Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations

Abhay Hukku,[1] Milton Pividori,[2] Francesca Luca,[3] Roger Pique-Regi,[3] Hae Kyung Im,[2] and Xiaoquan Wen[1,*]

Summary

Colocalization analysis has emerged as a powerful tool to uncover the overlapping of causal variants responsible for both molecular and complex disease phenotypes. The findings from colocalization analysis yield insights into the molecular pathways of complex diseases. In this paper, we conduct an in-depth investigation of the promise and limitations of the available colocalization analysis approaches. Focusing on variant-level colocalization approaches, we first establish the connections between various existing methods. We proceed to discuss the impacts of various controllable analytical factors and uncontrollable practical factors on outcomes of colocalization analysis through realistic simulations and real data examples. We identify a single analytical factor, the specification of prior enrichment levels, which can lead to severe inflation of false-positive colocalization findings. Meanwhile, the combination of many other analytical and practical factors all lead to diminished power. Consequently, we recommend the following strategies for the best practice of colocalization analysis: (1) estimating prior enrichment level from the observed data and (2) separating fine-mapping and colocalization analysis. Our analysis of 4,091 complex traits and the multi-tissue expression quantitative trait loci (eQTL) data from the GTEx (v.8) suggests that colocalizations of molecular QTLs and causal complex trait associations are widespread. However, only a small proportion can be confidently identified from currently available data due to a lack of power. Our findings set a benchmark for current and future integrative genetic association analysis applications.

Introduction

The advancements in genetic association analysis of complex and molecular traits have uncovered a large volume of putative causal genetic variants. Subsequently, utilizing genetic association discoveries to explore the molecular mechanisms of complex disease etiology has become a standard practice in human genetics research. Various types of analytical approaches designed for the integrative analysis of data from expression quantitative trait loci (eQTL) mapping and genome-wide association studies (GWASs) of complex traits have shown promise in implicating molecular pathways connecting genetic variations, molecular phenotype changes, and complex diseases.[1–6]

Colocalization analysis is an integrative analysis technique that aims to identify genetic variants introducing simultaneous phenotypic changes in multiple molecular and/or complex traits. Although colocalization analysis is not constrained by the types of phenotypes investigated, we focus our discussions in this paper on a single complex trait and one type of molecular trait, e.g., gene expressions. The discoveries from this class of analyses have resulted in molecular insights of complex diseases, e.g., atherosclerosis,[7] the age of onset of menarche and menopause,[8] and cardiovascular disease.[9]

There are two broad types of colocalization analysis approaches in the current literature. The first kind, represented by regulatory trait concordance (RTC)[2] and joint likelihood mapping (JLIM),[10] makes claims that causal GWAS hits and eQTL signals co-exist within a genomic region consisting of tightly linked genetic variants. We refer to this type as locus-level colocalization analysis. The second kind, illustrated by coloc,[11] eCAVIAR,[12] and ENLOC/fastENLOC,[6,13] attempts to uncover colocalization signals at the single SNP/variant resolution using probabilistic quantifications. We refer to this type as SNP-level colocalization analysis. The common obstacle for both types of colocalization analysis is linkage disequilibrium (LD) among candidate SNPs. Hypothetically, with complete linkage equilibrium, colocalization analysis becomes relatively trivial, and various approaches from both types converge. With the presence of LD, a SNP-level colocalization may not be identifiable. That is, multiple competing scenarios may be indistinguishable based solely on the observed association data. (See the example of two perfectly linked SNPs illustrated in Wen et al.[13]) Thus, the quantification of SNP-level colocalization evidence should acknowledge such uncertainty explicitly. Furthermore, relevant additional information, e.g., enrichment level of eQTLs in GWAS hits, should be incorporated to aid in the identification of more likely scenarios. Based on the above considerations, probabilistic analysis in the Bayesian framework becomes a natural choice for SNP-level colocalization analysis; all aforementioned SNP-level colocalization

[1]Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA; [2]Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL 60637, USA; [3]Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI 48201, USA
*Correspondence: xwen@umich.edu
https://doi.org/10.1016/j.ajhg.2020.11.012.

methods utilize Bayesian probabilistic modeling approaches. (In this paper, we use the terms "SNP-level colocalization analysis" and "probabilistic colocalization analysis" interchangeably.)

Probabilistic colocalization analysis at the variant level faces many challenges, both analytically and practically. Analytical factors—for example, prior specifications and model assumptions required for likelihood computation (e.g., consideration of allelic heterogeneity)—are known to have drastic impacts on analysis outcomes. Practically, even when the ideal analytical strategies are applied, the power of colocalization analysis can still be limited by the underlying association data, e.g., the power of marginal association analysis. In this paper, we take a divide-and-conquer strategy to systematically investigate various analytical and practical factors in probabilistic colocalization analyses. We attempt to isolate each factor through analytical derivation and numerical experiments and to quantify its effects on potential false positive and false negative findings. We seek to identify a set of best analytical strategies that enable robust and powerful probabilistic colocalization analysis. Also, we hope to practically illustrate the natural limitations of colocalization analysis based on the currently available data and establish a baseline for future development in integrative genetic research.

Material and methods

For probabilistic colocalization analysis, we focus our discussion on three representative methods: coloc,[11] eCAVIAR,[12] and ENLOC/fastENLOC.[6,13] A comprehensive overview of the three methods and additional approaches is provided in Section 1 of the supplemental material and methods.

Statistical framework of probabilistic colocalization analysis

Let the binary indicators γ and d denote the latent causal association status of a given variant with respect to the complex and gene expression traits of interest, respectively. The probabilistic quantification of colocalization for the variant is essentially to evaluate the conditional probability,

$$\Pr(\gamma = 1, d = 1 | \text{eQTL data}, \text{GWAS data}). \qquad \text{(Equation 1)}$$

All known probabilistic colocalization approaches aim to compute Equation 1, which is carried out by applying the Bayes rule, i.e.,

$$\Pr(\gamma = 1, d = 1 | \text{eQTL data}, \text{GWAS data}) \qquad \text{(Equation 2)}$$

$$\propto \Pr(\gamma = 1, d = 1) P(\text{eQTL data}, \text{GWAS data} \mid d = 1, \gamma = 1),$$

where eQTL data and GWAS data represent the genotype and phenotype data collected for eQTL mapping and GWAS analysis, respectively. Noticeably, the computation requires an explicit specification of the prior probability $\Pr(\gamma = 1, d = 1)$ and the likelihood function $P(\text{eQTL data}, \text{GWAS data} \mid d = 1, \gamma = 1)$.

The prior quantity, $\Pr(\gamma = 1, d = 1)$, reflects the frequency of colocalization sites in all interrogated variants and can be equivalently specified by the product of $p_d := \Pr(d = 1)$ and $\Pr(\gamma = 1 | d = 1)$, i.e.,

$$\Pr(\gamma = 1, d = 1) = \Pr(\gamma = 1 | d = 1) \Pr(d = 1). \qquad \text{(Equation 3)}$$

In ENLOC/fastENLOC, a set of parameters (α_0, α_1), referred to as "enrichment parameters," are introduced to parameterize $\Pr(\gamma = 1 | d)$, i.e.,

$$\text{logit}[\Pr(\gamma = 1 | d)] = \alpha_0 + \alpha_1 d, \qquad \text{(Equation 4)}$$

where α_1 is the log odds ratio quantifying the enrichment level of molecular QTLs in GWAS hits. Note that, with the specification of α_0, α_1 and the frequency of causal eQTLs, p_d, the frequency of the causal GWAS hits, $p_\gamma := \Pr(\gamma = 1)$, is induced.

In the implementation of coloc, the priors are defined by $p_1 := \Pr(\gamma = 0, d = 1)$, $p_2 := \Pr(\gamma = 1, d = 0)$, and $p_{12} := \Pr(\gamma = 1, d = 1)$. The equivalent parametrization by $(p_d, \alpha_0, \alpha_1)$ is given by

$$p_d = p_1 + p_{12}$$

$$\alpha_0 = \log\left[\frac{p_1}{1 - p_1 - p_2 - p_{12}}\right]$$

$$\alpha_1 = \log\left[\frac{p_{12}(1 - p_1 - p_2 - p_{12})}{p_1 p_2}\right]$$

The third approach, eCAVIAR, makes a simplifying assumption that the causal status of γ and d are *a priori* independent. In the ENLOC parameterization, this independence assumption implies that there is no enrichment of eQTLs in GWAS hits, i.e.,

$$\left(p_d, \alpha_0 = \log\left[\frac{p_\gamma}{1 - p_\gamma}\right], \alpha_1 \equiv 0\right).$$

Or equivalently,

$$\Pr(\gamma = 1, d = 1) = \Pr(\gamma = 1) \Pr(d = 1) = p_\gamma \cdot p_d.$$

Given the equivalence of different formulations in all methods, our subsequent discussion in the results section will focus on the ENLOC/fastENLOC parameterization because of its convenient interpretation.

Results

Analytical strategies in colocalization analysis

We consider two aspects of the analytical strategy in probabilistic colocalization approaches:the prior specification and the likelihood computation.

Specification of enrichment prior

We perform a series of numerical experiments to evaluate the sensitivity of colocalization analysis outcomes with respect to the enrichment parameter specification. In these experiments, we fix the frequencies of causal eQTLs and GWAS hits, p_d and p_γ, and consider α_1 the only free parameter. To isolate the effect of prior specification, we only consider computing the colocalization probability for a single SNP assumed to be in complete linkage equilibrium with other candidate variants. Particularly, we consider weak, modest, and strong association evidence from

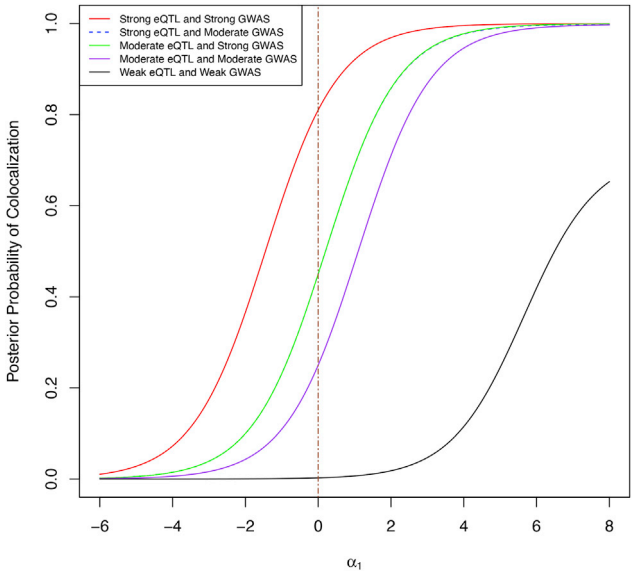

Figure 1. The impact of pre-defined enrichment level (α_1) on the SNP-level colocalization probability

The different curves represent different combined levels of marginal association evidence for a particular SNP from the eQTL and GWASs. Informed by our real data analysis of 4,091 GWAS traits and the eQTL data from the GTEx project, we consider $\alpha_1 \in [0, 5]$ a meaningful range for practical colocalization analysis. Within this range, all categories show different levels of sensitivity to the specification of the enrichment parameter.

respective eQTL or GWAS analysis for the variant and examine the effect of varying enrichment levels on the magnitude of resulting SNP-level colocalization probabilities. The details on the design of the numerical experiment are provided in Section 3 of the supplemental material and methods.

The results are summarized in Figure 1. Based on our real data analysis of 4,091 GWAS traits and the eQTL data from the GTEx project (Figure 5), we consider $\alpha_1 \in [0, 5]$ a meaningful range in practical colocalization analysis. Within this range, SNP-level colocalization probabilities are generally sensitive to the enrichment prior specification. However, depending on the combination of strength of evidence from individual association studies, different combination categories are differentially impacted. Specifically, when the eQTL and GWAS association evidence are both strong or weak, the resulting colocalization probabilities are relatively stable with respect to the changes of the enrichment prior. This phenomenon can be intuitively explained: when the marginal association evidence is highly informative (including the case that association evidence is weak, i.e., the evidence for no association is strong), the likelihood for colocalization is overwhelming, and the prior impact is diminished. On the other hand, SNPs with modest association evidence from either GWAS or eQTL analysis are most sensitive to the prior specification, as a strong enrichment assumption can significantly increase the corresponding probability of colocalization.

The sensitive nature of probabilistic colocalization analysis should caution practitioners. Because colocalization analysis is often treated as a discovery process similar to a hypothesis testing procedure, false positives, i.e., type I errors, should be carefully guarded against. From the numerical experiment, we observe that aggressively setting a high enrichment value tends to flag many more colocalization sites than setting a conservative value. However, this is also dangerous for inflating false-positive colocalization findings. Thus, we conclude that a conservative enrichment prior is, in principle, acceptable and may be preferable. Nevertheless, simply setting $\alpha_1 = 0$ in all circumstances can be too conservative and lead to a severe loss of power.

Given the observations from the above experiments, great care is warranted in prior specifications. For an ideal Bayesian analysis, the required prior information should be derived from historical analyses of similar types. In the case that such historical information is unavailable, we recommend to estimate required hyper-parameters from the observed data. One of the established estimation procedures is implemented in ENLOC/fastENLOC, where p_d, α_0, and α_1 are estimated by jointly analyzing the eQTL and GWAS data via a multiple imputation procedure[13] (a summary of the procedure is provided in Section 2 of the supplemental material and methods). This estimation procedure, designed specifically for dealing with the latent association status in both eQTLs and GWAS hits, has shown the ability to provide robust and reliable enrichment estimates in our simulation studies (see results). Recently, Wallace et al.[14] proposes to perform a sensitivity analysis of the priors for identified colocalization signals. While we completely agree that understanding prior sensitivity is critical for practitioners, it should be noted that sensitivity analysis alone does not justify selecting a specific set of priors.

Likelihood computation and accounting for allelic heterogeneity

The likelihood computation in probabilistic colocalization analysis refers to the evaluation of $P(\text{eQTL data}, \text{GWAS data} \mid d, \gamma)$ in Equation 2. In the current practice, eQTL and GWAS data are typically obtained from non-overlapping cohorts, and all methods compute the likelihood by

$$P(\text{eQTL data}, \text{GWAS data} \mid d, \gamma) = P(\text{eQTL data} \mid d) P(\text{GWAS data} \mid \gamma). \quad \text{(Equation 5)}$$

There are two different strategies for evaluating likelihood. The first strategy, adopted by eCAVIAR and fastENLOC, recovers the required likelihood information from multi-SNP fine-mapping analyses of eQTL and GWAS data. The second strategy, used by coloc, directly computes each trait's marginal likelihood from summary statistics under a simplifying assumption of no allelic heterogeneity. Allelic heterogeneity (AH) refers to the phenomenon of distinct genetic variants at a locus

Table 1. The impact of modeling consideration of AH on FDR and power in colocalization analysis

	Dataset			
	scenarios (1,2,3)		scenarios (1,2,4)	
AH modeling	FDR	power	FDR	power
Yes	0.047	0.997	0.041	0.971
No	0.039	0.937	0.044	0.411

fastENLOC and coloc are selected to represent the approaches with and without explicit AH modeling, respectively. Both approaches maintain the proper false discovery rate levels in all settings. However, the power difference is quite large when AH is indeed present in the molecular QTL data (i.e., scenarios (1,2,4)).

simultaneously affecting the same phenotype. Under the assumption of no AH, there is, at most, one causal SNP within a locus for a given trait. Henceforth, we refer to such an assumption as the "one causal variant" (OCV) assumption. The primary rationale for the OCV assumption is computational (rather than biological): if the assumption holds, the LD information within the locus of interest becomes obsolete for likelihood evaluation, and the computation can be carried out analytically based on single-variant association test statistics.[15–17]

In colocalization analysis of eQTLs and GWAS hits, a locus is typically defined as the *cis* region of a target gene, e.g., a 2 Mb window centered around the transcription start site.[18] At such a scale, AH is a widespread phenomenon in gene regulations based on overwhelming evidence from recent large-scale eQTL studies.[18,19] While it may work reasonably well for some complex traits, the OCV assumption is likely often violated in the analysis of molecular traits. Here, we are interested in exploring its implications on both false-positive and false-negative findings in colocalization analysis.

We first conduct simulation studies using real genetic data from the GTEx project and simulate gene expression and complex trait data based on linear regression models. To isolate the effect of AH, we focus on simulating strong genetic effects for all eQTLs and GWAS hits. (As shown in the previous section, these signals are robust to the misspecification of enrichment parameters.) Furthermore, both expression and complex trait data are independently generated from the same genotype data, which ensures LD mismatch is not a factor in the analysis. Our simulation considers the following scenarios for each gene-trait pair within a genomic locus:

1 single causal variants in both eQTL and GWAS data, no colocalization
2 AH in eQTL data (two causal eQTLs per gene), a single causal variant in GWAS data, no colocalization
3 single causal variants in both eQTL and GWAS data, single colocalization event
4 AH in eQTL data (two causal eQTLs per gene), a single causal variant in GWAS data, single colocalization event

The first two scenarios are designed to investigate potential false-positive findings, and the last two are for false-negative findings. The assembled data for analysis consist of different mixtures of the four scenarios, such that we could evaluate both the type I and the type II errors. The simulation details are provided in the Section 4 of the supplemental material and methods.

We use two different methods to analyze the simulated data. Particularly, we apply the default coloc method to represent methods making the OCV assumption and without explicit modeling of AH. We apply fastENLOC to represent methods that explicitly model potential AH. We provide the true priors for coloc analysis, and the fastENLOC analysis estimates the enrichment prior directly from the data. The results of our simulations are in Table 1. In all of our simulated scenarios, we find that both types of approaches control the false discovery rate (FDR). However, when AH is presented, the power of coloc is approximately half of the power of fastENLOC. The relative ratio of power between the two types of approaches is expected. When two independent eQTL signals co-exist, coloc identifies one signal with a stronger signal-to-noise ratio as the sole casual eQTL, which leads to a false-negative finding when the unselected eQTL overlaps the causal GWAS hit.

Even though the simulation study does not indicate apparent inflation of FDR, there are some theoretical concerns on potential false positives related to the OCV assumption. First, some implementations of the assumption, e.g., coloc, enumerate all possible causal association configurations from both traits to compute the normalizing constants and desired colocalization probabilities. Under the simplifying assumption of OCV, there are $(p + 1)^2$ possibilities precisely (where p represents the number of SNPs in the locus). If the assumption is violated, many more necessary scenarios are uncounted (the total possibilities without constraints are 2^p). This factor can lead to under-estimating the normalizing constants and over-estimating the colocalization probabilities. Second, false positives can be carried over from single-SNP analysis in the marginal studies. It is known that some non-causal SNPs that are in partial LD with multiple causal SNPs can generate the most significant single-SNP association evidence.[19] Such false-positive findings based on single-SNP association evidence are maintained and carried over into the colocalization analysis under the OCV assumption. Consequently, it becomes a source of false-positive colocalization findings. In contrast, methods that explicitly perform multi-SNP fine-mapping analysis can effectively dissect different scenarios by accounting for LD. Hence, they are unlikely to suffer from such false-positive findings. In summary, we conclude that the OCV assumption can, in theory, lead to anti-conservative quantifications of colocalization probabilities. Although the extent of the anti-conservativeness may not lead to observable inflations of FDR within our simulations, its effect can be observed in the numerical experiments examining the calibration of the reported colocalization probabilities

Table 2. Realized false discovery rates and power of colocalization analysis in simulated data

Method	Number of discoveries	Realized FDR	Power
fastENLOC (estimated prior)	406	0.034	0.186
fastENLOC (true prior)	458	0.046	0.208
coloc (default prior)	472	0.258	0.175
coloc (true prior)	200	0.045	0.095

The table shows an overall assessment of power and realized FDR (controlled at 5% level) for various colocalization approaches. Only coloc with its default subjective prior shows severely inflated type I errors. The enrichment priors are justified for the remaining approaches, and they properly control the FDR. The power (for methods properly controlling FDR) is overall quite low in this setting that resembles realistic applications.

(Section 5.1 of the supplemental material and methods and Figure S1).

It is worth emphasizing that the OCV assumption may not be invalid in all scenarios, but it is inappropriate at the scale of genomic regions commonly used in molecular QTL studies for molecular QTL mapping analysis. More generally, our recommendation for dealing with likelihood computation in colocalization analysis is to apply specialized multi-SNP association techniques. We note that state-of-the-art of fine-mapping approaches[8,17,20] all have the ability to account for AH without making the OCV assumption. It is also intuitive that colocalization analysis should be based on the best possible fine-mapping results. This is because the inaccuracy from poor likelihood computation will inevitably translate into inaccuracy in subsequent probabilistic quantification of colocalization. The emergence of fast and accurate Bayesian multi-SNP association analysis methods, e.g., FINEMAP,[21] DAP-G,[17] and SuSIE,[8] make it feasible to practically separate fine-mapping and colocalization analyses with affordable computational costs. The current implementation of fastENLOC can take the fine-mapping results from any of those Bayesian methods and perform colocalization analysis.

Practical factors in colocalization analysis

There are many practical factors in colocalization analysis that analysts have little control over. Nevertheless, their impacts on the outcomes are profound. Having established the fundamental inference principles, we proceed to assess the empirical performance of probabilistic colocalization analysis and investigate the other performance-impacting factors using realistically simulated eQTL and GWAS data.

Empirical assessment of probabilistic colocalization analysis

To construct a simulated dataset that resembles real applications of colocalization analysis, we simulate 20,000 non-overlapping genes with 1,500 SNPs within each *cis*-region. The scale of the simulated datasets resembles real applications of genome-wide colocalization analysis. We use the real genetic data from 400 participants of the GTEx project. In order to circumvent the issue of LD mismatch in this particular experiment, we use the same set of genotype data to simulate both the expression and complex trait phenotypes. Within each *cis* region, the causal eQTLs are randomly selected from a series of independent Bernoulli trials, such that on average there are three causal eQTLs per gene (i.e., $p_d = 2 \times 10^{-3}$). Similarly, we sample causal GWAS SNPs conditional on the simulated eQTL status using the probability model (Equation 1) with a true α_1 value of 4. As a result, the simulated dataset consists of 2,103 colocalized association signals distributed in 2,001 unique genes. Given the truly causal SNPs for each gene, we independently generate the molecular and complex trait data for the 400 individuals based on standard multiple linear regression models. Specifically, each causal variant-trait pair's genetic effects are independently drawn from the distribution $N(0,1)$. The residual errors for each trait and each individual are also independently generated from the standard normal distribution. Additional simulation details are provided in Section 5 of the supplemental material and methods.

To analyze the simulated dataset, we first perform separate Bayesian fine-mapping analyses for the simulated eQTL and GWAS datasets using the software package DAP-G.[17] Utilizing the resulting probabilistic annotations, we apply fastENLOC to estimate the enrichment parameters with the default shrinkage setting. As expected, the estimated enrichment parameter α_1 is slightly under-estimated, but reasonably close to the true value ($\widehat{\alpha}_1 = 3.644$ with the standard error 0.039). We then perform colocalization analysis using fastENLOC using the estimated and true enrichment parameters, respectively. For comparison, we also run coloc with its default prior model parameters and the true parameters, respectively. Note that the differences between coloc and fastENLOC results based on the true enrichment parameters should reflect the difference in fine-mapping analysis, including the consideration of AH.

We first examine the false positive rate and the power at 5% FDR level for different analysis settings (Table 2). We find that severe inflation of type I errors only occurs at the coloc run with its default model priors (which significantly exceeds the true enrichment parameter). All other analysis settings, including the coloc run with the true enrichment parameters, show proper control of the desired false discovery rate. For the methods that control the type I errors, the power seems low across the board (Table 2). The under-estimation of the enrichment parameter (α_1) due to shrinkage only explains a small fraction of the power loss. Although the power and type I error analysis only focus on high colocalization probability values, our conclusion extends to the full probability spectrum. Additional inspection of the calibration of the regional colocalization probabilities (RCPs) also confirms that various methods yield conservative colocalization results when supplied with

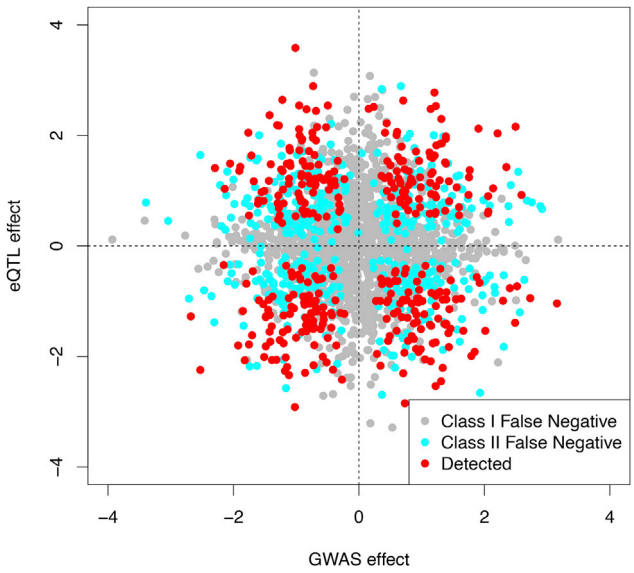

Figure 2. Classification of all colocalized SNPs
All truly colocalized variants from our realistic simulations, classified as either a class I false negative, class II false negative, or successfully detected by fastENLOC. We see the expected pattern that most points near one of the axes are class I false negatives, while points far away from both tend to be detected by fastENLOC.

the true enrichment prior (Figure S1 and Section 5.1 of the supplemental material and methods).

We identify two primary sources of false-negative errors by an in-depth examination of the simulated data and the corresponding analysis results. We refer to these two sources as class I and class II false-negative errors in colocalization analysis. Specifically, we define

1. Class I false negatives: lack of power in association analysis of individual traits
2. Class II false negatives: inaccurate quantification of association evidence at SNP level for individual traits

The class I false negatives (FNs) represent the cases of failure in detecting at least one type of association signals (eQTL or GWAS) in genetic association analysis. The class II FNs represent the scenarios where both types of association signals are correctly uncovered at locus level, but the inaccurate SNP-level quantifications imply that the causal variants for the two types of traits are unlikely overlapping. In our simulation studies, 59.0% (1,240) and 22.0% (463) of the true colocalization signals fall into the class I and II false negatives categories, respectively. To better visualize the two FN classes, we plot the true eQTL and GWAS effects of the colocalized signals with their corresponding labeled categories based on the fastENLOC results in Figure 2. Most points representing class I FNs (gray) are closely located around the axes, indicating at least one of the genetic effects (eQTL or GWAS) is too small to be detected by the corresponding association analysis. Marginally, at the 5% FDR level, the power for GWAS and eQTL is 44% and 62%, respectively. In comparison, most points representing class II FNs (cyan) and detected signals (red) are scattered around the two diagonals.

The impacts of class I FNs are easy to understand and well expected. As neither eQTL mapping nor genetic association analysis of complex traits achieves high power in practice, this class of FNs remains a primary source for failures in identifying colocalization sites.

A proportion of class II false negatives can be explained by a "threshold" effect. For example, some modest eQTL signals barely clear the bar to qualify for significant eQTL findings, but the underlying evidence is not strong enough to ensure a significant colocalization discovery. Additionally, the class II FNs can occur even when the association signals for both GWAS and eQTL can be narrowed down into the same genomic locus with high confidence. This is because, at the SNP level, it remains difficult to pinpoint the causal variants for both traits due to the combination of LD and insufficient sample size. The phenomenon of class II FNs is also closely related to a well-known fact in fine-mapping analysis: the lead (i.e., the most significant) SNPs that emerge from association analysis may not be the true causal SNPs.[22,23] Incidental correlation between the genotypes of non-causal SNPs (in LD with the true causal variant) and residual errors from the outcome variable could lead to stronger empirical correlation, especially with limited samples. There is generally a higher level of mismatching between lead and causal SNPs when the underlying studies are underpowered. In any association analysis, Bayesian or frequentist, the lead SNPs are always regarded as the most plausible causal SNPs from the data. When the mismatch of lead and causal SNPs occurs in at least one trait, all algorithms are led to believe there is a lack of evidence for SNP-level colocalization, even though the signal clusters for both traits are correctly identified.

To provide a visualization, we compute a ratio of posterior inclusion probabilities (PIPs) for the causal SNP versus the lead SNP (causal-versus-lead PIP ratio) in each signal cluster harboring a true colocalized signal for both simulated eQTL and GWAS data. The PIPs for a SNP, $\Pr(\gamma = 1|\text{GWAS data})$ and $\Pr(d = 1|\text{eQTL data})$, quantify the strength of association evidence in the GWAS and eQTL data, respectively. The PIP ratio $= 1$ indicates that the lead SNP is indeed the causal SNP (or they are in perfect LD). We further compute a combined ratio by multiplying the two trait-specific causal-versus-lead PIP ratios for each signal cluster. Note that the combined ratio $= 1$ suggests the causal SNPs are identified as lead SNPs in both traits, whereas the combined ratio < 1 indicates that in at least one trait, the lead SNP and the causal SNP do not match. The comparison of various PIP ratios between detected and class II false-negative signals is shown in the histograms of Figure 3. The overall patterns in Figure 3 indicates that in detected colocalization signals, the vast majority of causal SNPs are indeed lead SNPs in both traits; many mismatches between causal and lead SNPs lead to false negatives in identifying the colocalized signals.

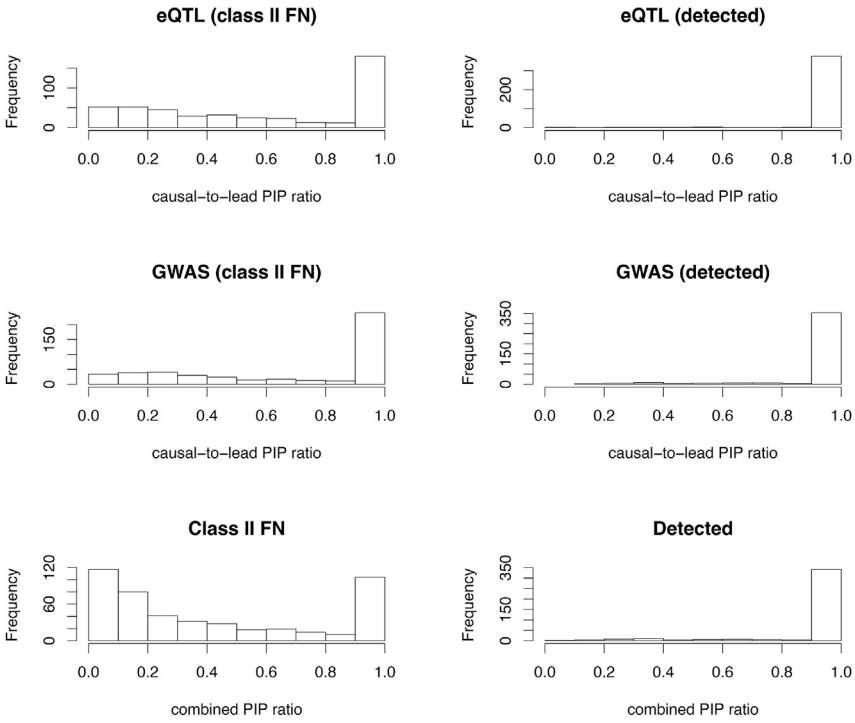

Figure 3. PIP ratios for all signal clusters
The first column of histograms represents the ratio of causal SNP PIP to lead SNP PIP for all class II false negatives from the eQTL simulated dataset, the GWAS dataset, then both combined. The second column represents the same ratio, but among all successfully detected colocalizations from the respective datasets.

Both classes of false-negative errors are intrinsic to genetic association analysis and are well known. However, it is somewhat surprising to observe that the combined effects from these factors have such a drastic effect on the power of colocalization analysis—even when the association analysis for individual traits is considered relatively well powered.

In practice, when the truth is unknown, we can still assess the relative power of the colocalization analysis based on the estimates of p_d, p_γ, and α_1. The expected number of colocalization sites based on the enrichment analysis can be computed by $Mp_\gamma/\left(1+\frac{p_d}{1-p_d}\exp(-\alpha_1)\right)$, where M represents the total number of genetic variants. In this simulation, the fastENLOC estimates

$$\widehat{p}_\gamma = 1.3\times 10^{-3}, \widehat{p}_d = 5.4\times 10^{-4}, \widehat{\alpha}_1 = 3.644,$$

and the expected number of the colocalization sites is ~800, which represents a lower bound estimate of true colocalization sites (due to the conservative estimate of α_1 and the class I FNs). The number of detected sites at the 5% FDR level is roughly half of the expected sites. Henceforth, we refer to this proportion as the rejection-to-expectation (rej-to-exp) ratio, representing an upper-bound estimate of the empirical power of the colocalization analysis.

Mismatching LD structures

Most existing colocalization analysis approaches are built on the experimental scheme known as the two-sample design, where the eQTL and GWAS data do not share common samples. While this design allows for using valuable eQTL resources, e.g., GTEx data, for analyzing a wide range of GWAS data collected from many different cohorts, it raises some practical concerns. To our knowledge, all existing methods implicitly assume that the LD structures between the two association samples are identical, which is at best questionable when the two sets of association data are collected from different cohorts. In general, there is a lack of empirical evaluation of how different levels of mismatching between LD structures affect colocalization analysis outcomes. To address this issue, we design a simulation experiment utilizing multi-population genetic data to quantify the effects of mismatching LD patterns on colocalization analysis with two-sample designs.

We take the genetic data from the GEUVADIS project, which consists of samples from four European populations—CEPH (CEU), Toscani (TSI), British (GBR), and Finnish (FIN)—and one African population, Yoruban (YRI) (Figure 4).[24] We select the SNPs located within a 200 kb *cis*-region from 6,977 protein-coding and lincRNA genes, each of which contains at least 500 candidate SNPs. For each gene, we first sample the causal eQTL and GWAS SNPs from its candidate *cis*-SNPs. We then simulate a single eQTL dataset using the FIN data. We subsequently generate 5 GWAS datasets using the genotype data and predetermined GWAS association status for all 5 population groups. Note that the LD patterns are perfectly matched for the Finnish population for GWAS and eQTL analysis, which forms a baseline for evaluating the effects of LD mismatching. Additional simulation details are provided in Section 6 of the supplemental material and methods.

We analyze the five pairs of eQTL-GWAS data using fastENLOC. Our comparisons focus on the enrichment estimates, false-positive colocalization findings, and power. The results are summarized in Table 3. In all cases, we do not observe any inflation of false-positive colocalization findings—the false discovery rates are properly controlled in all datasets. The impact of LD mismatch is reflected by the under-estimation of the enrichment parameters and the diminished power, especially in noting that the power of GWAS discovery (which is perfectly correlated with estimated p_γ) is not substantially different in all populations.

Table 3. LD mismatch impact on enrichment estimation, FDR, and power

Datasets	$\widehat{\alpha}_1$	$\widehat{p}_\gamma$	Realized FDR	Power
FIN versus FIN	4.076	1.25×10^{-3}	0.029 (0.036)	0.129 (0.129)
FIN versus GBR	3.964	1.23×10^{-3}	0.023 (0.027)	0.102 (0.103)
FIN versus TSI	3.935	1.27×10^{-3}	0.023 (0.028)	0.101 (0.099)
FIN versus CEU	3.842	1.10×10^{-3}	0.030 (0.047)	0.075 (0.077)
FIN versus YRI	3.438	1.19×10^{-3}	0.001 (0.006)	0.065 (0.067)

The enrichment estimates ($\widehat{\alpha}_1$) and from all combinations of eQTL and GWAS datasets we used for this analysis. The estimated frequency of GWAS hits ($\widehat{p}_\gamma$) reflects the GWAS power of each GWAS dataset. (Note that true $p_\gamma = 1.92\times10^{-3}$.) The quantities in parentheses show the realized FDR and power when using the true enrichment parameter in the colocalization analysis.

In the extreme case of mismatch, i.e., the analysis of YRI GWAS data and FIN eQTL data, we find that the enrichment parameter α_1 is most severely underestimated, and the resulting power is reduced to 50% of the perfectly matching association data (i.e., FIN GWAS and FIN eQTL). We also note that the underestimation of the enrichment parameter only explains a small proportion of the loss: even when the true enrichment parameter is used, the power of colocalization analysis from analyzing the YRI GWAS data remains significantly lower than the other European datasets. Within the European populations, the effects of LD mismatch on colocalization analysis are also noticeable. The comparison within the European populations may also be complicated by differences in sample sizes, where TSI and CEU have the largest (n = 92) and smallest (n = 78) sample sizes, respectively. The sample size difference is directly linked to the power of GWAS discovery.

Overall, in relative terms, our observation suggests that the power loss suffered from the LD mismatching is qualitatively less severe than from the imperfect power of individual association analysis—as long as the eQTL and the GWAS samples are from reasonably close populations. In addition, the power loss caused by LD mismatching may be compensated by increased power in single-trait association analysis.

Colocalization analysis of 4,091 GWAS datasets and GTEx eQTL data

To provide a comprehensive summary of colocalization analysis using the current available GWAS and eQTL data, we analyze 4,091 complex trait datasets and the final release of the GTEx data (v.8) from 49 tissues.[6,18] In total, we perform colocalization analysis on 200,459 trait-tissue pairs using fastENLOC. The biological implications from the colocalization analysis, coupled with PrediXcan analysis,[4] have been reported and discussed in Pividori et al.[6] In this section, we focus on the technical aspect of the colocalization analysis and provide a high-level summary of the colocalization results for a wide range of complex traits with currently available GWAS and eQTL datasets. The fastENLOC output for all trait-tissue combinations can be downloaded using the URLs in Table S1. Additional details of data processing and analysis are given in Section 7 of the supplemental material and methods.

We first examine the empirical distribution of the enrichment estimates, $\widehat{\alpha}_1$, over the 200,459 trait-tissue pairs. The histogram in Figure 5 shows the empirical distribution of the enrichment estimates. We observe a clear bimodal distribution: the estimates from the vast majority of the trait-tissue pairs are close to 0, and there is also a noticeable peak centered around $\alpha_1 = 4$. Upon close inspections, we find the vast majority of complex traits with near 0 eQTL enrichment have few significant GWAS hits, which may be attributed to the lack of power in the corresponding studies.

Next, we inspect the significant findings from the analysis of each individual trait-tissue pair. Based on the enrichment analysis results, we compute the expected number of colocalization sites for each trait-tissue pair. Additionally, we identify high-confidence colocalization sites at the 5% FDR level based on the output of RCP values using the Bayesian FDR control procedure. In total, 15,975 sites pass this type I error control threshold in all trait-tissue pairs. We consider the trait-tissue pairs with more than 50 expected colocalization sites as "well powered." For this set of trait-tissue pairs, we compare the expected colocalization sites and the identified high-confidence sites at the 5% FDR level (Figure 6). The average rej-to-exp ratio in this set is 10.7% (median = 9.43%). This result indicates that the current colocalization analysis is (severely) underpowered for most trait-tissue pairs.

Discussion

In this paper, we have systematically explored both the analytical and the practical factors that impact the performance of probabilistic colocalization analysis for a molecular and a complex trait. We identify a single analytical factor, i.e., the specification of prior enrichment levels, that can lead to a significant inflation of false-positive findings, and we recommend estimating the critical enrichment parameters directly from the data. On the other hand, we find that a combination of analytical and practical factors, including modeling considerations for AH, LD mismatch, and imperfect power in association analyses,

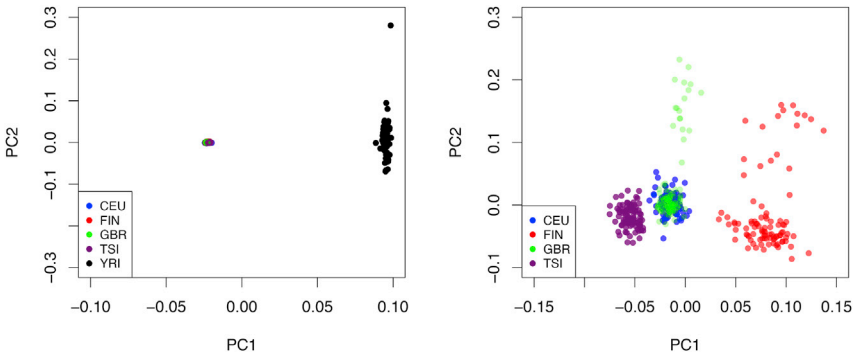

Figure 4. Population structures represented by PCA plots in GEUVADIS data
The left panel show the PCA plots (PC1 versus PC2) with the samples from all populations. The right panel shows the PCA plots using European samples only. Based on these plots, we expect maximum LD mismatch between YRI and any European population.

could severely diminish the power of SNP-level colocalization discoveries. As a result, current approaches often fail to identify the majority of colocalization signals in practical applications, even when they are appropriately applied. We argue that understanding the promise and limitations of the current state-of-the-art is critical for the practitioners to properly anticipate and correctly report their findings. For colocalization analysis of currently available molecular QTL and GWAS data, we may need to embrace the noticeable discrepancy between "expected colocalized signals" and the actual identified "significant colocalization findings."

There are many ways in which we can improve the power of existing colocalization methods based on our findings in this paper. For example, analytical strategies in improving the enrichment estimation, applying better fine-mapping methods, and explicitly modeling varying LD patterns across datasets will most likely result in enhanced power. Nevertheless, we suspect that improving the quality of the GWAS and molecular QTL datasets should have a more direct and visible impact. We note that most existing molecular QTL studies are limited by the high-throughput phenotyping cost and have modest sample sizes and relatively high experimental noise. The state-of-the-art eQTL annotations generated by the GTEx project are derived from bulk tissues of <1,000 samples. The current technology advancement, e.g., applying single-cell technology for molecular QTL mapping, combined with proven statistical strategies for data aggregation, e.g., a meta-analysis of molecular QTLs, could significantly enhance colocalization discoveries.

Another promising direction for improving colocalization analysis is to incorporate additional genomic information. This can be achieved by expanding the current prior model $\Pr(\gamma = 1, d = 1)$ to $\Pr(\gamma = 1, d = 1|$ additional genomic annotations$)$. The additional genomic features can be obtained from other relevant molecular phenotype studies, e.g., studies of methylation, chromatin accessibility, and histone modification. This added information provides a more relevant "local" genomic context for each candidate locus, hence improving both the sensitivity and specificity of the colocalization analysis.

Although our discussions in this paper are exclusively illustrated using two complex traits, the general principles extend to the analysis of multiple traits.[25,26] All of the analytical and practical factors that we have discussed impact SNP-level colocalization analysis for more than

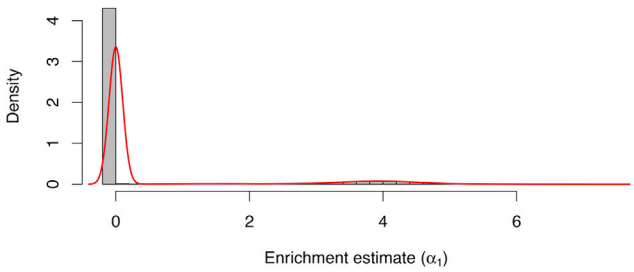

Figure 5. Enrichment estimates from all tissue-trait pairs in the phenomexcan analysis
The histogram displays a bimodal distribution, with a sharp peak at $\alpha_1 = 0$ and a wide peak centered around $\alpha_1 = 4$.

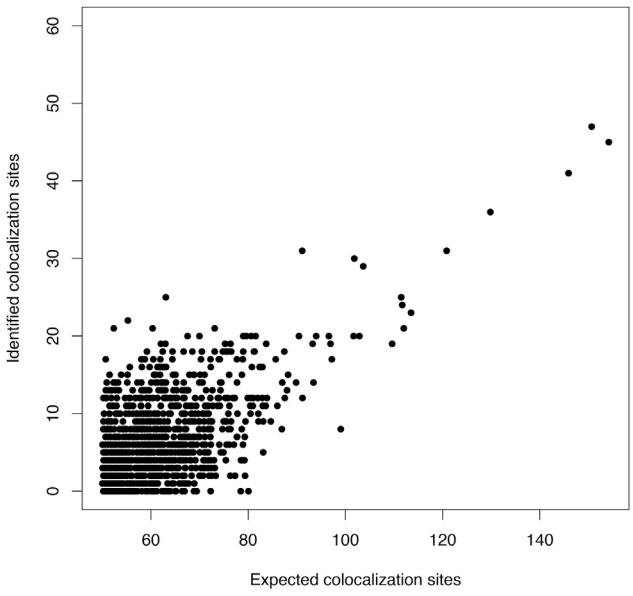

Figure 6. The comparison of the calculated expected colocalization sites and the detected high-confidence sites among well-powered trait-tissue pairs
This apparent lack of power falls in line with our findings from our simulations.

two traits. If not adequately dealt with, the resulting adverse effects can be even more severe. For example, both classes of false-negative errors discussed in the section of realistic power assessment increase with more traits considered. Additionally, the existing enrichment estimation procedure via multiple imputation does not scale well regarding a large number of traits (i.e., ≥ 5). Therefore, extending the best practice of colocalization analysis from two traits to multiple traits remains a critical challenge.

Colocalization analysis is also connected to other types of integrative analysis approaches, e.g., transcriptome-wide association studies (TWASs). In analyzing eQTL and GWAS data, a TWAS utilizes the same input data sources as the colocalization analysis. However, its results have some unique causal implications provided that a set of assumptions is met.[27] Despite the difference in their theoretical origins, positive findings from the two analyses can be driven by similar signals. Recent studies[6] find that integrating colocalization analysis into a TWAS can improve its sensitivity and specificity. More generally, the two prevailing types of integrative analysis approaches can complement each other. Thus, further exploration of their connections and distinctions becomes an important future direction.

Data and Code Availability

The harmonized summary statistics from 4,091 complex trait GWASs and the multi-tissue eQTL annotations derived from the GTEx (v.8) data are made publicly available (see web resources). The source code and scripts for data generation and data analysis in the numerical experiments (fastENLOC) are available in the Github repository.

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2020.11.012.

Acknowledgments

We thank the GTEx consortium. This work is supported by the NIH grants R35GM138121, R01GM109215.

Declaration of interests

The authors declare no competing interests.

Received: July 7, 2020
Accepted: November 17, 2020
Published: December 11, 2020

Web resources

coloc, https://cran.r-project.org/web/packages/coloc/index.html
eCAVIAR, http://genetics.cs.ucla.edu/caviar/
fastENLOC, https://github.com/xqwen/fastenloc
fastENLOC analysis of 4,091 complex traits and GTEx v8 data (detailed results), https://tinyurl.com/y6cx9ovm
fastENLOC data generating and analysis scripts/code for the numerical experiments, https://github.com/xqwen/fastenloc/tree/master/promise_and_limitation_paper/
GEUVADIS data, https://www.internationalgenome.org/data-portal/data-collection/geuvadis
GTEx, https://www.gtexportal.org/home/
Harmonized summary statistics for 4,091 complex-trait GWAS, https://zenodo.org/record/3629742/files/harmonized_imputed_gwas.tar
Multi-tissue eQTL annotations derived from the GTEx (v8) data, https://tinyurl.com/yyfser9a
phenomeXcan, http://apps.hakyimlab.org/phenomexcan/
Summary results from fastENLOC analysis of 4,091 complex traits and GTEx v8 data, https://doi.org/10.5281/zenodo.3530669

References

1. Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., and Cox, N.J. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. *6*, e1000888.
2. Nica, A.C., Montgomery, S.B., Dimas, A.S., Stranger, B.E., Beazley, C., Barroso, I., and Dermitzakis, E.T. (2010). Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. PLoS Genet. *6*, e1000895.
3. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W., Jansen, R., de Geus, E.J., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. *48*, 245–252.
4. Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., Nicolae, D.L., Cox, N.J., Im, H.K.; and GTEx Consortium (2015). A gene-based association method for mapping traits using reference transcriptome data. Nat. Genet. *47*, 1091–1098.
5. Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A.N., Knowles, D.A., Golan, D., Ermel, R., Ruusalepp, A., Quertermous, T., Hao, K., et al. (2019). Opportunities and challenges for transcriptome-wide association studies. Nat. Genet. *51*, 592–599.
6. Pividori, M., Rajagopal, P.S., Barbeira, A., Liang, Y., Melia, O., Bastarache, L., Park, Y., Consortium, G., Wen, X., and Im, H.K. (2020). PhenomeXcan: Mapping the genome to the phenome through the transcriptome. Sci. Adv. *6*, eaba2083.
7. Franceschini, N., Giambartolomei, C., de Vries, P.S., Finan, C., Bis, J.C., Huntley, R.P., Lovering, R.C., Tajuddin, S.M., Winkler, T.W., Graff, M., et al.; MEGASTROKE Consortium (2018). GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. Nat. Commun. *9*, 5141.
8. Wang, G., Sarkar, A.K., Carbonetto, P., and Stephens, M. (2019). A simple new approach to variable selection in regression, with application to genetic fine-mapping. J. R. Stat. Soc. B *82*, 1273–1300.
9. Taylor, K., Davey Smith, G., Relton, C.L., Gaunt, T.R., and Richardson, T.G. (2019). Prioritizing putative influential genes in cardiovascular disease susceptibility by applying tissue-specific Mendelian randomization. Genome Med. *11*, 6.
10. Chun, S., Casparino, A., Patsopoulos, N.A., Croteau-Chonka, D.C., Raby, B.A., De Jager, P.L., Sunyaev, S.R., and Cotsapas, C. (2017). Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. Nat. Genet. *49*, 600–605.

11. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. *10*, e1004383.
12. Hormozdiari, F., van de Bunt, M., Segrè, A.V., Li, X., Joo, J.W.J., Bilow, M., Sul, J.H., Sankararaman, S., Pasaniuc, B., and Eskin, E. (2016). Colocalization of gwas and eqtl signals detects target genes. Am. J. Hum. Genet. *99*, 1245–1260.
13. Wen, X., Pique-Regi, R., and Luca, F. (2017). Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. PLoS Genet. *13*, e1006646.
14. Wallace, C. (2020). Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. PLoS Genet. *16*, e1008720.
15. Veyrieras, J.-B., Kudaravalli, S., Kim, S.Y., Dermitzakis, E.T., Gilad, Y., Stephens, M., and Pritchard, J.K. (2008). High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet. *4*, e1000214.
16. Pickrell, J.K. (2014). Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. Am. J. Hum. Genet. *94*, 559–573.
17. Wen, X., Lee, Y., Luca, F., and Pique-Regi, R. (2016). Efficient integrative multi-snp association analysis via deterministic approximation of posteriors. Am. J. Hum. Genet. *98*, 1114–1129.
18. Aguet, F., Barbeira, A.N., Bonazzola, R., Brown, A., Castel, S.E., Jo, B., Kasela, S., Kim-Hellmuth, S., Liang, Y., Oliva, M., et al. (2019). The gtex consortium atlas of genetic regulatory effects across human tissues. bioRxiv. https://doi.org/10.1101/787903.
19. GTEx Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348*, 648–660.
20. Hormozdiari, F., Kostem, E., Kang, E.Y., Pasaniuc, B., and Eskin, E. (2014). Identifying causal variants at loci with multiple signals of association. Genetics *198*, 497–508.
21. Benner, C., Spencer, C.C., Havulinna, A.S., Salomaa, V., Ripatti, S., and Pirinen, M. (2016). FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics *32*, 1493–1501.
22. Schaid, D.J., Chen, W., and Larson, N.B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat. Rev. Genet. *19*, 491–504.
23. Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. Nat. Rev. Genet. *20*, 467–484.
24. Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A., Monlong, J., Rivas, M.A., Gonzàlez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al.; Geuvadis Consortium (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature *501*, 506–511.
25. Giambartolomei, C., Zhenli Liu, J., Zhang, W., Hauberg, M., Shi, H., Boocock, J., Pickrell, J., Jaffe, A.E., Pasaniuc, B., Roussos, P.; and CommonMind Consortium (2018). A Bayesian framework for multiple trait colocalization from summary association statistics. Bioinformatics *34*, 2538–2545.
26. Foley, C.N., Staley, J.R., Breen, P.G., Sun, B.B., Kirk, P.D., Burgess, S., and Howson, J.M. (2019). A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. bioRxiv. https://doi.org/10.1101/592238.
27. Chen, Y., Quick, C., Yu, K., Barbeira, A., Luca, F., Pique-Regi, R., Im, H.K., Wen, X.; and GTEx Consortium (2019). Investigating tissue-relevant causal molecular mechanisms of complex traits using probabilistic twas analysis. bioRxiv. https://doi.org/10.1101/808295.

The American Journal of Human Genetics, Volume *108*

Supplemental Data

Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations

Abhay Hukku, Milton Pividori, Francesca Luca, Roger Pique-Regi, Hae Kyung Im, and Xiaoquan Wen

Supplemental Figures

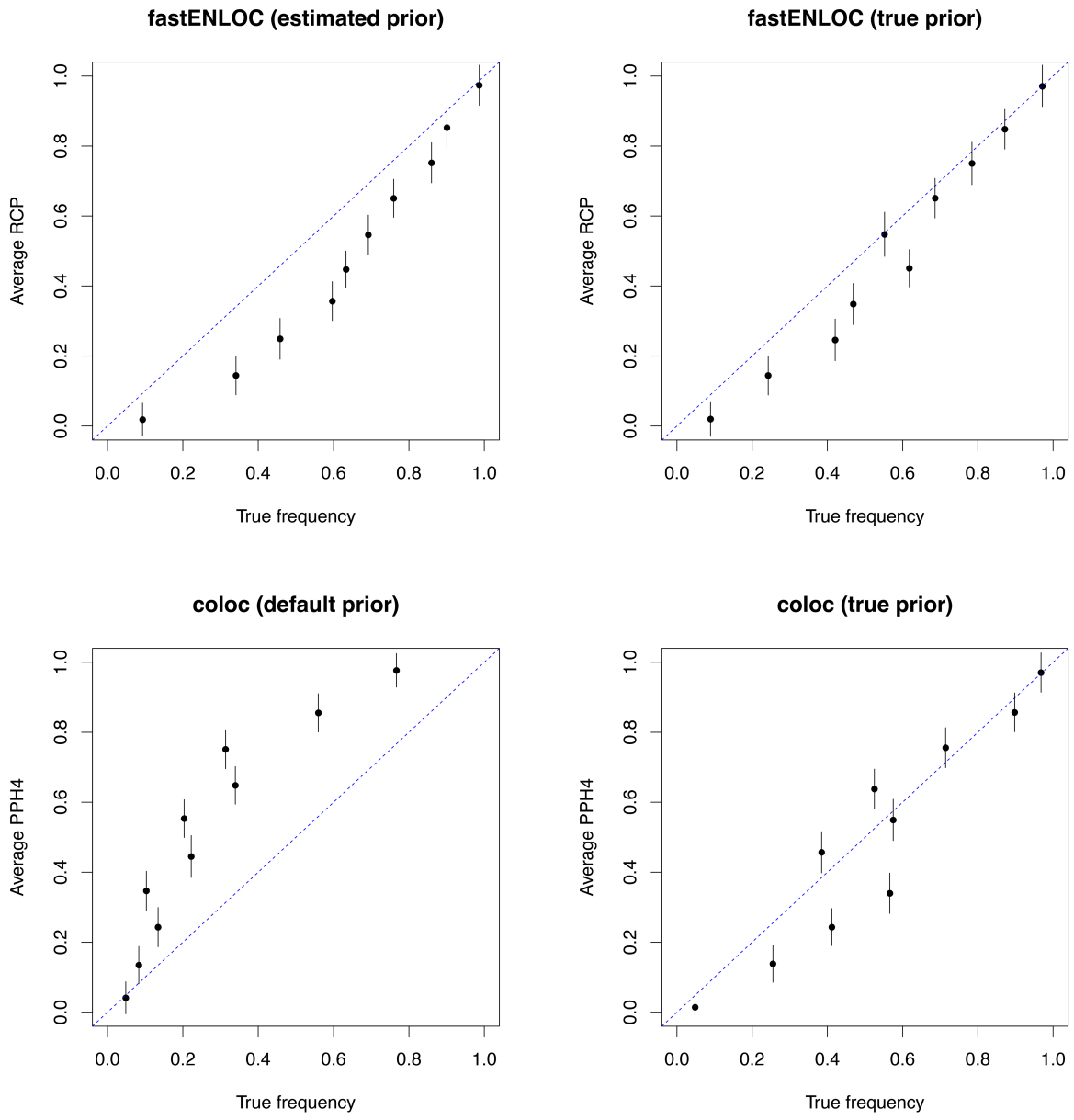

Figure S1: **Calibration of reported regional colocalization probabilities** The error bars represent 95% confidence intervals. The RCPs from the fastENLOC are conservative in comparison to the corresponding frequencies. The conservativeness is partially explained by the conservative estimation of the enrichment priors. In the case of *coloc*, the default prior leads to severely anti-conservative results. Even when the true priors are applied, the *coloc* results still show anti-conservativeness in some frequency bins.

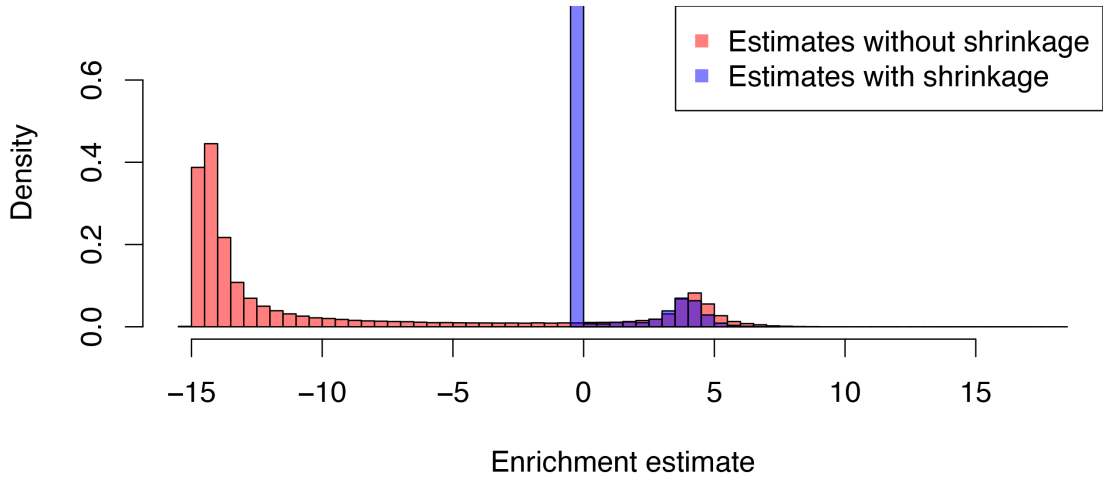

Figure S2: **Enrichment estimates with and without shrinkage from all tissue-trait pairs in the phenomexcan analysis**

Supplemental Material and Methods

1 Overview of existing probabilistic colocalization approaches

coloc *coloc* is first proposed in [1]. It makes the "one causal variant" (OCV) assumption for each trait in each candidate region. Under this specific assumption, it enumerates five possible distinct models/hypotheses within a region. These models are: 1) there are no causal variants from either trait (H_0); 2) there is only a causal eQTL variant but no causal GWAS variant (H_1); 3) there is only a causal GWAS variant but no causal eQTL variant (H_2); 4) there are distinct causal SNPs for both eQTL and GWAS, (H_3); and 5) there is a colocalized signal (H_4). For each hypothesis, the corresponding posterior probability is computed by considering all compatible latent association status configurations from GWAS and eQTL data via Bayesian model averaging (BMA). The colocalization within each region is quantified by the posterior probability of H_4 (PPH4).

Under the OCV assumption, the likelihood functions for all possible association scenarios can be analytically computed from the single-SNP association statistics, e.g., z-statistics, without the need of explicit modeling of LD [2]. The implementation of coloc allows user-specified priors for SNP i, i.e.,

$$p_1 = \Pr(d_i = 1, \gamma_i = 0)$$

$$p_2 = \Pr(d_i = 0, \gamma_i = 1)$$

$$p_{12} = \Pr(d_i = 1, \gamma_i = 1)$$

Recall that $d_i = 1$ indicates that SNP i is a causal eQTL and $\gamma_i = 1$ indicates that SNP i is a casual

GWAS variant. By default, p_1 and p_2 are set to 10^{-4}, and p_{12} is set to 10^{-5}, which corresponds to

$$\begin{aligned} p_d &= 1.1 \times 10^{-4} \\ \alpha_0 &= -9.23 \\ \alpha_1 &= 6.91 \end{aligned} \tag{1}$$

in the formulation of fastENLOC priors. From the observed data shown in Figure 5 of the main text, α_1 close to 7 represents a very high enrichment level and is rare in practice.

It is worth pointing out that *coloc* requires only summary-level statistics from single-SNP association analysis as input and performs fine-mapping internally assuming OCV.

eCAVIAR eCAVIAR, proposed in [3], is built upon a sophisticated Bayesian multi-variate fine-mapping algorithm, CAVIAR [4]. The CAVIAR algorithm enables computation of $\Pr(d_i = 1 \mid \text{eQTL data})$ and $\Pr(\gamma_i = 1 \mid \text{GWAS data})$ based only on summary-level statistics from single-SNP association analysis and an LD matrix between candidate SNPs. Based on the marginal fine-mapping result, eCAVIAR computes a SNP-level colocalization posterior probability, or CLPP, by

$$\begin{aligned} &\Pr(d_i = 1, \gamma_i = 1 \mid \text{eQTL data, GWAS data}) \\ &= \Pr(d_i = 1 \mid \text{eQTL data}) \Pr(\gamma_i = 1 \mid \text{GWAS data}). \end{aligned} \tag{2}$$

As we previously discussed, this equation corresponds to the special case that assumes $\alpha_1 = 0$, i.e., there is no enrichment of eQTL signals in causal GWAS hits. Based on the evidence observed from the real data, this assumption can be overly conservative.

Notably, eCAVIAR only provides a colocalization quantification at the SNP level but not for a genomic region. This is also a distinction between eCAVIAR and other approaches.

fastENLOC/ENLOC ENLOC is first proposed in [5]. An improved version, fastENLOC, with accelerated computation and enhanced precision, is recently described in [6]. A distinct feature of ENLOC/fastENLOC is its embedded function of estimating the enrichment level of eQTLs in GWAS

hits, i.e., (α_0, α_1), directly from the data. By treating the latent vectors of causal association status, $\boldsymbol{\gamma}$, and $\boldsymbol{d}$, as missing data, ENLOC employs a multiple imputation strategy and an EM algorithm for enrichment estimation. Given the enrichment estimates, ENLOC applies an empirical Bayes framework to compute the SNP-level colocalization probability strictly based on the Bayes rule, in which Eqn. (2) of eCAVIAR becomes a special case.

fastENLOC also takes advantage of the concept of credible sets of independent association signals inferred from multi-variant fine-mapping analysis to compute a colocalization probability for each LD region. A Bayesian credible set consists of a group SNPs in LD that represent the same underlying association signal. By utilizing credible sets, fastENLOC ensures correct dependence structures in the multiple imputation scheme's sampling procedures for enrichment estimation. Additionally, the credible sets naturally form regional LD units for regional-level colocalization analysis. This is because, even if the exact colocalized variant is uncertain, we can evaluate the probability that *one* of the member SNPs in the credible set is causal for both GWAS and expression traits by Bayesian model averaging.

Similar to eCAVIAR, fastENLOC requires a dedicated fine-mapping algorithm to generate required probability quantifications of marginal association evidence. Currently, DAP [7], SuSiE [8] and FINEMAP [9] are suitable methods for preparing fastENLOC input, because they can report signal clusters/credible sets in addition to traditional SNP-level posterior inclusion probabilities (PIPs). Additionally, all three work with summary-level statistics similar to CAVIAR (and eCAVIAR), which requires summary statistics from single-SNP association testing and LD matrices.

Other approaches There are other available methods proposed for general colocalization analysis between multiple complex traits. The majority of methods in this category deal with a slightly different problem that does not require quantification of colocalization at the SNP-level. For example, both RTC (Regulatory Trait Concordance) [10] and JLIM (Joint likelihood mapping) [11] use formal hypothesis testing procedures to examine if a group of SNPs in LD contain both a causal eQTL and a causal GWAS hit. However, within the SNPs that are tightly linked, they cannot distinguish if the two signals are overlapping at a single SNP. That is, there is no stringent distinction between the H_3 and the H_4 models in the formulation of *coloc*. Another approach, SMR [12], is built upon the framework of instrumental

variable analysis and identifies SNPs that are associated with both GWAS and expression traits. The authors attempt to further distinguish colocalization (H_4) from linkage (i.e., H_3) by a hypothesis testing procedure named HEIDI [13]. Although the idea is intuitive, the HEIDI procedure is set up to claim colocalization by *accepting* the null hypothesis. As a result, rigorous quantification or proper control of false-positive findings is lacking.

Finally, we want to emphasize that there is strong uniformity across all three probabilistic colocalization approaches despite the difference in their implementations: they share the same mathematical foundation and inference principles. Particularly, eCAVIAR can be viewed as a special case of fastENLOC with a special set of pre-defined enrichment parameters; the *coloc* algorithm converges to fastENLOC if the same enrichment parameters are supplied, and the OCV assumption is satisfied.

2 Overview of enrichment estimation in fastENLOC

The enrichment estimation procedure used in fastENLOC is detailed in [5, 6]. Here we give a summary of the procedure, serving as a quick reference.

The fundamental difficulty in estimating enrichment of causal eQTLs in causal GWAS hits lies in the fact that the true association status for each SNP i, γ_i and d_i, are not observed. If they are, α_0 and α_1 can be straightforwardly estimated by a 2×2 contingency table, or equivalently, by a simple logistic regression. fastENLOC solves this problem by utilizing the posterior inclusion probabilities (PIPs) for all p SNPs obtained from separate Bayesian association analysis of the complex and molecular traits, respectively (i.e., $\{\Pr(\gamma_i \mid \text{GWAS data}) : i = 1, \ldots, p\}$ and $\{\Pr(d_i \mid \text{eQTL data}) : i = 1, \ldots, p\}$.

[14, 7] propose an EM algorithm to estimate α_0 and α_1 when $\{d_i\}$ is observed (e.g., d_i is treated as an observed SNP-level binary annotation) and $\{\Pr(\gamma_i \mid \text{GWAS data})\}$ is available. [5] extends this idea, and propose sampling d_i from the corresponding posterior distributions, $\Pr(d_i \mid \text{eQTL data})$, via a multiple imputation scheme. Briefly, it first creates M sets of imputed $\{d_i\}$ values and runs the EM algorithm for each $(\{\Pr(\gamma_i \mid \text{GWAS data})\}, \{d_i\})$ pairs. The resulting estimates of (α_0, α_1) from M imputed datasets are combined using the multiple imputation formula. Guided by the principles of

multiple imputation, M is set to 25 by default [5].

It may seem intuitive to simply sample γ_i from $\Pr(\gamma_i \mid \text{GWAS data})$ and d_i from $\Pr(d_i \mid \text{eQTL data})$ independently, then fill in a 2×2 contingency table, and obtain the (α_0, α_1) estimate. But such a procedure completely ignores the potential dependence relationship between γ_i and d_i (i.e., the sampling processes for γ_i and d_i are independent), and it can lead to severe underestimation of α_1.

In fastENLOC, the sampling process of $\{d_i\}$ is sped up by taking advantage of the signal clusters reported by the software DAP-G. The signal clusters enable more efficient and precise samplings of independent causal associations. The credible sets output from methods SuSIE and FINEMAP can also be used to achieve similar performance improvement. Additionally, fastENLOC applies a shrinkage estimator that effectively reduces the variance of the estimate of α_1. The details of this shrinkage estimator are explained in Section 7.1 of this document.

3 Numerical experiment to examine sensitivity of prior impact

3.1 Constraints in prior model

In this numerical experiment, we fix the values of p_γ, p_d, and varying the value of α_1. Here, we show that α_0 is determined by $(p_\gamma, p_d, \alpha_1)$.

Note that

$$\begin{aligned} p_\gamma &= \Pr(\gamma = 1 \mid d = 0)\,(1 - p_d) + \Pr(\gamma = 1 \mid d = 1)\,p_d \\ &= \frac{\exp(\alpha_0)}{1 + \exp(\alpha_0)}\,(1 - p_d) + \frac{\exp(\alpha_0 + \alpha_1)}{1 + \exp(\alpha_0 + \alpha_1)}\,p_d \end{aligned} \tag{3}$$

Thus, α_0 can be solved from the above equation given p_γ, p_d, and α_1. Furthermore, an accurate approximation of α_0 can be analytically computed by noting that, in practice, $\Pr(\gamma = 1 \mid d = 0) \ll 1$ and $\Pr(\gamma = 1 \mid d = 1) \ll 1$ (i.e., GWAS hits for complex traits are sparse). It follows that

$$(1 + \exp(\alpha_0))\; p_\gamma \approx \exp(\alpha_0)\,(1 - p_d) + \exp(\alpha_0 + \alpha_1)\,p_d \tag{4}$$

and

$$\alpha_0 \approx \log\left(\frac{p_\gamma}{1 + p_d \exp(\alpha_1) - p_d - p_\gamma}\right) \tag{5}$$

3.2 Computing SNP-level Colocalization Probability

Here we derive the SNP-level colocalization probability (SCP) for an independent SNP given the posterior association probabilities, $q_d := \Pr(d = 1 \mid \text{eQTL data})$ and $q_\gamma = \Pr(\gamma = 1 \mid \text{GWAS data})$ for association analysis of individual traits. This result is used in the numerical experiment.

By the Bayes rule, the Bayes factor for eQTL association can be computed by

$$\begin{aligned}\text{BF}_d &= \frac{\Pr(\text{eQTL data} \mid d = 1)}{\Pr(\text{eQTL data} \mid d = 0)} \\ &= [q_d/(1 - q_d)] \Big/ [p_d/(1 - p_d)]\,.\end{aligned}$$

Similarly, the Bayes factor for GWAS association at the target SNP is given by

$$\text{BF}_\gamma = [q_\gamma/(1 - q_\gamma)] \Big/ [p_\gamma/(1 - p_\gamma)]\,.$$

Assuming a two-sample design, it follows that

$$\begin{aligned}&\Pr(\text{GWAS data, eQTL data} \mid d, \gamma) \\ &= \Pr(\text{eQTL data} \mid d)\ \Pr(\text{GWAS data} \mid \gamma)\end{aligned} \tag{6}$$

Applying the Bayes rule to compute SCP yields

$$\begin{aligned}&\Pr(d = 1, \gamma = 1 \mid \text{eQTL data, GWAS data}) \\ &= \frac{\Pr(d = 1, \gamma = 1)\text{BF}_d\text{BF}_\gamma}{\Pr(d = 0, \gamma = 0) + \Pr(d = 1, \gamma = 0)\text{BF}_d + \Pr(d = 0, \gamma = 1)\text{BF}_\gamma + \Pr(d = 1, \gamma = 1)\text{BF}_d\text{BF}_\gamma},\end{aligned} \tag{7}$$

where

$$\begin{aligned}
\Pr(d=1,\gamma=1) &= \frac{\exp(\alpha_0+\alpha_1)}{1+\exp(\alpha_0+\alpha_1)}\, p_d \\
\Pr(d=0,\gamma=0) &= \frac{1}{1+\exp(\alpha_0)}\,(1-p_d) \\
\Pr(d=1,\gamma=0) &= \frac{1}{1+\exp(\alpha_0+\alpha_1)}\, p_d \\
\Pr(d=0,\gamma=1) &= \frac{\exp(\alpha_0)}{1+\exp(\alpha_0+\alpha_1)}\,(1-p_d)
\end{aligned} \tag{8}$$

3.3 Experiment details

To isolate the impact of prior specification on colocalization quantifications, we consider a single SNP with various levels of association evidence from GWAS and eQTL analysis. In this experiment, we assume that priors for GWAS and eQTL associations are pre-determined at $p_d = 10^{-3}$ and $p_\gamma = 5\times10^{-5}$, respectively. The values roughly reflect the prevalence of eQTL and GWAS hits in practice. Given p_d and p_γ, the value of α_0 is determined by α_1 by Eqn. (5).

In this experiment, we consider only one independent genetic variant. We use the posterior inclusion probabilities, q_d and q_γ, to characterize each trait's association evidence. They are naturally scaled in $[0,1]$ and convenient to represent different association evidence magnitudes. (In comparison, because of the difference in the priors of marginal associations for different traits, the direct comparison of Bayes factors is not as straightforward.) In this experiment, we consider the values of q_d and q_γ $= 0.90, 0.50, 0.05$ to represent strong, modest, and weak evidence for association, respectively.

With $p_d, p_\gamma, q_d, q_\gamma$, and α_1 all fully specified, the SNP-level colocalization probability (SCP) can analytically computed by Eqn. (7).

To inspect the sensitivity of SCP with respect to the enrichment value, we vary α_1 over a wide range covering all values typically seen in data, from -6 to 8, and compute the corresponding SCP values.

4 Simulation to inspect effect of allelic heterogeneity

To investigate the impact of allelic heterogeneity in colocalization analysis, we design a simulation scheme with real genotype data obtained from the GTEx whole blood data. To isolate the impacting factors and simplify interpretations, we use the same genotype data to simulate gene expressions and complex traits. Thus, the LD patterns are perfectly matching between the two datasets. We use the genotypes from 400 individuals from the GTEx whole blood samples for this simulation. We create 4 sets of combined GWAS and eQTL datasets, corresponding to the 4 different scenarios described in the main text. These scenarios enable us to investigate potential false positive and false negative findings by different AH assumptions. We simulate gene expressions for 1,000 genes for each scenario and the corresponding complex trait data for 400 individuals using standard linear models. We particularly fix all gene expression effect sizes to 2.5, and the complex trait effect sizes to 1.5. The residual error variance is set to 1 for all simulations. Our consideration for this simulation setting is to ensure the generated association signals are strong. Hence the subsequent colocalization analysis can be robust to prior specifications (based on what we observe from the previous numerical experiment). We pool the 4 simulated scenarios into 2 different datasets. The first dataset consists of scenarios 1, 2, and 3, while the second dataset consists of scenarios 1, 2, and 4. Thus, each dataset has exactly 3,000 genes for analysis.

For fastENLOC analysis, we first run a multi-SNP association analysis of simulated GWAS and eQTL data using DAP-G for all simulated genes. The enrichment parameters are estimated from the data. For *coloc*, we directly supply the true priors and the summary statistics from the corresponding single-SNP analysis.

5 Simulation to benchmark power of practical colocalization analysis

To benchmark and investigate the performance of probabilistic colocalization analysis, we simulate complex trait and gene expression data that resemble observed data in practice. For this experiment, we use the real genotype data across 20,000 genes from 400 participants of the GTEx project. To

ensure the LD patterns in genotypes are perfectly matched in eQTL and GWAS data, we again use the identical genotype data to independently simulate expression and complex traits.

For each candidate gene, we select a fixed number of 1,500 *cis*-SNPs whose minor allele frequencies > 0.03. We use the following linear model to simulate the expression level of gene i for individual k,

$$y_{ik} = \mu_i + \sum_j \beta_{ij} g_{jk} + e_{ik},\ e_{ik} \sim \mathrm{N}(0,1), \tag{9}$$

where g_{jk} and β_{ij} represent the genotype and the genetic effect for SNP j ($j = 1, 2, ..., 1500$). Particularly, β_{ij} is independently drawn from a mixture distribution,

$$\beta_{ij} \sim \pi_0 \delta_0 + (1 - \pi_0)\mathrm{N}(0,1). \tag{10}$$

That is, with probability $\pi_0 = 1 - 3/1500 = 0.998$, the genetic effect of SNP j on expression is exactly 0; and with probability 0.002, the SNP has a non-zero random effect drawn from $\mathrm{N}(0,1)$ distribution. On average, the simulation scheme yields 3 causal eQTL SNPs per *cis* region.

To simulate the causal GWAS hits, we conduct an independent Bernoulli trial on each candidate SNP: if a SNP is not an eQTL SNP, its probability of being a causal GWAS SNP is set to 1/1500; otherwise, the corresponding probability increases to 0.035, or equivalently, $\alpha_1 = 4$ (which is similar to the enrichment level of blood eQTLs in causal GWAS hits of lipids traits [5]). Subsequently, we again simulate the complex trait using a multiple linear regression model. Specifically, the genetic effect of each causal GWAS SNP is drawn from the distribution $\mathrm{N}(0,1)$, and the residual error is also simulated from the same distribution.

Overall, across 20,000 genes, this scheme generates 59,937 causal eQTL SNPs and 22,054 casual GWAS hits. There are 2,103 instances that the causal variants for both traits are overlapped.

5.1 Calibration of Regional Colocalization Probabilities

In this simulation, we also examine the calibration of the regional colocalization probabilities (RCPs) reported by both fastENLOC and *coloc*. The concept of calibration refers to where reported Bayesian posterior probabilities represent the frequencies of events in multiple independent experiments, a frequentist property. To this end, we sort all RCP values into evenly-divided probability bins (e.g., $[0, 0.1), [0.1, 0.2), ..., [0.9, 1.0]$). Within each probability bin, we then compute the fractions of reported sites that truly harbor a colocalized signal. If the RCP values are calibrated, we expect that the mean RCP from each pre-defined bin aligns with the corresponding fraction of true colocalized signals. Overall, we find that the calibration of the computed RCPs is overly conservative (Figure S1). Applying the true enrichment parameters improves the overall calibration but does not completely resolve the issue of conservativeness, especially for the RCPs in the low to the modest range. This is also in agreement with our main conclusions from the power and type I errors of the colocalization analysis in Table 2 of the main text. We also conduct a similar analysis for the gene-level posterior probability of colocalization from the *coloc* analysis, which confirms that inaccurate prior information and/or ignoring potential AH could lead to an anti-conservative assessment of colocalization probabilities.

6 Simulation to investigate effect of LD mismatch

We construct these simulations based on the design of the GEUVADIS project, which studies the eQTLs across 5 different populations (FIN, CEU, GBR, TSI, and YRI). We use the real genotype data for the five populations that are originally genotyped in the 1000 Genome project Phase I [15]. We select 6,977 genes from the GEUVADIS project that contain at least 500 SNPs in a 200 KB *cis*-region. Note that we keep all SNPs in all studied populations, even though some are monomorphic or extremely rare in specific population groups. We apply a similar scheme as the previous simulation (Section 5) to sample the causal eQTL and GWAS SNPs based on independent Bernoulli trials and pre-specified enrichment parameters. We adjust the enrichment parameter α_1 from 4.0 to 4.5 to increase the instances of colocalizations and compensate for reduced numbers of *cis*-SNPs.

Given the causal eQTL and GWAS status, we simulate a single set of eQTL data using the genotypes from the FIN population. Next, we simulate 5 different sets of complex trait data for all populations. The sample sizes for each population are 89 (FIN), 78 (CEU), 84 (GBR), 92 (TSI), and 77 (YRI), respectively. The phenotype simulation is based on multiple linear regression models, and the residual error variance for each corresponding linear model is also set to 1. To compensate for the reduced sample size (compared to our previous simulations), we draw the genetic effects for both eQTLs and GWAS hits from a random effect model with variance = 2. In total, the simulated datasets contain 20,987 eQTLs and 9,378 GWAS hits with 2,488 colocalized signals.

7 Colocalization analysis of 4,091 complex trait and GTEx eQTL data

The eQTL data from 49 tissues are generated from the final version (v8) of the GTEx project. The eQTL data are processed based on the protocols established by the GTEx consortium. The processing pipeline and related software packages are provided in [16]. We collect the summary statistics in the form of single-SNP testing z-scores from 4,091 GWAS of complex traits. These summary statistics are harmonized to be compatible to be analyzed with the GTEx data. In particular, additional z-scores are imputed according to LD patterns observed in the GTEx data using the software package LDpred [17]. The technical details on the pre-processing of the complex data are also documented in [16, 6].

The multi-SNP fine-mapping analysis of the eQTL in each tissue is performed using DAP-G [18], and the resulting probabilistic annotations of eQTLs are available in the GTEx portal. Due to a lack of individual-level data or precise LD information for the complex trait GWAS, we choose to employ the fine-mapping algorithm similar to *fgwas* implemented in the software package TORUS [7]. Briefly, TORUS segments the genome into 1 to 2 Mb wide LD blocks [19] and performs fine-mapping analysis within each LD block assuming a single causal GWAS hit. Although the assumption is imperfect, we are much less likely to observe multiple strong GWAS hits within a single LD block in practice.

7.1 Shrinkage estimation of enrichment parameter

We estimate the enrichment levels of eQTLs in the GWAS hits for each tissue-trait pair using fastENLOC. Particularly, we experiment with two different strategies for estimation: with and without applying shrinkage to enrichment estimates.

Our observation is that in many cases where strong evidence for colocalization events is lacking, the direct estimates of α_1 without shrinkage are often unstable. That is, the point estimates can be wildly positive or negative, and they are always associated with substantially large variances. The corresponding z scores, which measure the signal-to-noise ratio, are rarely statistically significant. In these scenarios, direct use of the point estimates while ignoring the large standard errors is potentially dangerous for downstream colocalization analysis. Hence, we conclude it is necessary to apply shrinkage to the estimate of α_1. Although it is known that the shrinkage estimates are biased toward 0, the tradeoff to stabilize the estimates is necessary for this context.

Here, we detail the shrinkage estimation procedure. Let $\tilde{\alpha}_1$ and v_1 denote the point estimate of α_1 and the corresponding variance directly obtained from the multiple imputation procedure of ENLOC. In the case of a lack of strong evidence of colocalization events, $\tilde{\alpha}_1$ can be highly unstable, i.e., $\tilde{\alpha}_1$ can take extreme positive or negative values, and v_1 is also extremely large. The phenomenon is similar to the enrichment estimation using a 2×2 contingency table where at least one cell counts $\to 0$. To stabilize the estimate in the empirical Bayes framework, we consider a shrinkage prior, $\mathrm{N}(0, 1/\lambda)$, for α_1. The resulting shrinkage estimate of the α_1 is given by

$$\hat{\alpha}_1 = \frac{\tilde{\alpha}_1}{1 + \lambda v_1}, \tag{11}$$

and the corresponding variance is

$$\mathrm{Var}(\hat{\alpha}_1) = \frac{v_1}{1 + \lambda v_1} \tag{12}$$

The shrinkage parameter λ_1 defines the strength of the shrinkage. As $\lambda_1 \to 0$, it follows that $\hat{\alpha}_1 \to \tilde{\alpha}_1$ and $\mathrm{Var}(\hat{\alpha}_1) \to v_1$, i.e., there is no effect of shrinkage. On the other extreme, as $\lambda \to \infty$, it follows that $\hat{\alpha}_1 \to 0$ and $\mathrm{Var}(\hat{\alpha}_1) \to 0$. The default implementation of fastENLOC set $\lambda = 1$.

We observe that the shrinkage estimates of the enrichment parameter are effectively stabilized. The comparison of $\tilde{\alpha}_1$ and $\hat{\alpha}_1$ from analyzing 4,091 complex traits and GTEx eQTL data are shown in Figure S2.

References

[1] Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS genetics* **10** (2014).

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

[3] Hormozdiari, F. *et al.* Colocalization of gwas and eqtl signals detects target genes. *The American Journal of Human Genetics* **99**, 1245–1260 (2016).

[4] Hormozdiari, F., Kostem, E., Kang, E. Y., Pasaniuc, B. & Eskin, E. Identifying causal variants at loci with multiple signals of association. *Genetics* **198**, 497–508 (2014).

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

[6] Pividori, M. *et al.* Phenomexcan: Mapping the genome to the phenome through the transcriptome. *BioRxiv* 833210 (2019).

[7] Wen, X., Lee, Y., Luca, F. & Pique-Regi, R. Efficient integrative multi-snp association analysis via deterministic approximation of posteriors. *The American Journal of Human Genetics* **98**, 1114–1129 (2016).

[8] Wang, G., Sarkar, A. K., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine-mapping. *bioRxiv* 501114 (2019).

[9] Benner, C. *et al.* Finemap: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).

[10] Nica, A. C. *et al.* Candidate causal regulatory effects by integration of expression qtls with complex trait genetic associations. *PLoS Genet* **6**, e1000895 (2010).

[11] Chun, S. *et al.* Limited statistical evidence for shared genetic effects of eqtls and autoimmune-disease-associated loci in three major immune-cell types. *Nature genetics* **49**, 600 (2017).

[12] Zhu, Z. *et al.* Integration of summary data from gwas and eqtl studies predicts complex trait gene targets. *Nature genetics* **48**, 481 (2016).

[13] Wu, Y. *et al.* Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nature communications* **9**, 1–14 (2018).

[14] Wen, X., Luca, F. & Pique-Regi, R. Cross-population joint analysis of eqtls: fine mapping and functional annotation. *PLoS Genet* **11**, e1005176 (2015).

[15] Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013).

[16] Aguet, F. *et al.* The gtex consortium atlas of genetic regulatory effects across human tissues. *BioRxiv* 787903 (2019).

[17] Vilhjálmsson, B. J. *et al.* Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *The american journal of human genetics* **97**, 576–592 (2015).

[18] Wen, X., Lee, Y., Luca, F. & Pique-Regi, R. Efficient integrative multi-snp association analysis via deterministic approximation of posteriors. *The American Journal of Human Genetics* **98**, 1114–1129 (2016).

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