limited-memory BFGS algorithm [66], a quasi-Newton method that only requires the objective and the gradient as input [67]. As stated previously, we fix the non-centrality parameters, λ, and only optimize over γ due to the fact that our model would be over-specified otherwise. Specifically, we set the non-centrality parameters at each SNP to the observed Z-score if the absolute Z-score is greater than 3.7 (corresponding to a p-value of 10e-4) or the sign of the observed Z-score times 3.7 otherwise. Simulation results show that our strategy yields high accuracy to detect causal variants among several simulated approaches to approximate λ (Figure 2.8).

2.4.3 Simulation Framework

Starting from the 1000 Genomes (1KG) European samples, we used HAPGEN[68] to simulate fine-mapping data sets over 10Kb loci. We filtered monomorphic/rare SNPs (MAF < 0.01) and normalized genotypes to be mean-centered with unit variance. For each simulation we randomly chose one hundred 10 Kb loci and randomly assigned SNPs to binary annotations at a pre-specified proportion. We drew causal status for each SNP according to the logistic model above and varied γ to induce a desired prior probability for causality for SNPs part of the "functional" annotation, while maintaining an approximately fixed number of causals – typically one per locus in expectation. For example, to induce an 8-fold causal enrichment in a synthetic "functional" annotation that contained 1/3 of the SNPs, the (γ_0, γ_1) values were set to be (4.62, -2.15). We note that the random assignment of causal status would lead to loci with either zero (36), one (34), or multiple causal (30) variants on the average.

Once we established the causal SNPs, we used a linear model to simulate continuous phenotypes such that the causal SNPs aggregated to explain a fixed proportion of the phenotypic variance (h_g^2). This phenotypic variance was partitioned equally amongst all the causal SNPs (qualitatively similar results were obtained when phenotypic variance was unevenly partitioned among causal variants (see Figure 2.11)). In particular, the m^{th} individual's phenotype was drawn according to $Y_m = \sum_{i=1}^{N_c} \beta_i * G_{i,m} + \epsilon_m$, where N_c is the total number of causal variants, β_i is the effect size of the $i'th$ causal SNP, $G_{i,m}$ is number of copies of the risk allele i (randomly assigned as reference or alternate) for individual m, and $\epsilon_m \sim N(0, 1 - h_g^2)$. Finally, we calculated association Z-scores ($Z_{i,j}$) at each SNP i, j by taking the Wald statistic from the regression of the Y on $G_{i,j}$, where Y is a vector of phenotypes for M individuals and $G_{i,j}$ is the vector of corresponding genotypes for the i^{th} SNP at the j^{th} locus. For simulations that required loci greater than 10KB, we instead drew Z-scores from a Multivariate Normal distribution with covariance equal to LD based on the European 1KG and non-centrality parameters at causal sites drawn from a Normal distribution with mean 5 and standard deviation 0.2. When measuring performance of our simulations, we examine the proportion of causal SNPs identified as a function of the average number of SNPs per locus selected for follow-up restricted to loci that contain at least one causal variant (we show in Figure 2.14 that using Positive Predictive Value as a metric of accuracy attains qualitatively similar results).

2.4.4 Existing approaches for Fine Mapping

We compared our approach to a several of existing methods that can be used for fine-mapping[1, 41, 43, 44, 45]. To compute Maller et al.[1] posterior probabilities, we first calculated Bayes factors with the R package, BayesFactor, using the default settings. We converted the resultant Bayes factors into posterior probabilities of association using the following formula: $PPA_i = \frac{BF_i}{\sum_j BF_j}$. We show in Figure 2.6 and Supplementary Note S1 that posterior probabilities approximated directly from the Z-scores give virtually indistinguishable results. We downloaded fgwas[41] version 0.3.4 from GitHub and ran the software using the *-fine* flag. Due to the fact that we fit linear models to obtain Wald

statistics for each SNP, we were able to provide standard errors for the estimates of the prior variance. We segmented our input based on the individual loci as instructed in user manual, but provided a single file that contained all the fine-mapping SNPs so that fgwas could compute enrichment. The Guan and Stephens [45] method is implemented in the software *piMass* which we ran using the flags and MCMC parameters given in the user manual as defaults (burn-in =1000, samples = 100,000, step-size = 10). We used the posterior inclusion probabilities (PIPs) that had undergone Rao-Blackwellisation for prioritization due to the fact these had superior performance over naive PIPs. The R package implementing LLARRMA [44] was run using the default settings. Zuber et al. was implemented in the R package, care, which we also applied to the data using the default settings. We prioritized variants using the square of the CAT scores as described in [43]. We note that with exception fgwas, all the aforementioned methods were applied to each locus independently. Conditional analysis is a common procedure used to tease out secondary signals at associated loci[69]. For a single locus, we iteratively condition on the SNP most strongly associated with the simulated phenotype. We accomplish this in a step-wise fashion through marginal regression of the phenotype onto each SNP and subsequently conditioning on the one that is most significantly associated. At each iteration a new SNP will enter the regression model as a covariate until all the causal SNPs have been selected. The order in which the SNPs enter the model provides a natural ranking thus enabling us to compare iterative conditioning to other methods that rank SNPs probabilistically. As expected, we show in Figure 2.5 that conditioning is suboptimal for fine-mapping.

2.4.5 Functional information

We explored whether integration of the location of tissue-specific regulatory and coding DNA regions could increase resolution in statistical fine-mapping. The ENCODE[13] project provided a wealth of genomic landmarks that were systematically integrated to segment the genome into seven major classes: transcription start site and predicted promotor region (TSS) (1.2%), predicted promotor flanking region (PF) (0.7%) predicted en-

hancer (E) (1.8%) predicted weak enhancer (WE) (2.5%), CTCF-enriched element (CTCF) (0.1%) predicted transcribed region (T) (19.3%) and finally, predicted repressed or low-activity region (R) (69.6%). We examined these genomic segmentations for the six primary ENCODE cell lines: gm12878 (lymphoblastoid), h1hesc (embryonic stem cells), helas3 (cervical cancer), hepg2 (liver carcinoma), huvec(umbilical vein endothelial cells) , and k562 (chronic myelogenous leukemia). In addition, we also explored 403 DNase I Hypersensitivity Sites spanning numerous tissues and cell lines. Of these 403 DHS I maps, 349 came from Maurano et al. [48], 73 DHS I annotations from Thurman et al. [70], with the remaining DHS annotation being an overall DHS map derived from UCSC genome browser. These annotations have been used recently in the context of GWAS [41].

2.4.6 Measuring enrichment significance

Due to the fact that we fit our model using maximum likelihood, a natural way to ascribe statistical confidence to the inferred parameters is to use a likelihood ratio test. For example, to calculate the significance of a single annotation, we can compare a fitted null model to a model that contains the annotation under consideration using the following test statistic: $-2 * ln(L(Z; \hat{\gamma}_0, \lambda) + 2 * ln(L(Z; \hat{\gamma}_0, \hat{\gamma}_1, \lambda))$. We demonstrate in simulations that under the null, this test statistic follows approximately its theoretical $\chi^2_{df=1}$ distribution (see Figure 2.13).

In addition to a point estimate for the enrichment of functional annotation, it would be useful to derive an estimate of the variance. Unfortunately, the complex structure of the likelihood makes it difficult to derive an analytically tractable parameter covariance estimator. However, since we assume fine-mapping loci to be independent, we propose to use bootstrapping (i.e. re-sampling entire loci with replacement) and subsequently re-fitting the model (see Methods). We confirm that such a strategy does indeed reproduce a correct estimate of the parameter variance in simulations. We find that the mean bootstrap standard deviation largely mirrors the ”true” standard deviation of the parameter estimates (see Figure 2.12). As a result, a confidence interval based on the bootstrap stan-

dard deviation will attain desirable coverage properties due to the fact that estimation of the model parameters is unbiased.

2.4.7 An optimization framework for selecting the number of SNPs to follow-up

The budget of a fine mapping follow-up study constrains the total number of causal variants to be further examined for evidence of causality. This motivates approaches that, in addition to providing a prioritization of SNPs, also identify an optimal number of SNPs to be tested. We introduce here a benefit-to-cost framework for selecting the optimal number of SNPs for follow-up. Our framework assumes that every causal variant identified adds a benefit (B) while every selected variant is tested at a cost (C); therefore, the utility function we propose to maximize is U = B*Nc - C*Nt , where Nc is the total number of true causal variants identified from the total number of selected SNPs. A key parameter of the utility framework is the ratio of $r = B/C$ of benefit to utility.

2.4.8 Lipids Data Set

Publicly available GWAS summary data across four blood lipids phenotypes was downloaded from public access websites [71]. Data was part of a meta-analysis conducted in $> 100,000$ individuals of European ancestry that examined four plasma lipid traits (number of significant loci): LDL cholesterol (14 loci), HDL cholesterol (37 loci), trigylcerides (23 loci), and total cholesterol (24 loci). From the original 2.0M SNPs, we imputed an additional 5.3 million Z-scores using ImpG-Summary [62]. For each significant GWAS hit reported by Teslovich et al., we centered a 100KB window on the lead SNP and estimated LD from the European reference panel of the 1KG. We chose a conservative window of 50Kb on either side of the GWAS hit, as it has been previously shown that within European populations, average LD decays after 25KB [72]. These loci contained an average of 718 SNPs in the 1000 genomes reference panel, of which we were able to on average accurately impute 261 (see Table 2.16). This resulted in 2837 (10778), 1231 (3903), 1693 (5504), 1615 (5513) SNPs (with 1KG imputation) to which we fit our model for HDL, LDL,

TC, and TG respectively. In addition, we created the corresponding pool of functional annotations described above for every SNP in a window. We analyzed the dataset using PAINTOR in two phases. In the first phase we fit our model for each annotation independently to ascertain the functional annotations most phenotypically relevant. We did this for all four lipid traits for both the original and densely imputed data sets. After running PAINTOR marginally on each annotation, we selected the the top five most significant annotations for the final model (denoted with a * in Table: 2.5). We note that in the case of experimental replicates (i.e. same tissue and class), we only report the top replicate.

2.5 Appendix: Single Variant Fine mapping

Here, we demonstrate how single variant fine-mapping can be executed using only the vector of Z-scores. Under the assumption that the causal variant has been typed or imputed, the joint distribution of a set of association statistics Z given that the j'th SNP is causal follows a multivariate normal:

$$
\begin{aligned}
Z|C_j &\sim \mathcal{N}(\Sigma\Lambda_j, \Sigma) && (2.6)\\
&\propto \exp\left(-\frac{1}{2}(Z-\Sigma\Lambda_j)'\Sigma^{-1}(Z-\Sigma\Lambda_j)\right) && (2.7)\\
&= \exp\left(-\frac{1}{2}(Z-\Sigma\Lambda_j)'(\Sigma^{-1}Z-\Lambda_j)\right) && (2.8)\\
&= \exp\left(-\frac{1}{2}(Z'\Sigma^{-1}Z-2\Lambda_j'Z+\Lambda_j'\Sigma\Lambda_j)\right) && (2.9)
\end{aligned}
$$

Using the observed Z-score at SNP j to approximate the Non-Centrality Parameter, λ_j, it follows that

$$
\begin{aligned}
\Lambda_j'\Sigma\Lambda_j &= z_j^2 && (2.10)\\
\Lambda_j'Z &= z_j^2 && (2.11)
\end{aligned}
$$

Therefore the above expression reduces to

$$\exp\left(-\frac{1}{2}(Z'\Sigma^{-1}Z - z_j^2)\right) \tag{2.12}$$

$$= \exp\left(-\frac{1}{2}Z'\Sigma^{-1}Z\right)\exp\left(\frac{1}{2}z_j^2\right) \tag{2.13}$$

Placing a uniform prior that any SNP within a fine-mapping regions is causal, we have

$$P(C_j|Z) \propto P(Z|C_j) \implies \tag{2.14}$$

$$\tag{2.15}$$

$$P(C_j|Z) = \frac{\exp\left(-\frac{1}{2}Z'\Sigma^{-1}Z\right)\exp\left(\frac{1}{2}z_j^2\right)}{\sum_{j=1}^{M}\exp\left(-\frac{1}{2}Z'\Sigma^{-1}Z\right)\exp\left(\frac{1}{2}z_j^2\right)} \tag{2.16}$$

$$\tag{2.17}$$

$$= \frac{\exp\left(\frac{1}{2}z_j^2\right)}{\sum_{j=1}^{M}\exp\left(\frac{1}{2}z_j^2\right)} \tag{2.18}$$

Thus, under the assumption of a single causal variant, posterior probabilities can be obtained independent of LD.

2.6 Tables

Method	Summary Data	Incorporates Annotation	Assumed Num. of Causal Variants	Variant Ranking Across All Loci: Causals Identified 50%	Variant Ranking Across All Loci: Causals Identified 90%	Variant Ranking Independently at Each Locus: Causals Identified 50%	Variant Ranking Independently at Each Locus: Causals Identified 90%
P-value	Yes	No	n/a	5.74	12.60	2.94	19.15
CAT score [43]	No	No	n/a	2.04	13.29	2.56	17.80
LLARRMA [44]	No	No	Mult	1.98	21.93	2.46	23.11
piMass-RB [45]	No	No	Mult	2.83	16.31	2.18	15.15
Maller et al. [1]	Yes	No	Single	2.68	25.44	2.96	19.13
fgwas (no annot)	Yes	No	Single	2.69	25.48	2.95	19.11
PAINTOR (no annot,1CV)	Yes	No	Single	2.69	22.49	2.95	19.09
fgwas [41]	Yes	Yes	Single	1.95	24.77	2.05	17.37
PAINTOR (1CV)	Yes	Yes	Single	1.95	21.51	2.03	17.43
PAINTOR (no annot)	Yes	No	Mult	1.76	12.25	2.24	16.86
PAINTOR	Yes	Yes	Mult	1.26	10.42	1.61	13.68
PAINTOR True	Yes	Yes	Mult	1.23	10.22	1.59	13.48

Table 2.1: Summary of performance for various fine-mapping methods benchmarked using the average number of SNPs per locus selected to find (50%,90%) of all causal variants. We simulated a trait with 100 risk loci explaining $h_g^2 = 0.25$ fine-mapped through sequencing of N=10,000 samples and assessed accuracy only at loci that harbor at least one casual variant (64 loci on the average). We explored two methods to prioritizing variants: (1) "Variant Ranking Across All Loci" prioritizes SNPs across all loci while (2) "Variant Ranking Independently at Each Locus", first prioritizes variants at each risk locus followed by merging across all loci. We note that PAINTOR 1CV and/or no annot corresponds to running PAINTOR assuming a single causal variant and/or not providing access to annotations. PAINTOR True did not empirically estimate enrichment but used the true enrichment values for each functional annotation data

ρ-level	Method	Annotations	Causals Identified	SNPs Selected
90%	Maller et al.	-	64.2	265.0
	fgwas	+	64.5	209.6
	PAINTOR	-	91.9	510.3
	PAINTOR	+	91.2	393.7
95%	Maller et al.	-	69.6	343.7
	fgwas	+	70.2	290.8
	PAINTOR	-	97.2	687.8
	PAINTOR	+	97.0	567.2
99%	Maller et al.	-	77.7	506.6
	fgwas	+	77.9	457.3
	PAINTOR	-	102.6	1074.4
	PAINTOR	+	102.7	954.3

Table 2.2: Leveraging functional priors leads to improved fine-mapping resolution. We define an ρ-level confidence set as the number of SNPs we need to select in order to consume an ρ fraction of the total posterior probability mass over all loci. Results in the table correspond to averaging over 500 independent simulations where the average number of true causals SNPs per simulation was 109.2. The size of 90%, 95%, and 99% confidence sets are reduced by 22.8%, 17.5% and 11.1% when incorporating functional annotations as prior probabilities. Methods that assume one causal variant are miss-calibrated due to loci with multiple causals.

Locus Size	%Causal	p-value	Maller et al	PAINTOR
10Kb	10%	1.04	0.17	0.17
	50%	5.73	2.20	1.35
	90%	12.87	27.35	11.41
25Kb	10%	1.68	0.16	0.16
	50%	8.88	2.73	1.57
	90%	21.93	52.32	16.01
50Kb	10%	2.50	0.17	0.16
	50%	13.69	3.65	1.87
	90%	36.92	110.69	24.07

Table 2.3: Performance of PAINTOR compared to standard methodologies at variable sized loci. To expedite simulations, we used a modified version of the simulation setup. As before, causal SNPs were drawn according to a logistic prior such that in expectation there were a total of 100 causal variants – we did not enrich causal in any annotations. For this experiment, Z-scores were drawn directly from a multivariate normal distribution; this gave virtually identical results to using simulated genotypes derived from HAPGEN (see Methods).We find that PAINTOR increasingly outperforms existing methodologies as the size of the loci become larger.

Method	Causal Variants	N=1	N=5	N=10	N=25
PAINTOR	Typed	17.6 (17.6)	20.4 (3.2)	21.6 (1.6)	23.1(0.5)
	Masked	22.7 (22.7)	21.7(6.9)	22.0 (4.0)	23.3 (1.8)
Random	Typed	32.1 (32.1)	30.2 (8.9)	30.3 (4.5)	30.1 (1.7)
	Masked	32.0 (32.0)	30.6 (9.1)	30.6 (4.7)	30.2 (1.9)

Table 2.4: Average median (minimum) distance in Kb to true causal variant(s) for SNPs in the set of top N SNPs when causal variant is either present or absent from the fine-mapping data set. We simulated one 100Kb locus with causal status drawn from a uniform prior. We then masked the causal variant(s) to explore how this would effect fine-mapping resolution.

Phenotype	Cell Line	Type	Frequency	Log2 Relative Probability to be Causal	P.value
HDL	hepg2	TSS*	2.2%	3.46	1e-5
	-	Coding Exons*	1.4%	3.63	1e-3
	K562	Weak Enhancer*	0.7%	3.74	0.01
	MCF7	DHS*	30.5%	1.18	0.03
	gm12878	TSS*	1.8%	2.38	0.04
LDL	fKidney	DHS*	40.4%	2.23	6e-3
	fLung	DHS*	34.3%	1.99	7e-3
	-	Coding Exons*	3.9%	2.92	0.01
	Hepatocytes	DHS*	38.2%	1.97	0.02
	HAsp	DHS*	33.7%	1.76	0.02
TC	hepg2	Repressed*	53.9%	-1.87	6e-3
	fLung	DHS*	30.4%	2.18	6e-3
	fIntestine(Lg)	DHS*	18.8%	1.93	0.01
	hepg2	Transcribed*	31.2%	1.64	0.01
	NHDF_Neo	DHS*	26.0%	1.76	0.02
TG	-	Coding Exons*	1.5%	3.42	3e-3
	hepg2	Repressed*	57.9%	-1.63	6e-3
	GM19238	DHS*	22.4%	1.71	0.01
	fIntestine (Sm)	DHS*	29.9%	1.67	0.01
	-	Non-coding Exon*	2.5%	2.81	0.02
	-	DNASE UCSC	21.7%	1.72	0.02

Table 2.5: Top 5 most significant annotations for lipid traits. Displayed are the log2 relative probabilities of SNPs to be causal if they fall within the listed annotation. *Indicates use in final PAINTOR model for the phenotype.

rsID	Chrom	Pos	-Log10(P.value)	PAINTOR Probability	Annotations
rs1366544	chr16	56964719	43.86	> 0.99	K562 Weak Enchancers ,MCF7 DHS
rs1645788	chr19	54808174	10.79	> 0.99	MCF7 DHS
rs3136447	chr11	46744368	16.08	> 0.99	hepg2 TSS , MCF7 DHS
rs7241918	chr18	47160953	48.86	> 0.99	-
rs1077834	chr15	58723479	83.32	> 0.99	-
rs367070	chr19	54800500	14.69	> 0.99	-
rs3809630	chr16	67879400	32.32	0.99	hepg2 TSS , MCF7 DHS , gm12878 TSS
rs7239867	chr18	47164717	47.53	0.99	-
rs13107325$\star$	chr4	103188709	10.44	0.97	Coding Exons
rs7607980$\star$	chr2	165551201	9.71	0.96	hepg2 TSS ,Coding Exons,MCF7 DHS
rs4366775	chr17	76382079	8.50	0.93	gm12878 TSS
rs737337	chr19	11347493	8.81	0.92	hepg2 TSS ,Coding Exons , MCF7 DHS
rs4490057	chr17	76375095	8.20	0.90	hepg2 TSS ,MCF7 DHS , gm12878 TSS

Table 2.6: SNPs with posterior probability causality > 0.90 for HDL phenotype across the 37 risk loci (Results for TG/TC/LDL in Tables 2.13,2.14,2.15). $\star$ denotes a non-synonymous variant.

Phenotype	Total SNPs	SNPs with P-value <5e-8	Annotations	# SNPs	# Loci	# SNPS per locus
HDL	10778	1792	–	926	37	25.0
			+	778		21.0
LDL	3903	955	–	112	14	8
			+	83		5.9
TG	5513	975	–	488	23	20.3
			+	324		13.5
TC	5504	1381	–	390	24	17.0
			+	314		13.7
Average	6425	1276	–	479	24.5	17.5
			+	375		13.5

Table 2.7: Reduction in the number of SNPs in the 90% Credible Set after incorporating functional annotations. Show here are the number of SNPs in the 90% Confidence Set for each of the lipid phenotypes as estimated using PAINTOR. After marginally running PAINTOR on the entire pool of annotations, we selected the top five annotations for each trait and fit full trait-specific models on each of the densely imputed data sets. We compared PAINTOR with or without integration of functional annotation data. The magnitude in the reduction in the size of the confidence set approximately mirrors what we observe in simulations.

Parameter	Description
N_j	Number of SNPs at the the j^{th} locus
Z_j	Vector of Z-scores (1 x N_j)
Σ_j	Linkage disequilibrium matrix consisting of pairwise Pearson Correlation coefficients between SNPs (N_j x N_j)
$A_{i,j}$	Vector of annotations for the i^{th} SNP. $A_{ijk} = 1$ if member of annotation (1 x $(K+1)$)
λ_j	Vector of Non-centrality parameters (NCPs) (1 x N_j)
C_j	Indicator vector giving the causal status of all the SNPs at a locus. $C_{ij} = 1$ if the i^{th} SNP is causal (1 x N_j).
$\mathcal{C}_j$	Set of all possible causal configurations. ($\lvert\mathcal{C}_j\rvert = \sum_{i=0}^{S} \binom{N_j}{i}$, where S = number of causals one wants to consider at a locus).

Table 2.8: List of model parameters for the j^{th} locus ($j \in [1, L]$ where L is the total number of fine-mapping loci).

Method	Operate on Summary Data	Integrates Functional Priors	Can Handle Multiple Causal Variants
PAINTOR	Yes	Yes	Yes
Maller et al.	Yes	No	No
fgwas	Yes	Yes	No
piMass	No	No	Yes
LLARRMA	No	No	Yes
CAT score	No	No	Yes

Table 2.9: Basic summary of fine-mapping methods assessed. We highlight the key contribution of our approach is that we can use PAINTOR to do fine-mapping with functional priors while modeling multiple causal variants directly from summary association statistics (Z-scores).

Threshold on	PAINTOR			PAINTOR No Annotation		
Posterior	Total SNPs	Causals	Objective	TotalSNPs	Causals	Objective
0.10	232.85	78.61	1.00	265.16	73.79	1.00
0.20	132.59	65.52	0.93	128.20	56.09	0.90
0.30	93.51	57.09	0.84	80.28	45.68	0.77
0.40	72.61	50.84	0.76	57.92	39.20	0.68
0.50	57.96	45.00	0.68	43.03	33.23	0.59
0.60	48.78	40.54	0.62	35.88	30.06	0.53
0.70	41.33	36.30	0.56	31.23	27.59	0.48
0.80	34.20	31.60	0.48	27.16	25.17	0.44
0.90	27.56	26.51	0.40	22.95	22.07	0.38

Table 2.10: Performance of PAINTOR with and without integrating annotations if thresholding on the posterior probability (Average number of causals per simulation = 108). The objective function is given as ratio from the maximum objective at a cost to benefit ratio of 10.

% Causal	N=1000	N=2500	N=5000	N=10000
10%	0.66	0.25	0.17	0.16
50%	8.0	4.2	2.28	1.6
90%	25.9	19.0	12.5	10.8

Table 2.11: Performance of PAINTOR as a function of sample size. We fixed the proportion of phenotypic variance explained in a simulated trait to $h_g^2 = 0.25$ and selected a variable number of individuals to conduct fine-mapping experiments over. Displayed are the average number of SNPs per locus that need to be selected in order to identify the listed percentage of causals.

Significance Quartile	Fraction of Causals	PAINTOR	PAINTOR No Annot	Percent Change
Bottom 25%	0.50	17.98	26.37	31.79%
	0.90	130.32	165.83	21.41%
Top 25%	0.50	34.62	50.28	31.13%
	0.90	236.31	252.64	6.46%

Table 2.12: Incorporating prior probabilities provides larger benefit when Z-scores at the causal SNPs are smaller. Here, we illustrate the efficacy of fine-mapping at loci where the p-value at the causal SNPs fall in either the top or bottom quartile of significance (as indicated by the absolute z-score).

rsID	Chrom	Pos	-Log10(P.value)	PAINTOR Probability	Annotations
rs41290120	chr19	45382675	181.02	1.00	fKidney DHS, fLung DHS, Hepatocytes DHS, HAsp DHS
rs4420638	chr19	45422946	146.36	1.00	Hepatocytes DHS
rs5930	chr19	11224265	32.88	1.00	fKidney DHS, Coding Exons
rs4953023	chr2	44074000	32.78	1.00	fKidney DHS, fLung DHS, Hepatocytes DHS
rs6511720	chr19	11202306	116.67	1.00	fKidney DHS , fLung DHS, Hepatocytes DHS, HAsp DHS
rs7746081	chr6	16126934	13.26	0.98	fKidney DHS, fLung DHS, Hepatocytes DHS, HAsp DHS
rs629301	chr1	109818306	170.32	0.94	fKidney DHS ,fLung DHS, Hepatocytes DHS, HAsp DHS
rs1564348	chr6	160578860	17.07	0.93	fLung DHS ,HAsp DHS

Table 2.13: LDL SNPs attaining PAINTOR posterior probabiliites > 0.9 with functional annotations.

rsID	Chrom	Pos	-Log10(P.value)	PAINTOR Probability	Annotations
rs34006994	chr1	25780668	9.51	1.00	hepg2 Transcribed
rs2000999	chr16	72108093	23.79	1.00	fLung DHS, fIntestine(Lg) DHS, hepg2 Transcribed, NHDF_neo DHS
rs12916	chr5	74656539	46.35	0.97	fLung DHS, hepg2 Transcribed
rs6882076	chr5	156390297	27.43	0.95	hepg2 Repressed
rs7570971	chr2	135837906	8.02	0.95	hepg2 Transcribed

Table 2.14: TC SNPs attaining PAINTOR posterior probabiliites > 0.9 with functional annotations.

rsID	Chrom	Pos	-Log10(P.value)	PAINTOR Probability	Annotations
rs1260326 ⋆	chr2	27730940	132.55	1.00	Coding Exons ,Non-coding Exons
rs138022915	chr8	19885934	98.98	1.00	hepg2 Repressed
rs138570705	chr15	44266730	-28.19	1.00	hepg2 Repressed , GM19238 DHS
rs4665985	chr2	27753878	52.44	1.00	hepg2 Repressed
rs5110⋆	chr11	116691634	34.09	1.00	Coding Exons , hepg2 Repressed, GM19238 DHS ,fIntestine(Sm)
rs964184	chr11	116648917	227.68	1.00	fIntestine(Sm)
rs114366307	chr8	19885726	98.98	1.00	hepg2 Repressed
rs11743303	chr5	55859952	9.13	1.00	hepg2 Repressed , GM19238 DHS, fIntestine(Sm)
rs2412710	chr15	42683787	8.03	0.99	GM19238 DHS , fIntestine(Sm)

Table 2.15: TG SNPs attaining PAINTOR posterior probabiliites > 0.9 with functional annotations. $\star$ denotes a non-synonymous variant.

	HDL	LDL	TC	TG
1KG SNPS	766.81	736.71	677.75	694.57
Well-imputed	292.30	279.79	230.33	240.70

Table 2.16: Average number of SNPs that were well-imputed at the loci for the four lipid phenotypes. The top row corresponds to the average number of common SNPs in the 1000 Genomes reference panel at these loci. The bottom row corresponds to the average number of SNPs that were imputed with accuracy > 0.6 at these loci.

Percentile	HapMap	1000 Genomes
$\leq 10^{th}$	-8.73	-9.49
$\geq 90^{th}$	1.90	1.97

Table 2.17: Imputation boosts estimates of enrichment/depletion. The original data set was imputed up to the HapMap. Using ImpG-Summary we further imputed Z-scores up to the 1000 genomes reference panel. We combined enrichment estimates across all 4 phenotypes and examined the tails of log2 enrichment distributions.

2.7 Figures

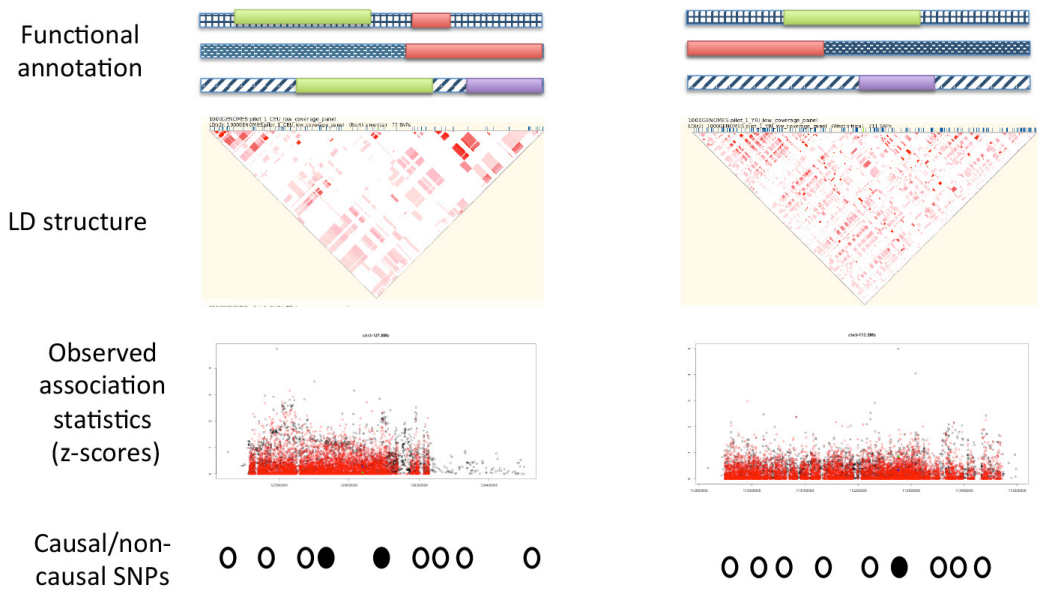

Figure 2.1: Illustration of model input. PAINTOR is a statistical model for incorporating functional annotations on top of association statistics to ascribe probabilistic confidence of causality to the SNPs at the loci. Depicted here are two loci with functional annotations from three different cell lines/tissues and three different classes. Causal variants are enriched within the green annotation class while depleted from others. PAINTOR is designed to upweight (with probability mass) SNPs residing in the green annotation while down-weighting SNPs residing in the red annotation.

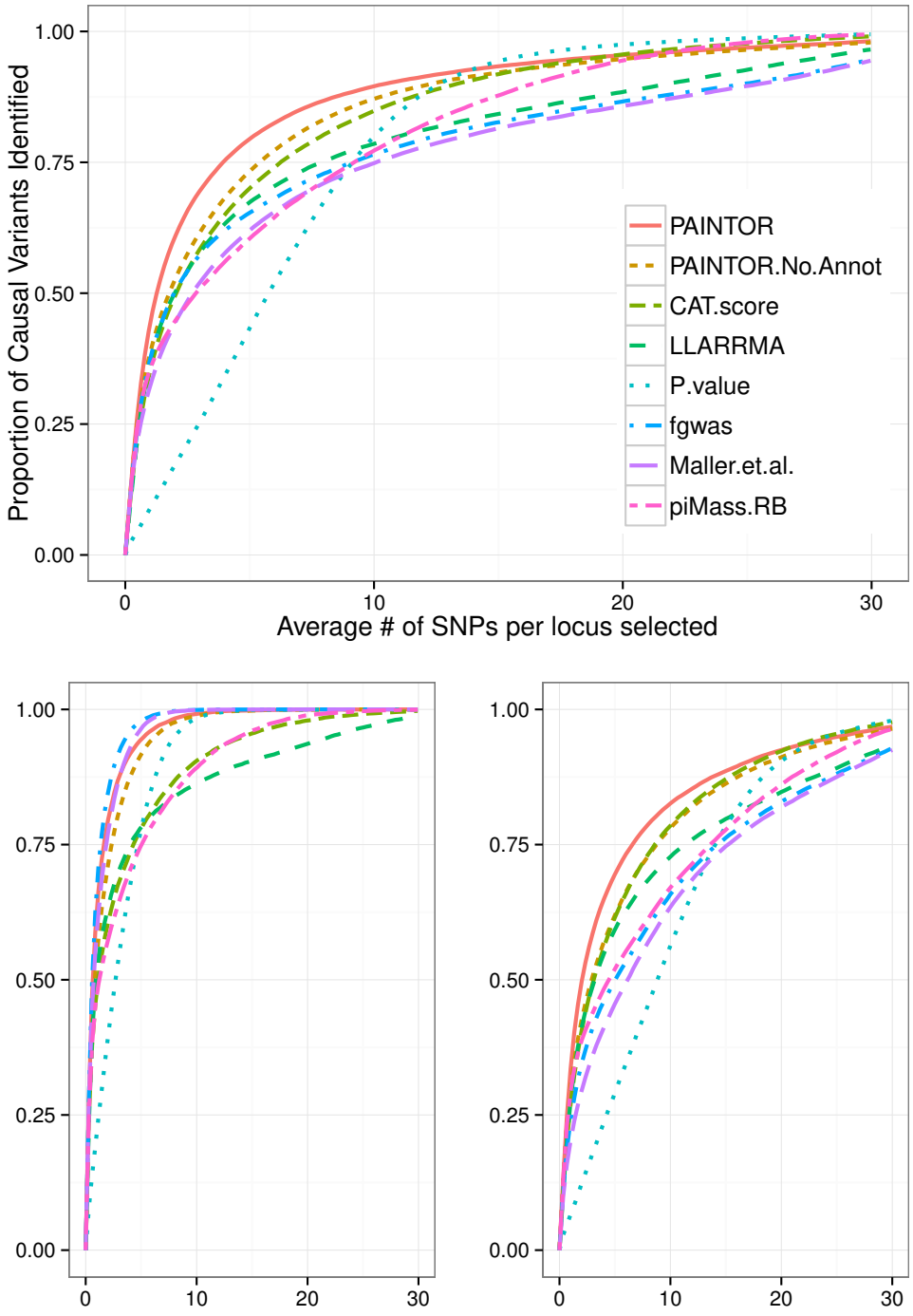

Figure 2.2: Accuracy of enrichment estimation for a synthetic annotation that contains 8-fold depletion to 8-fold enrichment of causal variants across simulations of fine-mapping data sets over 100 loci. Using a background and a synthetic functional annotation at a frequency of $1/3$ (A_0, A_1), we simulated with annotation effect sizes such that in expectation, we attained approximately 100 causal variants while maintaining enrichment at a fixed point.

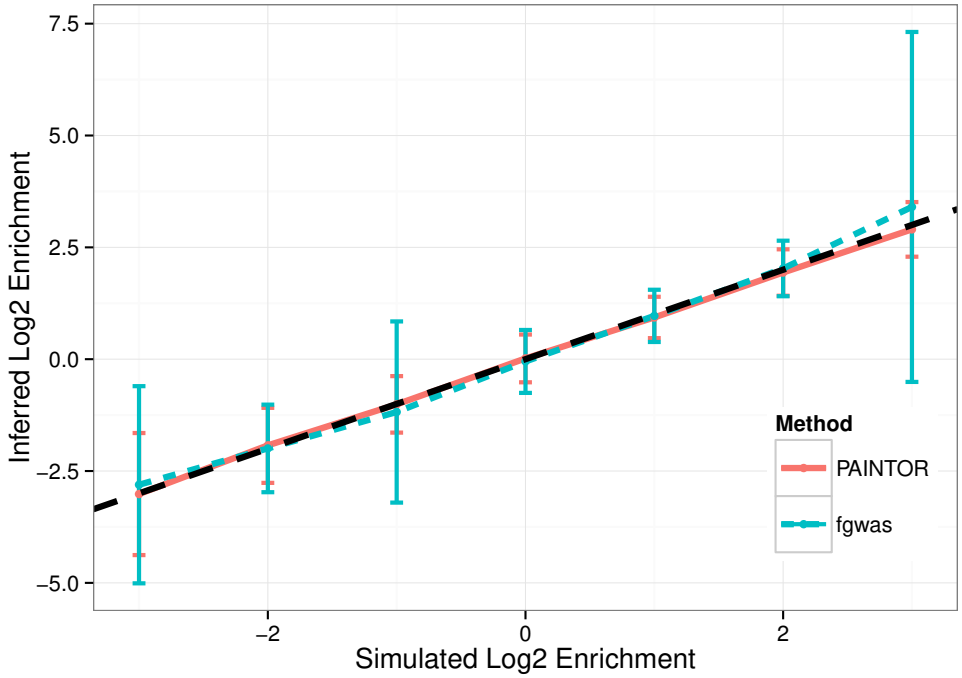

Figure 2.3: Accuracy of enrichment estimation for a synthetic annotation that contains 8-fold depletion to 8-fold enrichment of causal variants across simulations of fine-mapping data sets over 100 loci. Using a background and a synthetic functional annotation at a frequency of $1/3$ (A_0, A_1), we simulated with annotation effect sizes such that in expectation, we attained approximately 100 causal variants while maintaining enrichment at a fixed point. We used the standard simulation parameters, fixing the variance explained by these 100 loci to 0.25 and using $N = 10000$ genotypes. We discarded simulations where fgwas fails to converge (see Methods).

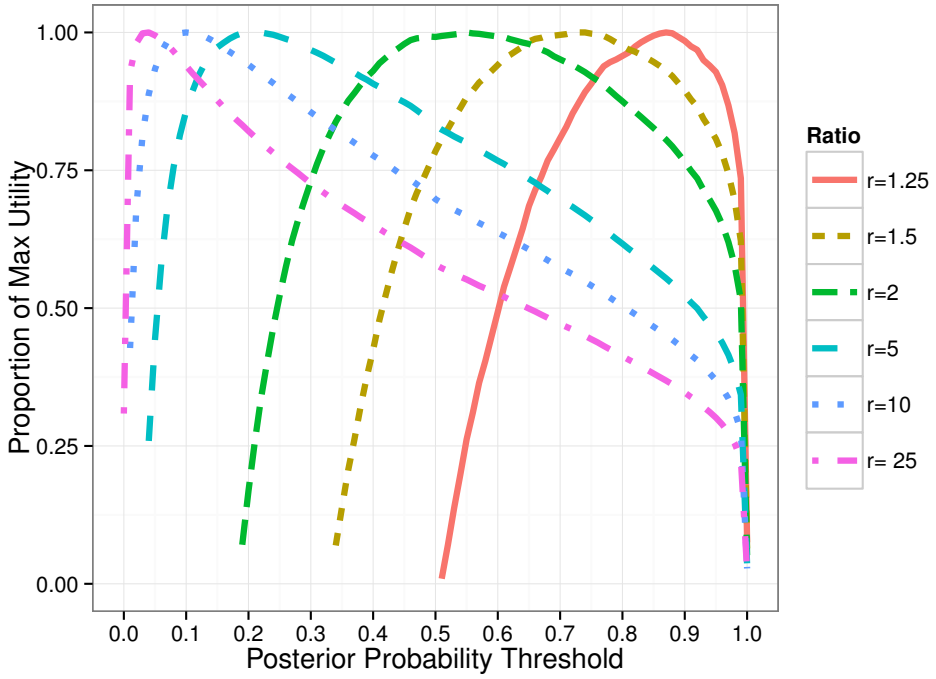

Figure 2.4: Thresholding on posterior probabilities provides a principled way to assess utility. We demonstrate how utility curves are optimized by selecting SNPs that achieve a minimum posterior probability threshold at various benefit-to-cost ratios (R). The total number of SNPs selected at the maximum utility for R = (1.25, 1.5, 2, 5, 10, 20) is (29.8, 39.2, 52.4, 119.1, 221.4, 405.4) which identifies approximately (29.8, 35.6, 43.4, 65.33, 79.9, 91.8) causal variants.

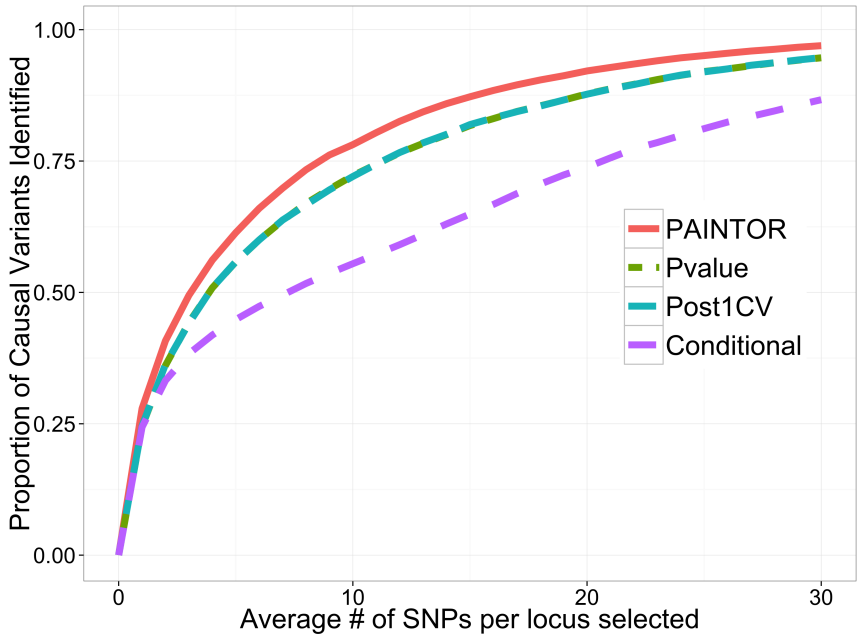

Figure 2.5: Single locus fine-mapping using four different prioritization strategies. Using HAPGEN-derived genotypes from a randomly selected a 10KB locus on chromosome 1, we simulated 10,000 fine-mapping data over N=2500 samples at a locus that explains 0.5% of variance in the phenotype. Each variant has a prior probability of 1/L (where L is the total number of variants at the locus) to be casual; the total variance was divided equally among variants when multiple causal variants were present. As previously observed, prioritization under the assumption of a single causal variant is identical to ranking based on p values at a single locus.

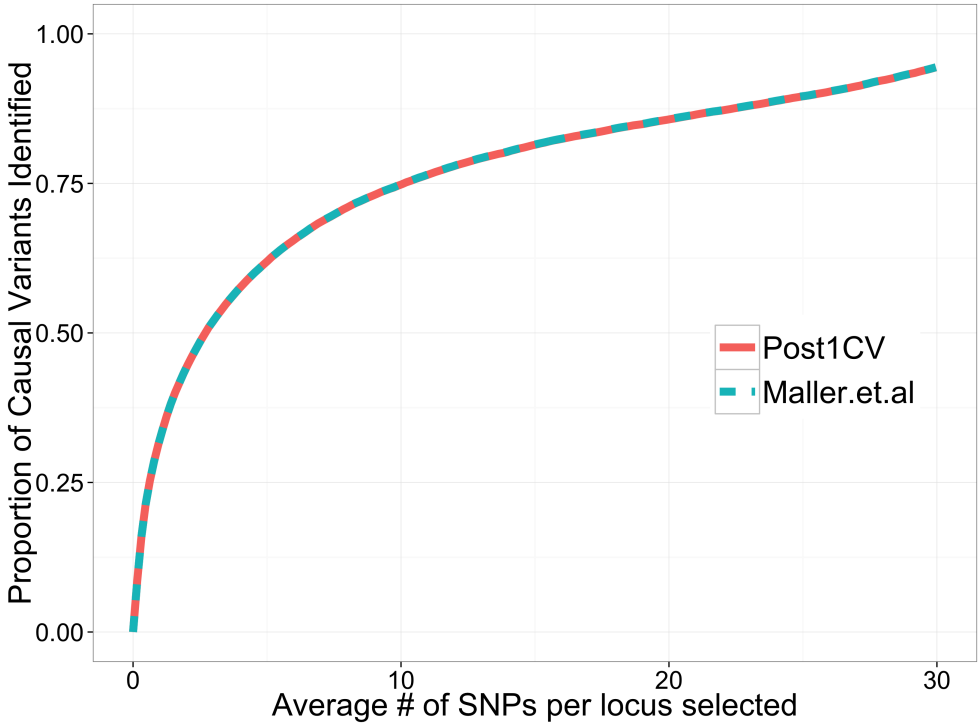

Figure 2.6: Posterior probabilities for causality under the assumption of a single causal variant approximated from z-scores give indistinguishable performance to that of the Bayesian approach described in Maller et al. [1]. Using the standard simulation framework ($h_g^2 = 0.25, N = 10,000$) we calculated posterior probabilities from either Bayes Factors computed using the R package BayesFactor or directly from the association statistics. We then used these posterior probabilities to rank SNPs across all causal loci. The average tau rank correlation between the resulting posterior probabilities is > 0.99.

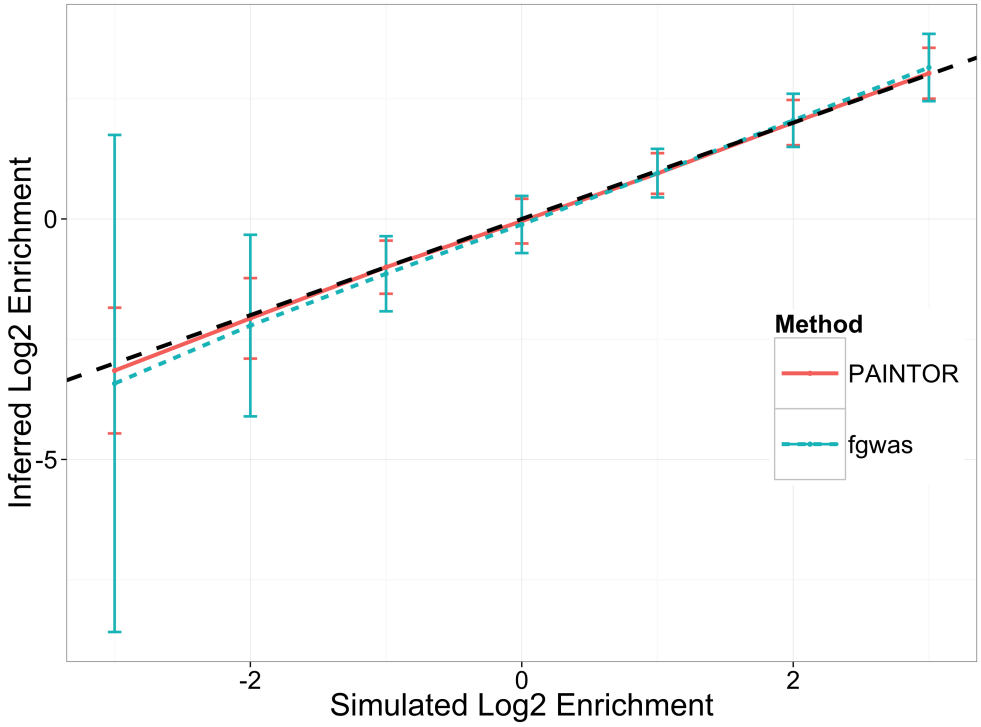

Figure 2.7: Contiguous annotations do not lead to appreciably different performance to randomly assigned annotations. Displayed here is the accuracy of enrichment estimation for a synthetic annotation that contains 8-fold depletion to 8-fold enrichment of causal variants across simulations of fine-mapping data sets over 100 loci. We enriched causal variants in an annotation that spanned a block 1/3 of the size of the locus and simulated with annotation effect sizes such that in expectation, we attained approximately 100 causal variants while maintaining enrichment at a fixed point. We used the standard simulation parameters, fixing the variance explained by these 100 loci to 0.25 and using $N = 10000$ genotypes. We discarded simulations where fgwas fails to converge (see Methods).

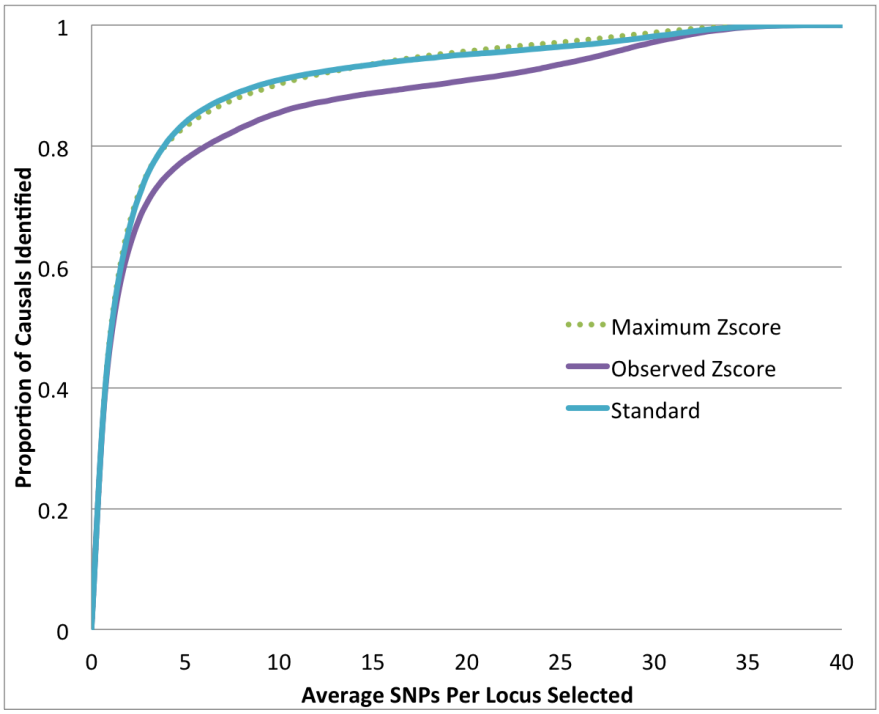

Figure 2.8: Performance using different strategies for approximating the non-centrality parameters λ. Observed Z-score corresponds to setting the $\lambda's$ to the observed z-score at that SNP. Maximum z-score corresponds to setting the NCPs to the maximum z-score at the locus times the sign of the observed z-score. Standard NCP's is the strategy described in the main Methods section wherein the NCP's are set to to the observed Z-score if the absolute Z-score is greater than 3.7 (corresponding to a p-value of 10e-4) or the sign of the observed Z-score times 3.7 otherwise.

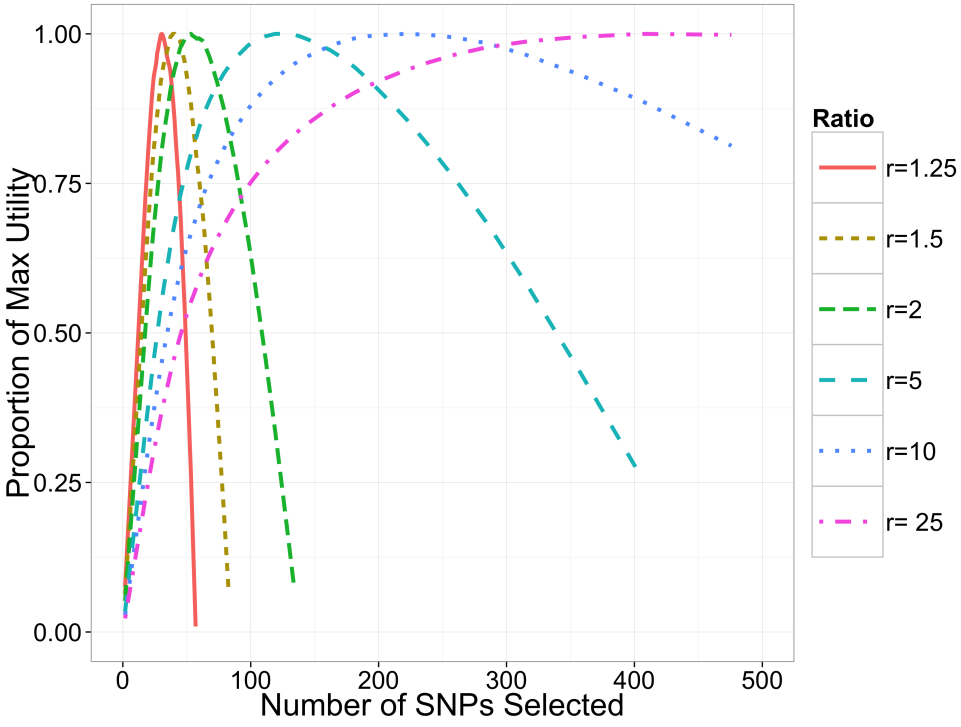

Figure 2.9: Selection of optimal fine-mapping set according to an utility function. Using our standard simulation parameters ($N = 10,000$ and $h_g^2 = 0.25$), causal variants were enriched in three functional annotations at relative marginal probabilities of 9.5, 5.7,and 3.65. Since different ratios will give different scales for the utility function, we normalize the output by the maximum utility.

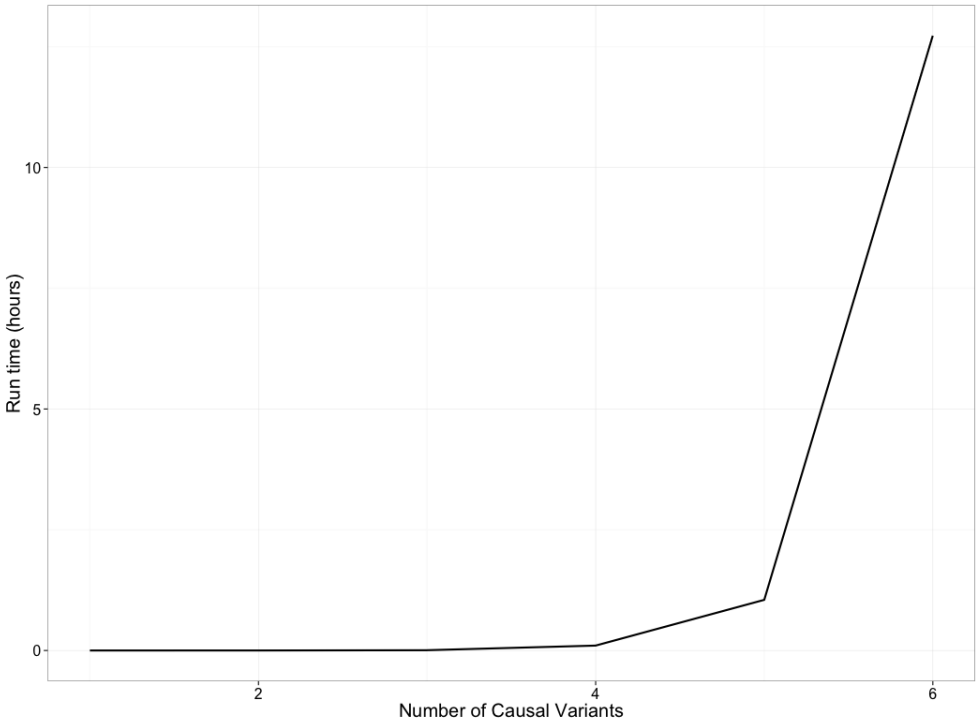

Figure 2.10: Runtime scales exponentially as the number of causal variants integrated over increases. We assesed run-time within the context of our standard simulation framework (ten simulations per point) and varied the number of causal variants PAINTOR integrated over. As the results suggest, we are required in practice to restrict the number of causal variants to a small fixed constant c in order to keep the computational burden reasonable.

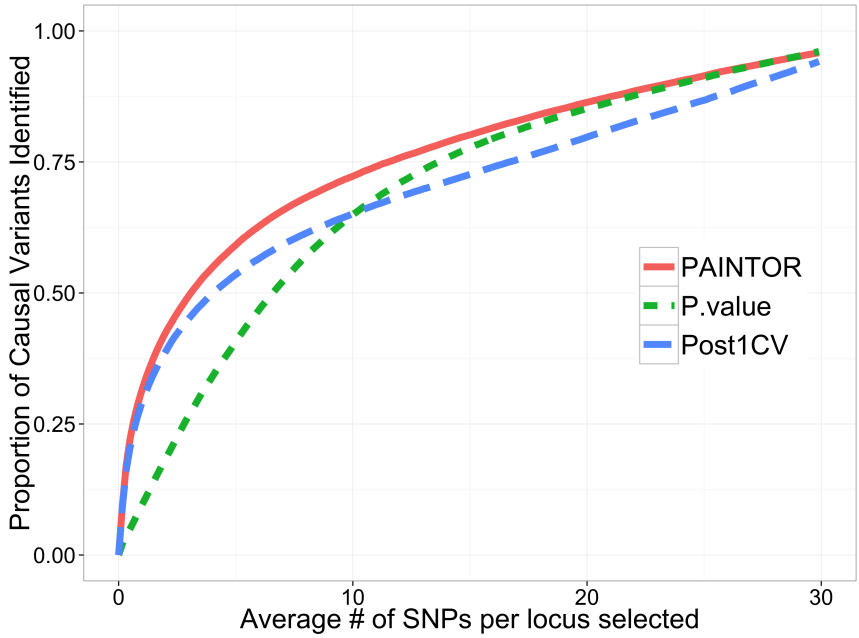

Figure 2.11: Overall performance with heterogenous SNPs effect sizes. To induce heterogeneity on SNPs, effect sizes of causal sites were drawn from an χ_1. These effect sizes were then normalized such that their aggregated effect summed up to a heritability of $h_g^2 = 0.25$. Other simulation parameters were equivalent to the standard framework (N=10,000, Loci = 100).

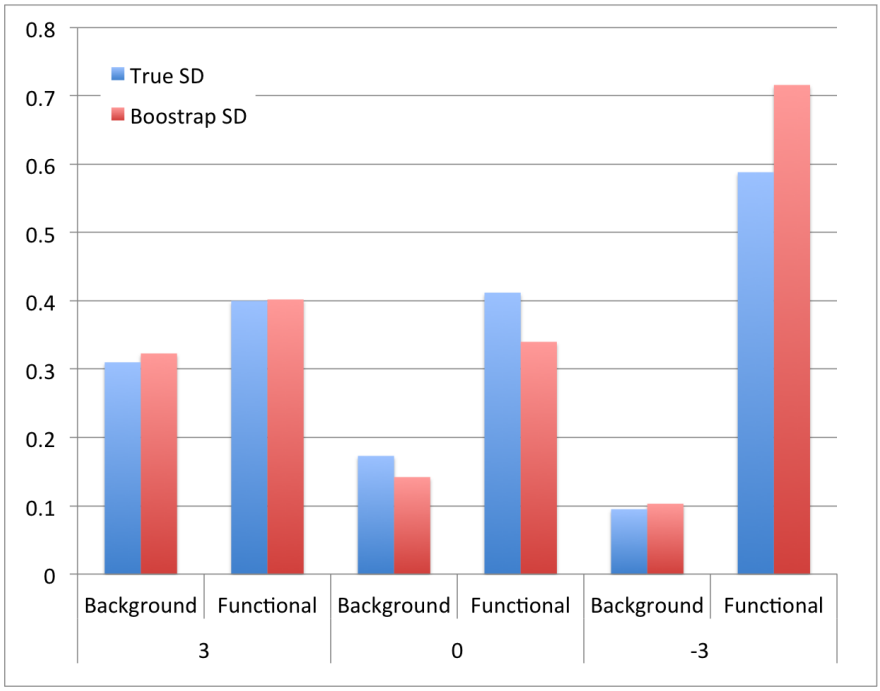

Figure 2.12: Bootstrap standard deviations for different log2 enrichment values. Using the standard simulation conditions (see Methods), we ran 100 simulations at three causal variant log2 enrichment values (-3,0,3) and for each of the simulations calculated 1000 bootstrap estimates. The standard deviations of the estimated γ coefficients were calculated across the 100 simulations (blue) and compared to the mean standard deviations of the bootstrap estimates(red). Background and functional refer to the whether the annotation represents the background SNPs or the synthetic functional annotation that we randomly assigned to 1/3 of the SNPs.

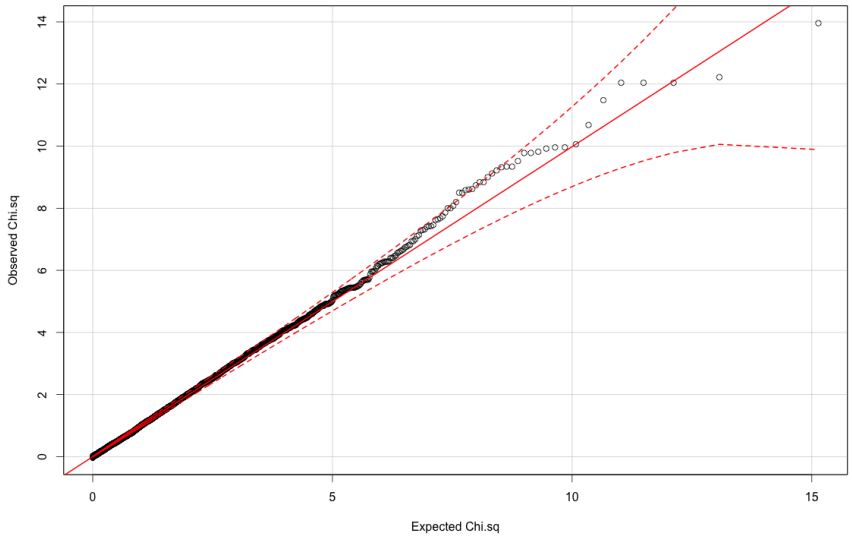

Figure 2.13: QQ Plot of likelihood ratio test statistics for a single annotation. Using the standard simulation conditions (see Methods), we ran 5000 null simulations wherein 1/3 of the SNPs were annotated to a "functional" annotation with zero effect size. We calculated LRT statistics (see Methods) from each simulation which are theoretically distributed χ^2 with df = 1 under the null. The resultant LRT statistics from the 5000 simulations have mean = 1.005, variance = 2.11, and median = 0.44, suggesting that our test statistic is well-calibrated.

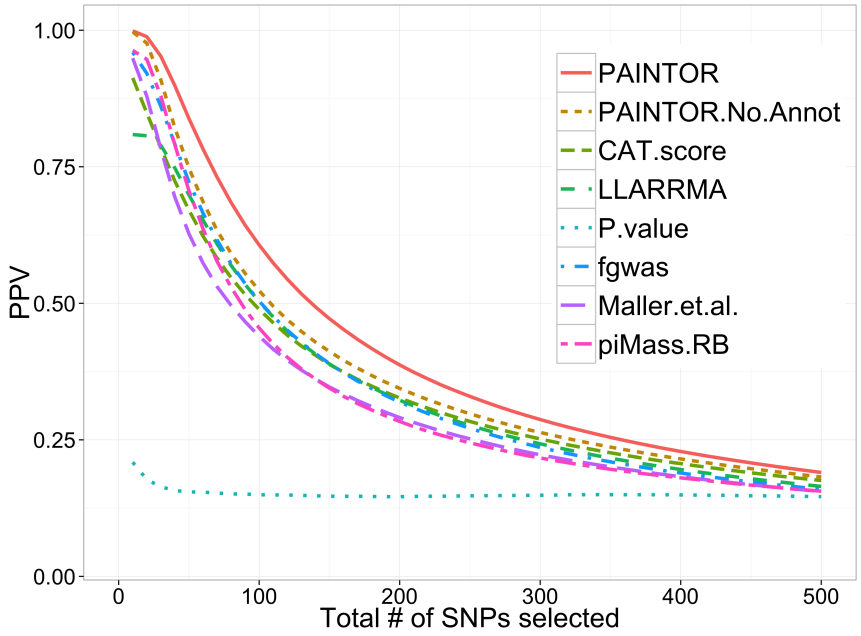

Figure 2.14: Comparison of current methodologies using positive predictive value (PPV) as the metric (defined as: $\frac{N_c}{N_t}$) . We find that the relative performance of all the methods investigated in this manuscript is maintained when assessing accuracy with the PPV.

CHAPTER 3

Trans-ethnic Probabilistic Annotation Integrator

3.1 Introduction

Genome-wide associations studies (GWAS) have reproducibly identified thousands of risk loci associated with complex traits and diseases [73, 74, 39, 75, 76, 77, 78]. Unfortunately the index variants reported in these studies are typically not biologically causal, but rather, correlated to the underlying causal variant through linkage disequilibrium (LD) [8]. Causal variants responsible for the GWAS signal are identified in fine-mapping experiments by first gathering dense genetic information, either through targeted sequencing or dense imputation, followed by statistical prioritization of variants which can subsequently be validated in functional studies [79, 39, 77, 80].

Divergent population histories due to various demographic forces such as bottlenecks and expansions have produced unique genomic landscapes across ethnicities [81, 82]. Such differences in LD patterns and variant frequencies across populations can increase the power of statistical fine-mapping if properly modeled [83, 84, 85, 86, 52, 87, 88]. Intuitively, if a locus contains a causal variant, the neighborhood of LD partners linked to this variant will be distinct in different populations. Thus, aggregating the strength of association across multiple populations may accentuate the signal from the true causal variant(s) while dampening the noise from correlated variants.

A common approach to combining information across multiple studies is through inverse variance fixed-effects meta-analysis [89] which assumes that effect sizes of causal variants are similar across studies or populations. This assumption can be relaxed using a random-effects strategy, though it has been observed that this usually results in

a decrease in statistical power [90]. A recent, and more robust, Bayesian meta-analysis framework [85] was proposed to reason over trans-ethnic studies with potential allelic heterogeneity. Both the fixed effects meta-analysis statistics as well as the Bayes Factors supplied by the latter approach can be readily converted into posterior probabilities of association to construct fine-mapping credible sets [39, 91]. However, these credible sets are commonly built under the assumption that a locus harbors at most a single causal variant [1, 40, 39, 92] which may be invalidated at many risk loci [52, 35, 53] leading to mis-calibrated credible sets [27, 20]. While conceptually it may be possible to create credible sets based on independent signals identified through conditional analysis, this strategy suffers from necessitating an ad-hoc re-definition of the fine-mapping region. Furthermore, multiple causal variants in LD can create synthetic associations at neighboring sites that are potentially stronger than the association at the true causal variants. The iterative conditioning strategy would necessarily select these synthetic SNPs first, thereby dissipating the signal from the true causal variants [27].

In addition to the strength of association between genotype and phenotype, an orthogonal source of information lies within a variant's functional genomic context. Projects such as the ENCODE/ROADMAP [13, 14] have provided a rich atlas of functional information, with numerous groups reproducibly demonstrating that disease-associated variants are systematically enriched within chromatin marks that delineate active regulatory regions in phenotypically relevant cell types [48, 49, 50, 51, 2]. While functional genomic data is often used as a post-hoc validation of association findings [80, 75, 93], a number of principled approaches have been proposed to jointly integrate functional and association data [20, 94, 95, 2]. In addition to increasing the accuracy of fine-mapping, these integrative approaches also provide insights into the genetic architecture of the trait by identifying relevant tissue-specific functional marks without making any prior assumptions. However, to the best of our knowledge, functional integrative approaches have not been extended to trans-ethnic fine-mapping and a rigorous assessment of trans-ethnic fine-mapping in the presence of multiple causal variants is currently lacking. While in principle the single population frameworks that allow for multiple causal variants [27, 20]

can operate directly on trans-ethnic meta-analysis statistics, they require ad-hoc averaging of trans-ethnic LD and do not properly account for heterogeneity by ancestry at causal variants.

In this work, we propose a statistical framework that integrates three sources of information to triangulate causal variants in fine-mapping studies: (1) the strength of association between genotype and phenotype, (2) differential genomic background across ethnic groups, and (3) tissue specific functional genomic annotations (Figure 3.1). Different allele frequencies (or sample sizes) across populations induce differential standardized effect sizes at all the variants in a region, even in the presence of no allelic effect size heterogeneity by ancestry. We model this induced heterogeneity across populations through a multi-variate normal (MVN) framework wherein the sets of population specific association statistics are realizations from population-specific MVN distributions. Similar to the case of a single population [27, 20], this allows us to consider multiple causal variants at any risk locus. We integrate functional genomic data using Empirical Bayes [20] which provides a means to select functional annotations most relevant to the trait of interest. Most importantly, our proposed approach requires only the summary association data for each population, thereby avoiding the many restrictions that may accompany analysis of individual level genotype data.

Through extensive simulations we show that our trans-ethnic framework significantly improves fine-mapping resolution relative to conventional meta-analysis strategies and demonstrate that considering multiple causal variants in multi-ethnic cohorts yields large gains in fine-mapping efficiency. We showcase our framework by reanalyzing empirical summary data from a large trans-ethnic Rheumatoid Arthritis (RA) GWAS [75] (OMIM: 180300). We first demonstrate that the functional architecture of RA is consistent across ethnicities and that there is a strong preponderance of immune-related functional classes that are enriched with causal variants. We then fine-map the RA GWAS loci using functional data and show that our method greatly outperforms current state-of-the art methodologies and uncovers a number of plausible functional variants.

3.2 Materials and Methods

3.2.1 Multi-population fine-mapping framework

Without loss of generality (as similar results can be derived for case/control traits), let y be a quantitative phenotype such that $y_i = g_i\beta + \epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0, \sigma_e^2)$ and g_i denotes a multi-SNP genotype containing $\{0, 1, 2\}$ counts of the reference allele at M SNPs for an individual i. The β vector represents allelic effects where the j'th entry will be non-zero only if SNP j is causal. Given genotype (G_p) and phenotype (Y_p) data over N_p individuals from population p, a standard approach to measure association strength at SNP j is through the Wald statistic $z_p^j = \frac{\hat{\beta}_p^j}{SE(\hat{\beta}_p^j)} = \frac{\text{Cov}(G_p^j, Y_p)\sqrt{N_p}}{\text{Var}(G_p^j)\sigma_e^2}$, which asymptotically follows a normal distribution $\mathcal{N}\left(\frac{\beta_p^j\sqrt{\text{Var}(G_p^j)}}{\sigma_e}\sqrt{N_p}, 1\right)$. We denote the non-centrality parameter (NCP) of the normal distribution as $\lambda_p^j = \frac{\beta_p^j\sqrt{\text{Var}(G_p^j)}}{\sigma_e}\sqrt{N_p}$. Under the null hypothesis that SNP j is not causal (or does not tag a causal variant, see below), $\beta_p^j = 0$ and thus $\lambda_p^j = 0$. If the SNP is causal, then $\beta_p^j \neq 0$ yielding a non-zero λ_p^j and governs the power of detecting this variant in an association study (i.e. rejecting the null at some confidence level). Importantly, even when the allelic effects at the causal variants are similar across populations (i.e. $\beta_p^j = \beta_{p'}^j$), different allele frequencies and sample sizes induce population-specific NCPs, yielding larger NCPs at more common SNPs and/or larger studies. This leads to the well-known result that causal variants are more readily detectable in populations in which they are present more frequently.

Pervasive LD at fine-scale resolutions induces correlations between tag SNPs and causal SNPs, thus creating an indirect association between tag SNP and trait [83]. More specifically, the LD-induced NCP at a SNP j (Λ_p^j), can be approximated as a linear combination of NCPs at the causal SNPs with LD-adjusted weights [83, 27, 20, 62] as

$$\Lambda_p^j = \sum_c r_p^{j,c}\lambda_p^c \tag{3.1}$$

where the sum is taken across all causal SNPs c and $r_p^{j,c}$ is the Pearson correlation coefficient between SNPs j and c in population p. We expand Equation 3.1 to include all SNPs

in the locus by incorporating an indicator variable C_p^k which is set to 1 if SNP k is causal in population p and 0 otherwise

$$\Lambda_p^j = \sum_{k=1}^{M} r_p^{j,k} \lambda_p^k C_p^k \tag{3.2}$$

In vector/matrix notation

$$\Lambda_p = \Sigma_p(\lambda_p \circ C_p) \tag{3.3}$$

where Σ_p is the LD matrix of Pearson correlations among the M SNPs, C_p is a binary vector indicating which SNPs are causal, and $\circ$ denotes the element-wise multiplication between two vectors. We can now write the probability of the data (i.e. the observed standardized effect sizes, z-scores) given the causal variants (C_p) in population p under a multi-variate normal assumption

$$Z_p|\lambda_p, C_p \sim \mathcal{N}(\Lambda_p, \Sigma_p) \tag{3.4}$$

This allows us to define the total likelihood of the data by marginalizing across all sets of causal SNPs ($\mathcal{C}$) as

$$L(Z_1, Z_2, \cdots, Z_P; \lambda_1, \lambda_2, \cdots, \lambda_P) = \prod_p \sum_{C_p \in \mathcal{C}} P(Z_p|\lambda_p, C_p) P(C_p) \tag{3.5}$$

which we simplify under the assumption that the causal vector set is identical across populations

$$L(Z_1, Z_2, \cdots, Z_P; \lambda_1, \lambda_2, \cdots, \lambda_P) = \sum_{C \in \mathcal{C}} \prod_p P(Z_p|\lambda_p, C) P(C) \tag{3.6}$$

Here, $P(Z_p|\lambda_p, C_p)$ is defined as the pdf of the multivariate normal distribution (see Equation 3.4) and $P(C)$ is the probability of a given causal set. Note that Equation 3.6 assumes the causal set is identical across populations but allows for different effect sizes at causal SNPs across populations.

3.2.2 Integration of functional annotation data

We assume that each variant can potentially have several phenotypically relevant genomic functional annotations (e.g. transcription factor binding site) which can be encoded

as binary variable A_{jk} for variant j and annotation k or as a continuous value (e.g. denoting a probabilistic membership of variants to different functional classes). We integrate the functional information through the probability of the causal set C as follows

$$P(C;\gamma) = \prod_j \left(\frac{1}{1+\exp(-\gamma^T A_j)}\right)^{C_j} \left(\frac{1}{1+\exp(\gamma^T A_j)}\right)^{1-C_j} \tag{3.7}$$

where γ is a vector containing prior log odds ratio of causality for every functional annotation. We extend the likelihood to incorporate functional data as

$$L(Z_1, Z_2, \cdots, Z_P; \lambda_1, \lambda_2, \cdots, \lambda_P, \gamma) = \sum_{C\in\mathcal{C}} \prod_p P(Z_p|\lambda_p, C) P(C;\gamma) \tag{3.8}$$

which can be further simplified by assuming data at different loci are independent:

$$L(Z_1, Z_2, \cdots, Z_P; \lambda_1, \lambda_2, \cdots, \lambda_P, \gamma) = \prod_l \sum_{C_l\in\mathcal{C}_l} \prod_p P(Z_{l,p}|\lambda_{p,l}, C_l) P(C_l;\gamma) \tag{3.9}$$

Finally, to obtain posterior probabilities for each SNP to be causal we use Bayes theorem to compute the joint posterior for each causal set

$$P(C_l|Z_{l,1}, Z_{l,2}, \cdots, Z_{l,P}; \lambda_{l,1}, \lambda_{l,2}, \cdots, \lambda_{l,P}, \gamma) = \frac{\prod_p P(Z_{l,p}|\lambda_{p,l}, C_l) P(C_l;\gamma)}{\sum_{C_l\in\mathcal{C}_l} \prod_p P(Z_{l,p}|\lambda_{p,l}, C_l) P(C_l;\gamma)} \tag{3.10}$$

and subsequently marginalize across all $C_l = (C_{1l}, C_{2l}, ..., C_{N_l})$ such that $C_{jl} = 1$

$$P(C_{jl}|Z_1, Z_2, \cdots, Z_P; \lambda_{l,1}, \lambda_{l,2}, \cdots, \lambda_{l,P}, \gamma) = \tag{3.11}$$
$$\sum_{C_l\in\mathcal{C}_l : C_{jl}=1} P(C_l|Z_1, Z_2, \cdots, Z_P; \lambda_{l,1}, \lambda_{l,2}, \cdots, \lambda_{l,P}, \gamma)$$

3.2.3 Model Fitting

Due to the finite nature of either the sample or reference panel, the LD matrix in practice may be ill-conditioned. We apply a Tikhonov Regularization [96] to all LD matrices to ensure their invertibility and as a result preserve the non-degeneracy and numerical stability of the MVN approximation. Furthermore, because we ensure that all $\Sigma's$ are positive

definite, there exists a Cholesky decomposition for each LD matrix and its corresponding inverse. Let $L_p = Chol(\Sigma_p)^{-1}$; it follows that $\tilde{Z}_p = L_p Z_p \sim \mathcal{N}(L_p \Lambda_p, I)$. In practice, we operate in the transformed Z-scores space ($\tilde{Z}_p$), as it improves numerical stability and reduces computational burden by removing a large, repetitive matrix multiplication when computing the MVN density.

We fit the parameters of the model to the data across all loci using a variant of the Expectation Maximization over the functional annotations ($\gamma's$) and approximate the NCP's using a simple function of the observed Z-scores (see Appendix). We note that since enumerating over all possible causal sets is combinatorially intractable, we typically restrict the number of causal variants per locus to two or three in practice.

3.2.4 Simulation Data

We benchmarked our proposed framework using simulations starting from real genotype data. Using the NHGRI catalog of GWAS variants on chromosome 1 [73], we centered 25KB windows on the lead SNP and used HAPGEN2[68] and the 1000 genomes [82] to simulate individuals from the Asian (N=286), African (N=246), and European (N=379) ancestries. SNPs that were polymorphic with a MAF ≥ 0.01 in at least one population were retained for analysis. For each simulation, we randomly chose 50 loci and simulated causal variants by drawing causal status according to the logistic prior model described above. Unless otherwise noted, we used the annotations (Coding, UTR, Promoter, DHS, Intronic, and Intergenic) and functional enrichments (13.8x, 8.4x, 2.8x, 5.1x, 0.1x) observed in Gusev et al. [2] for simulations below. We simulated phenotypes under a linear model such that for individual i of population p their phenotype Y was drawn as $Y_{i,p} = \sum_{j=1}^{N_c} \beta_j \cdot g_{j,i,p} + \epsilon_{i,p}$, where N_c is the total number of causal variants, β_j is the effect size of the j'th causal SNP, $g_{j,i,p}$ is number of copies of the risk allele j for individual i of population p. Following recent works, we simulated similar h^2 across populations [97]. The population-specific error term, $\epsilon_{i,p}$, was was drawn according to a $\mathcal{N}(0, \sigma^2_{e,p})$, where $\sigma^2_{e,p} = \frac{\sigma^2_{g,p} - h^2_g * \sigma^2_{g,p}}{h^2_g}$, $\sigma^2_{g,p} = \beta' \operatorname{Cov}(X_p)\beta$ and $\operatorname{Cov}(X_p)$ is the population specific covariance of the genotypes

(LD). The effect size, β_j, was set to be inversely proportional to the average standard deviation of the population allele frequencies; this is roughly equivalent to assuming that each causal SNP explains an equal proportion of the phenotypic variance [98].

3.2.5 Existing methods

We compared our proposed methods with other well-established probabilistic methods for fine-mapping. First we investigated MANTRA, a Bayesian trans-ethnic meta-analysis technique proposed in [85]. We obtained the software implementation from the author and ran it using the default settings, providing FST between the three populations as determined in [99] as the prior for the Bayesian Partition model. The output of MANTRA is a Bayes Factor which we subsequently converted to posterior probabilities of association (PPA, $PPA_i = \frac{BF_i}{\sum_k BF_k}$) as recommended in [1, 100, 39]. Similarly, we calculated posterior probabilities for SNPs to be causal based strictly on the inverse-variance fixed-effects [89] meta analysis using the CAVIARBF and PAINTOR frameworks described in [28, 20]. We note that the CAVIARBF and PAINTOR models require LD as input which we calculated as the average of the population-specific LD weighted by the sample size of each population. We assess accuracy by rank-ordering SNPs across all fine-mapping loci based on the output of each method, and then determined the proportion of causal variants that are identified as more SNPs are selected. We typically report the median number of SNPs one would need to validate in order to resolve 90% of the causal variants as our main metric of accuracy.

3.2.6 Rheumatoid Arthritis multi-ethnic fine-mapping data set

We downloaded summary statistics from a large trans-ethnic Rheumatoid Arthritis (RA) GWAS consisting of over 100,000 individuals ($\sim$68,000 of European ancestry and $\sim$36,000 of Asian ancestry) [75]. We used the reported genome-wide significant loci, excluding HLA regions, and centered 100KB windows around the top SNP yielding a total of 89 fine-mapping loci. For each of these regions, we estimated LD using the European and

Asian ancestry individuals from the 1000 Genomes [82]. We integrated 482 publicly available functional annotations comprising of 406 DNase I Hypersensitivity Sites spanning numerous cell types and tissues [70, 48], the seven genomic segmentations of the eight primary ENCODE cell lines [101], Fantom5 Enhancer and TSS regions [102], Immune-cell enhancers [80], genic elements derived from GenCode [103], and omnibus methylation/acetylation marks from the ENCODE [13]. The construction of a phenotypically-specific fine-mapping model requires two phases. First, we run the model marginally on each annotation and subsequently rank order all the annotations based on likelihood ratio statistics [94, 20]. Second, we select the top annotations that are minimally correlated with one another (usually no more than five) to enter a final model to estimate posterior probabilities for variants to be causal.

3.3 Results

3.3.1 Joint modeling of association statistics across populations increases fine-mapping performance

We used simulations to investigate the benefit of jointly modeling population-specific association statistics versus standard meta-analysis approaches. We simulated fine-mapping data sets over 10,000 individuals equally divided among European, Asian and African ancestries with total heritability of $h^2_g = 0.25$ across 50 loci with genetic architecture similar to Gusev et al [2]. The loci were simulated such that in expectation each locus harbored a single causal variant with allelic effects shared across populations (see Methods). This yielded 15 loci with a single causal and 13 loci with multiple causals on the average per simulation. In general, we find that trans-ethnic fine-mapping strategies that assume a single causal variant are sub-optimal compared to ones that allows for multiple causal variants (Table 3.1). For example, posterior probabilities based on MANTRA meta-analysis requires (1.9, 96.8) SNPs per locus in order to identify (50,90)% of the causal variant as opposed to (1.2, 7.0) SNPs per locus with methods that allow multiple causal vari-

ants but do incorporate functional data[28]. Existing integrative fine-mapping methods that leverage functional data [20] applied to fixed effects meta analysis statistics achieve accuracy of (1.0, 5.6) SNPs per locus. In contrast, our proposed framework resolves causal variants with the greatest efficiency (Figure 3.2), requiring only (0.9, 5.2) SNPs per locus (paired t-test $p < 0.001$) . Overall, this can be attributed to the fact that our approach models population-specific LD patterns while allowing for multiple causal variants in the presence of functional information.

Recent studies have shown that GWAS findings generally replicate across populations [25, 97] thus suggesting sharing of underlying causal variants. However, it is generally unknown if these variants contribute to disease risk uniformly across populations. We sought to assess performance of fine-mapping in the situation where the causal variants have either weak or strong heterogeneity by ancestry. In addition to the fine-mapping data sets where causal effects were similar across populations (no heterogeneity), we simulated allelic effects inversely proportional to the population-specific allele frequency standard deviation (weak heterogeneity) as well as normally distributed allelic effects for each ancestry independently (strong heterogeneity). We find that our framework significantly outperforms the fixed effects meta-analysis followed by probability estimation using existing methods. For example, in the case of weak heterogeneity, our approach requires 4.1 as opposed to 4.9 SNPs per locus (19.5% improvement); while in the presence of strong heterogeneity, our approach dramatically outperforms existing meta-analysis strategies, reducing the number of SNPs that need to be selected in order to identify 90% of the causal variants from 121.3 to 56.5 (214% improvement) (Table 3.1). The increase in performance is likely due to the fact that our framework makes no assumptions pertaining to the population-specific allelic effects at causal SNPs, as we allow the empirically observed z-scores in each population to dictate the effect size. This allows for arbitrary levels of heterogeneity in the effect size by population whereas fixed effects meta-analysis assumes similar effect sizes across populations.

3.3.2 Performance of trans-ethnic fine mapping

The benefit of trans-ethnic fine-mapping has been thoroughly documented both in simulations as well as in empirical data [83, 85, 88]. However, previous works have utilized the assumption of a single causal variant at a risk locus which is often invalidated in practice. Here, we sought to assess trans-ethnic fine-mapping in the presence of multiple causal variants at a risk locus while integrating functional annotation data. Consistent to previous works[83], we find that for the same sample size, multi-ethnic cohorts attain superior accuracy over single-population studies. However, allowing for multiple causal variants yields a much larger increase in performance of trans-ethnic versus single-population fine-mapping. We observe a near 3 to 4-fold increase in the median resolution for methods that model multiple causal variants as compared to only a 1.4 to 1.6-fold gain for methods that assume a single causal, see Table 3.2. We attribute this to the much smaller number of sets of causal variants (as proportion from the total possible sets) that are compatible to the observed association statistics. Diversity in LD patterns across populations additionally penalizes sets of variants that do not contain the true causal variants as they are unlikely to explain the observed data. Consequently, multi-ethnic cohorts will not only have proportionally more LD patterns than single population cohorts (therefore placing larger penalties on incorrect causal sets), but can also borrow power from populations where the causal variants are present more frequently.

3.3.3 Genetic trait architecture impacts fine-mapping performance

Functional information was demonstrated to improve fine-mapping resolution in a single population [20, 94, 95, 80] and we investigated the potential gains in a trans-ethnic setting. We simulated two disease architectures using five functional annotations where causal variants either localize predominantly within a single broad functional class as observed by Gusev et al. [2] (A1) versus a smaller, more diffuse localization within functionally specific cell types [20] (A2). For each class of disease architectures, we fit six trans-ethnic integrative models, with each successive model incorporating an additional

functional annotation into a joint framework. Not surprisingly, when the *true* genetic architecture of a trait at fine-mapping regions has a strong enrichment of causal variants within a common functional class (i.e. DNase Hypersensitivity Sites [2]), these functional annotations will be most informative for the purposes of fine-mapping (see Figure 3.3). On the other hand, more diffuse localization of causal variants requires multiple annotations to maximize the utility of functional data. For example, for genetic architecture A1, the addition of the DHS annotation yield a 70% increase in fine-mapping resolution whereas genetic architecture A2 required all five annotations to improve resolution by 18% (see Figure 3.3).

3.3.4 Integrative fine-mapping in a multi-ethnic rheumatoid arthritis data

We investigated whether similar results from simulations can be attained in empirical data of a trans-ethnic Rheumatoid Arthritis (RA) over 100,000 individuals [75] (see Methods). Since the functional genetic architecture of RA across different populations is unknown, we first quantified whether the enrichment of causal variants in various functional annotations is consistent across ancestries. Reassuringly, we see a strong correspondence in functional enrichment at the fine-mapping loci across all 482 functional categories we investigated ($r = 0.597$) (Figure 3.4). This provides evidence supporting the assumption that a single functional prior can be applied across populations uniformly when conducting trans-ethnic fine-mapping.

Next, we estimated trans-ethnic enrichment for each of the 482 annotations independently to allow the model to discern the most functionally relevant cell types and classes. The likelihood ratios for enrichment supplied by this procedure provide a natural way to prioritize functional annotations to move forward with fine-mapping [20]. We consistently find a strong and significant enrichment of causal variants within activity regulatory regions of immune-related cell types (see Figure 3.5) which is largely in-line with RA disease etiology (rank permutation $p < 0.001$). The final trans-ethnic integrative model included annotations of three cell-type specific DHS regions (Skin Keratinocytes,

T-h2, and B-lymphocytes); Immune Enhancer described in [80]; and GenCode defined exon regions. We find that simply applying existing multi-causal frameworks [27, 20] on the trans-ethnic meta-analysis statistics yields wider 90% credible sets, requiring approximately 28.5 SNPs per locus as opposed to 24.0 SNPs per locus for our proposed framework, thus demonstrating the benefit of modeling population-level LD. Furthermore, the integration of functional data additionally reduces the size of the credible set to 21.7 SNPs per locus (see Table 3.3), showing that leveraging functional annotations refines trans-ethnic fine-mapping signal.

Next, we explored the plausible causality of the SNPs that attained a high posterior probability under our framework (Table 3.4). For example, rs968567, a variant within the promoter region of the *FADS2* gene (OMIM: 606149) that was functionally validated to disrupt transcription factor binding and subsequent gene expression [104], achieved a trans-ethnic posterior probability of 0.29. However, this variant falls within all five functional annotations that our framework deemed important for this trait, and, upon appropriate re-weighting, achieved a posterior probability of 0.84. Alternatively, trans-ethnic association can be extremely beneficial on its own. For example, rs12693993, a variant within the coding region of *CD28* (OMIM: 186760), a gene implicated for it's importance for T-cell development, proliferation, and cytokine production achieved a posterior probability for causality of 0.34 and 0.02 in Europeans and Asians, independently. However upon integrating trans-ethnic association with functional data, it achieved a posterior probability for causality of 0.85. The identification of these two SNPs, among others, serve as important illustrations of the benefit of our proposed methodologies.

3.4 Discussion

In this work, we introduced a fine-mapping framework that bridges several sources of evidence to prioritize functional SNPs and demonstrated its efficacy in real and simulated data sets. As fine-mapping data becomes increasingly multi-ethnic [39, 75] and functional data becomes larger and more refined [14], we believe that our proposed methodology

will have increasing relevance. By operating exclusively on summary data, our approach reduces the need of individual data sharing that often prohibit large-scale analyses. In addition, a key advantage of our proposed methodology is that it provides an unbiased perspective on which functional genomic data is most relevant to the trait within an Empirical Bayes framework. Rather than relying on careful and manual selection of functional elements when conducting fine-mapping [80, 93], we allow the data to dictate the functional relevance of a particular annotation. As the catalog of functional data expands to encompass more diverse cell types and genomic signatures, a principled strategy to parsing these annotations is paramount.

We note that while our model does not assume a priori that there exists allelic heterogeneity by ancestry [85], by construction, it is capable of handling trans-ethnic heterogeneity whether it is due to a true difference in the per-allelic effects or simply a result of genetic drift that yielded distinct allele frequencies at the causal SNPs. We find that as the level of heterogeneity across populations increases, our framework increasingly out performs competing strategies. While extreme heterogeneity may be unlikely, gene by environment interactions in complex traits can manifest themselves as distinct allelic effects across populations [105].

We conclude with several limitations of our proposed framework. The efficacy of our proposed method is intimately connected to the underlying functional architecture of the trait being examined. In the scenario where the correct functional annotation is unavailable or if the distribution of casual variants is more or less uniform across the functional annotation categories, our method will likely underperform relative to fine-mapping strategies that either do not estimate parameters for functional enrichment [27, 28] or that pre-specify the correct enrichment parameters from other external analyses [2]. However, there is mounting evidence that suggests casual variants for most complex traits co-localize with epigenetic marks [80, 2, 94, 48] that are now available for the vast majority of human cell types [106]. Finally, additional improvements in performance could be made through a Bayesian treatment of non-centrality parameters within our framework [28] which we leave as a potential direction for future work.

3.5 Appendix

3.5.1 Optimization procedure

We optimize the parameters of our model using Expectation Maximization. First, we take expectations of the complete data log-likelihood with respect to the posterior distribution of causal sets and simplify to obtain a function, Q, that is readily optimized using standard techniques. Let $Z_{l,*}$ represent all P vectors of association statistics $(Z_{l,1}, Z_{l,2}, \cdots, Z_{l,P})$ at locus l and let $\lambda_{l,*}$ be the corresponding vectors of non-centrality parameters.

$$\begin{aligned}
Q(\gamma, \lambda|\gamma^{(t)}, \lambda) &= \sum_l \sum_{C_l} P(C_l|Z_{l,*}\lambda_{l,*}, \gamma^{(t)}) \ln P(Z_{l,*}; \lambda_{l,*}, \gamma^{(t)}) \\
&= \sum_l \sum_{C_l} P(C_l|Z_{l,*}\lambda_{l,*}, \gamma^{(t)}) \left(\ln P(C_l; \gamma^{(t)}) + \sum_p \ln P(Z_{l,p}|C_l, \lambda_{l,p}) \right) \\
&= \sum_l \sum_{C_l} P(C_l|Z_{l,*}\lambda_{l,*}, \gamma^{(t)}) \ln P(C_l; \gamma^{(t)}) \\
&\quad + \sum_l \sum_{C_l} P(C_l|Z_{l,*}, \gamma^{(t)}, \lambda_{l,*}) \sum_p \ln P(Z_{l,p}|C_l, \lambda_{l,p}) \\
&= Q(\gamma|\gamma^{(t)}) + Q(\lambda_p|\lambda_p)
\end{aligned}$$

thereby decoupling the prior from the likelihood. We simplify $Q(\gamma|\gamma^{(t)})$ to obtain

$$\begin{aligned}
Q(\gamma|\gamma^{(t)}, \lambda) &= \sum_l \sum_j \sum_{c_{jl} \in 0,1} P(c_{jl}|Z_{l,*}; \gamma^{(t)}, \lambda_{l,*}) \ln P(c_{jl}; \gamma^{(t)}) \\
&= -\sum_l \sum_j P(c_{jl} = 1|Z_{l,*}; \gamma^{(t)}, \lambda_{l,*}) \ln(1 + \exp(-\gamma^T A_{jl})) \\
&\quad - \sum_l \sum_j P(c_{jl} = 0|Z_{l,*}; \gamma^{(t)}, \lambda_{l,*}) \ln(1 + \exp(\gamma^T A_{jl}))
\end{aligned}$$

which is a concave function whose gradient is simply

$$\begin{aligned}
\frac{\partial Q(\gamma|\gamma^{(t)}, \lambda)}{\partial \gamma} &= \sum_j \sum_l P(c_{jl} = 1|Z_{l,*}; \gamma^{(t)}, \lambda_{l,*}) \frac{1}{1 + \exp(-\gamma^T A_{jl})} A_{jl} \\
&\quad - \sum_j \sum_l P(c_{jl} = 0|Z_{l,*}; \gamma^{(t)}, \lambda_{l,*}) \frac{1}{1 + \exp(\gamma^T A_{jl})} A_{jl}
\end{aligned}$$

To avoid potential numerical instability resulting from inverting a Hessian matrix as would be required for standard Newton-Raphson, we optimize this function Q using a limited-memory BFGS algorithm implemented in the NLopt library. Finally, as previously mentioned, the non-centrality parameter for SNP j at locus l from population p, $\lambda^j_{p,l}$, is set simply as:

$$f(Z^j_{p,l}) = \begin{cases} \arg\min(-3.7, Z^j_{p,l}) & \text{if } Z^j_{p,l} < 0 \\ \arg\max(3.7, Z^j_{p,l}) & \text{if } Z^j_{p,l} > 0 \\ 0 & \text{if } Z^j_{p,l} = 0 \text{ (SNP } j \text{ is monomorphic in population } p) \end{cases}$$

a strategy that was previously demonstrated to work well in practice [20]. This iterative algorithm is repeated until the change in the log-likelihood is less than 0.01.

3.6 Tables

		Single Causal Variant Per Locus		Multiple Causal Variants per Locus		
Heterogeneity Level	Proportion of causal variants identified	Fixed effects Single Causal Posterior	MANTRA[85] Posterior	CAVIARBF[28] Fixed Effects Posterior	PAINTOR[20] Fixed Effects Posterior[a]	PAINTOR Trans-Ethnic Posterior[a]
None	0.50	1.9	2.0	1.2	1.0	0.9
	0.75	29.8	30.3	2.9	2.1	1.9
	0.90	96.8	96.8	7.0	5.6	5.2
Weak	0.50	1.9	2.0	1.1	0.9	0.9
	0.75	62.3	62.7	2.9	2.0	1.8
	0.90	118.1	118.6	6.8	4.9	4.1
Strong	0.50	29.0	11.1	12.6	9.6	2.3
	0.75	105.0	92.7	68.6	58.4	19.7
	0.90	143.9	139.8	134.4	121.3	56.5

Table 3.1: Our trans-ethnic integrative framework is superior to conventional meta-analysis strategies as well as current-state-of the art methodologies. We simulated 1000 multi-ethnic fine-mapping data sets under various levels of allelic heterogeneity across populations. For the first two levels of heterogeneity (None and Weak), we invoked the standard infinitesimal assumption on allelic effects either globally or at the population level by setting effect sizes $(\beta_{c,p})$ at the causal snps inversely proportional to either the mean allele frequency standard deviation or the population-specific allele frequency standard deviation. To simulate strong heterogeneity across ancestries, we drew effect sizes from a standard normal for each population independently and added enough gaussian noise to maintain an $h_g^2 = 0.25$. Displayed here are the median number of SNPs selected per locus in order to identify a specified proportion of the causal variants. [a]Methods that also integrate functional data.

# Causals	Single		Multiple	
Functional Data	-	+	-	+
Asians	136.9	134.4	89.3	36.2
Europeans	135.0	130.9	82.9	33.5
Africans	104.0	95.0	34.4	14.7
Trans-ethnic	72.6	58.4	8.5	4.9
Relative	1.4	1.6	4.0	3.0

Table 3.2: Modeling multiple causal variants in multi-ethnic cohorts yields larger relative gains in fine-mapping efficiency. We simulated fine-mapping data sets with various ethnic compositions with allelic effects shared across populations. Displayed here are four fine-mapping strategies that consider either single or multiple causal variants at each risk locus with (+) and without (-) access to functional data across different ethnic study designs. The bottom row represents the relative gain in the median 90% causal variant resolution of trans-ethnic cohorts versus the next best-performing group.

Association	Annotations	
Statistics	-	+
Asian	35.2	31.9
European	32.0	28.7
Fixed Effects Meta Analysis	28.5	25.0
Trans Ethnic	**24.0**	**21.7**

Table 3.3: Integrative approaches that model population-level LD yield smallest credible sets in empirical data. Displayed here is the average number of SNPs per locus in the 90% credible sets for single and multi-population fine-mapping of rheumatoid arthritis loci. To compute credible sets we first order the SNPs across all 89 loci and then take the total number of ordered SNPs that consume 90% of the total posterior probability mass. Consistent with simulation findings, integrating multiple populations with functional data improves fine-mapping resolution.

rsID/ Chr:Pos	Euro Assoc. (Z-score)	Asian Assoc. (Z-score)	Posterior Probability	Posterior Probability[a]	Functional Annotations
rs2476601/ chr1:114377568	-26.04	NA	1.00	1.00	Coding Exon, Skin Keratinocytes DHS
rs7731626/ chr5:55444683	-9.84	NA	1.00	1.00	GM12865 DHS, hTH2 DHS, Immune Enhancers
NA/ chr1:2523878	-5.22	-4.18	1.00	1.00	Immune Enhancers
rs1893592/ chr21:43855067	-5.73	-4.01	1.00	1.00	Coding Exon, Immune Enhancers
NA/ chr19:10771941	-6.13	NA	1.00	1.00	Immune Enhancers
rs72767222/ chr5:55440788	5.11	NA	0.99	0.99	Skin Keratinocytes DHS, Immune Enhancers
rs12715125/ chr3:27763427	5.58	NA	0.95	0.99	Coding Exon, GM12865 DHS, hTH2 DHS, Skin Keratinocytes DHS, Immune Enhancers
rs71508903/ chr10:63779871	7.26	5.88	0.76	0.93	GM12865 DHS, Skin Keratinocytes DHS, Immune Enhancers
rs12693993/ chr2:204595597	-2.74	-1.76	0.68	0.88	hTH2 DHS, Skin Keratinocytes DHS, Immune Enhancers
rs968567/ chr11:61595564	-4.95	NA	0.29	0.85	Coding Exon, GM12865 DHS, hTH2 DHS, Skin Keratinocytes DHS , Immune Enhancers
rs909685/ chr22:39747671	6.29	4.62	0.65	0.84	hTH2 DHS, Skin Keratinocytes DHS, Immune Enhancers
rs657075/ chr5:131430118	2.54	4.46	0.73	0.82	Skin Keratinocytes DHS, Immune Enhancers

Table 3.4: Integrating trans-ethnic association strength with functional data promotes a number of SNPs to attain a high posterior probability for causality. We applied our framework across all 89 GWAS RA loci with relevant functional data. Displayed in this table are SNPs achieving a trans-ethnic posterior probability of greater than 0.8. [a]Probability estimation with relevant functional data that was identified by our framework.

3.7 Figures

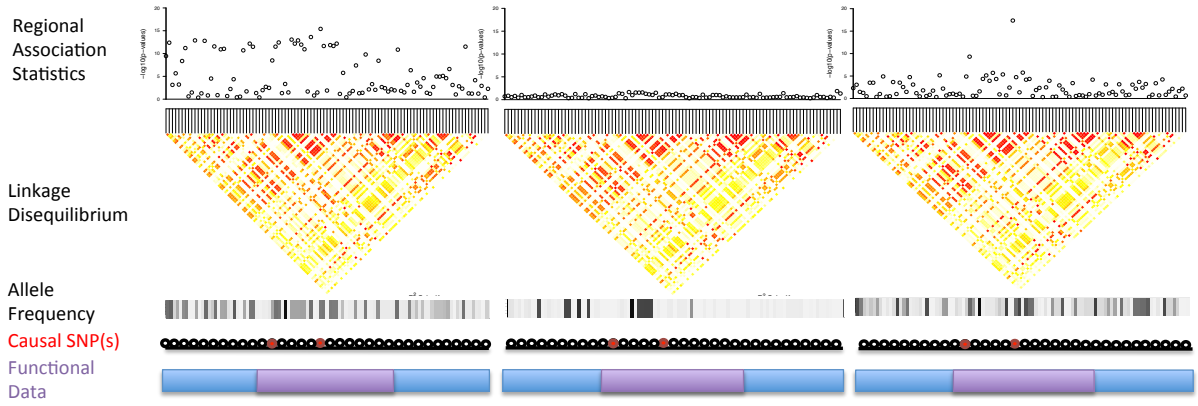

Figure 3.1: Example of a fine-mapping locus in three different populations. In Population 1 the causal variants are present but there is strong regional LD making it difficult to distinguish them from tagging SNPs. In Population 2 the causal variants both have very low frequency and/or are monomorphic resulting in no observable association between the SNPs and the trait. In population 3 the causal variants are common and have few tagging SNPs. Our framework jointly models population-specific LD structure and integrates functional genomic data to prioritize causal variants.

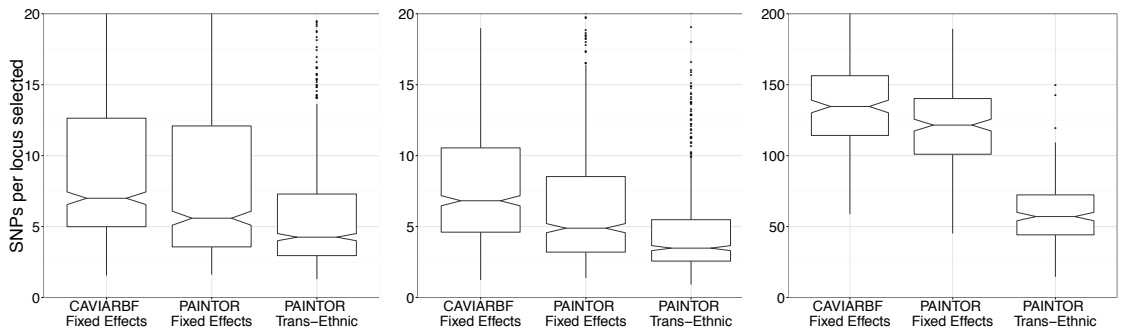

Figure 3.2: PAINTOR Trans-Ethnic is most efficient in identifying causal variants. The distributions of the number of SNPs required for follow-up in order to identify 90% of the causal variants across 1000 simulations are displayed as box plots. The different panels represent increasing levels of effect size heterogeneity by ancestry: none (left), weak (middle) and strong (right). The width of the notches in each box plot roughly correspond to 95% confidence intervals for the median number of SNPs required to resolve 90% of the causal variants. For the sake of clarity, we have cut the y-axis to emphasize the significant difference in performance across all three methods.

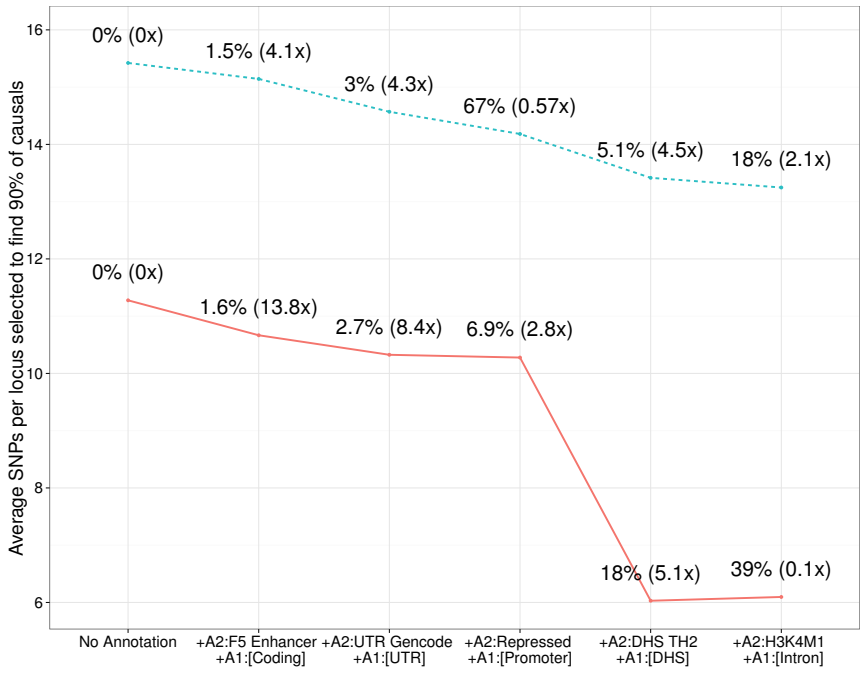

Figure 3.3: The underlying functional architecture of a trait impacts fine-mapping performance. We simulated two classes of disease architectures A1 (solid line) and A2 (dashed line). Architecture A1 was based on the functional enrichment observed in Gusev et al. [2] and had a strong enrichment within a single DHS class. Architecture A2 was simulated with a more diffuse enrichment in various cell types and classes and was based on what we empirically observe in the rheumatoid arthritis data set. Displayed on top of each point is the percentage of SNPs falling within that annotation and its corresponding enrichment.

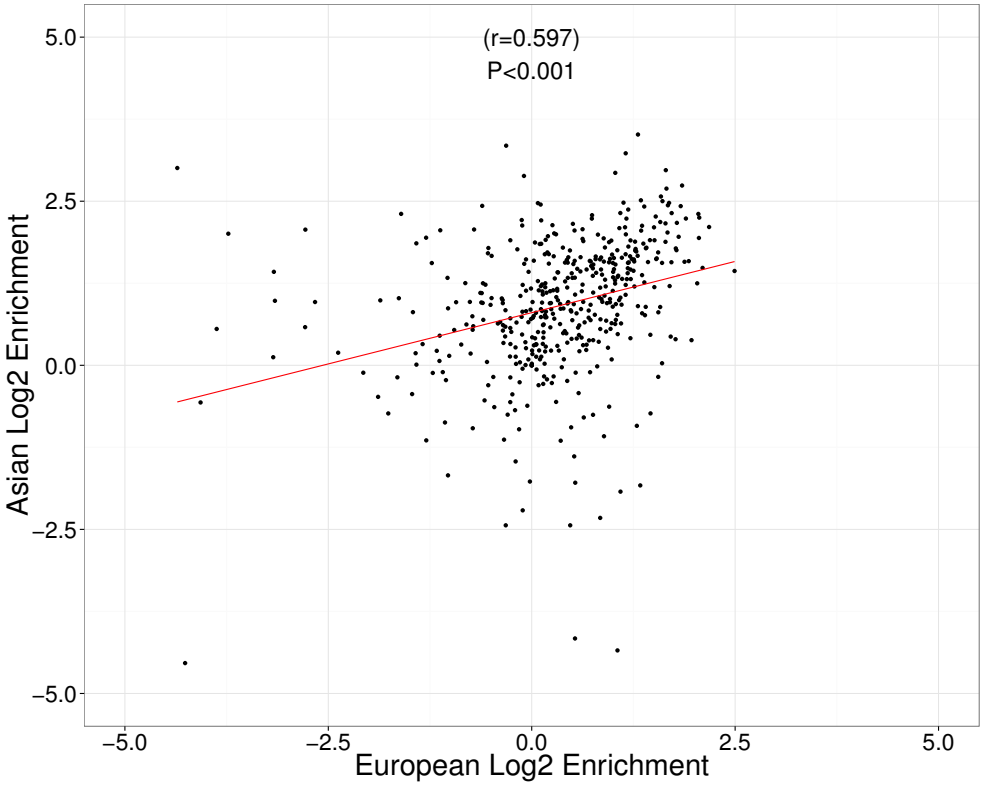

Figure 3.4: Functional enrichment is consistent across Europeans ($N \approx$ 68K) and Asians (N$\approx$ 36K). We compared the enrichment across 482 functional annotations at 89 rheumatoid arthritis loci in Europeans and Asians separately. Each point represents the estimated enrichment of an annotation in both European and Asian populations.

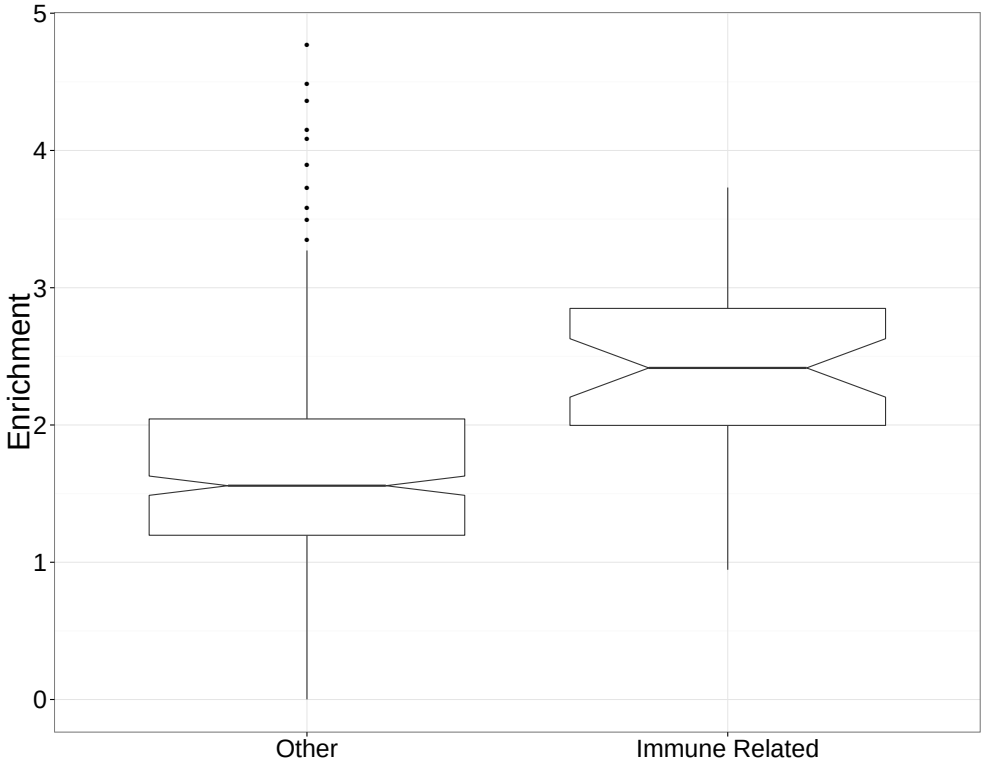

Figure 3.5: Trans-ethnic functional enrichment at rheumatoid arthritis GWAS loci indicate immune-related regulatory architecture. Here, we compare the enrichment of casual variants within 42 DNase Hypersensitivity sites of immune related cell-types (B-cells,T-cells, NK-cells, Keratinocytes, Monocytes, and Thymus) versus 354 DHS annotations of other cell types.

CHAPTER 4

Fast Probabilistic Annotation Integrator

Genome-wide association studies (GWAS) have identified thousands of regions in the genome containing risk variants for complex traits and diseases [74, 39, 75, 76, 77]. However, the vast majority of the GWAS reported variants are not biologically causal, but rather, correlated to the true causal variants through linkage disequilibrium (LD) [8, 27, 20]. Fine mapping studies gather detailed genetic information within the loci that have been implicated in GWAS [35, 36, 37] and statistically dissect these regions to prioritize variants according to probability of causality. The top variants resulting from this procedure may become candidates for functional validation [107, 108].

Many statistical methods for fine-mapping have been developed for the prioritization of causal variants. Standard approaches range from a simple ranking of SNPs based on their p-values to more sophisticated LD-aware ranking algorithms that quantify probabilities for variants to be causal [27, 28, 29, 20]. Initial probabilistic methods have assumed a simple model in which only one variant per locus is biologically causal [1], with more recent methods extending the statistical frameworks to accommodate multiple casual variants at risk regions [27, 28, 20, 21]. Although modeling multiple causal variants drastically increases performance, particularly at loci with evidence of multiple signals of association, it also presents a combinatorially challenging problem in performing inference in the model. That is, the likelihood formulation contains a model space size exponential in the number of variants at a locus, which clearly cannot be enumerated over for even a modestly-sized locus. To account for this combinatorial explosion, initial methods approximated the full likelihood by restricting the maximum number of causal variants allowed at a risk locus to a small number [27, 28, 20, 21]. More recent works [29] fur-

ther improved computational efficiency by sampling likely causal models using stochastic search, leveraging the intuition that most of the terms in the likelihood computation have near negligible contribution. The authors demonstrated that this achieves drastic reduction in runtime with comparable fine-mapping accuracy relative to enumerative methods [29]. However, this was done in the context of a single fine-mapping locus and did not integrate multiple sources of information.

In this work we propose a unified framework to perform fast, integrative fine-mapping across multiple traits. We integrate the strength of association across multiple traits or populations with functional annotation data to improve performance in the prioritization of causal variants. Our approach makes the assumption that the same variants at the risk loci impact both traits though with potentially distinct effect sizes. A key advantage of our approach is that it requires only summary association data for each trait, thus avoiding the restrictions that arise from the sharing of individual-level data. To balance computational efficiency and accuracy we propose an Importance Sampling technique that provides guarantees for convergence, while relaxing the assumption of the maximum number of causal variants allowed at each risk locus.

4.1 Methods

4.1.1 Fully Bayesian Statistical Fine-mapping

Chapters 1 and 2 demonstrated that given a set of causal variants C, the distribution of association statistics at a fine-mapping region is well-described the following Multivariate Normal Distribution:

$$Z \mid \lambda_C, \Sigma \sim \mathcal{N}(\Sigma\lambda_C, \Sigma) \tag{4.1}$$

However, the causal effect sizes (λ_C) are typically unknown apriori and must be either approximated [20, 21] or integrated out [27]. Leveraging the standard infinitesimal

model [98], Hormorzdiari et al. (2014) proposed to use a normal prior on the causal NCPs which, due to conjugacy, can be conveniently integrated analytically as follows:

$$\lambda_C \mid C, \sigma^2 \sim \mathcal{N}(0, \Sigma_C)) \tag{4.2}$$

$$\Sigma_C = \sigma_c^2 \text{Diag}(C) + \text{Diag}(\epsilon) \tag{4.3}$$

$$Z \mid \Sigma, C \sim \left(\int \mathcal{N}\left(\Sigma\lambda_C, \Sigma\right) \mathcal{N}\left(0, \Sigma_C\right) d\lambda_C \right) P(C) \tag{4.4}$$

$$= \mathcal{N}(0, \Sigma + \Sigma\Sigma_C\Sigma) P(C) \tag{4.5}$$

Here the prior probability of the causal set vector ($P(C)$) can be set to be uniform [1], hypergeometric [27], or can be estimated empirically using more sophisticated approaches that incorporate functional genomic data [20, 21]. Chen et al (2015) made the observation that the marginal likelihood in (eq. 4.5) is approximately proportional to a Bayes Factor comparing a causal and null model which depends on the Z-scores and LD only at the causal SNPs. This effectively reduces the computational burden from cubic in the number of SNPs to cubic in the number of causal variants considered at each likelihood evaluation. This not only improves efficiency, but also improves numerical stability since a much smaller matrix is inverted thus alleviating the need for stringent regularizations. In this work, we follow the Chen et al. implementation of the likelihood computations [28, 29]. Finally, the prior causal effect size variance σ_c^2 can be estimated using a fixed effect estimator [109]

$$\sigma_c^2 = Z'\Sigma^{-1}Z - k \tag{4.6}$$

Where k is the number of eigenvectors needed to accumulate 95% of the eigenvalue spectrum of the locus LD matrix Σ.

4.1.2 Incorporating functional genomic data

To integrate functional annotation data within this framework, we use a logistic function to connect a SNP's functional genomic context to its causal status as follows:

$$P(C^j = 1 \mid \gamma, A) = \frac{\exp(\gamma' A^j)}{1 + \exp(\gamma' A^j)} \tag{4.7}$$

$$P(C \mid \gamma, A) = \prod_{j=1}^{m} P(C^j \mid \gamma, A)^{C^j} \left(1 - P(C^j \mid \gamma, A)\right)^{1-C^j} \tag{4.8}$$

The vector A^j is the set of annotations corresponding to the j'th SNP and γ_k is the prior-log odds that a SNP in annotation k is causal. We note that γ can be estimated directly from the data through an Empirical Bayes approach first described in Chapters 1 and 2. However, this restricts functional enrichment estimation to only the fine-mapping loci under investigation. Alternatively, one could exploit potentially more powerful, genome-wide approaches such as stratified LD-score regression [110] that can infer global functional genomic enrichments using only summary data. Our framework is amenable to both approaches, and we allow the user to estimate γ from all the fine-mapping loci jointly using the EM algorithm proposed in [21] or supply it from external analyses.

4.1.3 Model Inference via Importance Sampling

The marginal likelihood in (eq. 4.5) requires enumeration of $O(2^m)$ possible causal sets $(\mathcal{C})$. This rapidly becomes intractable as the number of SNPs grows large, and strategies for dealing with this computational bottleneck need to be considered. Earlier frameworks [20, 28, 21] avoided this problem by simply restricting the total number of potential casual variants to a small number ($k << m$), thus reducing the computational burden to $O(m^k)$. However, even in this reduced model space, enumerating over all possible combinations is inefficient as most causal configurations will contribute minimally to the overall likelihood of the data. Recent works have shown that sampling can circumvent brute-force enumeration by efficiently exploring likely causal configurations through stochastic search [29] – though this still requires pre-specifying a subjective prior that explicitly upper-bounds the maximum number of causal variants considered at the locus.

In this work, we make use of Importance Sampling, a variance reduction technique commonly used in Monte Carlo integration [111], to provide an efficient approximation of the marginal likelihood (eq. 4.5). Unlike other recently proposed sampling techniques, Importance Sampling comes with asymptotic convergence guarantees and allows us to drop the hard cutoff on the maximum number of potential causal variants considered. The summation given in (eq. 4.5) could naively be approximated by sampling directly from the prior and computing a simple Monte Carlo average:

$$C^j \sim Bern\left(P(C^j \mid \gamma, A)\right) \tag{4.9}$$

$$L(Z_* \mid \Sigma, \sigma^2) \approx \frac{1}{S}\sum_{s=1}^{S}\prod_{p=1}^{P} P(Z_p \mid \Sigma, C^{(s)}, \sigma_p^2) \tag{4.10}$$

However, this is inefficient as highly probable causal sets in the posterior may not necessarily be reflected in the prior. To better guide the sampling of highly probable causal sets, we build off the intuition that SNPs with stronger associations (i.e. large Z-scores) are more likely to be casual than ones with weak associations. We can thus construct a discrete proposal distribution, G, to take this into account by simulating causal sets ($C^{(s)}$) at iteration s as independent Bernoulli draws with probabilities given by:

$$G(C^j \mid Z_*) \sim Bern\left(\frac{\sum_p (Z_p^j)^2}{\sum_i \sum_p (Z_p^i)^2}\right) \tag{4.11}$$

$$G(C^{(s)} \mid Z_*) = \prod_{j=1}^{m} G(C^j \mid Z_*)^{C^j}\left(1 - G(C^j \mid Z_*)\right)^{1-C^j} \tag{4.12}$$

Accumulating evidence across multiple traits/populations by summing the chi-square statistics (i.e $(Z^j)^2$), and normalizing by the total sum across all SNPs and traits, creates a probability distribution with the desirable property that it will favor selecting SNPs that have strong evidence of association in multiple traits/populations. By operating in the space of the chi-square statistics (as opposed to Z-scores), we have additional flexibility that allows for strongly associated SNPs to have opposing directional effects in different traits. We can then compute importance weights and re-adjust the bias introduced by sampling from G as follows:

$$L(Z_* \mid \Sigma, \sigma^2) \approx \frac{\sum_{s=1}^{S} \prod_{t=1}^{T} P(Z_t \mid \Sigma, C^{(s)}, \sigma_t^2) W(C^{(s)})}{\sum_{s=1}^{S} W(C^{(s)})} \tag{4.13}$$

$$W(C^{(s)}) = \frac{P(C^{(s)} \mid \gamma, A)}{G(C^{(s)} \mid Z_*)} \tag{4.14}$$

Which we can then use to approximate the per-SNP probabilities using the same S samples:

$$P(C^j = 1) \approx \frac{\sum_{s=1}^{S} 1(C^{j(s)} = 1) \prod_{t=1}^{T} P(Z_t \mid \Sigma, C^{(s)}, \sigma_t^2) W(C^{(s)})}{\sum_{s=1}^{S} \prod_{t=1}^{T} P(Z_t \mid \Sigma, C^{(s)}, \sigma_t^2) W(C^{(s)})} \tag{4.15}$$

4.1.4 Simulation Setup

For computational efficiency, we also performed simulations in which the vectors of association statistics where drawn directly from an MVN distribution (eq. 4.1). In this scenario the NCP (λ_C) was set to 5 at all causal SNPs.

4.1.5 Existing methods

We compared our approach to several existing fine-mapping methods. For single-trait fine-mapping, we compared to FINEMAP and CAVIARBF [29, 28], two methods based on the CAVIAR [27] model that do not incorporate functional annotation data. We ran CAVIARBF v1.4 using the default settings, setting prior variance explained to be 0.05 and the maximum number of causal variants in the model to 3. After CAVIARBF computed Bayes factors for each SNP, we ran their model search algorithm, which outputs posterior probabilities based on Bayes factors. In this step, we set the prior probability of each SNP being causal to $1/m$, where m is the number of variants in the locus. We ran the FINEMAP v1.1 software using default settings, allowing for 3 causal SNPs per locus with prior probabilities of (0.6, 0.3, 0.1) for 1, 2, and 3 causals respectively.

4.2 Results

4.2.1 Fast and reliable performance in single trait fine-mapping

We first sought to empirically assess how our sampling-based approach compared to fine-mapping methods CAVIARBF and FINEMAP. These previous approaches can model multiple causal variants, but were not designed to exploit pleiotropy. As such, in order to make the comparisons fair, we conducted our initial investigation in the context of a single trait. Furthermore, because these methods, as well as our proposed approach, are faster generalizations of the underlying CAVIAR model, we chose not to compare to CAVIAR nor PAINTOR, both of which would predictably have slower computational performance but similar accuracies.

We first assessed performance on the basis of CPU runtime. The number of samples that are drawn to approximate the posterior distribution is invariably connected to the resulting runtime for our method, fastPAINTOR. Therefore, we determined the number of samples required to yield approximately unbiased credible sets and find that one million samples was typically sufficient across a wide-range of locus sizes (see Figure 4.1). We then compared to existing approaches and, not surprisingly, discover that methods that approximate the posterior model space through sampling vastly outperform methods that enumerate over all possible combinations (see Figure 4.2). For example, both fastPAINTOR and FINEMAP scale favorably with the size of the locus, with average run times of (11.5s, 10.8s) per 25KB locus and (186s, 31s) per 250KB locus. The added computational overhead of fastPAINTOR is due to the fact that functional enrichments must be iteratively estimated using an EM-algorithm. If these estimates are supplied from external analyses, running fastPAINTOR* takes an average of 75s per 250KB locus to produce probabilities.

We next evaluated the accuracy of these methods in resolving causal variants to ensure that our sampling approximation did not deflate performance. We simulated 100KB regions with various levels of DHS enrichment to reflect a wide diversity of potential

functional genetic architectures. In general, we see that leveraging functional annotation data improves fine-mapping resolution relative to non-integrative approaches (Figure 4.3) – particularly as causal variants localize within smaller fractions of the genome (i.e. increasing enrichment). For example, the average rank of the causal SNPs was around 21.9 and 21.4 for CAVIARBF and FINEMAP across all functional genetics architectures. On the other hand, when causal variants are diffusely enriched within DHS, their average rank based on fastPAINTOR probabilities is 21.4 while strong functional enrichment yields an average rank of 15.0. Taken together, these results suggest that sampling-based, integrative methods are both scalable and achieve greater accuracy than current state-of-the-art methodologies.

4.3 Discussion

In this work, we introduced a fast fine-mapping method that integrates several sources of genetic data to efficiently and accurately prioritize causal variants. Our Importance Sampling strategy dramatically reduces runtime due to its ability to efficiently sample high probability causal configurations, demonstrating that enumerating over complex model spaces is not necessary for integrative fine-mapping. We conclude by highlighting some caveats and limitations of our proposed framework. Finally, while our Importance Sampling scheme does not explicitly upper-bound the number of causal variants at a fine-mapping regions, it favors exploring parsimonious models over complex ones. We therefore advocate that fine-mapping using our approach be undertaken where there is evidence of only moderate allelic heterogeneity.

4.4 Figures

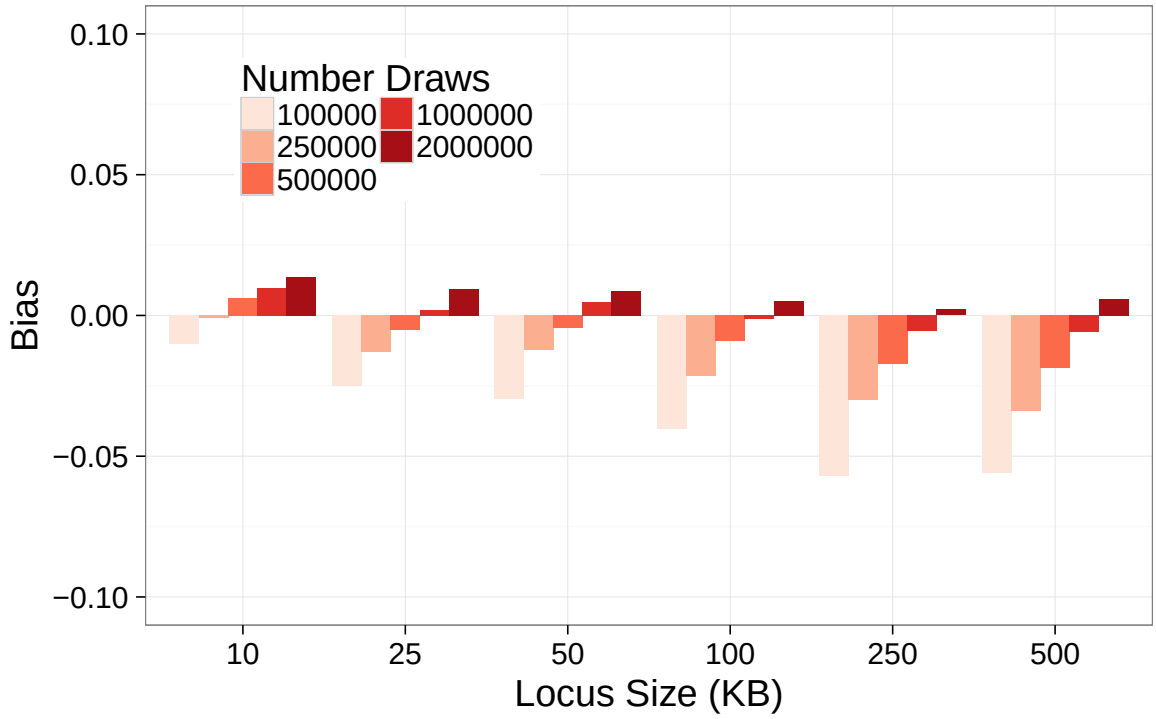

Figure 4.1: One million samples is sufficient to ensure approximately calibrated credible sets. We simulated variable sized regions by drawing from an MVN with reference LD given by the Europeans in the 1000 Genomes V3. We computed 95% credible sets for each simulated locus, and calculated the bias from defined as the difference between the proportion of simulated causal variants that were captured and the expected proportion (0.95). Here, negative bias represents a finding less causal variants than the credible set.

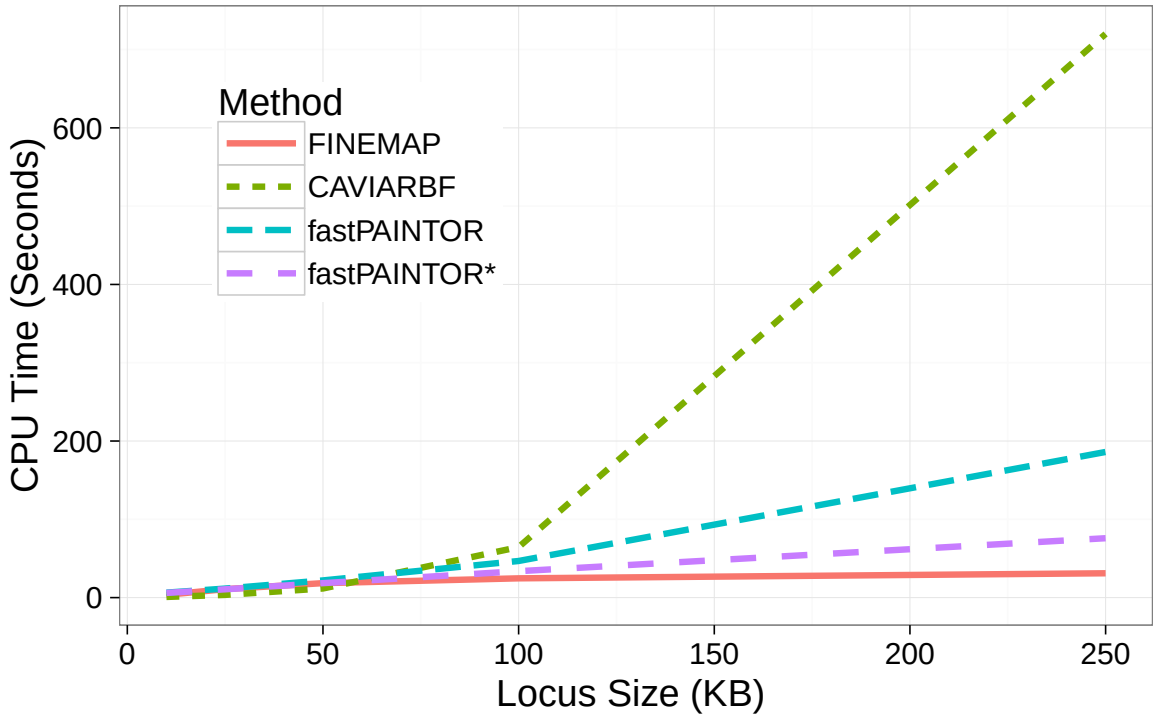

Figure 4.2: Importance sampling improves computational efficiency. Sampling approaches scale favorably with increasing number of SNPs being fine-mapped. We randomly selected 10 GWAS hits and centered increasingly large windows around them. For convenience, we simulated Z-scores by drawing from an MVN with reference LD given by the Europeans in the 1000 Genomes V3. Here, fastPAINTOR estimates functional enrichment empirically while fastPAINTOR* has it provided from external analyses.

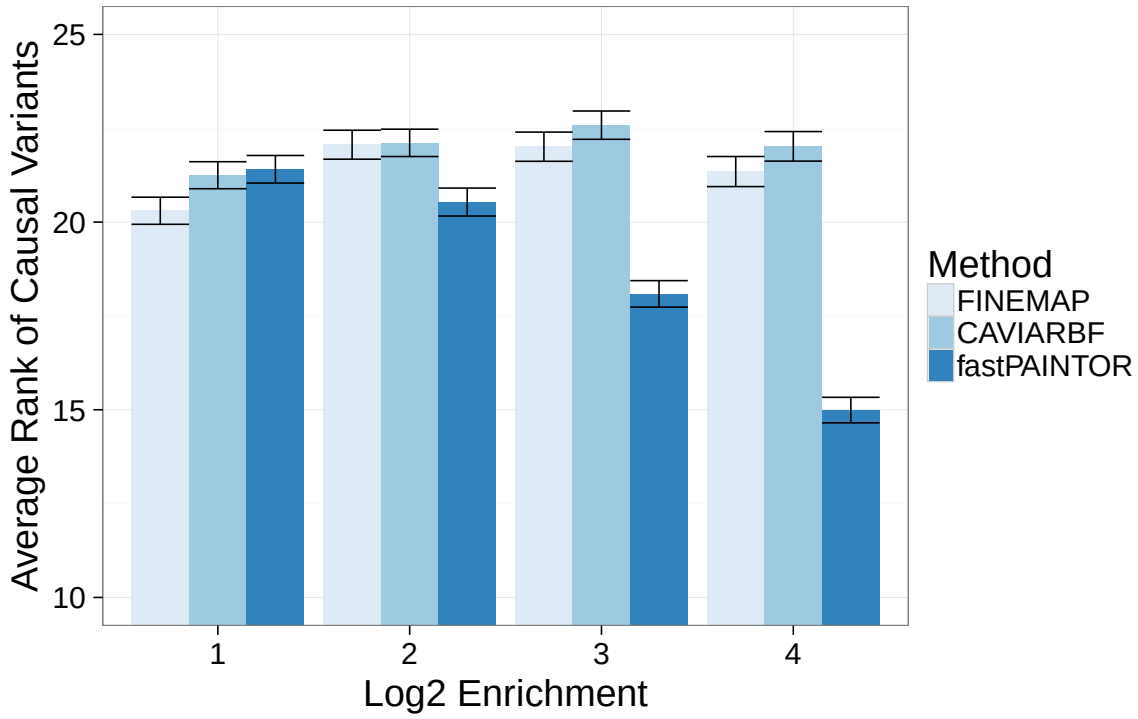

Figure 4.3: fastPAINTOR effectively leverages functional annotation data. We simulated fifty 100KB loci under various functional genetic architectures by drawing summary statistics directly from an MVN distribution. We applied all three methods using default settings and report the average ranks of the causal variants across all simulated loci.

CHAPTER 5

Functionally Informed Novel Discovery of GWAS Risk Regions

5.1 Introduction

Genome-wide Association Studies (GWAS) are the prevailing approach for identifying disease risk loci [9, 11], but the large number of statistical tests performed necessitates stringent p-value thresholds that can limit power. Emerging functional genomics data has revealed that certain categories of variants are enriched for disease heritability [13, **?**, 51, 2, 112, 110, 14, 113, 114, 3]. Thus, incorporating functional information into association analyses has the potential to increase GWAS power [115, 30, 116, 32, 33, 117, 118, 119, 120]. However, previous integrative methods for GWAS hypothesis testing either assume sparse genetic architectures when estimating functional enrichment [33, 119], require knowledge or approximation of the true effect size distribution [115, 30, 116], or do not produce p-values for each SNP as output [33, 117, 118, 120]. In addition, general-purpose methodologies for association testing that can integrate prior information [121, 122, 123] have not been thoroughly evaluated in the context of GWAS leveraging functional genomics data.

In this work, we propose an approach that uses polygenic modeling to weight SNPs according to how well they tag functional categories that are enriched for heritability. Our procedure takes as input summary association statistics along with pre-specified functional annotations (which can be overlapping and/or continuous-valued), and outputs well-calibrated p-values. We utilize a broad set of 75 coding, conserved, regulatory and LD-related annotations that have previously been shown to be enriched for disease

heritability [110, 3]. We incorporate the weights computed by our method using the weighted-Bonferroni procedure described by ref. [115], a theoretically sound approach that ensures proper null calibration and can improve power when employed with informative weights. Through extensive simulations and analysis of UK Biobank phenotypes [124, 125, 126], we demonstrate that our approach reproducibly identifies novel GWAS loci while controlling false positives.

5.2 Results

5.2.1 Overview of Methods

We propose an integrative GWAS framework for Functionally-Informed Novel Discovery of Risk loci (FINDOR). Our approach involves two steps. First, we use stratified LD score regression [110] to compute the expected χ^2 statistic of each SNP based on the functional annotations that it tags; we make use of a broad set of coding, conserved and regulatory annotations [110] as well LD-dependent annotations [3] (conditional on MAF, variants with lower LD have larger causal effect sizes). Second, we stratify SNPs into bins of expected χ^2 and estimate the proportion of null ($\hat{\pi}_0$) and alternative ($\hat{\pi}_1$) SNPs within each bin using the Storey π_0 estimator [127] to obtain bin-specific weights. We limit the number of bins to 100 and normalize the weights to have mean 1, ensuring proper null calibration [115]. We then divide the observed p-values within each bin by these weights to produce re-weighted p-values for each SNP. Bins with larger values of $\hat{\pi}_1$ will have larger weights, leading to more significant p-values. Details of the method are described in the Online Methods section; we have released open-source software implementing the method (see URLs).

5.2.2 Simulations assessing calibration and power

We assessed calibration and power via simulations using real genotypes from the UK Biobank interim release [124] ($N = 100K$ subsampled British-ancestry samples, $M =$

9.6M well-imputed SNPs; see Online Methods). We simulated polygenic traits with 10,000 or 20,000 causal variants and SNP-heritability (h_g^2) equal to 0.1 or 0.2. All causal variants were placed on odd chromosomes, with functional enrichment based on a meta-analysis of 31 traits using the baselineLD model described in ref. [3] (Table 5.2; see URLs), and even chromosomes served as null data. Weights were computed by running stratified LD score regression [110] on association statistics computed from simulated phenotypes, without knowledge of the true functional enrichment parameters used to generate the phenotypes. We compared FINDOR to three other methods that can incorporate auxiliary information for each SNP: Stratified False Discovery Rate (S-FDR) [121], Grouped Benjamini Hochberg (GBH) [122], and Independent Hypothesis Weighting (IHW) [123]. For each of the four methods, we considered four different criteria for stratifying SNPs into bins: predicted χ^2 statistics under the baselineLD model (baseLD); predicted χ^2 statistic under the baselineLD model trained using off-chromosome data via a Leave-One-Chromosome-Out approach (baseLD-LOCO); total LD score of a SNP (LDscore), motivated by a previous study reporting that simple LD information can be used to improve GWAS power [30]; and randomly chosen bins (Random). We also considered unweighted raw p-values (Unweighted), a natural benchmark. For both null (even) and causal (odd) chromosomes, the primary metric was the number of independent genome-wide significant associations identified. Throughout this work, we define an independent association as a SNP that exceeds a significance threshold (e.g., 5×10^{-8}), together with all linked SNPs that have an $r^2 > 0.01$ within 5Mb. We performed 1,000 simulations and averaged results across simulations. Further details of the simulation framework are provided in the Online Methods section.

We first assessed calibration on null chromosomes. We determined that FINDOR was well-calibrated, producing a similar number of false-positive (independent, genome-wide significant) associations at null loci as the Unweighted approach (see Figure 5.1 and Table 5.3). This remains true whether we infer functional enrichment and compute expected χ^2 statistics using all GWAS data (baseLD) or using off-chromosome data (baseLD-LOCO), motivating the use of the baseLD stratification criteria in the remainder of this work.

Similarly, FINDOR was well-calibrated at less stringent significance thresholds (see Table **??**). Although FINDOR makes multiple passes over the data, which in principle could overfit the data and produce false positives, this does not occur in practice, likely due to the small number of global parameters estimated (hundred) relative to the large number of hypothesis tests performed (millions).

On the other hand, S-FDR, GBH and IHW each exhibited moderate to severe increases in false-positive associations, particularly at higher polygenicity and lower SNP-heritability. For example, at a polygenicity of 20,000 causal variants and $h_g^2 = 0.1$, we observe an average (SE) of 0.10 (0.02) false positives per simulated GWAS using raw unweighted p-values and 0.06 (0.01) using FINDOR with baseLD criteria, while S-FDR, GBH, and IHW with baseLD yield 1.6 (0.2), 1.6 (0.2), and 1.3 (0.2) false positives, respectively (see Figure 5.1 and Table 5.3). This inflation is exacerbated at smaller sample sizes (see Figure 5.5). We hypothesize that this may be due to the fact that the theoretical guarantees provided by these procedures are unlikely to be valid when the auxiliary information incorporates the dependence structure between hypothesis tests; this limitation was previously noted by Ignatiadis et al. [123] and clearly affects both baseLD and LDscore stratifying criteria. Furthermore, while GBH and IHW were consistently well-calibrated under random stratification (see Figure 5.1, purple bars), S-FDR was not, perhaps because S-FDR requires additional adjustments for the number of strata used [128].

We next evaluated power to detect true associations on causal chromosomes. We restricted our assessment of power to Unweighted and FINDOR, as they were the only methods that were well-calibrated under the null for all stratification criteria. FINDOR attained an 8.6-38% increase in the number of true (independent, genome-wide significant) associations, depending on polygenicity (10,000 or 20,000 causal variants) and SNP-heritability (0.1 or 0.2) (see Figure 5.2 and Table 5.5). The relative improvement was smaller at lower polygenicity and larger SNP-heritability, each of which correspond to higher absolute power. Our method has a fixed budget of weights that it can allocate, and we hypothesize that when absolute power is high it is more likely to allocate weights to SNPs that are already genome-wide significant, explaining the smaller relative improve-

ment. In addition, the enrichment estimates provided by stratified LD score regression are expected to be less precise at lower polygenicity. However, the smaller relative improvement still translated into a larger absolute improvement in settings with higher absolute power.

5.2.3 Application to 27 UK Biobank traits

We applied FINDOR to the interim UKBiobank release [124], which includes N=145K European-ancestry samples and $M = 9.6M$ well-imputed SNPs. We analyzed 27 independent, highly heritable traits (average N=130K; see Table 5.1 and Online Methods). We computed summary association statistics using BOLT-LMM v2.1 [129] (Unweighted approach). We applied FINDOR to these summary statistics and compared the number of independent, genome-wide significant associations identified by FINDOR vs. the Unweighted approach. In total, FINDOR identified 207 more associations (see Table 5.1 and Table 5.6), a statistically significant improvement (block-jacknife SE = 20.4, $p < 1 \times 10^{-20}$). This corresponds to an average per-trait improvement of 13% (SE=2.5%) and an aggregate improvement of 6.8%; FINDOR identified more associations than the Unweighted approach for 24 out of 27 traits, and the same number of associations for the remaining three traits. The aggregate improvement was lower than the average per-trait improvement because the relative improvement was smaller for traits with higher power (i.e. more associations) (see Figure 5.3), consistent with simulations. In particular, disease traits exhibited a larger improvement (20% average per-trait, 22% aggregate, see Table 5.8), consistent with smaller effective sample size (i.e. smaller value of sample size * observed-scale SNP-heritability) due to the relatively small number of disease cases. Qualitatively similar results were obtained at a more stringent p-value threshold of 5×10^{-9} (see Table 5.9). We note that, compared to the 13% average per-trait improvement of FINDOR with the baselineLD model [3], FINDOR with the baseline model [110] (which excludes LD-related annotations) attained only a 7.1% average per-trait improvement and 4.3% aggregate improvement (72 fewer GWAS hits; jackknife SE on difference = 13.3, $p = 6.3 \times 10^{-8}$, see Table 5.6). This indicates that the LD-related annotations of the baselineLD model contain

valuable information for increasing association power; in particular, these annotations avoid the phenomenon of strong LD between in-annotation and out-annotation SNPs that may limit the potential of coding, conserved and regulatory annotations to increase association power despite their strong enrichments for trait heritability.

Next, we carried out a UK Biobank-based replication analysis for the 27 traits using non-overlapping samples in the full UK Biobank release. Starting with the 459K European-ancestry samples, we excluded the 145K samples that were present in the interim release and computed summary statistics using BOLT-LMM v2.3, a highly computationally efficient implementation for very large data sets [126]. This produced a well-powered replication data set (average N=283K). We evaluated strength of replication by computing the replication slope, defined as the slope of a regression of estimated standardized effect sizes in replication data vs. discovery data, restricting to lead SNPs at genome-wide significant loci from the discovery data (we excluded lead SNPs that were not present in the replication data). We computed replication slopes for three classes of loci: (1) those that were genome-wide significant only using the Unweighted approach, (2) only using FINDOR p-values, or (3) using both methods. The 49 loci that were significant only using the Unweighted approach produced a replication slope of 0.57 (SE=0.043). The 230 loci that were significant only using FINDOR (i.e. novel discoveries) produced a slightly stronger replication slope of 0.66 (SE=0.018); the difference was not statistically significant based on the small number of data points, particularly for Unweighted only. As expected, the 2766 loci that were significant using both methods produced the strongest replication slope of 0.91 (SE= 0.003), as this class of loci included the most significant associations (see Figure 5.4 and Table 5.10). We also performed a separate replication analysis for nine traits for which summary statistics from independent, non-UK Biobank GWAS were available (see Online Methods, Table 5.11). In this analysis, the 36 loci that were significant only using FINDOR (i.e. novel discoveries) produced a replication slope of 0.69 (SE=0.11) in non-UK Biobank data, which did not differ significantly from the replication slope for the 410 loci that were significant using both methods (0.66, SE=0.012, see Figure 5.4 and Table 5.12). Only a single locus was significant only using

Unweighted p-values in this analysis, therefore we do not report a replication slope for this class. Overall, these results confirm that the novel loci identified by FINDOR robustly replicate in independent samples.

Finally, we applied FINDOR to the 27 traits using the full set of 459K European-ancestry samples (average N=416K), analyzing summary statistics computed using BOLT-LMM v2.3 [126]. The Unweighted approach identified 13,283 independent genome-wide significant associations in this data. FINDOR identified 583 more associations (see Table 5.13, Jackknife SE = 40.6, $p < 1 \times 10^{-20}$), corresponding to an average per-trait improvement of 6.9% (SE = 0.66%) and an aggregate improvement of 4.1% (see Table 5.1); FINDOR identified more associations than the Unweighted approach for all 27 traits. Once again, the relative improvements decreased as a function of sample size times observed-scale SNP-heritability (see Figure 5.3, Table 5.1), with larger relative improvements for disease traits (10% average per-trait, 10% aggregate) and smaller relative improvements in the 459K release vs. the 145K release, consistent with simulations. We further characterized Unweighted-only and FINDOR-only loci by contrasting their overlap with molecular QTL 95% causal sets [4] (which are weakly correlated with the baselineLD model annotations used by FINDOR: $|r| \approx 0.05$ for most annotations, see ref. [4]). The lead SNPs at FINDOR-only loci had substantial overlap with molecular QTL 95% causal sets (and substantially larger molecular QTL causal posterior probabilities on average), compared to Unweighted-only loci (see Table 5.14); this implies that loci identified by FINDOR are not only more numerous, but also more amenable to biological interpretation and mechanistic insights. Overall, these results indicate that FINDOR can provide a substantial increase in power – particularly for studies with smaller effective sample sizes, such as studies of disease traits.

5.3 Discussion

We have introduced a p-value weighting approach that leverages polygenic functional enrichment to improve association power. We demonstrated in simulations that our

FINDOR framework is properly calibrated under the null and improves power to detect causal loci. We reproducibly identified hundreds of new loci across a broad set of UK Biobank traits, with increased prospects for biological interpretation (see Table 5.14). We achieved this by using a multi-faceted functional enrichment model that includes coding, conserved, regulatory and LD-related annotations [110, 3].

Previous studies that assumed sparse genetic architectures achieved 3-5% increases in association power [33, 119]. In detail, ref. [33] reported a 5.0% increase in power (average N=57K for 18 traits) and ref. [119] reported a 2.7% increase in power ($P < 1 \times 10^{-8}$; median $N_{eff} = 4/(1/N_{case} + 1/N_{control}) =$ 6K for 123 binary traits, median N=23K for 96 quantitative traits). (Ref. [119] also reported a 13.7% increase in the number of "unsettled" associations ($1 \times 10^{-10} < P < 1 \times 10^{-8}$), a metric that yields much larger increases.) In contrast, our polygenic approach achieved a 7% increase in association power (or 13% increase in power averaged across traits) in the interim UK Biobank analysis despite the larger sample size analyzed (average N=130K), which corresponds to smaller increases in power (see Figure 5.3). Ideally, we would have assessed those previous methods in the current study; however, we were unable to do so, either because no software implementation was available [119], or because the available output (Bayes factors and posterior probabilities of association) was not directly comparable to the p-value thresholds used to assess significance in our study (and most GWAS studies) [33, 117, 118, 120]. We instead elected to assess previous methods that could incorporate information from our polygenic functional enrichment model and produce p-value thresholds for hypothesis testing: Stratified FDR (S-FDR) [121], Grouped Benjamini Hochberg (GBH) [122], and Independent Hypothesis Weighting (IHW) [123].

Stratifying SNPs based on predicted (tagged) variance was previously proposed by ref. [32] (incorporating 10 functional annotations), which made a key contribution to the literature by highlighting the potential of this approach. The study demonstrated that this criteria improved replication rates, and also reported that it increased power when applying S-FDR [121]. However, S-FDR did not achieve proper null calibration in our simulations, even under random stratification, perhaps because S-FDR requires additional

adjustments for the number of strata used [128]. Furthermore, S-FDR, GBH, and IHW were all unable to correctly control false positives when LD-dependent stratification criteria (LDscore or BaseLD) were employed; as noted above, theoretical guarantees about false positives are unlikely to be valid when the stratification criteria incorporate the dependence structure between hypothesis tests [123]. Our approach bears some similarity to the multi-threshold association tests proposed by ref. [30, 116], which use knowledge of the true effect size distribution to solve a convex optimization problem to determine appropriate thresholds. Given knowledge of the true effect size distribution, this approach is theoretically optimal [115, 30]; however, this information is rarely available in practice and must be fixed a priori or approximated from the data [115, 30, 116]. Finally, although we employ a fundamentally different weighting strategy, our method draws on insights from ref. [115], which established the theoretical basis for data-driven p-value weighting.

We conclude with several limitations of our work. First, previous studies have demonstrated that complex traits often exhibit cell-type specific functional enrichments [**?**, 51, 2, 112, 14, 113, 114, 110, 33, 130, 131], which we did not incorporate in this study. Incorporating cell-type-specific functional enrichments may further increase power, although care will be required to avoid overfitting since identifying critical cell types requires extensive model selection. Second, our modeling of MAF-dependent architectures is limited; while our baselineLD functional model includes MAF-bin annotations for common SNPs (MAF $> 5\%$), it does not model MAF-dependent architectures for rare and low-frequency variants. A possible future direction would be to incorporate MAF-dependent annotations, e.g., via the widely used α model [132, 133, 134]. Third, we anticipate that GWAS will grow larger and more powerful in the years ahead, but the relative improvement of our method decreases as a function of absolute power. However, we anticipate that our method will continue to produce large relative improvements for disease phenotypes (as in Table 5.1), for which the ongoing challenge of recruiting disease cases will continue to limit effective sample size. Fourth, our UK Biobank replication of novel loci from the interim UK Biobank release could in principle be inflated by relatedness within the UK Biobank; however, our non-UK Biobank replication produced a concordant replication

slope, suggesting that this effect is limited. Fifth, we evaluated our method only using European-ancestry samples. Although our previous work has provided evidence that functional enrichment is consistent across populations [113, 21], generalizing our results to non-European samples is currently an open question, as it is unclear whether functional enrichments inferred in large European samples should be incorporated. Despite these limitations, we anticipate that FINDOR will be a valuable and practical tool for leveraging polygenic functional enrichment to improve GWAS power.

5.4 Online Methods

5.4.1 FINDOR method

The aim of our method is to re-weight SNPs according to how well they tag heritability enriched categories. This is accomplished in two steps. First, we estimate a function that predicts the χ^2 statistic (i.e. tagged variance) at each SNP using a comprehensive assortment of functional annotations which include coding, conserved and regulatory annoations [110], as well as LD-dependent annotations [3]. The stratified LD score regression [110, 3] framework is a natural choice for this task. In stratified LD score regression, the association statistic at SNP j measured (or imputed) in N_j individuals is expressed in terms of its tagging of studied annotations. Specifically,

$$\mathrm{E}(\chi_j^2) = N_j \sum_C \tau_C \ell(j, C) + N_j \alpha + 1 \tag{5.1}$$

where α represents confounding biases [135], τ_C is the effect size on per-SNP heritability of annotation C, and $\ell(j, C)$ is the LD score which indicates the degree to which SNP j tags annotation C:

$$\ell(j, C) = \sum_k C(k) r_{k,j}^2 \tag{5.2}$$

Here, $C(k)$ is the value of annotation C at SNP k and $r^2_{k,j}$ signifies the squared Pearson correlation coefficient between SNPs k and j [110, 3] (computed from 503 European individuals of the 1000 Genomes (V3) reference panel [82]). In a typical analysis, the quantity of interest is an estimate of τ_C ($\widehat{\tau_C}$) which can be interpreted as the strength of enrichment (or depletion) of heritability within annotation C. These values are obtained through a multivariate (weighted) regression of the observed χ^2 statistics at HapMap3 SNPs against the corresponding values of $\ell(j, C)$. In this work, we use $\widehat{\tau_C}$ to predict the expected χ^2 statistic at all GWAS SNPs. For a given SNP j, we have:

$$\widehat{\chi^2_j} = N_j \sum_C \widehat{\tau_C} \ell(j, C) + N_j \hat{\alpha} + 1$$

The $\widehat{\tau_C}$ parameters can either be global estimates that are learned from the entire GWAS data set (restricted to HapMap3 SNPs), or chromosome-specific estimates that are learned from the remaining off-chromsome data. Empirically, we find that using the entire genome does not introduce false positives (see Figure 5.1).

Next, we stratify SNPs based on their expected χ^2 into B distinct, evenly-sized bins. In practice, to ensure a sufficiently coarse partitioning of the data we set $B = 100$. For densely imputed data such as the UK Biobank this results in each bin b containing $\approx$ 100K SNPs. We then estimate the proportion of null ($\hat{\pi}_{0,b}$) and alternate SNPs ($\hat{\pi}_{1,b}$) by fitting a cubic spline to the histogram of p-values as proposed by ref. [127] and implemented in the q-value package [136]. Following ref. [122] we weight each p-value by dividing the nominal p-value by the ratio of $\hat{\pi}_{1,b}$ to $\hat{\pi}_{0,b}$. Intuitively, bins with higher proportion of true alternates will have their p-value weighted downward (i.e. made more significant). However, unlike ref. [122], we normalize these weights to have mean one:

$$\hat{w}_b = \frac{\frac{\hat{\pi}_{1,b}}{\hat{\pi}_{0,b}}}{\frac{1}{B}\sum_{b=1}^{B} \frac{\hat{\pi}_{1,b}}{\hat{\pi}_{0,b}}} \tag{5.3}$$

Theory developed in ref. [115] suggests that despite the fact that $\hat{w}_b$ is learned in a data-dependent manner, a weighting scheme with this property preserves control of type I

error since the number of weights we learn (i.e. 100) is significantly less than number of hypothesis test we perform.

5.4.2 S-FDR, GBH and IHW methods

We adapted three previously proposed methodologies that leverage prior information to serve as comparators to our approach: Stratified False Discovery Rate (S-FDR) [121], Grouped Benjamani Hochberg (GBH) [122], and Independent Hypothesis Weighting (IHW) [123]. Because these are FDR-controlling procedures, we calibrate the expected level of FDR control required to match the more traditional criteria for genome-wide significance ($p \leq 5 \times 10^{-8}$). We refer to this level of genome-wide FDR control as q_{GW}, which we estimate as the maximum q-value [127] amongst SNPs with p-values $\leq 5 \times 10^{-8}$. We implemented S-FDR by binning SNPs according to various criteria used in this study. We then computed q-values for each bin and rejected all SNP within the bin whose q-value was less than q_{GW}. This stratified FDR strategy is similar to Schork et al. [32]. GBH and IHW were implemented in the IHW (v1.1.3) and IHWpaper (v1.0.2) packages [123] which we ran using the default setting and specified the level of FDR control to be q_{GW}. GBH takes as input group labels which were identical to the groupings used with FINDOR and S-FDR, while IHW handled raw measurements of the auxiliary information (e.g., each SNP had its own unique value of predicted tagged variance under BaseLD).

5.4.3 Functional Annotations

We employed the 75 functional annotations of the baselineLD model, which were previously demonstrated to be enriched for heritability across a wide variety of complex traits [3] (see Table 5.2). For clarity, we provide a brief description of the model's contents below. This model is an extension of the 53 annotation baseline model developed in ref. [110]. Briefly, the initial baseline model consisted of 24 main annotations to which 500bp flanking windows were added to create secondary annotations. These include histone modifications H3K4me1, H3K4me3, H3K4ac, H3K9ac, and H3K27ac that span mul-

tiple cell types; genic elements describing coding, 3′ UTR, 5′ UTR, promoter, and intronic regions; combined chromeHMM and Segway segmentations (7 states); Digital genomic footprint and transcription factor binding sites; DNase Hypersensitivity I sites; Super enhancers and FANTOM5 enhancers; and sites conserved across mammals (see ref. [110] and references therein). The baseline model was augmented in ref. [3] by adding four more binary annotations based on super-enhancers and typical enhancers, as well as two conserved annotations based on GERP++ scores. The baselineLD model was then created by adding ten common MAF bin annotations and six LD-related annotations (predicted allele age, LLD-AFR, recombination rate, nucleotide diversity, background selection statistic, and CpG content).

5.4.4 Simulations

Simulations were based on real imputed genotypes of British ancestry individuals from the UK Biobank interim release (N=113K). We removed poorly imputed SNPs whose INFO score was less than 0.6, filtered out rare variants whose minor allele count was less than five in European individuals of the 1000 genomes, and additionally excluded the MHC region on chromosome six. This resulted in 9.6M SNPs for analysis. We randomly subsampled N individuals from this data set (in our main simulations, N=100K) and simulated continuous phenotypes under a polygenic model with normally distributed causal effect sizes and a specified number of causal variants. Genotypes were standardized so that each causal variant explained an equal proportion of the phenotypic variance. To induce functional enrichment, we altered the prior probability that a SNP was selected to be causal, setting this to be proportional to $\mathrm{Var}(\beta_j) = \sum_C C(j)\tau_C$. Empirically estimated enrichment parameters (τ's) were obtained from a meta-analysis of the 31 traits reported in ref. [3] (see Table 5.2). This allowed our simulations to more closely reflect the complex, multi-faceted genetic architectures observed in real data. We note that functional enrichment was estimated without knowledge of the true functional enrichment used to simulate phenotypes. To obtain the baseLD-LOCO criteria, we estimated chromosome specific $\tau's$ using off-chromosome data. Finally, we used PLINK v1.9 [137] to compute as-

sociation statistics for each SNP. The primary metric of interest in both real and simulated data was the number of independent GWAS hits (at a level of $p < 5 \times 10^{-8}$) that the various methodologies identified. We conservatively define independent hits using PLINK's clumping algorithm with 5MB window and an r^2 threshold of 0.01. Reference LD for this procedure was based on the same 113K British ancestry individuals for both simulations and real data analysis. To avoid over-counting loci where allelic heterogeneity was likely present in real data, we collapsed independent signals that were within 100KB of one another into a single locus.

5.4.5 UK Biobank data set

We used BOLT-LMM [129, 126] to compute mixed model association statistics. A key advantage of this approach that it allowed us to retain related individuals in this dataset, thereby maximizing power and data usage [126]. We performed basic QC on each trait following standard GWAS practices (see ref. [126] for details). For each phenotype, we generated three sets of summary statistics based on individuals of self-reported European ancestry. The first set of summary statistics consisted of 145K individuals from the interim UK Biobank release [124, 129]. This served as our "discovery" dataset and had mean sample size of $\approx 130K$ across 27 independent traits (see below). We then created two additional sets of summary statistics derived from the full UKBiobank release [125]. Our "replication" dataset consisted of 314K individuals in the final release that were not present in the interim release (mean sample size = 283K). This dataset was used to verify findings in the discovery sample. Our "full" dataset was the entire compendium of 459K individuals (average N= 416K). While we computed summary statistics at 20 million SNPs which passed filtering and QC thresholds (see ref. [126]), to ensure compatibility with simulations, we ran association analyses restricting to the same set of well-imputed $\approx$ 9.6M biallelic SNPs which, upon intersection, resulted in 9.6M SNPs for the interim release and 8.9M in the full release.

To avoid over-representation of certain phenotypic classes in our real data analysis

that may bias our results, we constructed a set of 27 (roughly) independent and heritable traits, only retaining traits that exhibited a phenotypic correlation $r^2 < 0.1$. To ensure adequate power to estimate functional enrichment, we also required that the traits have a heritability Z-score that was greater six in the 145K dataset to be included in our analysis [110]. An overview of the phenotypes analyzed in this work can be found in Table 5.1.

5.4.6 Independent Non-UK Biobank data

To confirm the robustness of our findings we sought to replicate them in non-UK Biobank GWAS. We were able to obtain publicly available GWAS summary statistics for nine GWAS traits that were part of the 27 trait analysis (see Table 5.11). As SNP coverage was not uniform, we intersected the data sets and only examined significant findings that were present both GWAS. When per-SNP sample sizes were unavailable, we used the max N obtained from the corresponding publication (see Table 5.11). External GWAS alleles were polarized to the UK Biobank and standardized effect sizes were compared ($\frac{Z}{\sqrt{N}}$).

5.4.7 Replication Analysis

We carried out replication analysis in independent UK Biobank (27 traits; 3307 loci) and non-UK Biobank data (9 traits; 446 loci). To ensure compatibility across all traits and data sets, standardized effect sizes were computed by dividing Z-scores by the square root of the study sample size. To quantify replication, we computed the replication slope, defined as the slope resulting from a regression of the standardized effect sizes in the replication data versus the discovery data. We restricted our analysis to lead SNPs at independent, genome-wide significant loci in the discovery data that were also present in the replication data. We defined three class of loci: those that were genome-wide significant only using the Unweighted approach, only using FINDOR p-values, or using both methods. Because re-weighting could result in different lead SNPs at the same locus, we desig-

nated a locus as genome-wide significant using both methods if the lead SNP discovered by unweighted p-values had an $r^2 > 0.01$ with the lead SNP discovered by FINDOR.

5.5 Tables

				145K			459K		
Class	Trait	N	h_g^2	Unweighted	FINDOR	%Improve	Unweighted	FINDOR	%Improve
Anthropometric	Balding Type I	68K/208K	0.21	96	100	4.2%	334	346	3.6%
	Body Mass Index	145K/458K	0.28	117	132	12.8%	908	950	4.6%
	Heel T Score	141K/446K	0.33	300	308	2.7%	1130	1149	1.7%
	Height	145K/458K	0.64	674	690	2.4%	2395	2402	0.3%
	Waist-hip Ratio	145K/458K	0.17	98	104	6.1%	460	506	10.0%
Blood Cell	Eosinophil Count	140K/440K	0.21	187	200	7.0%	699	731	4.6%
	Mean Corpular Hemoglobin	141K/443K	0.22	237	248	4.6%	765	791	3.4%
	Red Blood Cell (RBC) Count	141K/445K	0.25	192	206	7.3%	840	885	5.4%
	RBC Distribution Width	141K/445K	0.20	198	212	7.1%	652	674	3.4%
	White Blood Cell Count	131K/444K	0.21	148	165	11.5%	713	750	5.2%
Disease	Auto Immune Traits	145K/459K	0.04	14	18	28.6%	75	86	14.7%
	Cardiovascular Diseases	145K/459K	0.12	38	49	28.9%	285	314	10.2%
	Eczema	145K/459K	0.08	35	46	31.4%	181	198	9.4%
	Hypothyroidism	145K/459K	0.05	27	30	11.1%	139	153	10.1%
	Respiratory Diseases	145K/459K	0.06	24	29	20.8%	104	109	4.8%
	Type 2 Diabetes	145K/459K	0.05	14	14	0.0%	76	86	13.2%
Other	Age at Menarche	75K/242K	0.25	52	56	7.7%	318	338	6.3%
	Age at Menopause	44K/143K	0.11	18	18	0.0%	85	91	7.1%
	FEV1-FVC Ratio	124K/370K	0.27	174	185	6.3%	684	714	4.4%
	Forced Vital Capacity (FVC)	124K/372K	0.23	90	99	10.0%	544	565	3.9%
	Hair Color	143K/452K	0.14	140	143	2.1%	428	436	1.9%
	Morning Person	130K/410K	0.11	14	14	0.0%	156	165	5.8%
	Neuroticism	124K/372K	0.11	11	16	45.5%	128	149	16.4%
	Smoking Status	145K/458K	0.10	18	24	33.3%	154	178	15.6%
	Sunburn Occasion	109K/344K	0.07	23	25	8.7%	78	82	5.1%
	Systolic Blood Pressure	134K/422K	0.22	98	106	8.2%	666	703	5.6%
	Years of Education	144K/455K	0.14	17	24	41.2%	286	315	10.1%
	Overall	145K/459K	NA	3054	3261	6.8%	13283	13866	4.4%
	Average Per-Trait	130K/409K	0.18	113	120	13%	491	513	6.9%

Table 5.1: **FINDOR increases power across 27 UK Biobank traits**. For each trait, we report the number of independent, genome-wide significant loci identified by the Unweighted approach and by FINDOR in the 145K and 459K UK Biobank releases.

5.6 Figures

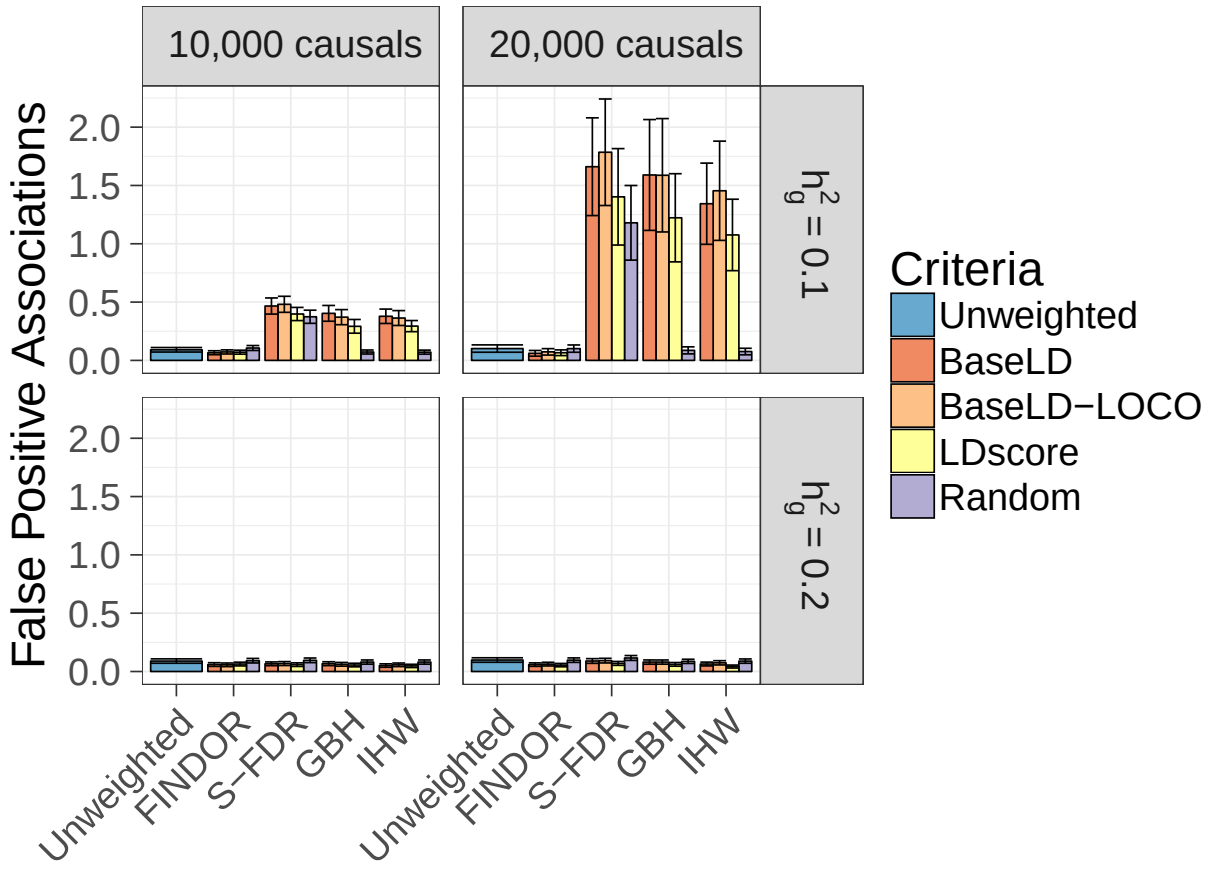

Figure 5.1: **FINDOR is well-calibrated in simulations of null loci.** We report the average number of independent, genome-wide significant ($p < 5 \times 10^{-8}$) associations on null chromosomes. Results are averaged across 1000 simulations. Error bars represent 95% confidence intervals. Numerical results are reported in Table 5.3.

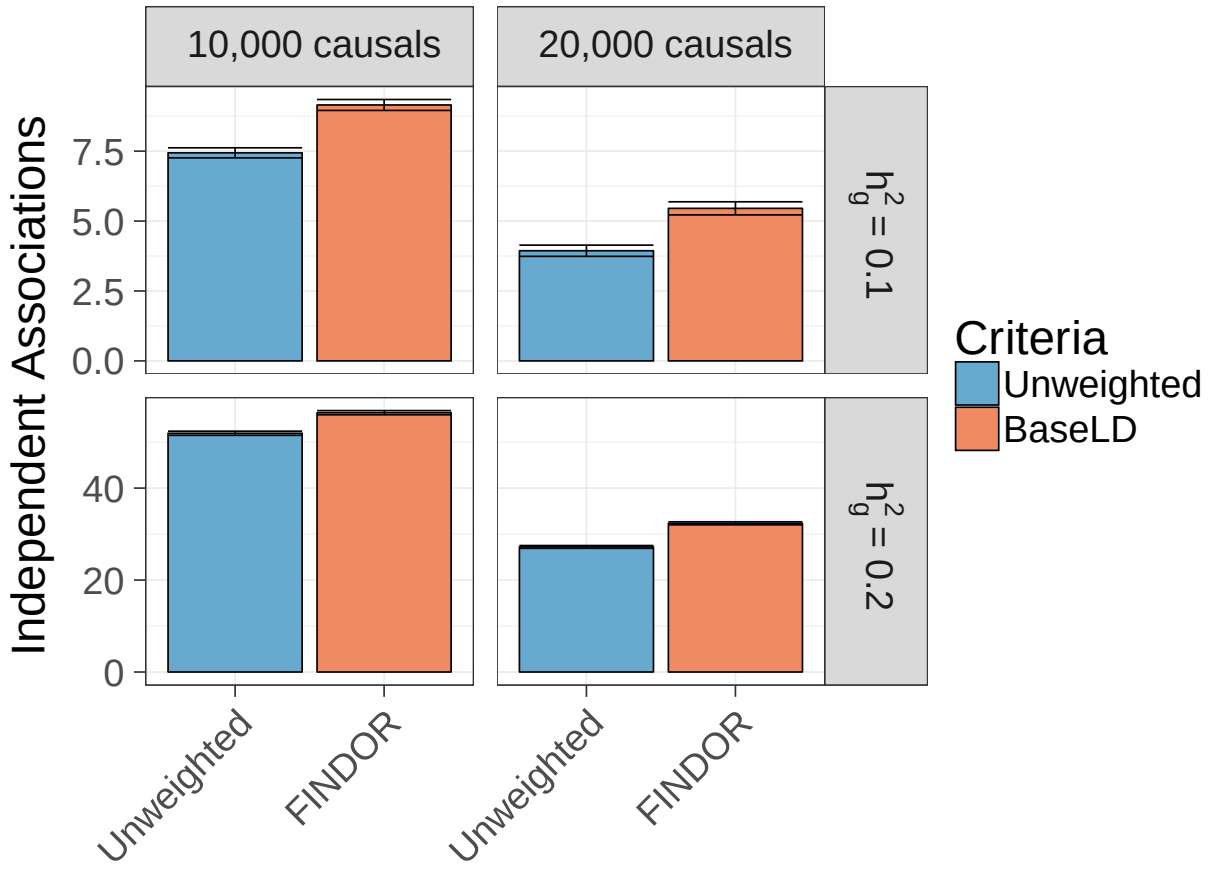

Figure 5.2: **FINDOR increases power in simulations of causal loci.** We report the average number of independent, genome-wide significant ($p < 5 \times 10^{-8}$) associations on causal chromosomes. Results are averaged across 1000 simulations. Error bars represent 95% confidence intervals. Numerical results are reported in Table 5.5.

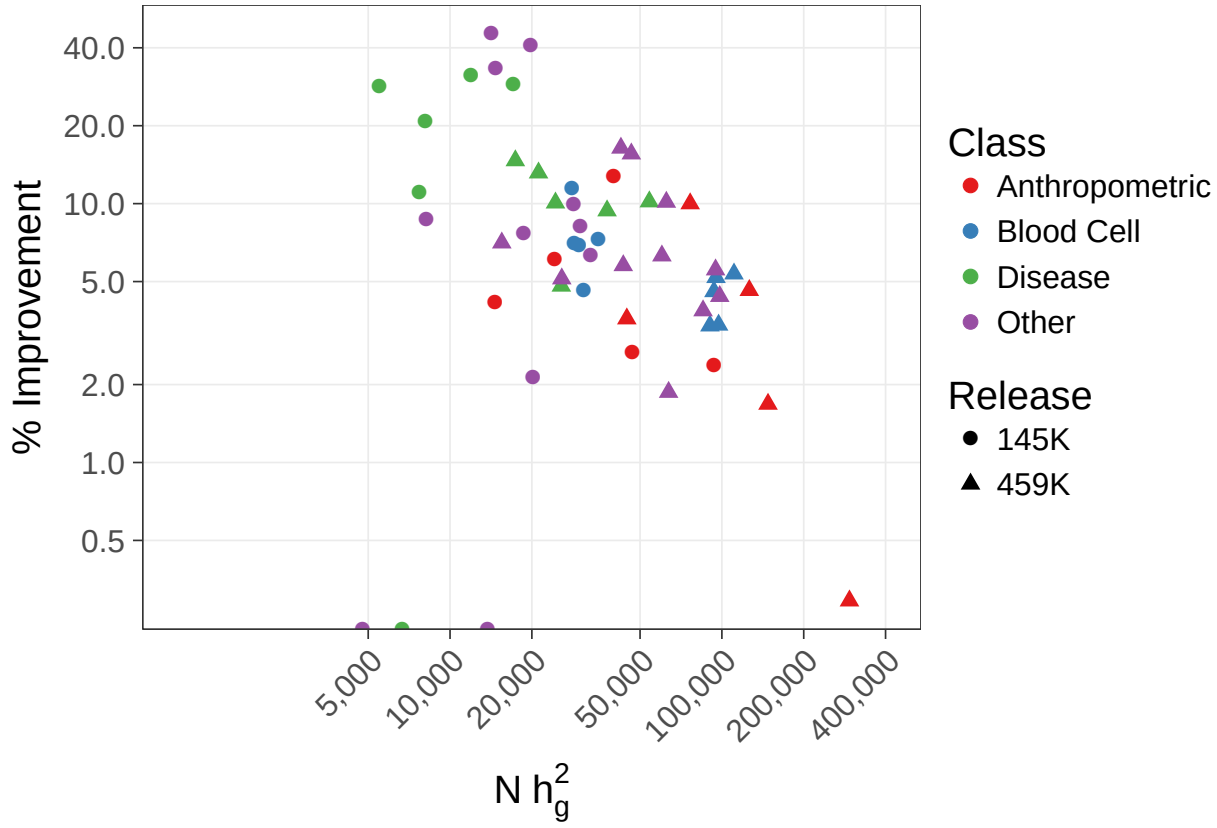

Figure 5.3: **Relative improvement of FINDOR in real UK Biobank phenotypes decreases as a function of absolute power**. We plot the relative improvement in the number of independent GWAS loci identified by FINDOR compared to Unweighted p-values vs. sample size times observed-scale SNP-heritability, using log scales. The three circles at the bottom of plot correspond to traits where the number of loci was identical for FINDOR compared to Unweighted p-values (0% improvement). Numerical results are reported in Table 5.1 and Tables 5.6 and 5.13.

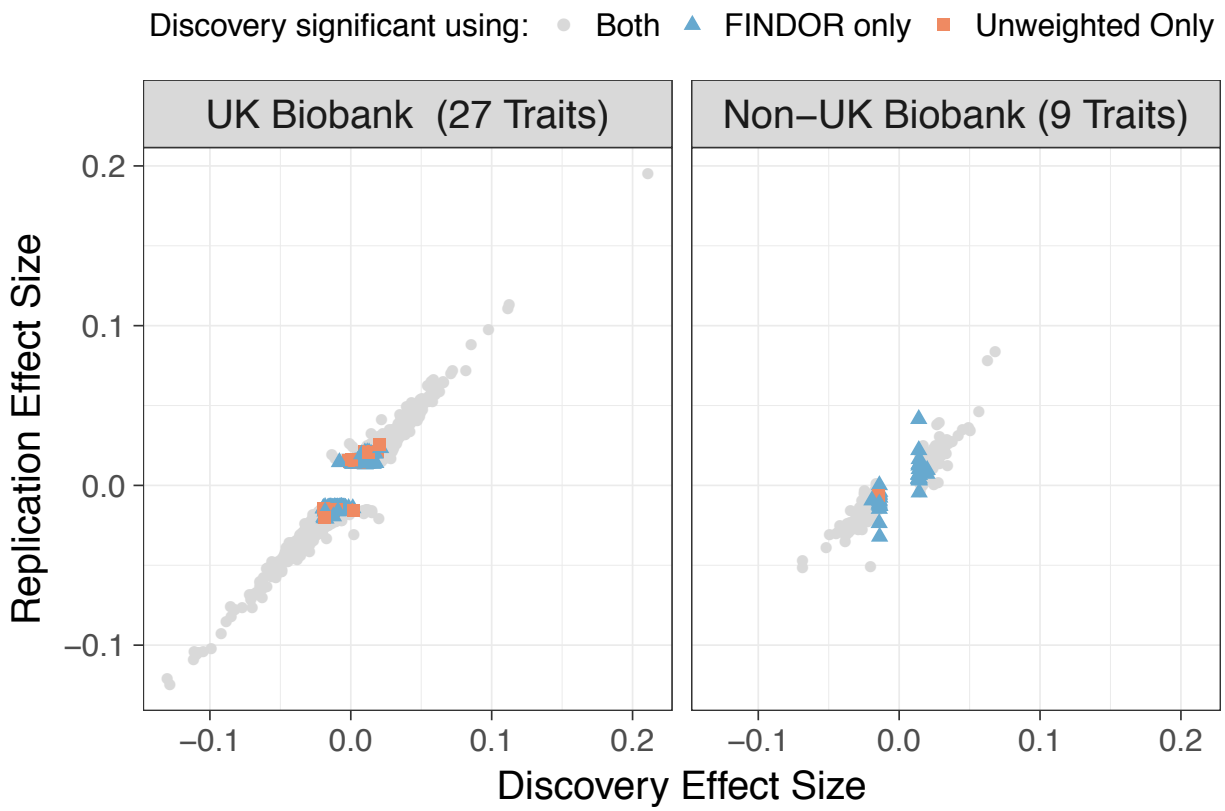

Figure 5.4: **Novel loci identified by FINDOR replicate in independent samples.** We plot the standardized effect sizes ($\frac{Z}{\sqrt{N}}$) in the UK Biobank replication sample (average N = 283K, left panel) and non-UK Biobank replications sample (average N = 158K, right panel) vs. the UK Biobank discovery sample (average N = 132K). For novel loci identified by FINDOR (blue triangles), the replication slope was positive and highly significant in both cases (UK Biobank = 0.66, Non-UK Biobank = 0.69). Numerical results are reported in Tables 5.10 and 5.12

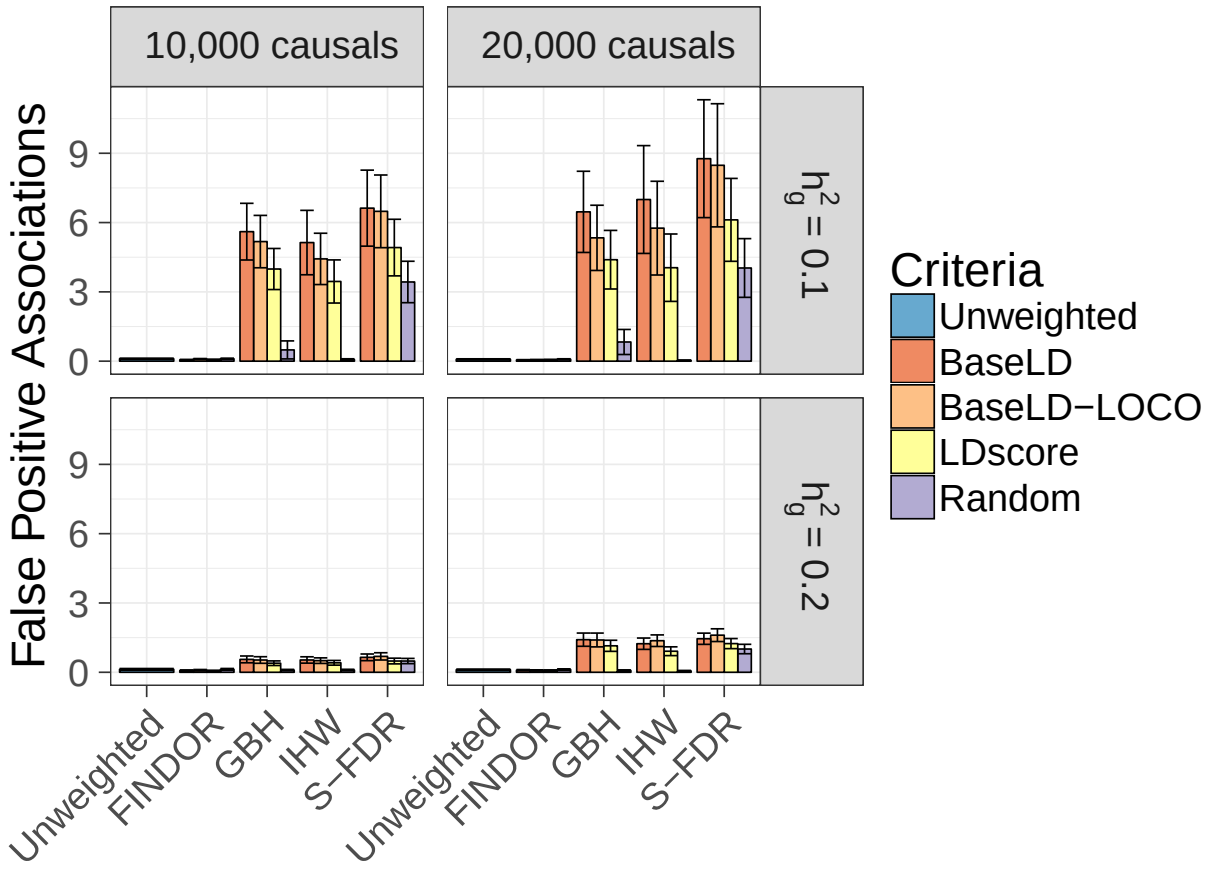

Figure 5.5: **Null mis-calibration for GBH, IHW and S-FDR is worse at lower effective sample size (50K)**. We report the average number of independent, genome-wide significant ($p < 5 \times 10^{-8}$) associations on null chromosomes in simulations with 50K individuals (vs. 100K in Figure 1). Results are averaged across 500 simulations. Error bars represent 95% confidence intervals.

Tables

Annotation	Proportion of SNPs	h_g^2 Enrichment	Tau
All SNPs	1	NaN	-1.011
Coding	0.014	4.634	1.248
Coding + 500bp	0.064	1.523	-1.303
Conserved (GERP NS)	-	1.801	0.3244
Conserved (GERP RS >= 4)	0 8	12.99	7.121
Conserved (Lindblad-Toh)	0.026	9.353	6.166
Conserved (Lindblad-Toh) + 500bp	0.33	1.669	-0.3263
CTCF	0.024	0.3465	-1.074
CTCF + 500bp	0.071	0.7298	-0.3127
DGF	0.136	2.062	0.1708
DGF + 500bp	0.538	1.367	0.05873
DHS Peaks	0.111	2.272	0.08154
DHS	0.166	2.017	-0.5108
DHS + 500bp	0.496	1.3	-0.1541
FANTOM5 Enhancer	0 4	1.296	-1.391
FANTOM5 Enhancer + 500bp	0.019	1.723	-0.3048
Enhancer	0.042	2.724	0.6413
Enhancer + 500bp	0.09	2.113	-0.1927
Fetal DHS	0.084	2.493	0.1489
Fetal DHS + 500bp	0.283	1.581	-0.1934
H3K27ac (Hnisz)	0.389	1.526	-0.4566
H3K27ac (Hnisz) + 500bp	0.42	1.534	0.5073
H3K27ac (PGC2)	0.269	1.716	-0.546
H3K27ac (PGC2) + 500bp	0.335	1.708	0.5095
H3K4me1 Peaks	0.17	2.254	0.5072
H3K4me1	0.424	1.71	0.5112
H3K4me1 + 500bp	0.606	1.338	-0.2583
H3K4me3 Peaks	0.042	2.936	0.3316
H3K4me3	0.133	2.378	0.3035
H3K4me3 + 500bp	0.255	1.756	-0.08552
H3K9ac Peaks	0.038	3.261	0.5507
H3K9ac	0.125	2.592	0.6866
H3K9ac + 500bp	0.23	1.915	-0.232
Intron	0.387	1.11	2.252
Intron + 500bp	0.397	1.177	-2.379

Annotation	Proportion of SNPs	h_g^2 Enrichment	Tau
Promoter Flanking	0 8	1.797	-0.06156
Promoter Flanking + 500bp	0.033	1.373	-0.8093
Promoter	0.031	1.961	1.882
Promoter + 500bp	0.038	1.5	-2.017
Repressed	0.461	0.7185	0.05893
Repressed + 500bp	0.719	0.7835	0.1844
Super Enhancer (Vahedi)	0.021	2.076	2.19
Super Enhancer (Vahedi) + 500bp	0.021	2.017	-2.265
Super Enhancer (Hnisz)	0.167	1.814	-1.809
Super Enhancer (Hnisz) + 500bp	0.17	1.878	1.835
Typical Enhancer	0.022	1.698	0.995
Typical Enhancer + 500bp	0.026	1.653	-0.9213
TFBS	0.131	2.439	0.9492
TFBS + 500bp	0.341	1.499	-0.1196
Transcribed	0.346	1.173	0.309
Transcribed + 500bp	0.762	0.965	-0.09606
TSS	0.018	3.469	0.7095
TSS + 500bp	0.034	2.916	0.1618
3 UTR	0.011	2.905	0.3657
3 UTR + 500bp	0.026	1.991	-0.04583
5 UTR	0 5	3.271	0.6998
5 UTR + 500bp	0.027	1.581	-0.3904
Weak Enhancer	0.021	1.69	-0.4957
Weak Enhancer + 500bp	0.089	1.41	-0.4051
MAF bin 1	0.102	0.6522	0.5037
MAF bin 2	0.1	0.7087	0.5734
MAF bin 3	0.1	0.8438	0.7306
MAF bin 4	0.101	0.7483	0.6469
MAF bin 5	0.098	0.9875	0.9017
MAF bin 6	0.1	1.088	0.985
MAF bin 7	0.1	1.093	1.028
MAF bin 8	0.1	1.168	1.102
MAF bin 9	0.101	1.18	1.109
MAF bin 10	0.098	1.291	1.222
MAF-Adjusted Allele Age	-	NA	-0.2584
LLD-AFR	-	NA	-0.2012
Recombination Rate (10kb)	-	0.891	-0.08077
Nucleotide Diversity (10kb)	-	0.8324	-0.05295
Background Selection Statistic	-	1.238	0.6152
CpG-Content (50kb)	-	1.162	41.6

Table 5.2: **Generative τ values used to simulate BaseLD enrichment (continued).** Values are derived from a meta-analysis of 31 traits (see ref. [3]).

h_g^2	# Causals	Criteria	FINDOR	GBH	IHW	S-FDR	Unweighted
0.1	10,000	BaseLD	0.066 (0.0086)	0.4 (0.035)	0.38 (0.031)	0.47 (0.035)	NA
		BaseLD-LOCO	0.073 (0.009)	0.37 (0.033)	0.36 (0.033)	0.48 (0.035)	NA
		LDscore	0.07 (0.0087)	0.29 (0.03)	0.29 (0.024)	0.4 (0.029)	NA
		Random	0.11 (0.011)	0.072 (0.0088)	0.071 (0.0089)	0.37 (0.029)	NA
		Unweighted	NA	NA	NA	NA	0.091 (0.0099)
0.1	20,000	BaseLD	0.061 (0.013)	1.6 (0.24)	1.3 (0.18)	1.7 (0.21)	NA
		BaseLD-LOCO	0.073 (0.014)	1.6 (0.25)	1.5 (0.22)	1.8 (0.23)	NA
		LDscore	0.066 (0.012)	1.2 (0.19)	1.1 (0.16)	1.4 (0.21)	NA
		Random	0.1 (0.016)	0.086 (0.015)	0.076 (0.014)	1.2 (0.16)	NA
		Unweighted	NA	NA	NA	NA	0.1 (0.016)
0.2	10,000	BaseLD	0.06 (0.008)	0.067 (0.0084)	0.051 (0.0075)	0.066 (0.0082)	NA
		BaseLD-LOCO	0.058 (0.0079)	0.062 (0.0081)	0.058 (0.0077)	0.068 (0.0085)	NA
		LDscore	0.064 (0.0081)	0.055 (0.0078)	0.048 (0.0071)	0.059 (0.0078)	NA
		Random	0.093 (0.0099)	0.081 (0.0092)	0.081 (0.0092)	0.096 (0.01)	NA
		Unweighted	NA	NA	NA	NA	0.09 (0.0096)
0.2	20,000	BaseLD	0.059 (0.0076)	0.081 (0.0093)	0.064 (0.0083)	0.09 (0.01)	NA
		BaseLD-LOCO	0.063 (0.008)	0.081 (0.0091)	0.075 (0.0093)	0.093 (0.01)	NA
		LDscore	0.055 (0.0073)	0.061 (0.008)	0.042 (0.0065)	0.069 (0.0084)	NA
		Random	0.097 (0.0098)	0.087 (0.0092)	0.089 (0.0094)	0.12 (0.011)	NA
		Unweighted	NA	NA	NA	NA	0.098 (0.0098)

Table 5.3: **Numerical results for simulations of null loci (Figure 1).** We report the average number of independent, genome-wide significant ($p < 5 \times 10^{-8}$) associations on null chromosomes. Results are averaged across 1000 simulations. Standard errors are reported in parentheses.

h_g^2	# Causals	log10 threshold	BaseLD	BaseLD (LOCO)	LDscore	Random	Unweighted
0.1	10000	-8	1.07 (0.79)	0.96 (0.74)	0.89 (0.68)	0.52 (0.32)	0.51 (0.32)
		-7	3.9 (1.14)	3.99 (1.05)	3.66 (1.03)	3.45 (0.8)	3.49 (0.82)
		-6	27.77 (1.77)	28.27 (1.72)	25.91 (1.61)	26.16 (1.34)	26.27 (1.35)
		-5	257.06 (4.68)	260.59 (4.47)	246.15 (4.18)	243.82 (2.94)	244.22 (2.98)
		-4	2497.25 (16.01)	2534.26 (15.07)	2471.05 (15.68)	2425.97 (9.73)	2421.79 (9.93)
0.1	20000	-8	0.23 (0.08)	0.26 (0.08)	0.29 (0.09)	0.29 (0.08)	0.26 (0.07)
		-7	2.22 (0.35)	2.12 (0.31)	2.02 (0.31)	2.48 (0.28)	2.45 (0.28)
		-6	25.41 (1.73)	25.34 (1.67)	23.31 (1.43)	23.86 (1.07)	24.01 (1.1)
		-5	253.37 (6.05)	258.33 (6)	239.84 (5.62)	241.9 (3.75)	241.82 (3.8)
		-4	2505.8 (25.29)	2552.36 (24.37)	2450.59 (23.69)	2429.73 (15.65)	2424.49 (15.96)
0.2	10000	-8	0.57 (0.25)	0.59 (0.26)	0.51 (0.26)	0.33 (0.09)	0.31 (0.08)
		-7	2.89 (0.4)	2.81 (0.4)	2.57 (0.39)	2.67 (0.35)	2.65 (0.35)
		-6	27.3 (1.85)	27.16 (1.64)	26.69 (1.92)	24.99 (0.81)	25.05 (0.82)
		-5	253.01 (4.7)	258.05 (4.63)	260.4 (5)	247.23 (2.89)	247.72 (2.93)
		-4	2433.43 (14.02)	2485.62 (13.81)	2467.72 (14.75)	2444.87 (10.06)	2439.6 (10.15)
0.2	20000	-8	0.24 (0.06)	0.25 (0.06)	0.2 (0.06)	0.22 (0.05)	0.22 (0.05)
		-7	2.13 (0.23)	2.25 (0.24)	2.15 (0.2)	2.29 (0.17)	2.31 (0.18)
		-6	24.09 (1.01)	24.73 (0.99)	24.37 (1.03)	24.28 (0.69)	24.36 (0.71)
		-5	241.74 (3.9)	249.05 (3.76)	246.46 (4.15)	245.44 (2.67)	246.35 (2.71)
		-4	2414.67 (14.36)	2468.78 (13.93)	2468.45 (14.56)	2431.91 (9.74)	2426.62 (9.81)

Table 5.4: **FINDOR is well-calibrated at less stringent significance thresholds in simulations of null loci.** We report the average *total number of associated SNPs* on null chromosomes at various significance thresholds. (In contrast to our main simulations, we do not report the average number of independent associations, due to problems with clumping using PLINK at less significant thresholds.) Results are averaged across 1000 simulations. Standard errors are reported in parentheses.

??

h_g^2	# Causals	Criteria	FINDOR	Unweighted	% Improve
0.1	10,000	Baseline	8.92 (0.1)	NA	20.00
		BaseLD	9.15 (0.1)	NA	23.00
		LDscore	7.95 (0.095)	NA	6.90
		Random	7.59 (0.092)	NA	2.00
		Unweighted	NA	7.44 (0.092)	0
0.1	20,000	Baseline	5.28 (0.12)	NA	34.00
		BaseLD	5.45 (0.12)	NA	38.00
		LDscore	4.5 (0.11)	NA	14.00
		Random	4.02 (0.1)	NA	2.00
		Unweighted	NA	3.94 (0.1)	0
0.2	10,000	Baseline	55.4 (0.23)	NA	6.70
		BaseLD	56.4 (0.23)	NA	8.70
		LDscore	52.2 (0.23)	NA	0.58
		Random	52.3 (0.23)	NA	0.77
		Unweighted	NA	51.9 (0.23)	0
0.2	20,000	Baseline	31.5 (0.16)	NA	16.00
		BaseLD	32.3 (0.17)	NA	19.00
		LDscore	29.1 (0.16)	NA	7.00
		Random	27.4 (0.16)	NA	0.74
		Unweighted	NA	27.2 (0.15)	0

Table 5.5: **Numerical results for simulations of causal loci (Figure 2)**. We report the average number of independent, genome-wide significant ($p < 5 \times 10^{-8}$) associations on causal chromosomes. Results are averaged across 1000 simulations. Standard errors are reported in parentheses.

Trait	Baseline	BaseLD	LDscore	Random	Unweighted
Eosinophil Count	198	200	189	188	187
Mean Corpular Hemoglobin	247	248	233	237	237
Red Blood Cell Distribution Width	205	212	201	199	198
Red Blood Cell Count	201	206	191	192	192
White Blood Cell Count	158	165	148	148	148
Heel T Score	308	308	295	302	300
Balding Type I	96	100	92	96	96
Body Mass Index	126	132	119	117	117
Height	685	690	668	675	674
Waist-hip Ratio	102	104	100	99	98
Systolic Blood Pressure	105	106	98	96	98
Years of Education	19	24	18	17	17
Smoking Status	22	24	21	19	18
Auto Immune Traits	15	18	14	14	14
Eczema	43	46	39	34	35
Cardiovascular Diseases	47	49	39	41	38
Hypothyroidism	27	30	27	27	27
Respiratory and Disease	26	29	24	25	24
Type 2 Diabetes	16	14	13	15	14
FEV1-FVC Ratio	178	185	172	174	174
Forced Vital Capacity (FVC)	99	99	94	91	90
Neuroticism	15	16	10	11	11
Morning Person	12	14	13	14	14
Hair Color	142	143	139	139	140
Sunburn Occasion	24	25	23	22	23
Age at Menarche	56	56	54	52	52
Age at Menopause	17	18	17	17	18
Total	3189	3261	3051	3061	3054
Difference	135	207	-3	7	0
Jacknife SE	17.22	20.41	14.56	5.88	0

Table 5.6: **Results for FINDOR with different stratification criteria in the 145K UK Biobank release.** For each trait, we report the number of independent, genome-wide significant loci identified by the Unweighted approach and by FINDOR with various stratification criteria in the 145K UK Biobank release.

Table 5.7: **List of independent, genome-wide significant loci for all 27 traits in 145K and 460K UK Biobank releases.** We report independent, genome-wide significant loci for both Unweighted and FINDOR. See Excel file.

	145K				459K			
Class	# Loci Unweighted	# Loci FINDOR	Overall % Increase	Average % Increase	# Loci Unweighted	# Loci FINDOR	Overall %. Increase.	Average % Increase
Anthropometric	1285	1334	4%	6%	5227	5353	2%	4%
Blood Cell	962	1031	7%	8%	3669	3831	4%	4%
Disease	152	186	22%	20%	860	946	10%	10%
Other	655	710	8%	15%	3527	3736	6%	7%
Overall	3054	3261	7%	13%	13283	13866	4%	7%

Table 5.8: **Results for each phenotype class in 145K and 459K UK Biobank releases.** For each phenotype class, we report the number of independent, genome-wide significant loci identified by the Unweighted approach and by FINDOR in the 145K and 459K UK Biobank releases.

Trait	Baseline	BaseLD	LDscore	Random	Unweighted
Eosinophil Count	166	164	157	159	159
Mean Corpular Hemoglobin	209	210	200	203	203
Red Blood Cell Distribution Width	165	170	159	160	160
Red Blood Cell Count	164	167	153	153	154
White Blood Cell Count	119	120	112	114	112
Heel T Score	244	250	238	240	239
Balding Type I	79	83	75	75	76
Body Mass Index	79	82	76	78	78
Height	563	568	539	538	538
Waist-hip Ratio	79	75	74	71	70
Systolic Blood Pressure	75	75	73	73	71
Years of Education	10	11	10	10	10
Smoking Status	11	13	8	7	7
Auto Immune Traits	11	11	10	9	10
Eczema	27	28	27	24	24
Cardiovascular Diseases	30	30	29	28	29
Hypothyroidism	22	23	23	23	24
Respiratory and Disease	21	22	18	20	19
Type 2 Diabetes	9	9	7	10	10
FEV1-FVC Ratio	135	138	131	134	134
Forced Vital Capacity (FVC)	62	63	58	57	57
Neuroticism	5	8	5	5	5
Morning Person	6	5	5	5	5
Hair Color	117	122	119	121	121
Sunburn Occasion	20	20	18	18	18
Age at Menarche	39	39	37	39	38
Age at Menopause	14	14	13	14	14
Total	2545	2582	2438	2452	2450
Difference	95	132	-12	2	0
Jacknife SE	15.83	16.72	10.14	5.28	0

Table 5.9: **Results for FINDOR with different stratification criteria with p-value threshold of 5×10^{-9} in the 145K UK Biobank release.** For each trait, we report the number of independent, $p < 5 \times 10^{-9}$ loci identified by the Unweighted approach and by FINDOR with various stratification criteria in the 145K UK Biobank release.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.0000	0.0001	-0.46	0.6468
Both Methods	0.9114	0.0039	234.92	$< 1 \times 10^{-20}$

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.0005	0.0003	-1.83	0.0685
FINDOR only	0.6613	0.0179	37.04	$< 1 \times 10^{-20}$

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.0015	0.0007	-2.30	0.0262
Unweighted Only	0.5724	0.0427	13.42	$< 1 \times 10^{-20}$

Table 5.10: **Numerical results for UK Biobank replication analysis (Figure 4, left panel).** For loci detected using Both Methods, FINDOR only, or Unweighted only, respectively, we report results of a regression of standardized effect sizes ($\frac{Z}{\sqrt{N}}$) at lead SNPs in UK Biobank replication data vs. UK Biobank discovery data.

Phenotype	Replication Study	Discovery N	Replication N (max)
BMI	Locke et al. (Nature 2015) [77]	145,209	322,091
Height	Wood et al. (Nature 2014) [76]	145,368	253,237
WHR adjusted BMI	Shugin et al (Nature 2015) [78]	145,375	210,039
Edu Years	Rietveld et al. (Science 2013) [138]	144,204	126,559
Eczema	Paternoster et al. (Nature Genetics 2015) [139]	145,416	40,835
Type II Diabetes	Morris et al. (Nature Genetics 2012) [140]	145,298	29,842
Ever Smoked	Furberg et al. (Nature Genetics 2010) [141]	145,227	74,035
Age at Menarche	Perry et al. (Nature 2014) [142]	74,944	182,416
Age at Menopause	Day et al (Nature Genetics 2015) [143]	44,410	69,360

Table 5.11: **List of nine traits used for non-UK Biobank replication analysis.** We report the non-UK Biobank replication reference, UK Biobank discovery sample size and non-UK Biobank replication sample size for each trait.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.0001	0.0003	-0.39	0.6970
Both Methods	0.6717	0.0126	53.39	$< 1 \times 10^{-20}$

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.0006	0.0015	-0.38	0.7054
FINDOR only	0.6884	0.1021	6.74	9.52×10^{-8}

Table 5.12: **Numerical results for non-UK Biobank replication analysis (Figure 4, right panel)**. For loci detected using Both Methods or FINDOR only, respectively, we report results of a regression of standardized effect sizes ($\frac{Z}{\sqrt{N}}$) at lead SNPs in non-UK Biobank replication data vs. UK Biobank discovery data. We do not report results for Unweighted only, which contained only a single locus.

Trait	Baseline	BaseLD	LDscore	Random	Unweighted
Eosinophil Count	710	731	686	700	699
Mean Corpular Hemoglobin	791	791	758	766	765
Red Blood Cell Distribution Width	677	674	641	651	652
Red Blood Cell Count	878	885	834	839	840
White Blood Cell Count	744	750	710	713	713
Heel T Score	1148	1149	1113	1127	1130
Balding Type I	346	346	335	335	334
Body Mass Index	930	950	913	907	908
Height	2397	2402	2354	2395	2395
Waist-hip Ratio	496	506	472	458	460
Systolic Blood Pressure	694	703	661	664	666
Years of Education	302	315	293	287	286
Smoking Status	169	178	164	154	154
Auto Immune Traits	84	86	72	75	75
Eczema	191	198	179	182	181
Cardiovascular Diseases	304	314	286	285	285
Hypothyroidism	151	153	141	140	139
Respiratory and Disease	108	109	98	105	104
Type 2 Diabetes	87	86	80	78	76
FEV1-FVC Ratio	703	714	684	684	684
Forced Vital Capacity (FVC)	559	565	541	543	544
Neuroticism	143	149	136	128	128
Morning Person	161	165	159	156	156
Hair Color	433	436	427	429	428
Sunburn Occasion	82	82	74	79	78
Age at Menarche	326	338	318	318	318
Age at Menopause	89	91	85	86	85
Total	13703	13866	13214	13284	13283
Difference	420	583	-69	1	0
Jacknife SE	39.95	40.64	33.81	10.02	0

Table 5.13: **Results for FINDOR with different stratification criteria in the 459K UK Biobank release.** For each trait, we report the number of independent, genome-wide significant loci identified by the Unweighted approach and by FINDOR with various stratification criteria in the 459K UK Biobank release.

		% of lead SNPs within 95% credible set		
Data	Method	GTEx_GE	BLUEPRINT_GE	BLUEPRINT_H3K27ac
145K	FINDOR-only	17.6% (2.4%)	10.2% (1.9%)	21.2% (2.6%)
	Unweighted-only	0.0% (0.0%)	0.0% (0.0%)	2.0% (1.9%)
459K	FINDOR-only	13.6% (1.2%)	7.1% (0.9%)	23.4% (1.5%)
	Unweighted-only	5.4% (1.7%)	3.0% (1.3%)	10.7% (2.4%)

		Average Posterior Probability		
Data	Method	GTEx_GE	BLUEPRINT_GE	BLUEPRINT_H3K27ac
145K	FINDOR-only	0.052 (0.012)	0.029 (0.009)	0.039 (0.009)
	Unweighted-only	0 (0)	0 (0)	0.015 (0.015)
459K	FINDOR-only	0.037 (0.006)	0.025 (0.005)	0.055 (0.006)
	Unweighted-only	0.025 (0.011)	0.008 (0.006)	0.032 (0.011)

Table 5.14: **Novel loci identified by FINDOR are more likely to be molecular QTL**. Top panel: for lead SNPs at loci detected using FINDOR only or Unweighted only, in both 145K and 459K UK Biobank releases, we report the % of lead SNPs that lie inside 95% causal sets for three molecular QTL, as described in ref. [4]. Bottom panel: for lead SNPs at loci detected using FINDOR only or Unweighted only, in both 145K and 459K UK Biobank releases, we report the average causal posterior probabilities for three molecular QTL, as described in ref. [4].

REFERENCES

[1] Julian B Maller, Gilean McVean, Jake Byrnes, Damjan Vukcevic, Kimmo Palin, Zhan Su, Joanna MM Howson, Adam Auton, Simon Myers, Andrew Morris, et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nature genetics*, 44(12):1294–1301, 2012.

[2] Alexander Gusev, S Hong Lee, Gosia Trynka, Hilary Finucane, Bjarni J Vilhjálmsson, Han Xu, Chongzhi Zang, Stephan Ripke, Brendan Bulik-Sullivan, Eli Stahl, et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *The American Journal of Human Genetics*, 95(5):535–552, 2014.

[3] Steven Gazal, Hilary K Finucane, Nicholas A Furlotte, Po-Ru Loh, Pier Francesco Palamara, Xuanyao Liu, Armin Schoech, Brendan Bulik-Sullivan, Benjamin M Neale, Alexander Gusev, et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nature Genetics*, 49(10):1421–1427, 2017.

[4] Farhad Hormozdiari, Steven Gazal, Bryce van de Geijn, Hilary Finucane, Chelsea J-T Ju, Po-Ru Loh, Armin Schoech, Yakir Reshef, Xuanyao Liu, Luke O'Connor, et al. Leveraging molecular qtl to understand the genetic architecture of diseases and complex traits. *bioRxiv*, page 203380, 2017.

[5] Tinca JC Polderman, Beben Benyamin, Christiaan A De Leeuw, Patrick F Sullivan, Arjen Van Bochoven, Peter M Visscher, and Danielle Posthuma. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature genetics*, 47(7):702, 2015.

[6] Eric S Lander and Nicholas J Schork. Genetic dissection of complex traits. *Science*, 265(5181):2037–2048, 1994.

[7] Mark I McCarthy, Gonçalo R Abecasis, Lon R Cardon, David B Goldstein, Julian Little, John PA Ioannidis, and Joel N Hirschhorn. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature reviews genetics*, 9(5):356, 2008.

[8] Peter M Visscher, Matthew A Brown, Mark I McCarthy, and Jian Yang. Five years of gwas discovery. *The American Journal of Human Genetics*, 90(1):7–24, 2012.

[9] Alkes L Price, Chris CA Spencer, and Peter Donnelly. Progress and promise in understanding the genetic basis of common diseases. In *Proc. R. Soc. B*, volume 282, page 20151684. The Royal Society, 2015.

[10] Lucia A Hindorff, Praveen Sethupathy, Heather A Junkins, Erin M Ramos, Jayashri P Mehta, Francis S Collins, and Teri A Manolio. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*, 106(23):9362–9367, 2009.

[11] Peter M Visscher, Naomi R Wray, Qian Zhang, Pamela Sklar, Mark I McCarthy, Matthew A Brown, and Jian Yang. 10 years of gwas discovery: biology, function, and translation. *The American Journal of Human Genetics*, 101(1):5–22, 2017.

[12] Frank W Albert and Leonid Kruglyak. The role of regulatory variation in complex traits and disease. *Nature Reviews Genetics*, 16(4):197, 2015.

[13] ENCODE Project Consortium et al. An integrated encyclopedia of dna elements in the human genome. *Nature*, 489(7414):57–74, 2012.

[14] Anshul Kundaje, Wouter Meuleman, Jason Ernst, Misha Bilenky, Angela Yen, Alireza Heravi-Moussavi, Pouya Kheradpour, Zhizhuo Zhang, Jianrong Wang, Michael J Ziller, et al. Integrative analysis of 111 reference human epigenomes. *Nature*, 518(7539):317–330, 2015.

[15] Kristin G Ardlie, Leonid Kruglyak, and Mark Seielstad. Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics*, 3(4):299, 2002.

[16] Sarah L Spain and Jeffrey C Barrett. Strategies for fine-mapping complex traits. *Human molecular genetics*, 24(R1):R111–R119, 2015.

[17] Bertrand Servin and Matthew Stephens. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS genetics*, 3(7):e114, 2007.

[18] Matthew Stephens and David J Balding. Bayesian statistical methods for genetic association studies. *Nature Reviews Genetics*, 10(10):681–690, 2009.

[19] Farhad Hormozdiari, Emrah Kostem, Eun Yong Kang, Bogdan Pasaniuc, and Eleazar Eskin. Identifying causal variants at loci with multiple signals of association. *Genetics*, pages genetics–114, 2014.

[20] Gleb Kichaev, Wen-Yun Yang, Sara Lindstrom, Farhad Hormozdiari, Eleazar Eskin, Alkes L Price, Peter Kraft, and Bogdan Pasaniuc. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS genetics*, 10(10):e1004722, 2014.

[21] Gleb Kichaev and Bogdan Pasaniuc. Leveraging functional-annotation data in trans-ethnic fine-mapping studies. *The American Journal of Human Genetics*, 97(2):260–271, 2015.

[22] Gleb Kichaev, Megan Roytman, Ruth Johnson, Eleazar Eskin, Sara Lindstroem, Peter Kraft, and Bogdan Pasaniuc. Improved methods for multi-trait fine mapping of pleiotropic risk loci. *Bioinformatics*, 33(2):248–255, 2017.

[23] Orli Bahcall. Fine mapping with function. *Nature genetics*, 46(12):1257, 2014.

[24] Lucia A Hindorff, Vence L Bonham, Lawrence C Brody, Margaret EC Ginoza, Carolyn M Hutter, Teri A Manolio, and Eric D Green. Prioritizing diversity in human genomics research. *Nature Reviews Genetics*, 2017.

[25] Urko M Marigorta and Arcadi Navarro. High trans-ethnic replicability of gwas results implies common causal variants. *PLoS genetics*, 9(6):e1003566, 2013.

[26] Jimmy Z Liu, Suzanne van Sommeren, Hailiang Huang, Siew C Ng, Rudi Alberts, Atsushi Takahashi, Stephan Ripke, James C Lee, Luke Jostins, Tejas Shah, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature genetics*, 47(9):979, 2015.

[27] Farhad Hormozdiari, Emrah Kostem, Eun Yong Kang, Bogdan Pasaniuc, and Eleazar Eskin. Identifying causal variants at loci with multiple signals of association. *Genetics*, 198(2):497–508, 2014.

[28] Wenan Chen, Beth R Larrabee, Inna G Ovsyannikova, Richard B Kennedy, Iana H Haralambieva, Gregory A Poland, and Daniel J Schaid. Fine mapping causal variants with an approximate bayesian method using marginal test statistics. *Genetics*, pages genetics–115, 2015.

[29] Christian Benner, Chris CA Spencer, Samuli Ripatti, and Matti Pirinen. Finemap: Efficient variable selection using summary data from genome-wide association studies. *bioRxiv*, page 027342, 2015.

[30] Eleazar Eskin. Increasing power in association studies by using linkage disequilibrium structure and molecular function as prior information. *Genome research*, 18(4):653–660, 2008.

[31] Buhm Han, Hyun Min Kang, and Eleazar Eskin. Rapid and accurate multiple testing correction and power estimation for millions of correlated markers. *PLoS Genet*, 5(4):e1000456, 2009.

[32] Andrew J Schork, Wesley K Thompson, Phillip Pham, Ali Torkamani, J Cooper Roddey, Patrick F Sullivan, John R Kelsoe, Michael C O'Donovan, Helena Furberg, Nicholas J Schork, et al. All snps are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated snps. *PLoS Genet*, 9(4):e1003449, 2013.

[33] Joseph K Pickrell. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *The American Journal of Human Genetics*, 94(4):559–573, 2014.

[34] Evan A Boyle, Yang I Li, and Jonathan K Pritchard. An expanded view of complex traits: from polygenic to omnigenic. *Cell*, 169(7):1177–1186, 2017.

[35] Kerstin B Meyer, Martin OReilly, Kyriaki Michailidou, Saskia Carlebur, Stacey L Edwards, Juliet D French, Radhika Prathalingham, Joe Dennis, Manjeet K Bolla, Qin Wang, et al. Fine-scale mapping of the fgfr2 breast cancer risk locus: putative functional variants differentially bind foxa1 and e2f1. *The American Journal of Human Genetics*, 93(6):1046–1060, 2013.

[36] Zsofia Kote-Jarai, Edward J Saunders, Daniel A Leongamornlert, Malgorzata Tymrakiewicz, Tokhir Dadaev, Sarah Jugurnauth-Little, Helen Ross-Adams, Ali Amin Al Olama, Sara Benlloch, Silvia Halim, et al. Fine-mapping identifies multiple prostate cancer risk loci at 5p15, one of which associates with tert expression. *Human molecular genetics*, 22(12):2520–2528, 2013.

[37] Ying Wu, Lindsay L Waite, Anne U Jackson, Wayne HH Sheu, Steven Buyske, Devin Absher, Donna K Arnett, Eric Boerwinkle, Lori L Bonnycastle, Cara L Carty, et al. Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. *PLoS genetics*, 9(3):e1003379, 2013.

[38] Laura L Faye, Mitchell J Machiela, Peter Kraft, Shelley B Bull, and Lei Sun. Re-ranking sequencing variants in the post-gwas era for accurate causal variant identification. *PLoS genetics*, 9(8):e1003609, 2013.

[39] Asian Genetic Epidemiology Network Type, South Asian Type, Diabetes SAT2D Consortium, Mexican American Type, Diabetes MAT2D Consortium, Anubha Mahajan, Min Jin Go, Weihua Zhang, Jennifer E Below, Kyle J Gaulton, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nature genetics*, 46(3):234–244, 2014.

[40] ' International Multiple Sclerosis Genetics Consortium et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nature genetics*, 45(11):1353–1360, 2013.

[41] Joseph K. Pickrell. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *The American Journal of Human Genetics*, 94(4):559–573, 2014.

[42] Daniel J Gaffney, Jean-Baptiste Veyrieras, Jacob F Degner, Roger Pique-Regi, Athma A Pai, Gregory E Crawford, Matthew Stephens, Yoav Gilad, Jonathan K Pritchard, et al. Dissecting the regulatory architecture of gene expression qtls. *Genome Biol*, 13(1):R7, 2012.

[43] Verena Zuber, A Pedro Duarte Silva, and Korbinian Strimmer. A novel algorithm for simultaneous snp selection in high-dimensional genome-wide association studies. *BMC bioinformatics*, 13(1):284, 2012.

[44] William Valdar, Jeremy Sabourin, Andrew Nobel, and Christopher C Holmes. Reprioritizing genetic associations in hit regions using lasso-based resample model averaging. *Genetic epidemiology*, 36(5):451–462, 2012.

[45] Yongtao Guan, Matthew Stephens, et al. Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *The Annals of Applied Statistics*, 5(3):1780–1815, 2011.

[46] Su-In Lee, Aimée M Dudley, David Drubin, Pamela A Silver, Nevan J Krogan, Dana Pe'er, and Daphne Koller. Learning a prior on regulatory potential from eqtl data. *PLoS genetics*, 5(1):e1000358, 2009.

[47] Peter Carbonetto and Matthew Stephens. Integrated enrichment analysis of variants and pathways in genome-wide association studies indicates central role for il-2 signaling genes in type 1 diabetes, and cytokine signaling genes in crohn's disease. *PLoS genetics*, 9(10):e1003770, 2013.

[48] Matthew T Maurano, Richard Humbert, Eric Rynes, Robert E Thurman, Eric Haugen, Hao Wang, Alex P Reynolds, Richard Sandstrom, Hongzhu Qu, Jennifer Brody, et al. Systematic localization of common disease-associated variation in regulatory dna. *Science*, 337(6099):1190–1195, 2012.

[49] Gosia Trynka and Soumya Raychaudhuri. Using chromatin marks to interpret and localize genetic associations to complex human traits and diseases. *Current opinion in genetics & development*, 23(6):635–641, 2013.

[50] Konrad J Karczewski, Joel T Dudley, Kimberly R Kukurba, Rong Chen, Atul J Butte, Stephen B Montgomery, and Michael Snyder. Systematic functional regulatory assessment of disease-associated variants. *Proceedings of the National Academy of Sciences*, 110(23):9607–9612, 2013.

[51] Gosia Trynka, Cynthia Sandor, Buhm Han, Han Xu, Barbara E Stranger, X Shirley Liu, and Soumya Raychaudhuri. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nature genetics*, 45(2):124–130, 2013.

[52] Miriam S Udler, Kerstin B Meyer, Karen A Pooley, Eric Karlins, Jeffery P Struewing, Jinghui Zhang, David R Doody, Stewart MacArthur, Jonathan Tyrer, Paul D Pharoah, et al. Fgfr2 variants and breast cancer risk: fine-scale mapping using african american studies and analysis of chromatin conformation. *Human molecular genetics*, 18(9):1692–1703, 2009.

[53] Gosia Trynka, Karen A Hunt, Nicholas A Bockett, Jihane Romanos, Vanisha Mistry, Agata Szperl, Sjoerd F Bakker, Maria Teresa Bardella, Leena Bhaw-Rosun, Gemma Castillejo, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nature genetics*, 43(12):1193–1201, 2011.

[54] Nikolaos A Patsopoulos, Lisa F Barcellos, Rogier Q Hintzen, Catherine Schaefer, Cornelia M van Duijn, Janelle A Noble, Towfique Raj, Pierre-Antoine Gourraud, Barbara E Stranger, Jorge Oksenberg, et al. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: Hla and non-hla effects. *PLoS genetics*, 9(11):e1003926, 2013.

[55] Jimmy Z Liu, Mohamed A Almarri, Daniel J Gaffney, George F Mells, Luke Jostins, Heather J Cordell, Samantha J Ducker, Darren B Day, Michael A Heneghan,

James M Neuberger, et al. Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. *Nature genetics*, 44(10):1137–1141, 2012.

[56] Jacques Fellay, Alexander J Thompson, Dongliang Ge, Curtis E Gumbs, Thomas J Urban, Kevin V Shianna, Latasha D Little, Ping Qiu, Arthur H Bertelsen, Mark Watson, et al. Itpa gene variants protect against anaemia in patients treated for chronic hepatitis c. *Nature*, 464(7287):405–408, 2010.

[57] Juan Pablo Lewinger, David V Conti, James W Baurley, Timothy J Triche, and Duncan C Thomas. Hierarchical bayes prioritization of marker associations from a genome-wide association scan for further investigation. *Genetic epidemiology*, 31(8):871–882, 2007.

[58] MA Quintana and DV Conti. Integrative variable selection via bayesian model uncertainty. *Statistics in medicine*, 32(28):4938–4953, 2013.

[59] Miriam S Udler, Jonathan Tyrer, and Douglas F Easton. Evaluating the power to discriminate between highly correlated snps in genetic association studies. *Genetic epidemiology*, 34(5):463–468, 2010.

[60] Bradley P Carlin and Thomas A Louis. *Bayes and empirical Bayes methods for data analysis*. CRC Press, 2000.

[61] Tanya M Teslovich, Kiran Musunuru, Albert V Smith, Andrew C Edmondson, Ioannis M Stylianou, Masahiro Koseki, James P Pirruccello, Samuli Ripatti, Daniel I Chasman, Cristen J Willer, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, 466(7307):707–713, 2010.

[62] Bogdan Pasaniuc, Noah Zaitlen, Huwenbo Shi, Gaurav Bhatia, Alexander Gusev, Joseph Pickrell, Joel Hirschhorn, David P Strachan, Nick Patterson, and Alkes L Price. Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics*, page btu416, 2014.

[63] Buhm Han, Hyun Min Kang, and Eleazar Eskin. Rapid and accurate multiple testing correction and power estimation for millions of correlated markers. *PLoS genetics*, 5(4):e1000456, April 2009.

[64] Karen N Conneely and Michael Boehnke. So many correlated tests, so little time! Rapid adjustment of P values for multiple correlated tests. *American journal of human genetics*, 81(6):1158–1168, December 2007.

[65] Noah Zaitlen, Bogdan Pasaniuc, Tom Gur, Elad Ziv, and Eran Halperin. Leveraging genetic variability across populations for the identification of causal variants. *American journal of human genetics*, 86(1):23–33, 01 2010.

[66] Dong C Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1-3):503–528, 1989.

[67] Steven G Johnson. The nlopt nonlinear-optimization package, 2010.

[68] Zhan Su, Jonathan Marchini, and Peter Donnelly. Hapgen2: simulation of multiple disease snps. *Bioinformatics*, 2011.

[69] Jian Yang, Teresa Ferreira, Andrew P Morris, Sarah E Medland, Pamela AF Madden, Andrew C Heath, Nicholas G Martin, Grant W Montgomery, Michael N Weedon, Ruth J Loos, et al. Conditional and joint multiple-snp analysis of gwas summary statistics identifies additional variants influencing complex traits. *Nature genetics*, 44(4):369–375, 2012.

[70] Robert E Thurman, Eric Rynes, Richard Humbert, Jeff Vierstra, Matthew T Maurano, Eric Haugen, Nathan C Sheffield, Andrew B Stergachis, Hao Wang, Benjamin Vernot, et al. The accessible chromatin landscape of the human genome. *Nature*, 489(7414):75–82, 2012.

[71] Tanya M. Teslovich et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, 466(7307):707–713, 08 2010.

[72] Sagiv Shifman, Jane Kuypers, Mark Kokoris, Benjamin Yakir, and Ariel Darvasi. Linkage disequilibrium patterns of the human genome across populations. *Human molecular genetics*, 12(7):771–776, 2003.

[73] Danielle Welter, Jacqueline MacArthur, Joannella Morales, Tony Burdett, Peggy Hall, Heather Junkins, Alan Klemm, Paul Flicek, Teri Manolio, Lucia Hindorff, et al. The nhgri gwas catalog, a curated resource of snp-trait associations. *Nucleic acids research*, 42(D1):D1001–D1006, 2014.

[74] ' Global Lipids Genetics Consortium et al. Discovery and refinement of loci associated with lipid levels. *Nature genetics*, 45(11):1274–1283, 2013.

[75] Yukinori Okada, Di Wu, Gosia Trynka, Towfique Raj, Chikashi Terao, Katsunori Ikari, Yuta Kochi, Koichiro Ohmura, Akari Suzuki, Shinji Yoshida, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*, 506(7488):376–381, 2014.

[76] Andrew R Wood, Tonu Esko, Jian Yang, Sailaja Vedantam, Tune H Pers, Stefan Gustafsson, Audrey Y Chu, Karol Estrada, Jian'an Luan, Zoltán Kutalik, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature genetics*, 46(11):1173–1186, 2014.

[77] Adam E Locke, Bratati Kahali, Sonja I Berndt, Anne E Justice, Tune H Pers, Felix R Day, Corey Powell, Sailaja Vedantam, Martin L Buchkovich, Jian Yang, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*, 518(7538):197–206, 2015.

[78] Dmitry Shungin, Thomas W Winkler, Damien C Croteau-Chonka, Teresa Ferreira, Adam E Locke, Reedik Mägi, Rona J Strawbridge, Tune H Pers, Krista Fischer, Anne E Justice, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature*, 518(7538):187–196, 2015.

[79] Manuel A Rivas, Mélissa Beaudoin, Agnes Gardet, Christine Stevens, Yashoda Sharma, Clarence K Zhang, Gabrielle Boucher, Stephan Ripke, David Ellinghaus, Noel Burtt, et al. Deep resequencing of gwas loci identifies independent rare variants associated with inflammatory bowel disease. *Nature genetics*, 43(11):1066–1073, 2011.

[80] Kyle Kai-How Farh, Alexander Marson, Jiang Zhu, Markus Kleinewietfeld, William J Housley, Samantha Beik, Noam Shoresh, Holly Whitton, Russell JH Ryan, Alexander A Shishkin, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, 2014.

[81] D Altshuler et al. Integrating common and rare genetic variation in diverse human populations. *Nature*, 467(7311):52–8, 2010.

[82] ' 1000 Genomes Project Consortium et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422):56–65, 2012.

[83] Noah Zaitlen, Bogdan Pasaniuc, Tom Gur, Elad Ziv, and Eran Halperin. Leveraging genetic variability across populations for the identification of causal variants. *The American Journal of Human Genetics*, 86(1):23–33, 2010.

[84] Rick Twee-Hee Ong, Xu Wang, Xuanyao Liu, and Yik-Ying Teo. Efficiency of transethnic genome-wide meta-analysis and fine-mapping. *European Journal of Human Genetics*, 20(12):1300–1307, 2012.

[85] Andrew P Morris. Transethnic meta-analysis of genomewide association studies. *Genetic epidemiology*, 35(8):809–822, 2011.

[86] Yik-Ying Teo, Rick TH Ong, Xueling Sim, E Tai, Kee-Seng Chia, et al. Identifying candidate causal variants via trans-population fine-mapping. *Genetic epidemiology*, 34(7):653–664, 2010.

[87] Simon N Stacey, Patrick Sulem, Carlo Zanon, Sigurjon A Gudjonsson, Gudmar Thorleifsson, Agnar Helgason, Aslaug Jonasdottir, Soren Besenbacher, Jelena P Kostic, James D Fackenthal, et al. Ancestry-shift refinement mapping of the c6orf97-esr1 breast cancer susceptibility locus. *PLoS genetics*, 6(7):e1001029, 2010.

[88] Diabetes SAT2D Consortium, Diabetes MAT2D Consortium, Anubha Mahajan, Min Jin Go, Weihua Zhang, Jennifer E Below, Kyle J Gaulton, Teresa Ferreira, Momoko Horikoshi, Andrew D Johnson, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nature genetics*, 46(3):234–244, 2014.

[89] Evangelos Evangelou and John PA Ioannidis. Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics*, 14(6):379–389, 2013.

[90] Xu Wang, Hui-Xiang Chua, Peng Chen, Rick Twee-Hee Ong, Xueling Sim, Weihua Zhang, Fumihiko Takeuchi, Xuanyao Liu, Chiea-Chuen Khor, Wan-Ting Tay, et al.

Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies. *Human molecular genetics*, page ddt064, 2013.

[91] Ching-Ti Liu, Martin L Buchkovich, Thomas W Winkler, Iris M Heid, Ingrid B Borecki, Caroline S Fox, Karen L Mohlke, Kari E North, L Adrienne Cupples, et al. Multi-ethnic fine-mapping of 14 central adiposity loci. *Human molecular genetics*, page ddu183, 2014.

[92] Suna Onengut-Gumuscu, Wei-Min Chen, Oliver Burren, Nick J Cooper, Aaron R Quinlan, Josyf C Mychaleckyj, Emily Farber, Jessica K Bonnie, Michal Szpak, Ellen Schofield, et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nature genetics*, 2015.

[93] Dennis J Hazelett, Suhn Kyong Rhie, Malaina Gaddis, Chunli Yan, Daniel L Lakeland, Simon G Coetzee, Brian E Henderson, Houtan Noushmehr, Wendy Cozen, Zsofia Kote-Jarai, et al. Comprehensive functional annotation of 77 prostate cancer risk loci. *PLoS genetics*, 10(1):e1004102, 2014.

[94] Joseph K. Pickrell. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *The American Journal of Human Genetics*, 94(4):559–573, 2014.

[95] Dongjun Chung, Can Yang, Cong Li, Joel Gelernter, and Hongyu Zhao. Gpa: a statistical approach to prioritizing gwas results by integrating pleiotropy and annotation. *PLoS genetics*, 2014.

[96] Andrej Nikolaevich Tikhonov and Vasiliy Yakovlevich Arsenin. Solutions of ill-posed problems. 1977.

[97] Marc A Coram, Qing Duan, Thomas J Hoffmann, Timothy Thornton, Joshua W Knowles, Nicholas A Johnson, Heather M Ochs-Balcom, Timothy A Donlon, Lisa W Martin, Charles B Eaton, et al. Genome-wide characterization of shared and distinct genetic components that influence blood lipid levels in ethnically diverse human populations. *The American Journal of Human Genetics*, 92(6):904–916, 2013.

[98] Jian Yang, Teri A Manolio, Louis R Pasquale, Eric Boerwinkle, Neil Caporaso, Julie M Cunningham, Mariza de Andrade, Bjarke Feenstra, Eleanor Feingold, M Geoffrey Hayes, et al. Genome partitioning of genetic variation for complex traits using common snps. *Nature genetics*, 43(6):519–525, 2011.

[99] Mari Nelis, Tõnu Esko, Reedik Mägi, Fritz Zimprich, Alexander Zimprich, Draga Toncheva, Sena Karachanak, Tereza Piskáčková, Ivan Balaščák, Leena Peltonen, et al. Genetic structure of europeans: a view from the north–east. *PloS one*, 4(5):e5472, 2009.

[100] Nora Franceschini, Frank JA van Rooij, Bram P Prins, Mary F Feitosa, Mahir Karakas, John H Eckfeldt, Aaron R Folsom, Jeffrey Kopp, Ahmad Vaez, Jeanette S

Andrews, et al. Discovery and fine mapping of serum protein loci through transethnic meta-analysis. *The American Journal of Human Genetics*, 91(4):744–753, 2012.

[101] Jason Ernst, Pouya Kheradpour, Tarjei S Mikkelsen, Noam Shoresh, Lucas D Ward, Charles B Epstein, Xiaolan Zhang, Li Wang, Robbyn Issner, Michael Coyne, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345):43–49, 2011.

[102] The FANTOM Consortium et al. A promoter-level mammalian expression atlas. *Nature*, 507(7493):462–470, 2014.

[103] Fiona Cunningham, M Ridwan Amode, Daniel Barrell, Kathryn Beal, Konstantinos Billis, Simon Brent, Denise Carvalho-Silva, Peter Clapham, Guy Coates, Stephen Fitzgerald, et al. Ensembl 2015. *Nucleic acids research*, 43(D1):D662–D669, 2015.

[104] E Lattka, S Eggers, G Moeller, K Heim, M Weber, D Mehta, H Prokisch, T Illig, and J Adamski. A common fads2 promoter polymorphism increases promoter activity and facilitates binding of transcription factor elk1. *Journal of lipid research*, 51(1):182–191, 2010.

[105] Malaria Genomic Epidemiology Network. Reappraisal of known malaria resistance loci in a large multicenter study. *Nature Genetics*, 2014.

[106] Jason Ernst and Manolis Kellis. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nature biotechnology*, 33(4):364–376, 2015.

[107] Melina Claussnitzer, Simon N Dankel, Kyoung-Han Kim, Gerald Quon, Wouter Meuleman, Christine Haugen, Viktoria Glunk, Isabel S Sousa, Jacqueline L Beaudry, Vijitha Puviindran, et al. Fto obesity variant circuitry and adipocyte browning in humans. *New England Journal of Medicine*, 373(10):895–907, 2015.

[108] Kiran Musunuru, Alanna Strong, Maria Frank-Kamenetsky, Noemi E Lee, Tim Ahfeldt, Katherine V Sachs, Xiaoyu Li, Hui Li, Nicolas Kuperwasser, Vera M Ruda, et al. From noncoding variant to phenotype via sort1 at the 1p13 cholesterol locus. *Nature*, 466(7307):714–719, 2010.

[109] Huwenbo Shi, Gleb Kichaev, and Bogdan Pasaniuc. Contrasting the genetic architecture of 30 complex traits from summary association data. *The American Journal of Human Genetics*, 99(1):139–153, 2016.

[110] Hilary K Finucane, Brendan Bulik-Sullivan, Alexander Gusev, Gosia Trynka, Yakir Reshef, Po-Ru Loh, Verneri Anttila, Han Xu, Chongzhi Zang, Kyle Farh, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics*, 47(11):1228–1235, 2015.

[111] Peter W Glynn and Donald L Iglehart. Importance sampling for stochastic simulations. *Management Science*, 35(11):1367–1392, 1989.

[112] Hong-Hee Won, Pradeep Natarajan, Amanda Dobbyn, Daniel M Jordan, Panos Roussos, Kasper Lage, Soumya Raychaudhuri, Eli Stahl, and Ron Do. Disproportionate contributions of select genomic compartments and cell types to genetic risk for coronary artery disease. *PLoS genetics*, 11(10):e1005622, 2015.

[113] Alexander Gusev, Huwenbo Shi, Gleb Kichaev, Mark Pomerantz, Fugen Li, Henry W Long, Sue A Ingles, Rick A Kittles, Sara S Strom, Benjamin A Rybicki, et al. Atlas of prostate cancer heritability in european and african-american men pinpoints tissue-specific regulation. *Nature Communications*, 7, 2016.

[114] Qiongshi Lu, Ryan Lee Powles, Qian Wang, Beixin Julie He, and Hongyu Zhao. Integrative tissue-specific functional annotations in the human genome provide novel insights on many complex traits and improve signal prioritization in genome wide association studies. *PLoS genetics*, 12(4):e1005947, 2016.

[115] Kathryn Roeder, B Devlin, and Larry Wasserman. Improving power in genome-wide association studies: weights tip the scale. *Genetic Epidemiology*, 31(7):741–747, 2007.

[116] Gregory Darnell, Dat Duong, Buhm Han, and Eleazar Eskin. Incorporating prior information into association studies. *Bioinformatics*, 28(12):i147–i153, 2012.

[117] Qiongshi Lu, Xinwei Yao, Yiming Hu, and Hongyu Zhao. Genowap: Gwas signal prioritization through integrated analysis of genomic functional annotation. *Bioinformatics*, 32(4):542–548, 2015.

[118] Xiaoquan Wen, Yeji Lee, Francesca Luca, and Roger Pique-Regi. Efficient integrative multi-snp association analysis via deterministic approximation of posteriors. *The American Journal of Human Genetics*, 98(6):1114–1129, 2016.

[119] Gardar Sveinbjornsson, Anders Albrechtsen, Florian Zink, Sigurjón A Gudjonsson, Asmundur Oddson, Gísli Másson, Hilma Holm, Augustine Kong, Unnur Thorsteinsdottir, Patrick Sulem, et al. Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nature genetics*, 48(3):314–317, 2016.

[120] Jingjing Yang, Lars G Fritsche, Xiang Zhou, Goncalo Abecasis, International Age-Related Macular Degeneration Genomics Consortium, et al. A scalable bayesian method for integrating functional information in genome-wide association studies. *The American Journal of Human Genetics*, 101(3):404–416, 2017.

[121] Lei Sun, Radu V Craiu, Andrew D Paterson, and Shelley B Bull. Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies. *Genetic epidemiology*, 30(6):519–530, 2006.

[122] James X Hu, Hongyu Zhao, and Harrison H Zhou. False discovery rate control with groups. *Journal of the American Statistical Association*, 105(491):1215–1227, 2010.

[123] Nikolaos Ignatiadis, Bernd Klaus, Judith B Zaugg, and Wolfgang Huber. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nature methods*, 13(7):577–580, 2016.

[124] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, et al. Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*, 12(3):e1001779, 2015.

[125] Clare Bycroft, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T Elliott, Kevin Sharp, Allan Motyer, Damjan Vukcevic, Olivier Delaneau, Jared O'Connell, et al. Genome-wide genetic data on˜ 500,000 uk biobank participants. *bioRxiv*, page 166298, 2017.

[126] Po-Ru Loh, Gleb Kichaev, Steven Gazal, Armin P Schoech, and Alkes L Price. Mixed model association for biobank-scale data sets. *bioRxiv*, page 194944, 2017.

[127] John D Storey and Robert Tibshirani. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100(16):9440–9445, 2003.

[128] Daniel Yekutieli. Hierarchical false discovery rate–controlling methodology. *Journal of the American Statistical Association*, 103(481):309–316, 2008.

[129] Po-Ru Loh, George Tucker, Brendan K Bulik-Sullivan, Bjarni J Vilhjalmsson, Hilary K Finucane, Rany M Salem, Daniel I Chasman, Paul M Ridker, Benjamin M Neale, Bonnie Berger, et al. Efficient bayesian mixed-model analysis increases association power in large cohorts. *Nature genetics*, 47(3):284–290, 2015.

[130] Gleb Kichaev, Wen-Yun Yang, Sara Lindstrom, Farhad Hormozdiari, Eleazar Eskin, Alkes L Price, Peter Kraft, and Bogdan Pasaniuc. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS genetics*, 10(10):e1004722, 2014.

[131] Hilary Finucane, Yakir Reshef, Verneri Anttila, Kamil Slowikowski, Alexander Gusev, Andrea Byrnes, Steven Gazal, Po-Ru Loh, Giulio Genovese, Arpiar Saunders, et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nature Genetics (in press)*, 2017.

[132] Doug Speed, Gibran Hemani, Michael R Johnson, and David J Balding. Improved heritability estimation from genome-wide snps. *The American Journal of Human Genetics*, 91(6):1011–1021, 2012.

[133] Jian Zeng, Ronald de Vlaming, Yang Wu, Matthew Robinson, Luke Lloyd-Jones, Loic Yengo, Chloe Yap, Angli Xue, Julia Sidorenko, Allan McRae, et al. Widespread signatures of negative selection in the genetic architecture of human complex traits. *bioRxiv*, page 145755, 2017.

[134] Armin Schoech, Daniel Jordan, Po-Ru Loh, Steven Gazal, Luke O'Connor, Daniel J Balick, Pier F Palamara, Hilary Finucane, Shamil R Sunyaev, and Alkes L Price. Quantification of frequency-dependent genetic architectures and action of negative selection in 25 uk biobank traits. *bioRxiv*, page 188086, 2017.

[135] Brendan K Bulik-Sullivan, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, Nick Patterson, Mark J Daly, Alkes L Price, Benjamin M Neale, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics*, 47(3):291–295, 2015.

[136] Alan Dabney, John D Storey, and GR Warnes. qvalue: Q-value estimation for false discovery rate control. *R package version*, 1(0), 2010.

[137] Christopher C Chang, Carson C Chow, Laurent CAM Tellier, Shashaank Vattikuti, Shaun M Purcell, and James J Lee. Second-generation plink: rising to the challenge of larger and richer datasets. *Gigascience*, 4(1):7, 2015.

[138] Cornelius A Rietveld, Sarah E Medland, Jaime Derringer, Jian Yang, Tõnu Esko, Nicolas W Martin, Harm-Jan Westra, Konstantin Shakhbazov, Abdel Abdellaoui, Arpana Agrawal, et al. Gwas of 126,559 individuals identifies genetic variants associated with educational attainment. *science*, 340(6139):1467–1471, 2013.

[139] Lavinia Paternoster, Marie Standl, Johannes Waage, Hansjörg Baurecht, Melanie Hotze, David P Strachan, John A Curtin, Klaus Bønnelykke, Chao Tian, Atsushi Takahashi, et al. Multi-ethnic genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nature genetics*, 47(12):1449, 2015.

[140] AP Morris, BF Voight, TM Teslovich, T Ferreira, AV Segre, V Steinthorsdottir, RJ Strawbridge, H Khan, H Grallert, A Mahajan, I Prokopenko, HM Kang, C Dina, T Esko, RM Fraser, S Kanoni, A Kumar, V Lagou, C Langenberg, J Luan, CM Lindgren, M Muller-Nurasyid, S Pechlivanis, NW Rayner, LJ Scott, S Wiltshire, L Yengo, L Kinnunen, EJ Rossin, S Raychaudhuri, AD Johnson, AS Dimas, RJ Loos, S Vedantam, H Chen, JC Florez, C Fox, CT Liu, D Rybin, DJ Couper, WH Kao, M Li, MC Cornelis, P Kraft, Q Sun, RM van Dam, HM Stringham, PS Chines, K Fischer, P Fontanillas, OL Holmen, SE Hunt, AU Jackson, A Kong, R Lawrence, J Meyer, JR Perry, CG Platou, S Potter, E Rehnberg, N Robertson, S Sivapalaratnam, A Stancakova, K Stirrups, G Thorleifsson, E Tikkanen, AR Wood, P Almgren, M Atalay, R Benediktsson, LL Bonnycastle, N Burtt, J Carey, G Charpentier, AT Crenshaw, AS Doney, M Dorkhan, S Edkins, V Emilsson, E Eury, T Forsen, K Gertow, B Gigante, GB Grant, CJ Groves, C Guiducci, C Herder, AB Hreidarsson, J Hui, A James, A Jonsson, W Rathmann, N Klopp, J Kravic, K Krjutskov, C Langford, K Leander, E Lindholm, S Lobbens, S Mannisto, G Mirza, TW Muhleisen, B Musk, M Parkin, L Rallidis, J Saramies, B Sennblad, S Shah, G Sigurethsson, A Silveira, G Steinbach, B Thorand, J Trakalo, F Veglia, R Wennauer, W Winckler, D Zabaneh, H Campbell, C van Duijn, AG Uitterlinden, A Hofman, E Sijbrands, GR Abecasis,

KR Owen, E Zeggini, MD Trip, NG Forouhi, AC Syvanen, JG Eriksson, L Peltonen, MM Nothen, B Balkau, CN Palmer, V Lyssenko, T Tuomi, B Isomaa, DJ Hunter, L Qi, AR Shuldiner, M Roden, I Barroso, T Wilsgaard, J Beilby, K Hovingh, JF Price, JF Wilson, R Rauramaa, TA Lakka, L Lind, G Dedoussis, I Njolstad, NL Pedersen, KT Khaw, NJ Wareham, SM Keinanen-Kiukaanniemi, TE Saaristo, E Korpi-Hyovalti, J Saltevo, M Laakso, J Kuusisto, A Metspalu, FS Collins, KL Mohlke, RN Bergman, J Tuomilehto, BO Boehm, C Gieger, K Hveem, S Cauchi, P Froguel, D Baldassarre, E Tremoli, SE Humphries, D Saleheen, J Danesh, E Ingelsson, S Ripatti, V Salomaa, R Erbel, KH Jockel, S Moebus, A Peters, T Illig, U de Faire, A Hamsten, AD Morris, PJ Donnelly, TM Frayling, AT Hattersley, E Boerwinkle, O Melander, S Kathiresan, PM Nilsson, P Deloukas, U Thorsteinsdottir, LC Groop, K Stefansson, F Hu, JS Pankow, J Dupuis, JB Meigs, D Altshuler, M Boehnke, and MI McCarthy. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*, 44(9):981–990, 2012.

[141] Tobacco, Genetics Consortium, et al. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature genetics*, 42(5):441–447, 2010.

[142] John RB Perry, Felix Day, Cathy E Elks, Patrick Sulem, Deborah J Thompson, Teresa Ferreira, Chunyan He, Daniel I Chasman, Tõnu Esko, Gudmar Thorleifsson, et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature*, 514(7520):92–97, 2014.

[143] Felix R Day, Katherine S Ruth, Deborah J Thompson, Kathryn L Lunetta, Natalia Pervjakova, Daniel I Chasman, Lisette Stolk, Hilary K Finucane, Patrick Sulem, Brendan Bulik-Sullivan, et al. Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and brca1-mediated dna repair. *Nature genetics*, 47(11):1294–1303, 2015.