A flexible summary statistics-based colocalization method with application to the mucin cystic fibrosis lung disease modifier locus

Authors

Fan Wang, Naim Panjwani, Cheng Wang, Lei Sun,
Lisa J. Strug

Correspondence

sun@utstat.toronto.edu (L.S.),
lisa.strug@utoronto.ca (L.J.S.)

CellPress

A flexible summary statistics-based colocalization method with application to the mucin cystic fibrosis lung disease modifier locus

Fan Wang,[1,2] Naim Panjwani,[2] Cheng Wang,[2] Lei Sun,[1,3,*] and Lisa J. Strug[1,2,3,4,5,*]

Abstract

Mucus obstruction is a central feature in the cystic fibrosis (CF) airways. A genome-wide association study (GWAS) of lung disease by the CF Gene Modifier Consortium (CFGMC) identified a significant locus containing two mucin genes, *MUC20* and *MUC4*. Expression quantitative trait locus (eQTL) analysis using human nasal epithelia (HNE) from 94 CF-affected Canadians in the CFGMC demonstrated *MUC4* eQTLs that mirrored the lung association pattern in the region, suggesting that *MUC4* expression may mediate CF lung disease. Complications arose, however, with colocalization testing using existing methods: the locus is complex and the associated SNPs span a 0.2 Mb region with high linkage disequilibrium (LD) and evidence of allelic heterogeneity. We previously developed the Simple Sum (SS), a powerful colocalization test in regions with allelic heterogeneity, but SS assumed eQTLs to be present to achieve type I error control. Here we propose a two-stage SS (SS2) colocalization test that avoids *a priori* eQTL assumptions, accounts for multiple hypothesis testing and the composite null hypothesis, and enables meta-analysis. We compare SS2 to published approaches through simulation and demonstrate type I error control for all settings with the greatest power in the presence of high LD and allelic heterogeneity. Applying SS2 to the *MUC20/MUC4* CF lung disease locus with eQTLs from CF HNE revealed significant colocalization with *MUC4* ($p = 1.31 \times 10^{-5}$) rather than with *MUC20*. The SS2 is a powerful method to inform the responsible gene(s) at a locus and guide future functional studies. SS2 has been implemented in the application LocusFocus.

Introduction

Cystic fibrosis (CF [MIM: 219700]) is a life-limiting genetic disease caused by mutations in the CF transmembrane conductance regulator (*CFTR* [MIM: 602421]). Multiple organs are affected in CF with variation in disease severity influenced by *CFTR* genotype, environmental factors, and modifier genes.[1] The majority of morbidity and mortality in CF results from lung disease which is heritable beyond the contributions of *CFTR*.[2] Mucus pathology is a hallmark of CF airway disease, and thus the mucin family of genes have been hypothesized to contribute to lung disease severity in CF.[3] A genome-wide association study (GWAS; n = 6,365) of CF lung disease from the International CF Gene Modifier Consortium (ICFGMC) identified an associated locus on chromosome 3 in an intergenic region between two mucin genes—mucin 4, tracheobronchial (*MUC4* [MIM: 158372]) and mucin 20, cell surface-associated (*MUC20* [MIM: 610360])—providing support for the mucin hypothesis but leading to uncertainty around the responsible gene(s) at the locus. Given the associated variants are not tagging protein coding variation, the assumption is that the associated locus is marking gene regulation.

Colocalization analysis using GWAS and gene expression quantitative trait locus (eQTL) summary statistics in a CF airway model can test this hypothesis and provide sta-

tistical support for the most probable gene(s) at the locus. However, complications arise when we try to formally test colocalization at this locus using published tools.[4–15] First, the null hypothesis of no colocalization is a composite null hypothesis that consists of several different null scenarios, e.g., the null scenario where there are significant GWAS and eQTL SNPs at a locus, but the two do not colocalize, or that there is a GWAS significant SNP but no eQTL at the locus (see Material and methods for the four different null scenarios). Type I error rate control under all possible null scenarios can be challenging for any method, especially in the presence of multiple hypothesis testing across multiple genes or tissues. Second, the associated SNPs at the *MUC4/MUC20* locus span a 0.2 Mb region with evidence of allelic heterogeneity and high linkage disequilibrium (LD); these are factors that can substantially reduce the power of existing methods.[4,5] Third, in our study, the participants in the eQTL analysis overlap with a subset of the participants in the GWAS; this induces correlation between the eQTL and GWAS summary statistics that could lead to an increased false positive rate of colocalization depending on the degree of overlap.[16] Lastly, our GWAS summary statistics are obtained from a meta-analysis with related individuals within sub-studies, further complicating LD adjustment in colocalization inference.

Several statistical methods have been developed to test for colocalization between GWAS and eQTL summary

[1]Department of Statistical Sciences, University of Toronto, Toronto, ON M5G 1Z5, Canada; [2]Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada; [3]Biostatistics Division, Dalla Lana School of Public Health, University of Toronto, Toronto, ON M5T 3M7, Canada; [4]Department of Computer Science, University of Toronto, Toronto, ON M5S 2E4, Canada; [5]The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada
*Correspondence: sun@utstat.toronto.edu (L.S.), lisa.strug@utoronto.ca (L.J.S.)

statistics, but none of the existing methods accommodate all the aforementioned complications, and we were surprised to see that the other published colocalization methods[4,11,12,14] did not conclude strong colocalization for *MUC4* eQTLs and the GWAS summary statistics. Bayesian colocalization approaches that are amenable to the use of summary statistics include eCAVIAR,[5] CO-LOC,[4] ENLOC,[9] GWAS-PW,[6] and COLOC2.[14] These methods aim to identify whether there is a shared *causal* genetic variant that contributes to both the disease outcome and gene expression variation. Both eCAVIAR and ENLOC compute SNP-level posterior probabilities for being causal and a regional colocalization probability (RCP) by summing up the SNP-level probabilities. The RCP is computed for a locus that has been identified by GWASs, where the evidence under the composite null hypothesis is not explicitly calculated so it is unknown whether this method will have reliable operating characteristics more generally. COLOC, in contrast, computes the posterior probability incorporating subjective priors under five scenarios: the four null scenarios and a specific alternative when one single causal variant is shared. Co-localization is concluded using a recommended value for the posterior colocalization probability (e.g., 0.8).[14] Although the threshold for the posterior probability can be modified to account for multiple hypothesis testing by adjusting the prior probability and other factors in the calculation of false positive report probability (FPRP),[17,18] explicit recommendations to control the false positive rate when testing multiple genes and/or tissues at a locus for these Bayesian colocalization methods have not been detailed. GWAS-PW extends COLOC by empirically estimating priors from the genome-wide data for the five scenarios to investigate genetic variants that influence a pair of traits and provides software functions to account for overlapping participants in the two studies.[6] Similarly, COLOC2 incorporates features implemented in GWAS-PW and provides an updated version of CO-LOC.[14] However, COLOC, GWAS-pw, and COLOC2 all assume one single causal variant, and it has been shown that COLOC can have substantial loss in power when there are multiple, independent GWAS or eQTL signals at a locus (allelic heterogeneity).[8] To gain additional power for a locus with multiple independent eQTL signals, Dobbyn et al.[14] proposed a forward stepwise conditional analysis before conducting the COLOC2 analysis. For each identified eQTL signal, COLOC2 integrates the GWAS result with eQTL evidence conditional on all the other eQTL signals. However, the conditional analysis requires individual-level data and can be computationally intensive. An alternative approach adopts the SuSiE[19] framework to distinguish multiple signals for a given trait, and then conducts COLOC analysis on all possible pairs of signals between the traits.[20] This method has been shown to provide more accurate inference than COLOC based on conditional analysis; however, the identification of different signals relies on the power of SuSiE.[20]

There are several frequentist-based methods that also calculate colocalization evidence, for example the gene expression imputation approaches such as TWAS[13] and PrediXcan.[21] These methods first use reference expression databases to define a set of genetic variants predictive of gene expression levels, then they impute gene expression levels in a study sample and test for association with a disease phenotype. The reliability of these methods depends on their prediction accuracy which can be limited.[22–24] Extensions that enable the use of external summary statistics such as S-PrediXcan,[7] S-TWAS,[13] and S-MultiXcan[10] and pre-computed parameters are available. However, in our case we already have gene expression data from individuals with CF in a relevant airway model. Integration methods based on Mendelian randomization also *indirectly* address the question of colocalization. Two such approaches, SMR[11] and Multi-SNP-based SMR (SMR-multi),[12] derive their test statistics by assuming independence between GWASs and eQTL studies and use SNPs with eQTL p values $< 5 \times 10^{-8}$ as instrumental variables to test whether gene expression differences cause the disease phenotype. SMR and SMR-multi restrict their analyses to regions with significant eQTLs, and they can accommodate meta-analysis but their robustness to related or overlapping samples is unknown. Lastly, JLIM[15] evaluates whether there is a shared causal variant between eQTL and GWASs by developing a test statistic that contrasts the joint likelihood assuming one shared causal variant with that assuming distinct causal variants between eQTL and GWAS. To calculate accurate colocalization p values, JLIM requires individual-level gene expression data for permutation.

We previously developed an alternative frequentist and summary statistics-based colocalization method called Simple Sum (SS).[8] It too does not address all of the four factors necessary for reliable colocalization analysis at our chr3q29 locus. SS performs well in the presence of linkage disequilibrium and allelic heterogeneity, but at a given GWAS region with significant association but no eQTLs, the SS could have type I error inflation under this null scenario of no colocalization. Thus, we recommended an *ad hoc* approach that restricted SS analyses to regions with significant eQTL evidence, similar to those imposed by other methods (e.g., SMR and SMR-multi). Here we extend the SS and propose a principled two-stage SS colocalization method (SS2) that controls the type I error rate for the composite null hypothesis, even in the presence of multiple hypothesis testing across multiple genes and/or tissues, and remains powerful in the presence of LD and heterogeneity. The SS2 can also accommodate summary statistics calculated from meta-analysis with related samples within sub-studies, and it is robust to the presence of a subset of overlapping samples between the two sets of summary statistics. The SS2 has been implemented in the application LocusFocus.[25]

For method comparison, we choose COLOC and CO-LOC2 from the class of Bayesian methods, and SMR and SMR-multi from the class of frequentist methods. Among

the three Bayesian approaches that consider the issue of the composite null hypothesis (COLOC, GWAS-PW, and COLOC2), COLOC has been shown to have better performance than GWAS-PW when a single causal variant is shared between the two studies,[15] while GWAS-PW can have better performance when eQTL and GWAS causal variants are distinct. We thus implement COLOC2 that incorporates features of both GWAS-PW and COLOC. Among the frequentist colocalization approaches that are applicable to summary statistics, S-PrediXcan, S-TWAS, and S-MultiXcan use prediction weights inferred from a publicly available source (see PredictDB Data Repository in web resources) to estimate the association between the predicted transcriptome and phenotypes, which can lead to biased or anti-conservative results.[22–24] In our application, we already have expression from CF tissue (human nasal epithelia). Therefore, we implement SMR and SMR-multi which can be directly applied to summary statistics from any eQTL study without the need for training on a new sample to predict weights.

We first conduct extensive simulation studies to compare the proposed SS2 method with COLOC, COLOC2, SMR, and SMR-multi. We then apply these colocalization methods to the GWAS summary statistics from the *MUC4/MUC20* (chr3q29) CF lung function locus[26] and the eQTL summary statistics, respectively, for *MUC20* and *MUC4* in CF primary human nasal epithelia (HNE). Finally, for completeness we extend the colocalization analyses to the genes within 1 Mb of the locus using CF primary HNE, as well as other CF-related tissues from the genotype tissue expression project (GTEx).[27] We demonstrate statistical support for *MUC4* as the responsible gene at the locus. The *MUC4* impact appears to be relevant in the lungs, both from GTEx and CF HNE, even after adjustment for multiple hypothesis testing of all 564 gene-by-tissue pairs evaluated at this locus.

Material and methods

Notation and model

For a locus of interest (e.g., chr3q29), here we assume that available information includes summary statistics from a phenotype-SNP association analysis and an expression-SNP eQTL analysis for a gene and tissue of interest (e.g., *MUC20* in human nasal epithelia), although this is generalizable to any SNP-level summary data. Assume there are $j = 1,\ldots, m$ SNPs at the locus, let $\mathbf{Z} = (Z_1,\ldots Z_j, \ldots, Z_m)'$ be the vector containing the GWAS summary statistics, and $\mathbf{T} = (T_1,\ldots T_j, \ldots, T_m)'$ containing the eQTL summary statistics for a specific gene-by-tissue pair. $\mathbf{T}$ can be obtained from, for example, GTEx[27] or one's own expression study, although the latter requires additional care if the GWAS and eQTL study samples are related or overlap, which we discuss later.

Our alternative of interest is that the GWAS signal and the eQTL signal coincide. As the complement to this alternative, the null hypothesis, H_0, that there is no colocalization is composite, including four different scenarios:[4]

H_{01}: no SNP-phenotype association and no eQTL,

H_{02}: no SNP-phenotype association but eQTL present,
H_{03}: SNP-phenotype association present but no eQTL,
H_{04}: both SNP-phenotype association and eQTL present but occurring at two independent variants.

That is, $H_0 = H_{01} \cup H_{02} \cup H_{03} \cup H_{04}$.

To test the composite H_0, the Simple Sum test statistic is defined as

$$SS = \sum_j Z_j^2 \left(\frac{t_j - \bar{t}}{\sum_j t_j^2 - m\bar{t}^2} \right) = \mathbf{Z}'A\mathbf{Z}, \qquad \text{(Equation 1)}$$

where $A = diag(a_j)$, $a_j = \frac{t_j - \bar{t}}{\sum_j t_j^2 - m\bar{t}^2}$, and t_j is a continuous eQTL evidence measure that can be defined as T_j^2 or $-\log_{10}(\text{eQTL p})$.[8] In practice, T_j^2 and $-\log_{10}(\text{eQTL p})$ produce SS p values that provide the same colocalization conclusion.

Under the null hypothesis scenarios, $H_{01} \cup H_{02}$, that there is no SNP-phenotype association, $\mathbf{Z} \sim N(0, \Sigma)$, where Σ captures the LD structure of the region of interest (e.g., chr3q29).[28–32] Thus, SS is distributed as $\sum_j d_j \chi_1^2$ under $H_{01} \cup H_{02}$, where the d_j's are the eigenvalues of $\left(\Sigma^{\frac{1}{2}} \right)' A \Sigma^{\frac{1}{2}}$. Implementing a one-sided test based on SS will then control the type I error not only under H_{01} or H_{02}, but also under H_{04} as SS tends to be negative when the SNP-phenotype association and eQTL signals occur at two independent SNPs. However, this test could have type I error inflation under H_{03}, and therefore we previously used caution to interpret colocalization findings when calculated from a small observed eQTL signal.

SS2 colocalization test and type I error control

Here we develop SS2 to test the composite null hypothesis that there is no colocalization, controlling the type I error rate under all four null scenarios. SS2 is a two-stage testing procedure that formally evaluates the eQTL evidence at the region of interest prior to conducting the colocalization analysis. We further extend the method for use with summary statistics obtained from meta-analysis and related individuals within sub-studies, and we investigate the robustness of SS2 to GWAS and eQTL summary statistics calculated using overlapping samples.

If we let the matrix A in Equation 1 be the identity matrix, the SS test statistic is simplified to $\sum_{j=1}^{m} Z_j^2$, which has been used as a gene-based association test statistic.[33–35] Here we replace Z_j^2's with the eQTL summary statistics T_j^2's to evaluate eQTL evidence at the locus, as stage 1 of the SS2 test

$$\sum_{j=1}^{m} T_j^2, \qquad \text{(Equation 2)}$$

for a given gene in a tissue of interest. Alternative gene-based eQTL tests (e.g., maximum of T_j^2)[36,37] can be implemented, depending on factors such as the genetic architecture at the locus of interest.

Recall that under the null scenario of H_{03} where there is GWAS association but no eQTL, the original SS test has inflated type I error rate because the assumption of $\mathbf{Z} \sim N(0, \Sigma)$ is violated. However, $\mathbf{T} \sim N(0, \Sigma)$ in this case, so $\sum_{j=1}^{m} T_j^2$ is distributed as $\sum_j f_j \chi_1^2$, where f_j's are the eigenvalues of Σ.[35] Thus, the SS2 stage 1 test can control the type I error rate of SS2 under H_{03}: if stage 1

Table 1. Empirical type I error rates of SS2, SMR, and SMR-multi and false positive rates of COLOC2 and COLOC for a single hypothesis test

Locus	Null scenarios	Type I error of SS2	Type I error of SMR	Type I error of SMR-multi	False positive rate of COLOC	False positive rate of COLOC2
MUC20/MUC4	H_{01}	0.0021	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0011
	H_{02}	0.0450	0.0377	0.0332	1.00×10^{-4}	0.0245
	H_{03}	0.0229	$<10^{-4}$	$<10^{-4}$	4.00×10^{-4}	0.0858
	H_{04}	0.0262	0.0389	0.0367	$<10^{-4}$	1.00×10^{-4}
SLC6A14	H_{01}	0.0025	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0022
	H_{02}	0.0507	0.0372	0.0342	5×10^{-4}	0.0468
	H_{03}	0.0099	$<10^{-4}$	$<10^{-4}$	9×10^{-4}	0.088
	H_{04}	0.0133	0.0368	0.0336	$<10^{-4}$	0.0012

The LD pattern at the simulated region follows that at the *MUC20/MUC4* and *SLC6A14* loci, respectively. Each row corresponds to a specific null scenario when there is no co-localization. H_{01} represents the scenario when there are no SNP-phenotype associations and no eQTL; H_{02} represents the scenario when there are no SNP-phenotype associations but eQTLs are present; H_{03} represents the scenario where SNP-phenotype associations are present but no eQTL; H_{04} represents the scenario where both SNP-phenotype association and eQTLs are present, but occurring at two independent SNPs. For SS2, SMR,[11] and SMR-multi,[12] the nominal type I error was set at $\alpha = 0.05$. SMR and Multi-SNP-based SMR test (SMR-multi) are conducted under the default setting such that a SNP is picked only if the eQTL p value is less than 5×10^{-8}. For COLOC2[14] and COLOC,[4] the false positive rates are calculated by applying the 0.8 threshold (as recommended by Dobbyn et al.[14]) for the colocalization posterior probability. In total, 10^4 replications are simulated for each null scenario.

of the SS2 test is not significant, we conclude that there is no evidence of an eQTL and thus there is no evidence of colocalization; if stage 1 of the SS2 test is significant, we then implement the SS test statistic of Equation 1 in stage 2.

In terms of the overall type I error control of the proposed two-stage SS2 test under the composite null, $H_{01} \cup H_{02} \cup H_{03} \cup H_{04}$, intuitively, under the null scenarios of H_{01} or H_{03} when there are no eQTLs, stage 1 already controls the false positives. Under the null scenarios of H_{02} or H_{04}, even if the power of detecting the eQTL evidence is 100% in stage 1, stage 2 then provides the control of false positives. In the supplemental material and methods, we show the independence between the stage 1 and stage 2 tests when there is no overlapping or related samples between the GWAS and eQTL study and demonstrate analytically the type I error rate control of SS2.

To ensure type I error rate control under the composite null hypothesis across all four different null scenarios, SS2 is conservative for certain null scenarios. For example, under H_{01} where there is no SNP-phenotype association or eQTL evidence, both stage 1 and stage 2 tests control the false positive rate at the nominal α level; by the independence of the two tests, the overall type I error rate is α^2 (Table 1). This conservativeness is necessary so that the type I error rate under other null scenarios (e.g., H_{03}) is not inflated. Furthermore, this conservativeness is necessary for conducting multiple colocalization hypothesis tests where the family of tests could consist of different null scenarios, which we discuss in the section on Colocalization testing in the presence of multiple hypothesis testing.

Meta-analyses with related individuals within sub-studies

Many GWASs, such as our CF lung function GWAS, use a fixed or random effects meta-analysis to combine association evidence across multiple studies,[26,38] and these multiple studies may contain related individuals within the sub-studies. To implement a colocalization analysis using meta GWAS Z-scores, COLOC[4] assumes unrelated individuals, while COLOC2[14] incorporates features from GWAS-PW[6] that consider sample overlap between studies but not sample relatedness within a study. SMR[11] is derived

based on the assumption that samples are independent between two studies, and its operating characteristics are unknown when there are related samples in a component study. Other methods such as eCAVIAR,[5] ENLOC,[9] SMR-multi,[12] and JLIM[15] compute their statistics under the assumption that GWAS Z-scores follow a multivariate normal distribution with covariance matrix Σ, where Σ is estimated assuming an independent sample using either one's own data or external data such as that from the 1000 Genomes Project.[39] However, accounting for the sample relatedness in the LD matrix, Σ, is important for valid colocalization analysis which we address below.

Assume the GWAS meta-analysis consists of C studies with sample sizes n_c, $c = 1, ..., C$. Let $(Z_{1,j}, ..., Z_{C,j})'$ and $(\widehat{\beta}_{1,j}, ..., \widehat{\beta}_{C,j})'$ denote, respectively, the vector of Z scores and the vector of estimated effect sizes for SNP j from the C studies. Let $Z_{meta,j}$ denote the Z score from the meta-analysis for SNP j, then

$$Z_{meta,j} = \frac{\sum_{c=1}^{C} w_{c,j} \widehat{\beta}_{c,j}}{\sqrt{\sum_{c=1}^{C} w_{c,j}}}, \quad \text{(Equation 3)}$$

where $w_{c,j}$ represents the weight for study c, which can take different forms.[38] For the traditional fixed-effect approach, $w_{c,j}$ is the inverse variance of $\widehat{\beta}_{c,j}$ or estimated by $n_c \widehat{p}_{c,j}(1 - \widehat{p}_{c,j})$, where $\widehat{p}_{c,j}$ represents the estimated minor allele frequency for SNP j in study c. For the random-effect approach, the estimated between-study variance is incorporated into $w_{c,j}$ to account for heterogeneity between studies.[38] Note that Equation 3 is often applied for the scenario with no related or overlapping samples between sub-studies. The meta-analysis with presence of related or overlapping samples between sub-studies can be conducted by alternative approaches such as the method proposed in Zhu et al.,[40] which is not the scenario in the CF lung GWAS and thus is not applied here.

Based on Equation 3, we can show that

$$cov(Z_{meta,j}, Z_{meta,l}) = \frac{\sum_{c=1}^{C} \sqrt{w_{c,j}} \sqrt{w_{c,l}} cov(Z_{c,j}, Z_{c,l})}{\sqrt{\sum_{c=1}^{C} w_{c,j}} \sqrt{\sum_{c=1}^{C} w_{c,l}}} \quad \text{(Equation 4)}$$

for SNPs j and l. When individuals in study c are independent of each other and a simple linear model is used for GWAS,

$cov(Z_{c,j}, Z_{c,l}) = r_{c,jl}$, where $r_{c,jl}$ is the standard Pearson correlation coefficient that represents the LD between SNPs j and l in study c.[5] In this case, the covariance of meta-Z scores between the two SNPs is a weighted sum of study-specific LD measures.

In the presence of related individuals within sub-studies, assume that the GWAS is conducted using a linear mixed effect model, Z scores and their asymptotic covariance matrix can be written in a closed form that is equivalent to using generalized least-squares (supplemental material and methods). For a GWAS that only contains the genotypes as predictors, $cov(Z_{c,j}, Z_{c,l}) = r^*_{c,jl}$, where $r^*_{c,jl}$ can be viewed as the pairwise Pearson correlation coefficient derived from the Cholesky-transformed genotype matrix. The supplemental material and methods provides a general form for $cov(Z_{c,j}, Z_{c,l})$ in the presence of additional covariates (e.g., age and sex) and alternative ways to approximate $cov(Z_{c,j}, Z_{c,l})$ using (for example) the R package nlme or GMMAT.[41]

The eQTL summary statistics are typically obtained from a single study. If the eQTL summary statistics are also obtained from a meta-analysis, the covariance adjustment is also needed when conducting the stage 1 eQTL testing, following the principle in Equation 4. This adjustment, however, does not influence the covariance adjustment in the stage 2 colocalization testing, where the inference is conditional on the observed eQTL evidence.

Overlapping or related samples between GWASs and eQTL studies

The presence of overlapping or related samples can induce correlation between summary statistics even when there are no shared genetic effects, which may bias the model in favor of the alternative hypothesis.[6,42] Let Z_1 and T_1 denote the summary statistics for a variant from a GWAS of sample size n_{GWAS} and an eQTL study for a given gene-by-tissue pair of sample size n_{eQTL}, respectively. In particular when there is no relatedness between those non-overlapping samples,

$$cov(Z_1, T_1) = \frac{n_0}{\sqrt{n_{GWAS} n_{eQTL}}} * \rho, \quad \text{(Equation 5)}$$

where n_0 represents the number of overlapping samples; ρ represents the correlation between phenotypes for the overlapping samples due to (for example) shared environmental factors, and ρ equals 0 if there is no sample overlap.[6] In the presence of related individuals between the GWAS and eQTL study, the correlation between summary statistics could be more complicated than Equation 5,[40] but could be calculated by using methods proposed in Zhu et al.[40] or Province and Borecki.[43] Several published approaches have addressed the influence of overlapping or related samples in the context of two GWASs, and they have implemented decorrelation approaches that we repurpose for our SS2 colocalization test in Colocalization analysis at the MUC4/MUC20 CF lung disease modifier locus.[16,40,43,44]

Colocalization testing in the presence of multiple hypothesis testing

So far we have investigated the properties of SS2 when testing colocalization for a single gene in a specified tissue. In the present study, we are interested in determining whether the SNP-phenotype association evidence is colocalizing with gene expression of *MUC4* or *MUC20* in HNE, an established CF airway model.[3] In fact, there are 50 genes annotated to a 1 Mb region surrounding the top GWAS signal. Moreover, even though the *MUC4/MUC20* locus was identified as associated with CF lung disease, it would be of in-terest to investigate colocalization evidence in other tissues that may be affected in CF and for which the GTEx consortium provides eQTL summary statistics (GTEx V8).

To evaluate type I error rate control of the proposed SS2 colocalization test in the presence of multiple hypothesis testing, we consider the family-wise error rate (FWER). To maintain FWER control of SS2 under the composite null hypothesis for testing multiple genes and tissues, we implement stage 1 of the SS2 test of Equation 2 for all the genes in each tissue and adjust the α for the total number of tests by Bonferroni correction, $\frac{\alpha}{K}$. We then implement stage 2 of the SS2 test of Equation 1 only for those K_2 significant stage 1 eQTL tests and adjust for the corresponding multiple hypothesis testing by $\frac{\alpha}{K_2}$.

The two-stage Bonferroni correction, $\frac{\alpha}{K}$ followed by $\frac{\alpha}{K_2}$, intuitively should control FWER at or below the nominal level. Indeed, we can show (supplemental material and methods) that when there is no GWAS association, the upper bound of the FWER is α when the K tests are a mixture of H_{01} and H_{02}, or all under one of the null scenarios (H_{01} or H_{02}). However, complications arise when there is GWAS association at the locus but the K tests are a mixture of H_{03} and H_{04}; some genes/tissues do not have eQTLs (H_{03}) while the remaining ones have eQTLs (H_{04}) but do not overlap with the GWAS signal. To see this, first, when all K tests are under H_{03}, even though the colocalization test in stage 2 may have inflated type I error, say as large as 1, the FWER is controlled at α via stage 1, as each eQTL test in stage 1 is controlled at $\frac{\alpha}{K}$. Second, when all K tests are under H_{04}, even though all K tests pass stage 1 due to strong eQTLs, the corresponding colocalization tests in stage 2 have properly controlled type I error. However, when there is a mixture of H_{03} and H_{04}, the eQTL test in stage 1 is no longer controlled at $\frac{\alpha}{K}$ due to the presence of H_{04}. This, combined with inflated false positives in stage 2 for H_{03}, can lead to increased FWER, which has not been investigated before by us or by others.

In the supplemental material and methods, we provide a specific example and show that a crude upper bound for FWER is $2\alpha - \alpha^2$. However, this bound assumes the empirical type I error rate of the colocalization test for H_{03} in stage 2 is 1, which is unrealistic. Our empirical studies below show that we did not observe a single iteration with empirical FWER greater than the specified significance level under H_{03} and H_{04} (Tables 2 and S8), and the proposed two-stage SS2 testing procedure tends to be conservative under H_{01} and H_{02}.

Simulations

To evaluate the performance of the proposed SS2 colocalization test, we conduct extensive simulation studies using the LD pattern observed at the *MUC4/MUC20* locus and for completeness also LD at *SLC6A14*, a locus previously investigated.[8] For method comparison, we choose four alternative colocalization procedures, SMR, SMR-multi, COLOC, and COLOC2. For SS2, SMR,[11] and SMR-multi,[12] the nominal significance level is set at $\alpha = 0.05$. For CO-LOC2[14] and COLOC,[4] colocalization is concluded for each gene and/or tissue if the colocalization posterior probability > 0.8 (as recommended by Dobbyn et al.[14]), although there is no theoretical reason to expect that this threshold will correspond to a 0.05 false positive rate. We outline the simulation study design here and provide additional simulation details in the supplemental material and methods.

Simulation for a single gene-by-tissue pair

We first consider a simulation study assessing colocalization at a locus between SNP-phenotype association (GWAS) p values and SNP-expression (eQTL) p values for a single gene in a given tissue

Table 2. Empirical family-wise error rates (FWERs) of SS2, SMR, and SMR-multi and false positive rates of COLOC and COLOC2 for multiple genes

Locus	Proportion of genes with eQTL association but do not colocalize	FWER of SS2	FWER of SMR	FWER of SMR-multi	False positive rate of COLOC	False positive rate of COLOC2
MUC20/MUC4	0%	0.0377	0.0010	0.0010	0.1307	0.0088
	20%	0.0294	0.0003	0.0002	0.1089	0.0023
	40%	0.0297	0.0002	0.0001	0.0854	0.0027
	60%	0.0300	9.00×10^{-5}	5.00×10^{-5}	0.0606	0.0034
	80%	0.0317	8.00×10^{-5}	4.00×10^{-5}	0.0351	0.0040
	100%	0.0346	6.00×10^{-5}	5.00×10^{-5}	0.0043	0.0044
SLC6A14	0%	0.0054	0.0008	0.0008	0.1626	0.0003
	20%	0.0016	0.0008	0.0004	0.1360	0.0049
	40%	0.0011	0.0004	0.0002	0.1103	0.0056
	60%	0.0010	0.0003	0.0002	0.0831	0.0070
	80%	0.0008	0.0002	0.0001	0.0542	0.0080
	100%	0.0007	0.0001	9.00×10^{-5}	0.0234	0.0089

The height of the GWAS peak is set at 5.06 on the $-\log 10 p$ scale such that 10% power is achieved to detect the GWAS association at significance level of 10^{-8}. In total, 600 genes are simulated based on the LD pattern at the *MUC20/MUC4* locus or the *SLC6A14* locus, respectively. Each row corresponds to a different proportion of genes that have eQTL association (0%, 20%, 40%, 60%, 80%, and 100%). The eQTL peaks are randomly generated from 6 different intervals (50%–60%, 60%–70%, 70%–80%, 80%–90%, 90%–95%, 95%–100% power is achieved to detect the eQTL association at the significance level of 10^{-8}) with probabilities according to the proportion of the -log10(maximum eQTL p value) within each interval observed at the corresponding locus. None of the eQTL peak colocalizes with the GWAS peak for FWER evaluation. SMR and Multi-SNP-based SMR test (SMR-multi) are conducted under the default setting such that a SNP is picked only if the eQTL p value is less than 5×10^{-8}. COLOC2 is conducted by using the algorithm implemented in GWAS-PW, where the posterior probability is calculated based on the likelihood of all gene-by-tissue pairs. In total, 10^5 replications are simulated to evaluate FWER of 0.05 and the false positive rates by applying the 0.8 threshold (as recommended by Dobbyn et al.[14]) for the colocalization posterior probability. The empirical FWER (or false positive rates for COLOC and COLOC2) is calculated by counting the proportion of 10^5 replications where at least one gene has a false colocalization claim.

type. Following the simulation procedure in Hormozdiari et al.[5] and Gong et al.,[8] we focus on SNPs 0.1 Mb on either side of the top-associated SNP from the CF lung GWAS[26] at the *MUC4/MUC20* locus and at the *SLC6A14* locus, respectively, to generate data for the two loci with different LD distributions. For type I error evaluation, we generate GWAS and eQTL summary statistics from a multivariate normal distribution based on the LD pattern of SNPs at each of the two loci and simulate null scenarios from H_{01} to H_{04} under the composite null hypothesis. Simulation details are provided in the supplemental material and methods and their corresponding parameter values are provided in Table S6.

We then assess the power of different methods by considering six alternative colocalization scenarios (Figures 1, 2, and S1), which go beyond the simple alternative when only one causal variant is shared by the GWAS and eQTL studies. Simulation details are provided in the supplemental material and methods, and their corresponding parameter values are provided in Table S7.

Simulation for multiple gene-by-tissue pairs

We next evaluate the performance of the methods when studying colocalization evidence across many gene-by-tissue pairs. To be consistent with the presence of a GWAS signal and 564 gene-by-tissue pairs in the CF application, we simulate a locus, based on the LD pattern at the *MUC20/MUC4* locus, with one GWAS signal and 600 sets of independent eQTL summary statistics. We also considered 100, 200, 300, 400, or 500 colocalization tests simultaneously, varying the composition of the different alternative and null scenarios. Finally, we repeat the analysis for *SLC6A14*, a locus we studied previously[8] with a LD pattern different from *MUC20/MUC4*. COLOC2 is conducted by using the algorithm imple-

mented in GWAS-PW, where the posterior probability is calculated based on the likelihood of all gene-by-tissue pairs.

To evaluate the empirical FWER control, the simulated locus of interest has GWAS association evidence as in the CF application, while the eQTL summary statistics are simulated either under the null scenario of no eQTL (H_{03}) or with an eQTL but does not colocalize with the GWAS signal (H_{04}). When none (0%) of the 600 genes have eQTL evidence, all 600 tests are under H_{03}. We also consider five different proportions of genes (20%, 40%, 60%, 80%, and 100%) with eQTL summary statistics but under H_{04}. When these proportions are varied, the 600 tests are a mixture of H_{03} and H_{04}, which is challenging in terms of type I error control as discussed earlier.

To evaluate power across different amounts of eQTL evidence while the locus has the same SNP-phenotype association evidence as in practice, we vary the colocalized eQTL evidence among the genes analyzed. In 5% of the 600 genes, we simulate the eQTL summary statistics that colocalize with the SNP-phenotype association evidence (i.e., simulated under the alternative). For the remaining 95% of genes, we simulate a proportion of the eQTL summary statistics to have no eQTL signal (under H_{03}) while the remaining to have eQTL signals distinct from the SNP-phenotype association signal (under H_{04}). We calculate power (or true positive rates for COLOC and COLOC2) by determining the proportion of 10^5 simulated replications where at least one gene from the alternative is correctly identified.

Simulation for overlapping samples

In the presence of overlapping samples, the summary statistics for a variant from the GWAS and eQTL studies are correlated, which can lead to increased false positives in theory.[6] To evaluate the practical
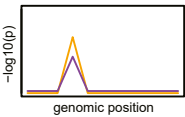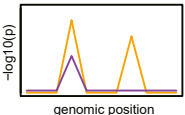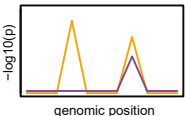
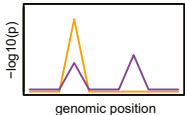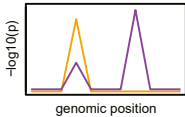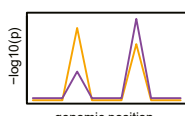| Scenarios	Illustration
	GWAS (orange) / eQTL (purple)
Scenario 1: one shared SNP for GWAS and eQTL study (Figure 2A and Figure S1A)	
Scenario 2: two independent GWAS SNPs, and one eQTL SNP is shared with the first GWAS SNP (Figure 2B and Figure S1B)	
Scenario 3: two independent GWAS SNPs, and one eQTL SNP is shared with the second GWAS SNP (Figure 2C and Figure S1C)	
Scenario 4: two independent eQTL SNPs, one GWAS SNP colocalizes with the eQTL SNP, given the non-overlapped eQTL association is low (Figure 2D and Figure S1D)	
Scenario 5: two independent eQTL SNPs, one GWAS SNP colocalizes with the eQTL SNP, given the non-overlapped eQTL association is strong (Figure 2E and Figure S1E)	
Scenario 6: two independent GWAS SNPs, two independent eQTL SNPs and both SNPs are shared (Figure 2F and Figure S1F)	

Figure 1. Overview and illustration of the six alternative scenarios we simulated to assess the power/true positive rate of different methods

The general visualization of GWAS and eQTL colocalization patterns (on the −log10 p scale) in a region of interest are illustrated by lines in orange and purple, respectively.

is approved by the Research Ethics Board of the Hospital for Sick Children for the usage of public and external data.

Sample source and collection

We conducted RNA sequencing of HNE cells (n = 94) collected as part of the CGMS and the CF Canada Sick Kids Program in Individual CF Therapy (CFIT).[45] The HNE samples are collected using a 3-mm diameter sterile cytology brush (MP Corporation) or Rhino-probe curette on either inferior turbinate. Sequencing was performed in two rounds with Illumina HiSeq 2000 and HiSeq 2500 platforms (Illumina Inc.), respectively. The HiSeq 2000 round was sequenced with 25 million paired-end reads (49 base pairs in length) per sample, and the samples processed using the HiSeq 2500 platform had average library size of 35 mill paired-end reads (124 base pairs in length).

RNA-seq data processing and analysis

Quality of sequencing reads was assessed using FastQC (v.0.11.5; web resources) before and after trimming by Trim Galore (v.0.4.4; web resources). Processed reads were aligned to human reference genome hg38 with GENCODE comprehensive gene annotation (release 29) using STAR (v.2.5.4b).[46] The reference genome included alternative haplotype contigs to account for the sequence diversity in the mucin gene locus. Expression quantification was performed by RNA-SeQC (v.2.0.0), which generated both read counts and normalized transcripts per million (TPM) measures.[47] Normalized trimmed mean of M values (TMM) measures were obtained for a sub-sample of genes with ≥0.1 TPM and ≥6 read counts in more than 20% of the sample.[48]

Expression quantitative trait loci were analyzed by conducting differential gene expression analysis of the effect of SNP genotypes on TMM-normalized expression level. eQTL analysis was carried out using FastQTL (v.2.0)[49] with RNA-sequencing (RNA integrity number ≥ 7) of HNE from 94 Canadians with CF enrolled in the ICFGMC; these 94 individuals were also included in the CF lung GWAS. Additional covariates adjusted for in the model include the top 3 principal components, 15 probabilistic estimates of expression residuals (PEER) factors, study sites, sex, genotyping platform, RNA integrity number (RIN), and PTPRC/CD45 gene expression (immune cell content adjustment). R packages GENESIS (v.2.14.3) and peer (v.1.0) were used to generate genotype principal components and PEER factors, respectively.[50–52]

Analysis of colocalization at the *MUC4/MUC20* locus

The CGMS participants were included in a genome-wide association study of CF lung disease by the International Cystic Fibrosis Gene

impact of sample overlap on SS2, we consider the scenario where half or all of the participants whose genotypes used to compute the eQTL summary statistics are also included in the GWAS. Mimicking the scenario in the CF application, we simulate 100 participants in the eQTL study and to be conservative we include an additional 1,900 participants in the GWAS (i.e., the GWAS sample size is 2,000 of which 100 overlap with the eQTL study). We evaluate the empirical type I error rate of SS2 under the composite null hypothesis from H_{01} to H_{04}, with four different phenotypic correlations, 0.3, 0.5, 0.7, and 0.9. For comparison, we also demonstrate the type I error rate when there are no samples overlapping using the same simulation procedure. In addition, considering a fixed level of phenotypic correlation (0.5 or 0.9), we also demonstrate the empirical type I error rate control of SS2 as we vary sample size for the eQTL study, from 100, 200, 300, 400, to 500. Lastly, we simulate the scenario where all GWAS samples are overlapping with eQTL samples (both with 2,000 individuals) and demonstrate the type I error rate of SS2 under the composite null hypothesis.

EQTL analysis

Informed consent

The Canadian Gene Modifier Study (CGMS) was approved by the Research Ethics Board of the Hospital for Sick Children (# 0020020214 from 2012 to 2019 and #1000065760 from 2019 to the present) and all participating sub-sites. Written informed consent was obtained from all participants or parents/guardians/substitute decision makers prior to inclusion in the study. The CGMS
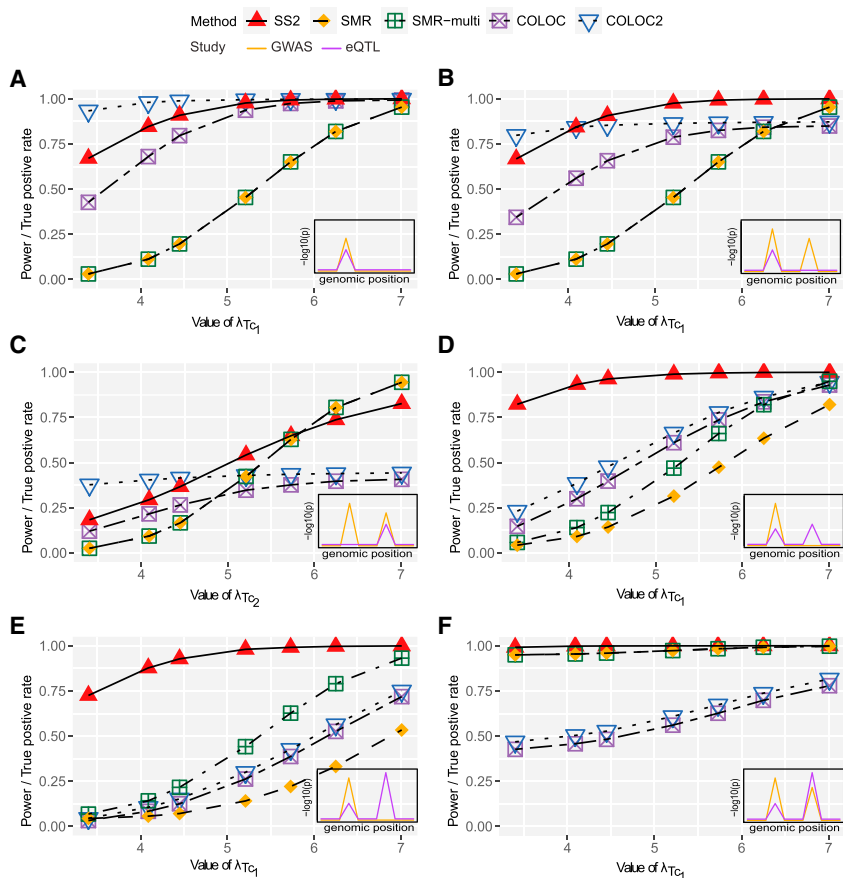
Figure 2. **Empirical power of SS2, SMR, and SMR-multi and true positive rates of COLOC and COLOC2 for testing a single gene for colocalization under six alternative scenarios:** (A) one shared SNP for GWAS and eQTL study, (B) two independent GWAS SNPs, and one eQTL SNP is shared with the first GWAS SNP, (C) two independent GWAS SNPs, and one eQTL SNP is shared with the second GWAS SNP, (D) two independent eQTL SNPs, one GWAS SNP colocalizes with the eQTL SNP, given the non-overlapped eQTL association is low, (E) two independent eQTL SNPs, one GWAS SNP colocalizes with the eQTL SNP, given the non-overlapped eQTL association is strong, and (F) two independent GWAS SNPs, two independent eQTL SNPs and both SNPs are shared.

The LD pattern at the simulated region follows that at the *MUC20/MUC4* locus at chromosome 3. For SS2, SMR,[11] and SMR-multi,[12] the nominal type 1 error rate is set at $\alpha = 0.05$. For COLOC2[14] and COLOC,[4] the false positive rates are controlled by applying the default 0.8 threshold (as recommended in Dobbyn et al.[14]) for the colocalization posterior probability. In total, 10^4 replications are simulated to obtain the empirical power and true positive rates. The x-axis demonstrates the parameter values for $\lambda_{T_{c1}}$ or $\lambda_{T_{c2}}$ which is set to be 3.4, 4.09, 4.45, 5.21, 5.73, or 7.01 such that 0.01, 0.05, 0.1, 0.3, 0.5, or 0.9 power is achieved to detect the eQTL association at the significance level of 10^{-8}. The inserted figure in the right bottom of each plot provides the general visualization of GWAS (orange line) and eQTL (purple line) colocalization patterns in a region of interest.

Modifier Consortium, comprised of 6,365 individuals (including 1,443 sib-pairs) with CF.[26] The lung disease severity was measured as a percentile from a CF reference population of forced expiratory volume in 1 s that is survival adjusted.[53] Individuals on CFTR-modulator treatment were not included. The GWAS summary statistics are publicly available (see data and code availability) and we use them here in the implementation of the SS2. The summary statistics were constructed from a meta-analysis of 13 studies, including siblings and individuals who ranged in age from 6 to 63.3 years. A detailed description of the studies and participants included can be found in Corvol et al.[26] Among other loci, 24 genome-wide significant SNPs were identified and annotated between two mucin genes: *MUC4* and *MUC20* on chromosome 3.

To determine whether gene expression variation at the chr3q29 locus could influence lung disease severity and which genes the associated SNPs impact, we first focused on colocalization analysis for *MUC4* and *MUC20* given their biologic plausibility. For completeness, we subsequently expanded our analysis to investigate eQTLs of 50 genes in a 1 Mb region on either side of the peak and in 14 CF-relevant tissues from GTEx V8. We focus on SNPs in a 0.1 Mb region on either side of the lead GWAS SNP for colocalization analysis of all 564 gene-by-tissue pairs for which there is gene expression data.

To conduct the SS2 stage 1 eQTL test using Equation 2, we computed the LD matrix for the eQTL analysis in HNE by calculating the Pearson correlation coefficient using the 94 CF independent samples, while for the eQTL studies in other CF-related tissues we

use the GTEx resource. To apply the SS2 stage 2 colocalization test to the GWAS meta-summary statistics while accounting for sample relatedness, we estimated the LD matrix using Equation 4, where the covariance of summary statistics for the sub-studies were calculated from Cholesky-transformed genotype data. The sample size of the CF lung GWAS is 6,365 while the sample size of the eQTL study is 94. Among the 94 CF samples with eQTL data, there are 85 participants included in both the GWASs and eQTL studies, but otherwise the participants across the two studies were unrelated. To account for the overlapping samples in the analysis, we note that $\frac{n_0}{\sqrt{n_{GWAS}} \, n_{eQTL}} = 0.11$ and so even under the extreme case that $\rho = 1$, $cov(Z_1, T_1) < 0.11$, which would have a negligible impact on our inference. We demonstrate this empirically through a comprehensive simulation study (Tables S1–S5) and apply the decorrelation approach[16] in our CF SS2 implementation assuming that the lung function distribution in the 85 overlapping samples is representative of the full CF sample included in the GWAS.

Results

Simulation results
Simulation results for a single gene-by-tissue pair
Table 1 demonstrates the empirical type I error rates of SS2, SMR,[11] and SMR-multi[12] and the false positive rates of COLOC[4] and COLOC2[14] for each of the four null scenarios of

the composite null hypothesis, based on the LD pattern observed at the *MUC20/MUC4* locus or the *SLC6A14* locus. The SS2, SMR, and SMR-multi all control the type I error rate at or below the nominal $\alpha = 0.05$ significance level under all four null scenarios. When there are no eQTLs, under H_{01} or H_{03}, SMR and SMR-multi are extremely conservative (the empirical $\alpha < 10^{-4}$). This is due to the recommended pre-screening step whereby colocalization analysis can only be conducted when there is an eQTL p value less than 5×10^{-8} at the locus under investigation. Other eQTL p value thresholds such as 0.01 or 0.05 could be adopted when conducting the SMR or SMR-multi; however, there is type I error inflation under H_{03} when one chooses to do so (Table S18).

Among all the methods, COLOC shows the most conservative results under the different null scenarios. The false positive rate for COLOC2 is controlled except for the H_{03} scenario. When there is SNP-phenotype association but no eQTLs, the empirical false positive rate is 0.09 for the nominal level of 0.05. The empirical false positive rate of COLOC2 decreases as one increases the sample size of the eQTL study (Table S17). For example, with 2,000 participants in the GWAS, the empirical false positive rate of COLOC2 is below 0.05 when the sample size for the eQTL study is larger than 1,000. However, in practice, the sample size of an eQTL study (e.g., GTEx)[27] is usually much smaller than 1,000, and in our CF study we only have 94 participants included in the eQTL analysis in HNE. We observed qualitatively similar results with the different LD patterns between the *MUC20* and *SLC6A14* loci.

The effect of either half or all of the participants in the eQTL study (100 individuals) overlapping with the GWAS study (2,000 individuals) on the type I error rate of the SS2 is demonstrated in Tables S2–S5. The type I error rate of the SS2 remains controlled with increasing phenotypic correlation and increasing eQTL sample sizes, which demonstrates that the overlapping samples have minimal impact on the inference in practice.

We evaluate the power of SS2, SMR, and SMR-multi at the $\alpha = 0.05$ significance level and the true positive rate of COLOC2 and COLOC using a posterior probability of colocalization cut-off of 0.8 (Figure 2). Figure 2A illustrates the power under the simplest alternative scenario 1, where one single SNP-phenotype association signal colocalizes with one single eQTL signal. In that case, COLOC2 has the highest true positive rate across different levels of eQTL evidence, but its false positive rate is inflated under H_{03} as demonstrated in Table 1. Among the methods that control the false positive rates across all null scenarios, namely SS2, SMR, SMR-multi, and COLOC, the proposed SS2 method is the most powerful.

When there are two GWAS peaks under the alternative scenarios 2 and 3 (Figures 2B and 2C, respectively) and the eQTL evidence is weak ($\lambda_{T_{c1}} < 5.21$ in Figure 2B and $\lambda_{T_{c2}} < 5.21$ in Figure 2C; i.e., less than 30% power to detect the eQTL evidence at $\alpha = 0.05$), SS2 has less power than COLOC2; this represents a trade-off to achieve type I error

control under H_{03}. However, SS2 is more powerful than the other three methods that do not have inflated false positive rates. SMR and SMR-multi are very conservative in this case because there are few SNPs with eQTL p values less than 5×10^{-8}. As the eQTL evidence gets stronger ($\lambda_{T_{c1}} > 5.21$ in Figure 2B and $\lambda_{T_{c2}} > 5.21$ in Figure 2C; eQTL power greater than 30%), the power of SS2 could reach a similar level or even exceed the true positive rate of COLOC2. The power of SMR and SMR-multi are similar when there is only one eQTL SNP in the region (Figures 2A–2C) and increase rapidly as the size of the eQTL signal increases. Under the alternative scenario 3, when the eQTL evidence is strong ($\lambda_{T_{c2}} > 6.25$; eQTL power greater than 70% power) and colocalizes with the second SNP-phenotype association peak (Figure 2C), SMR and SMR-multi are more powerful than SS2.

Figures 2D and 2E demonstrate the alternative scenarios 4 and 5, respectively, when there are two independent eQTL SNPs but only one eQTL SNP colocalizes with the SNP-phenotype association. In this case, the SS2 is more powerful than all of the other methods across all levels of eQTL evidence considered. The power advantage of SS2 over other methods is especially notable when, between the two eQTL SNPs, the one with weaker signal colocalizes with the SNP-phenotype associated variant. Finally, under the alternative scenario 6 when two independent GWAS SNPs colocalize with two independent eQTL SNPs, methods SS2, SMR, and SMR-multi show equally high power, while COLOC and COLOC2 have reduced power due to the allelic heterogeneity. Qualitatively similar results based on the LD pattern at the *SLC6A14* locus are evident in Figure S1.

Simulation result for multiple gene-tissue pairs

Table 2 demonstrates the impact of multiple hypothesis testing on the family-wise error rate, where a locus with 600 genes and LD modeled after the *MUC20/MUC4* and *SLC6A14* loci were evaluated. Similar to the single hypothesis testing result in Table 1, SS2, SMR, and SMR-multi control the FWER with the SMR-based tests being the most conservative. However, COLOC now has inflated false positive rates (>0.08) for multiple scenarios, for example when 60% of genes at the locus have no eQTLs (H_{03}) while the remaining 40% of genes have eQTLs but these are distinct from the SNP-phenotype association signal (H_{04}). As the proportion of H_{04} genes decrease from 40% to 0% (or the proportion of H_{03} genes increase from 60% to 100%), the empirical false positive rate of COLOC increases from 0.08 to 0.13. Although COLOC2 shows an inflated false positive rate under H_{03} for our single hypothesis test investigation (Table 1), the false positive rate of COLOC2 is controlled after applying the algorithm implemented in GWAS-PW where the posterior probability is calculated based on the likelihood of all gene-by-tissue pairs (Table 2).

We also investigated type I error rate control of the five methods when the number of genes tested at the locus was varied from 100 to 500 (Tables S8–S12). Overall, SS2,

Table 3. Power of SS2, SMR, and SMR-multi and true positive rate of COLOC2 for multiple genes

Locus	EQTL height for the 5% genes have colocalization	Power of SS2	Power of SMR	Power of SMR-multi	True positive rate of COLOC2
MUC20/MUC4	5.48–5.73	0.8181	0.6826	0.6059	0.6218
	5.73–5.98	0.8178	0.6935	0.6119	0.6355
	5.98–6.25	0.8171	0.7033	0.6169	0.6447
	6.25–6.57	0.8161	0.7144	0.6216	0.6507
	6.57–7.01	0.8143	0.7270	0.6254	0.6536
SLC6A14	5.48–5.73	0.6855	0.6323	0.5531	0.6605
	5.73–5.98	0.6992	0.6462	0.5614	0.6684
	5.98–6.25	0.7076	0.6588	0.5679	0.6741
	6.25–6.57	0.7138	0.6706	0.5745	0.6786
	6.57–7.01	0.7184	0.6831	0.5817	0.6821

The LD pattern at the simulated region follows that at the *MUC20/MUC4* and *SLC6A14* loci, respectively. The height of the GWAS peak is set at 5.06 on the $-\log10p$ scale such that 10% power is achieved to detect the GWAS association at significance level of 10^{-8}. Each row corresponds to a different range of the eQTL height for the 5% genes that have colocalization ([5.48, 5.73], [5.73, 5.98], [5.98, 6.25], [6.25, 6.57], and [6.57, 7.01]). The eQTL peaks are set with 5 different intervals such that 40%–50%, 50%–60%, 60%–70%, 70%–80%, 80%–90% power is achieved to detect the eQTL association at the significance level of 10^{-8}. For the remaining 95% of genes, there is eQTL evidence with a mixture of null cases under H_{03} and H_{04} and details are demonstrated in the supplemental material and methods. SMR and Multi-SNP-based SMR test (SMR-multi) are conducted under the default setting such that a SNP is picked only if the eQTL p value is less than 5×10^{-8}. In total, 10^5 replications are simulated to evaluate power at 0.05 significance level and the true positive rates by applying the 0.8 threshold (as recommended by Dobbyn et al.[14]) for the colocalization posterior probability. The power (or true positive rate for COLOC and COLOC2) is calculated by counting the proportions of 10^5 replications where at least one gene is correctly identified with colocalization.

SMR, and SMR-multi show conservative FWER (<0.05; Table S8, S9, and S10, respectively). As the number of genes increases, the FWER of SS2 decreases when the LD is modeled after the *SLC6A14* locus and moderately increases when the LD is modeled after the *MUC20* locus. The increase in FWER tapers off as the number of genes tested at the locus increase. This is because, with more genes passing the first stage test, the stage 2 colocalization test requires a more stringent significance level, resulting in the overall two-stage test being conservative enough to control the FWER. The false positive rate of COLOC increases as the number of genes evaluated increases, with inflation observed when the number of genes tested exceeds 200 (Table S11). This is due to the subjective choice of priors and cut-off (i.e., 0.8) for the colocalization posterior probability without explicit adjustment for multiple genes. In contrast, COLOC2 is conservative and stays conservative as one increases the total number of genes evaluated at a locus from 100 to 500 (Table S12).

Given the inflation in false positives for COLOC (Table S11), we compare power only between COLOC2, SMR, SMR-multi, and SS2. Keeping the eQTL evidence constant and with 600 genes, SS2 demonstrates the greatest power among the four methods; SMR has more power than SMR-multi (Table 3). The power of the four methods for testing 100 to 500 genes is provided in Tables S13–S16. SS2 shows consistently higher power compared to SMR when the number of genes is greater than 200, and SS2 shows higher power than SMR-multi, COLOC, and COLOC2 across all gene numbers investigated. Interestingly, however, we observe different power trends with increasing numbers of genes analyzed at the locus. The power of SS2, SMR, and SMR-multi decrease as the number of genes increases due to the Bonferroni correction, while the true positive rate of COLOC2 increases as the number of genes increases.

Colocalization analysis at the *MUC4/MUC20* CF lung disease modifier locus

Figure 3 shows that there are multiple SNPs with similar GWAS p values as the lead SNP, suggesting the presence of strong LD in the locus. Given the same LD structure, the eQTL peak is much wider than the GWAS peak, suggesting allelic heterogeneity for the eQTL summary statistics. There is a clear GWAS signal around the lead GWAS SNP visually coinciding with one of the eQTL signals, suggesting that MUC4 expression may mediate CF lung disease. These characteristics suggest this locus is similar to scenario 4 in Figure 1.

We first assess the colocalization evidence for *MUC20* and *MUC4* using the CF lung GWAS meta-analysis summary statistics[26] with eQTLs calculated from the HNE gene expression and genotype data of 94 individuals with CF (Table 4). For *MUC20* in the HNE, the stage 1 test does not provide evidence of an eQTL at the 5% level (uncorrected p value = 0.083), suggesting that the eQTL evidence is not strong enough to move on to stage 2 colocalization analysis. For the *MUC4* eQTLs, both stage 1 and stage 2 tests are significant with p value = 3.16×10^{-7} and 1.31×10^{-5}, respectively, providing statistical evidence of colocalization consistent with the visualization (Figure 3). We applied the decorrelation approach
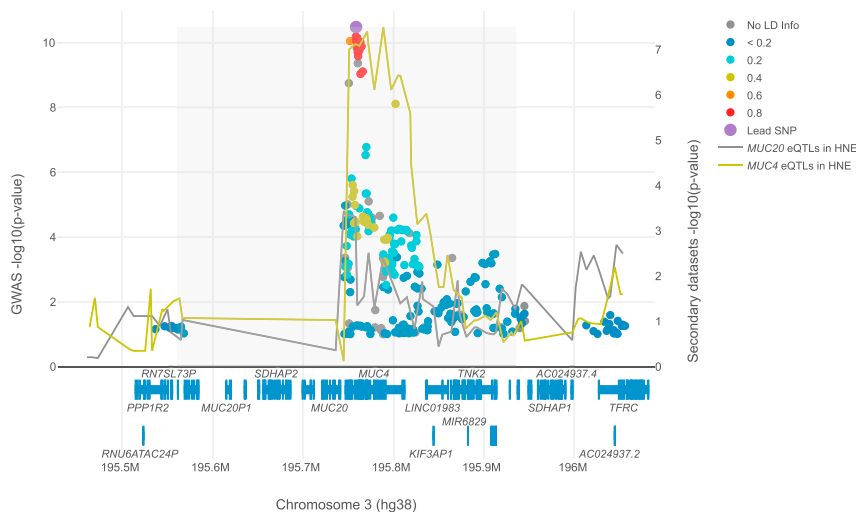
Figure 3. LocusFocus visualization of the CF lung disease GWAS summary statistics and eQTL summary statistics from primary human nasal epithelia (HNE) at the MUC20/MUC4 locus

Overlay of p values (on the $-\log_{10}$ scale) from the lung function GWAS (green/blue palette of colored dots) and eQTL association for *MUC20* and *MUC4* expression (colored lines) for HNE.

demonstrated in LeBlanc et al.[16] to account for the 85 individuals included in both the GWASs and eQTL studies. The resulting stage 1 and stage 2 p values were 2.62×10^{-7} and 1.19×10^{-5}, respectively, assuming the worst case scenario that ρ is 1. As expected, these p values are very similar to the p values without sample overlap correction.

To be comprehensive, we apply the SS2 to all genes annotated by GENCODE v.26 for hg38 GTEx V8 to the 1 Mb region encompassing the peak CF lung-associated variant at the *MUC4/MUC20* GWAS locus. The colocalization evidence for each gene is calculated for the set of SNPs within 0.1 Mb of the peak GWAS variant. Using the cross-tissue eQTLs from GTEx,[27] we select the tissues that are relevant to CF and remove genes with low or no expression in a given tissue; this results in 564 gene-by-tissue pairs available for colocalization analysis. We apply the stage 1 set-based test on all the gene-by-tissue pairs, using a Bonferroni corrected significance level of $\frac{0.05}{564} = 8.87 \times 10^{-5}$. This results in 114 gene-by-tissue pairs providing evidence of significant eQTLs to move to the stage 2 test for colocalization. Stage 2 requires a significance level of $\frac{0.05}{114} = 0.00044$ for each gene-by-tissue pair to conclude colocalization; 39 colocalization tests exceed this threshold. We present the SS2 cross-tissue and gene colocalization results in heatmaps (Figures 4A and S2). For *MUC20* and *MUC4* we provide the stage 1 p value adjusted by 564 gene-by-tissue pairs and the stage 2 p value adjusted by 114 significant gene-by-tissue pairs in Table 4. *MUC4* remains significant after correction for the multiple tests at this locus, with stage 1 adjusted p value of 1.78×10^{-4} and stage 2 adjusted p value of 1.49×10^{-3}, respectively. Interestingly, no other gene shows significant evidence of colocalization in HNE with the SS2. *MUC4* does show evidence of colocalization across several tissues, although these tissues are likely not relevant to lung disease. Several genes in the region also appear regulated by the GWAS-associated SNPs such as the pseudogene *SDHAP1*, but not in lung-relevant tissues (Figure S2).

For comparison, we also implement SMR, SMR-multi, COLOC, and COLOC2. Among the 564 gene-by-tissue pairs, SMR and SMR-multi are calculated on 143 gene-by-tissue pairs with top eQTL p value less than 5×10^{-8}, then applying a Bonferroni corrected significance level of $\frac{0.05}{143} = 0.00035$. We provide both raw p values and multiple testing adjusted p values for SMR and SMR-multi analyses in Table 4. For *MUC20*, there are no SNPs with eQTL p value smaller than 5×10^{-8} and therefore the SMR and SMR-multi test would not be applied. For *MUC4*, SMR provides a raw colocalization p value of 2.65×10^{-3}, but the multiple testing adjusted p value = 0.379. SMR-multi demonstrates an association between gene-expression and the lung GWAS statistics with multiple testing adjusted colocalization p value of 0.0296. Overall, the SMR and SMR-multi tests identified, respectively, 16 and 34 significant genes and tissues with colocalization evidence (Figures 4B, 4C, and S2).

The colocalization posterior probability of COLOC2 for *MUC20* and *MUC4* are both small with 0.000472 and 0.000712, respectively. In this case, the empirical estimation of the prior for colocalization is low, which drags down the colocalization evidence for this locus. At this locus, there are no gene-by-tissue pairs with COLOC2 posterior probability higher than 0.8. In contrast, if we apply COLOC2 with the GWAS-PW algorithm ignoring the multiple hypothesis testing and focusing only on the likelihood from the single gene *MUC4*, the colocalization posterior probability is high (0.9482). However, the colocalization posterior probability for *MUC2*0 is low (0.0666; Table 4). A COLOC2 analysis based on the likelihood from each single gene-by-tissue pair at this locus provides 39 gene-by-tissue pairs with posterior probabilities higher than 0.8 (Figure S3).

The COLOC2 results at this locus are consistent with our simulation study, where COLOC2 has inflation of false positives when the GWAS-PW algorithm is implemented based on the likelihood from a single gene (Table 1). Yet, the test becomes over-conservative when priors are empirically estimated from the likelihood of multiple gene-by-tissue pairs (Table 2).

COLOC does not implement an approach to adjust for multiple hypothesis testing, but the empirical posterior probabilities of colocalization are <0.8 (Table 4) for both *MUC20* (0.1057) and *MUC4* (0.7573). The results of

Table 4. Results of SS2, SMR, SMR-multi, COLOC, and COLOC2 applied to the *MUC20/MUC4* locus in primary human nasal epithelia (HNE) at chromosome 3

Gene and tissue	SS2 stage 1 test		SS2 stage 2 test		SMR		SMR-multi		COLOC	COLOC2	
	p value	Adjusted p value	p value	Adjusted p value	p value	Adjusted p value	p value	Adjusted p value	CLPP	CLPP1[a]	CLPP2
MUC20 and HNE	0.083	1	N/A	N/A[b]	N/A	N/A	N/A	N/A	0.1057	0.0666	0.0005
MUC4 and HNE	3.16×10^{-7}	1.78×10^{-4}	1.31×10^{-5}	1.49×10^{-3}	2.65×10^{-3}	0.379	2.07×10^{-4}	0.0296	0.7573	0.9482[c]	0.0007

Colocalization analyses are conducted for all genes within a 1 Mb region on either side of the peak lung GWAS-associated variant and 14 CF-related tissues. In total, there are 564 gene-by-tissue pairs. Raw p values and adjusted p values by 564 gene-by-tissue pairs are both demonstrated for the SS2, SMR, and SMR-multi. The eQTL evidence for conducting the SS2 is the eQTL p value based on the $-\log10$(eQTL p) scale for a specified gene and tissue. SMR and Multi-SNP-based SMR test (SMR-multi) are conducted under the default setting such that a SNP is picked only if the eQTL p value is less than 5×10^{-8}. N/As are listed for *MUC20* since no SNP has eQTL p value less than 5×10^{-8}. For COLOC and COLOC2, the colocalization posterior probability (CLPP) is calculated, and a high posterior probability (>0.8) suggests strong colocalization evidence. For COLOC2, we show both the CLPP calculated based on the likelihood from the single gene and tissue (CLPP1) and the CLPP calculated based on the likelihood from 564 gene-by-tissue pairs (CLPP2).
[a]This method has been shown with inflation in Table 1. CLPP1 > 0.8 suggests colocalization.
[b]The stage 1 SS2 test p value (0.083) for *MUC20* and HNE does not pass the significant threshold (0.05) and therefore, the stage 2 SS2 test p value is not applicable (N/A).
[c]This method has been shown with inflation in Table 1.

COLOC applied to all 564 gene-by-tissue pairs are shown in Figures 4D and S2.

Discussion

The majority of associated genetic variants identified through GWASs fall in non-coding regions of the genome, and thus the underlying mechanism by which the associated variants contribute to disease remains unclear but may point to gene regulation. The associations identified in the largest GWAS of CF lung disease to date[26] are of no exception, with none of the five genome-wide significant loci tagging protein-coding variation. One locus at chr3q29 is especially noteworthy as it encompasses *MUC4* and *MUC20,* members of a gene family that encode membrane-spanning "tethered" mucins.[54–56] These mucins prevent mucus penetration into the periciliary space and are present in the airway mucus, possibly contributing to mucociliary host defense. Mucus pathology is a defining characteristic of CF, with mucus hyperproduction and plugging, most notably in the CF airways.[3] It has been presumed that mucus pathology is a downstream consequence of CFTR dysfunction,[8] but GWAS identification at this locus suggests the possibility that polymorphisms impacting gene regulation of mucins may, themselves, impact the severity of CF lung disease.

The scenario of the chr3q29 locus in CF highlights the challenges we commonly face while conducting colocalization analysis, where both high LD and allelic heterogeneity are present at a locus with significant SNP-phenotype association as in scenario 4 in Figure 1. The purpose of colocalization analysis is to inform the causal mechanism and guide future functional investigations. Since GWAS identifies loci and many genes could be annotated to a GWAS locus and the tissue of action may not always be obvious, colocalization analysis informs the gene and cell type for future study.

Application of the SS2 to the CF lung disease associated locus at ch3q29 with eQTLs from CF HNE support that the associated lung disease variants colocalize with eQTLs for *MUC4*, prioritizing *MUC4* at the locus for further functional investigation. However, it should be noted that *MUC4* and *MUC20* are localized to a highly polymorphic region[26] with several tandem repeats including a 48 bp repeat region ranging from 7 to 19 kb. The GWAS array data suggests a high frequency of large copy number variants around the clustered mucin region, but highly variable across individuals. This complex genomic context requires further consideration when differentiating between the two mucin genes at the locus.

There are several published colocalization methods, but the chr3q29 locus in a region of high LD with evidence of allelic heterogeneity poses challenges for existing procedures.[4–15] Furthermore, the CF lung disease GWAS summary statistics are derived from a meta-analysis that includes sub-studies with related individuals; current methods cannot accommodate this scenario. We therefore developed the frequentist two-stage colocalization test, SS2. The SS2 integrates GWAS summary statistics with eQTL summary statistics across any number of gene-by-tissue pairs, is applicable when there are overlapping participants in the two studies and can be applied to GWAS summary statistics computed through meta-analysis, even with related individuals. Through simulation we demonstrate that the SS2 controls the type I error rate under the composite null hypotheses and is powerful in regions with high LD and allelic heterogeneity.

Bayesian colocalization approaches aim to identify a shared causal variant between two studies and to differentiate between distinct causal variants in LD. Similarly, Zhu et al. implemented the HEIDI test[11] after the SMR test to further differentiate distinct causal variants in LD if the SMR test suggests significant association between two sets of summary statistics. Previous studies show that the statistical power of COLOC and HEIDI for differentiating distinct causal variants decreases as the LD between causal variants increase.[11] For a fair comparison, we calculated the FWER of SMR and SMR-multi tests without
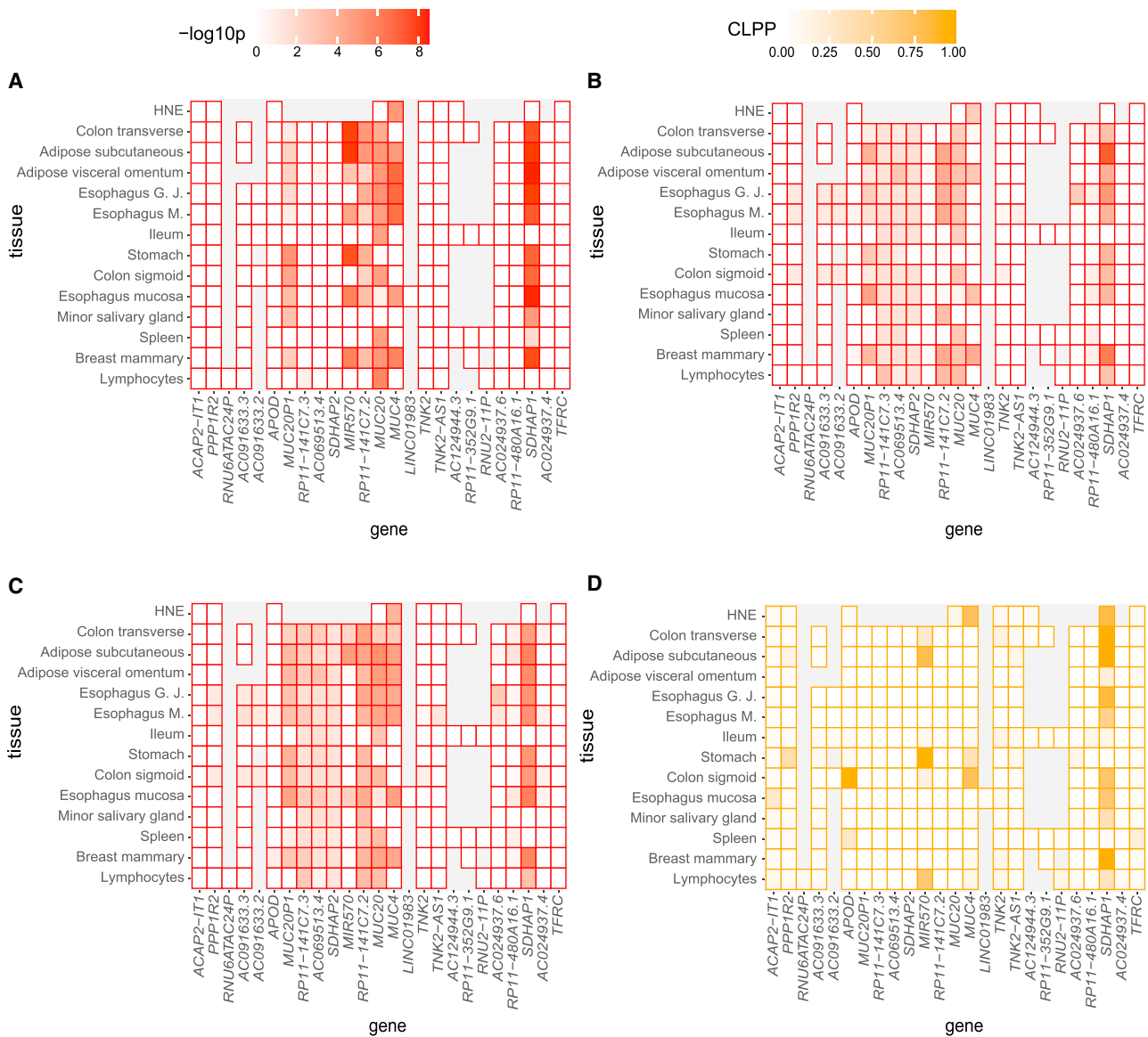
Figure 4. Heatmaps of colocalization evidence across genes and tissues at chromosome 3

In each panel, each cell shows the colocalization evidence for the specified tissue and gene calculated from SNPs within 0.1 Mb of the GWAS peak variant. Genes on the x axis are annotated by GENCODE v.26 for hg38 GTEX V8 and are ordered by their chromosomal positions. For illustration purposes, only the closest annotated genes centered around the GWAS peak variant are shown, and all genes analyzed within 1 Mb around the GWAS peak variant is shown in Figure S2. Grey indicates insufficient expression levels attained for the gene in the tissue under study.

(A) The color intensity corresponds to the SS2 colocalization evidence as measured by $-\log_{10}$(SS2 p value), with red representing $-\log_{10}(p)$ = 8.5 and white representing eQTL evidence for the corresponding gene and tissue does not pass the stage 1 test.

(B and C) The color intensity corresponds to the SMR and SMR-multi colocalization evidence as measured by $-\log_{10}$(SMR p value) and $-\log_{10}$(SMR-multi p value), respectively; with red representing $-\log_{10}(p)$ = 8.5 and white representing eQTL evidence for the corresponding gene and tissue does not pass the eQTL p value threshold (5×10^{-8}).

(D) The color intensity corresponds to the COLOC colocalization evidence as measured by colocalization posterior probability (CLPP) ranging from 0 to 1. The eQTL analyses used for all gene/tissue pairs are those conducted by GTEx[27] v.8 release, except for the HNE eQTLs. eQTL analyses in HNE is conducted using FastQTL[49] with RNA-sequencing of HNE from 94 CF-affected Canadians enrolled in the Canadian CF Gene Modifier study. Esophagus G.J represents Esophagus Gastresophageal Junction; Esophagus M. represents Esophagus Muscularis; Ileum represents Small Intestine Terminal Ileum; Lymphocytes represents cells EBV-transformed lymphocytes.

taking into account the results from the HEIDI test since the SS2 test does not try to differentiate between distinct causal variants if they are in LD. In contrast, the SS2 aims to identify the association between two studies by leveraging the LD in the region and making inference based on the pattern similarity between summary statistics. Therefore, the SS2 can provide reliable inference even when the causal variant is not contained in the analysis set, as long as the LD pattern with the missing variant is retained.

The SS2 is a two-stage framework, designed to accomplish type 1 error control over the complex, composite null hypothesis. Although we implemented the $\sum_{j=1}^{m} T_{jk}^2$ gene/set-based test as the first stage test of the SS2, in practice, the method does not require use of one gene-based test over another. The alternative gene/set-based tests include versions with weighted sums of summary statistics, known as gene set analysis (GSA) tests or burden tests for rare variants.[57–59] Summary statistics can also be decorrelated before being summed together, which is powerful under heterogeneity of effect sizes and variation between pairwise LD patterns.[57] The stage 1 test implemented here has the same functional form as that used in VEGAS[34] and fastBAT.[35] This set-based test can be more powerful than the max-of-chi-square approach (an approach implemented in GATES[36] and Pascal-Max)[37] when there are multiple independent association signals, but can be less powerful when there is a single causal variant present at the locus.[35] In contrast, SMR and SMR-multi use a p value threshold (i.e., 5×10^{-8}) to screen regions for analysis, presumably to ensure they are not using a weak instrument (this is similar to the approach implemented when the SS colocalization statistic was first defined).[8] This stringent screening step results in power loss compared to SS2 when the eQTL association is only moderate which can be a function of several factors including sample size (Figure 2).

We modeled our simulation studies after the CF application and demonstrated that the SS2 has type I error rate control when 85 samples are included in both the GWAS and HNE eQTL studies. For other applications where there is a higher proportion of overlapping samples, the SS2 could have type I error inflation (Table S19) due to the correlation induced by the sample overlap. To address the effect of overlapping samples on statistical inference, several methods propose ways to estimate the correlation using summary statistics.[6,16,40,43] We implemented one such approach[16] to decorrelate the summary statistics before applying the SS2 framework, although it was not necessary for our CF application as we demonstrated through simulation and application.

When the eQTL summary statistics are replaced by GWAS evidence from a second phenotype, the SS2 framework enables the study of genetic overlap of the two traits. Similarly, the SS2 could assess colocalization using any SNP-level data including DNA methylation (meQTLs), protein QTLs (pQTLs), or metabolites (metQTLs). The SS2 framework as delineated here does not integrate summary statistics from greater than two studies, although there would be value in colocalizing GWAS summary data with multiple molecular phenotypes and multiple GWAS traits simultaneously as proposed in Giambartolomei et al.[60] This will be addressed in future work. SS2 is implemented in a web-based colocalization tool, LocusFocus,[25] which enables integration of GWAS summary statistics with any secondary SNP-level dataset by using p values and LD for the region of interest. The eQTL summary statistics from GTEx are made available for selection within the web server to test colocalization with tissues and genes from GTEx. All code and sample datasets are publicly available via GitHub under the MIT license.

Data and Code Availability

Summary statistics from the GWAS are available at https://strug.research.sickkids.ca/GWAS_Summary_Public/gwas.public.txt.gz. The calculated covariance matrix for summary statistics from the GWAS is available at https://strug.research.sickkids.ca/GWAS_Summary_Public/ld_matrix_chr3_195560_195935Kbp.tar.gz. R scripts to enable the extensions here can be found at https://github.com/FanWang0216/SimpleSum2Colocalization and https://github.com/naim-panjwani/LocusFocus. Access to the RNA-sequencing data from the nasal epithelial are available through the CF Canada-SickKids Program for Individualized Therapy Biobank https://lab.research.sickkids.ca/cfit/.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.12.012.

Web resources

FastQC (ver. 0.11.5), https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
GTEx Portal (release v.8), https://www.gtexportal.org

LocusFocus, https://locusfocus.research.sickkids.ca

Online Mendelian Inheritance in Man, https://www.omim.org/

PredictDB Data Repository, https://predictdb.org

Trim Galore (ver. 0.4.4), https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

References

1. Cutting, G.R. (2010). Modifier genes in Mendelian disorders: the example of cystic fibrosis. Ann. N Y Acad. Sci. *1214*, 57–69.

2. Vanscoy, L.L., Blackman, S.M., Collaco, J.M., Bowers, A., Lai, T., Naughton, K., Algire, M., McWilliams, R., Beck, S., Hoover-Fong, J., et al. (2007). Heritability of lung disease severity in cystic fibrosis. Am. J. Respir. Crit. Care Med. *175*, 1036–1043.

3. Kreda, S.M., Davis, C.W., and Rose, M.C. (2012). CFTR, mucins, and mucus obstruction in cystic fibrosis. Cold Spring Harb. Perspect. Med. *2*, a009589. https://doi.org/10.1101/cshperspect.a009589.

4. 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. https://doi.org/10.1371/journal.pgen.1004383.

5. 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. https://doi.org/10.1016/j.ajhg.2016.10.003.

6. Pickrell, J.K., Berisa, T., Liu, J.Z., Ségurel, L., Tung, J.Y., and Hinds, D.A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. Nat. Genet. *48*, 709–717. https://doi.org/10.1038/ng.3570.

7. Barbeira, A.N., Dickinson, S.P., Bonazzola, R., Zheng, J., Wheeler, H.E., Torres, J.M., Torstenson, E.S., Shah, K.P., Garcia, T., Edwards, T.L., et al.; GTEx Consortium (2018). Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat. Commun. *9*, 1825. https://doi.org/10.1038/s41467-018-03621-1.

8. Gong, J., Wang, F., Xiao, B., Panjwani, N., Lin, F., Keenan, K., Avolio, J., Esmaeili, M., Zhang, L., He, G., et al. (2019). Genetic association and transcriptome integration identify contributing genes and tissues at cystic fibrosis modifier loci. PLoS Genet. *15*, e1008007. https://doi.org/10.1371/journal.pgen.1008007.

9. 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. https://doi.org/10.1371/journal.pgen.1006646.

10. Barbeira, A.N., Pividori, M., Zheng, J., Wheeler, H.E., Nicolae, D.L., and Im, H.K. (2019). Integrating predicted transcriptome from multiple tissues improves association detection. PLoS Genet. *15*, e1007889. https://doi.org/10.1371/journal.pgen.1007889.

11. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., and Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. *48*, 481–487. https://doi.org/10.1038/ng.3538.

12. Wu, Y., Zeng, J., Zhang, F., Zhu, Z., Qi, T., Zheng, Z., Lloyd-Jones, L.R., Marioni, R.E., Martin, N.G., Montgomery, G.W., et al. (2018). Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. Nat. Commun. *9*, 918. https://doi.org/10.1038/s41467-018-03371-0.

13. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. *48*, 245–252. https://doi.org/10.1038/ng.3506.

14. Dobbyn, A., Huckins, L.M., Boocock, J., Sloofman, L.G., Glicksberg, B.S., Giambartolomei, C., Hoffman, G.E., Perumal, T.M., Girdhar, K., Jiang, Y., et al.; CommonMind Consortium (2018). Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-localization with Schizophrenia GWAS. Am. J. Hum. Genet. *102*, 1169–1184. https://doi.org/10.1016/j.ajhg.2018.04.011.

15. 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. https://doi.org/10.1038/ng.3795.

16. LeBlanc, M., Zuber, V., Thompson, W.K., Andreassen, O.A., Frigessi, A., Andreassen, B.K.; and Schizophrenia and Bipolar Disorder Working Groups of the Psychiatric Genomics Consortium (2018). A correction for sample overlap in genome-wide association studies in a polygenic pleiotropy-informed framework. BMC Genomics *19*, 494. https://doi.org/10.1186/s12864-018-4859-7.

17. Consortium, W.T.C.C.; and Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447*, 661–678.

18. Wacholder, S., Chanock, S., Garcia-Closas, M., El Ghormli, L., and Rothman, N. (2004). Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J. Natl. Cancer Inst. *96*, 434–442.

19. Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A simple new approach to variable selection in regression, with application to genetic fine mapping. J. R. Stat. Soc. Series B Stat. Methodol. *82*, 1273–1300.

20. Wallace, C. (2021). A more accurate method for colocalisation analysis allowing for multiple causal variants. PLoS Genet. *17*, e1009440.

21. 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. https://doi.org/10.1038/ng.3367.

22. Fryett, J.J., Morris, A.P., and Cordell, H.J. (2020). Investigation of prediction accuracy and the impact of sample size, ancestry, and tissue in transcriptome-wide association studies. Genet. Epidemiol. *44*, 425–441. https://doi.org/10.1002/gepi.22290.

23. 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. https://doi.org/10.1038/s41588-019-0385-z.

24. Wang, S., McCormick, T.H., and Leek, J.T. (2020). Post-prediction inference. bioRxiv. https://doi.org/10.1101/2020.01.21.914002.

25. Panjwani, N., Wang, F., Wang, C., He, G., Mastromatteo, S., Bao, A., Gong, J., Rommens, J.M., Sun, L., and Strug, L.J. (2020). LocusFocus: A web-based colocalization tool for the annotation and functional follow-up of GWAS. bioRxiv. https://doi.org/10.1101/2020.01.02.891291.

26. Corvol, H., Blackman, S.M., Boëlle, P.-Y., Gallins, P.J., Pace, R.G., Stonebraker, J.R., Accurso, F.J., Clement, A., Collaco, J.M., Dang, H., et al. (2015). Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. Nat. Commun. 6, 8382. https://doi.org/10.1038/ncomms9382.

27. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al.; GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45, 580–585. https://doi.org/10.1038/ng.2653.

28. He, X., Fuller, C.K., Song, Y., Meng, Q., Zhang, B., Yang, X., and Li, H. (2013). Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS. Am. J. Hum. Genet. 92, 667–680.

29. 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.

30. Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E., et al. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics 21, 650–659.

31. Kryuchkova-Mostacci, N., and Robinson-Rechavi, M. (2017). A benchmark of gene expression tissue-specificity metrics. Brief. Bioinform. 18, 205–214.

32. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G.R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. 44, 955–959. https://doi.org/10.1038/ng.2354.

33. Sun, L., Rommens, J.M., Corvol, H., Li, W., Li, X., Chiang, T.A., Lin, F., Dorfman, R., Busson, P.F., Parekh, R.V., et al. (2012). Multiple apical plasma membrane constituents are associated with susceptibility to meconium ileus in individuals with cystic fibrosis. Nat. Genet. 44, 562–569. https://doi.org/10.1038/ng.2221.

34. Liu, J.Z., McRae, A.F., Nyholt, D.R., Medland, S.E., Wray, N.R., Brown, K.M., Hayward, N.K., Montgomery, G.W., Visscher, P.M., Martin, N.G., Macgregor, S.; and AMFS Investigators (2010). A versatile gene-based test for genome-wide association studies. Am. J. Hum. Genet. 87, 139–145. https://doi.org/10.1016/j.ajhg.2010.06.009.

35. Bakshi, A., Zhu, Z., Vinkhuyzen, A.A.E., Hill, W.D., McRae, A.F., Visscher, P.M., and Yang, J. (2016). Fast set-based association analysis using summary data from GWAS identifies novel gene loci for human complex traits. Sci. Rep. 6, 32894. https://doi.org/10.1038/srep32894.

36. Li, M.X., Gui, H.S., Kwan, J.S., and Sham, P.C. (2011). GATES: a rapid and powerful gene-based association test using extended Simes procedure. Am. J. Hum. Genet. 88, 283–293. https://doi.org/10.1016/j.ajhg.2011.01.019.

37. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z., and Bergmann, S. (2016). Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. PLoS Comput. Biol. 12, e1004714. https://doi.org/10.1371/journal.pcbi.1004714.

38. Han, B., and Eskin, E. (2011). Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. Am. J. Hum. Genet. 88, 586–598. https://doi.org/10.1016/j.ajhg.2011.04.014.

39. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R.; and 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature 526, 68–74. https://doi.org/10.1038/nature15393.

40. Zhu, X., Feng, T., Tayo, B.O., Liang, J., Young, J.H., Franceschini, N., Smith, J.A., Yanek, L.R., Sun, Y.V., Edwards, T.L., et al.; COGENT BP Consortium (2015). Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. Am. J. Hum. Genet. 96, 21–36. https://doi.org/10.1016/j.ajhg.2014.11.011.

41. Chen, H., Conomos, M.P., and Chen, M.H. (2019). Package 'GMMAT'.

42. Park, H., Li, X., Song, Y.E., He, K.Y., and Zhu, X. (2016). Multivariate analysis of anthropometric traits using summary statistics of genome-wide association studies from GIANT Consortium. PLoS ONE 11, e0163912.

43. Province, M.A., and Borecki, I.B. (2013). A correlated meta-analysis strategy for data mining "OMIC" scans. Pac. Symp. Biocomput. 2013, 236–246.

44. Lin, D.-Y., and Sullivan, P.F. (2009). Meta-analysis of genome-wide association studies with overlapping subjects. Am. J. Hum. Genet. 85, 862–872. https://doi.org/10.1016/j.ajhg.2009.11.001.

45. Eckford, P.D.W., McCormack, J., Munsie, L., He, G., Stanojevic, S., Pereira, S.L., Ho, K., Avolio, J., Bartlett, C., Yang, J.Y., et al. (2019). The CF Canada-Sick Kids Program in individual CF therapy: A resource for the advancement of personalized medicine in CF. J. Cyst. Fibros. 18, 35–43.

46. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

47. DeLuca, D.S., Levin, J.Z., Sivachenko, A., Fennell, T., Nazaire, M.-D., Williams, C., Reich, M., Winckler, W., and Getz, G. (2012). RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics 28, 1530–1532.

48. Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11, R25.

49. Ongen, H., Buil, A., Brown, A.A., Dermitzakis, E.T., and Delaneau, O. (2016). Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics 32, 1479–1485. https://doi.org/10.1093/bioinformatics/btv722.

50. Gogarten, S.M., Sofer, T., Chen, H., Yu, C., Brody, J.A., Thornton, T.A., Rice, K.M., and Conomos, M.P. (2019). Genetic association testing using the GENESIS R/Bioconductor package. Bioinformatics 35, 5346–5348.

51. Conomos, M.P., Miller, M.B., and Thornton, T.A. (2015). Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genet. Epidemiol. 39, 276–293.

52. Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat. Protoc. 7, 500–507.

53. Taylor, C., Commander, C.W., Collaco, J.M., Strug, L.J., Li, W., Wright, F.A., Webel, A.D., Pace, R.G., Stonebraker, J.R., Naughton, K., et al. (2011). A novel lung disease phenotype adjusted for mortality attrition for cystic fibrosis genetic modifier studies. Pediatr. Pulmonol. *46*, 857–869.

54. Kesimer, M., Ehre, C., Burns, K.A., Davis, C.W., Sheehan, J.K., and Pickles, R.J. (2013). Molecular organization of the mucins and glycocalyx underlying mucus transport over mucosal surfaces of the airways. Mucosal Immunol. *6*, 379–392. https://doi.org/10.1038/mi.2012.81.

55. Ali, M., Lillehoj, E.P., Park, Y., Kyo, Y., and Kim, K.C. (2011). Analysis of the proteome of human airway epithelial secretions. Proteome Sci. *9*, 4.

56. Reid, C.J., Gould, S., and Harris, A. (1997). Developmental expression of mucin genes in the human respiratory tract. Am. J. Respir. Cell Mol. Biol. *17*, 592–598.

57. Vsevolozhskaya, O.A., Shi, M., Hu, F., and Zaykin, D.V. (2020). DOT: Gene-set analysis by combining decorrelated association statistics. PLoS Comput. Biol. *16*, e1007819. https://doi.org/10.1371/journal.pcbi.1007819.

58. Zhao, Y., and Sun, L. (2019). On set-based association tests: Insights from a regression using summary statistics. Can. J. Stat. *49*, 754–770.

59. Derkach, A., Lawless, J.F., and Sun, L. (2014). Pooled Association Tests for Rare Genetic Variants: A Review and Some New Results. Stat. Sci. *29*, 302–321. https://doi.org/10.1214/13-STS456.

60. 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.

Supplemental information

A flexible summary statistics-based colocalization method with application to the mucin cystic fibrosis lung disease modifier locus

Fan Wang, Naim Panjwani, Cheng Wang, Lei Sun, and Lisa J. Strug

Supplemental Figures
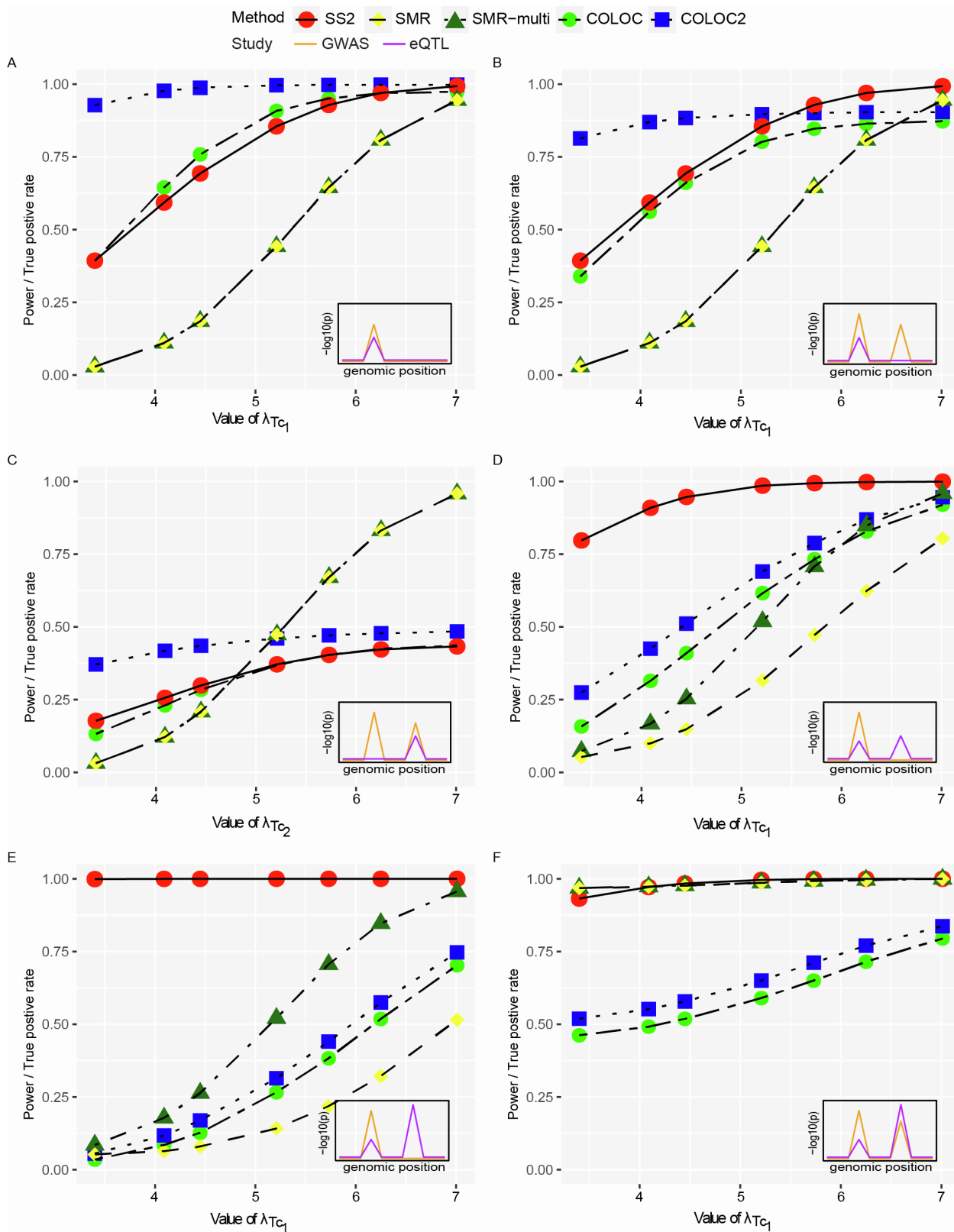
Figure S1. Empirical power of SS2, SMR, SMR-multi and true positive rates of COLOC and COLOC2 for testing a single gene for colocalization under six alternative scenarios: (A) one shared SNP for GWAS and eQTL study, (B) two independent GWAS SNPs, and one eQTL SNP is shared with the first GWAS SNP, (C) two independent GWAS SNPs, and one eQTL SNP is shared with the second GWAS SNP, (D) two independent eQTL SNPs, one GWAS SNP colocalizes with the eQTL SNP, given the non-overlapped eQTL association is low, (E) two independent eQTL SNPs, one GWAS SNP colocalizes with the eQTL SNP, given the non-overlapped eQTL association is strong, (F) two independent GWAS SNPs, two independent eQTL SNPs and both SNPs are shared. The LD pattern is simulated following the *SLC6A14* locus at chromosome X. For SS2, SMR[1] and SMR-multi,[2] the nominal type 1 error rate is set at alpha = 0.05. For COLOC[3] and COLOC2,[4] the false positive rates are controlled by applying the default 0.8 threshold (as recommended in [4]) for the colocalization posterior probability. In total, 10^4 replications are simulated to obtain the empirical power and true positive rates. The x-axis represents parameter values for $\lambda_{T_{c1}}$ or $\lambda_{T_{c2}}$ (3.4, 4.09, 4.45, 5.21, 5.73 and 7.01) such that 0.01, 0.05, 0.1, 0.3, 0.5 and 0.9 power is achieved to detect the eQTL association at significance level 10^{-8}. The subplot in the bottom right of each plot provides a general visualization of GWAS (orange line) and eQTL (purple line) colocalization patterns in a region of interest.
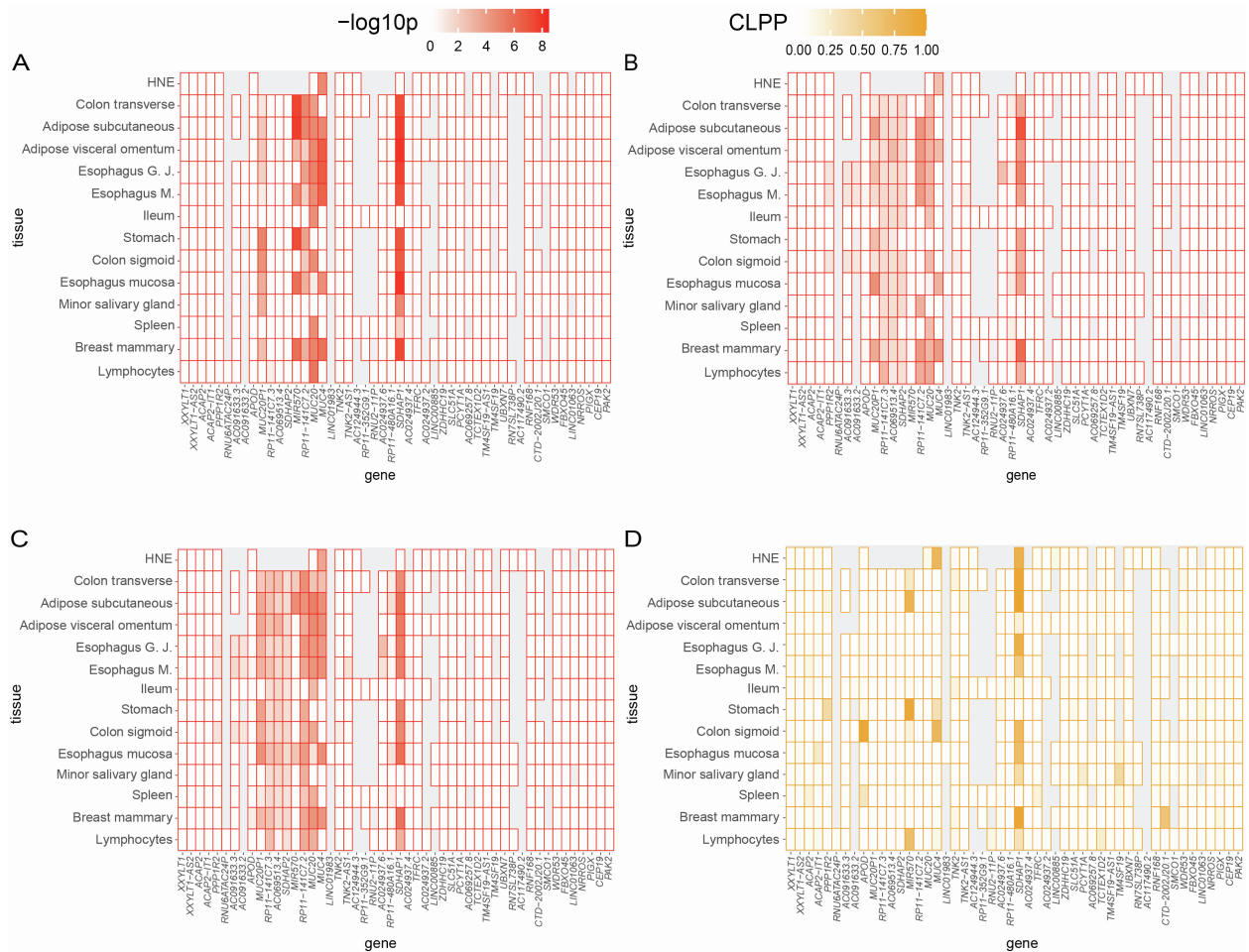
Figure S2. Heatmaps of colocalization evidence across genes and tissues for a 1Mb region encompassing the peak lung-associated variants at chromosome 3: (A) SS2, (B) SMR, (C) SMR-multi, and (D) COLOC. In each panel, each cell shows the colocalization evidence for the specified tissue and gene calculated from SNPs within 0.1Mb of the peak GWAS variant. The genes on the x-axis are annotated by GENCODE version 26 for hg38 GTEX V8 to 1Mb on either side of the peak GWAS variant and are ordered by their chromosomal positions. Grey indicates insufficient expression levels attained for the gene in the tissue under study. (A): the color intensity corresponds to the SS2 colocalization evidence as measured by –log$_{10}$(SS2 p-

value), with red representing $-\log_{10}(p) = 8.5$ and white representing eQTL evidence for the corresponding gene and tissue does not pass the stage 1 test. (B) and (C): the color intensity corresponds to the SMR and SMR-multi colocalization evidence as measured by $-\log_{10}(\text{SMR p-value})$ and $-\log_{10}(\text{SMR-multi p-value})$, respectively; with red representing $-\log_{10}(p) = 8.5$ and white representing eQTL evidence for the corresponding gene and tissue does not pass the eQTL p-value threshold (5×10^{-8}). (D): the color intensity corresponds to the COLOC colocalization evidence as measured by colocalization posterior probability (CLPP) ranging from 0 to 1. The eQTL analyses used for all gene/tissue pairs are those conducted by GTEx[5] version 8 release, except for the HNE eQTLs. eQTL analyses in HNE is conducted using FastQTL[6] with RNA-sequencing of HNE from 94 CF Canadians enrolled in the Canadian CF Gene Modifier study. Esophagus G.J represents Esophagus Gastroesophageal Junction; Esophagus M. represents Esophagus Muscularis; Ileum represents Small Intestine Terminal Ileum; Lymphocytes represents Cells EBV-transformed lymphocytes.
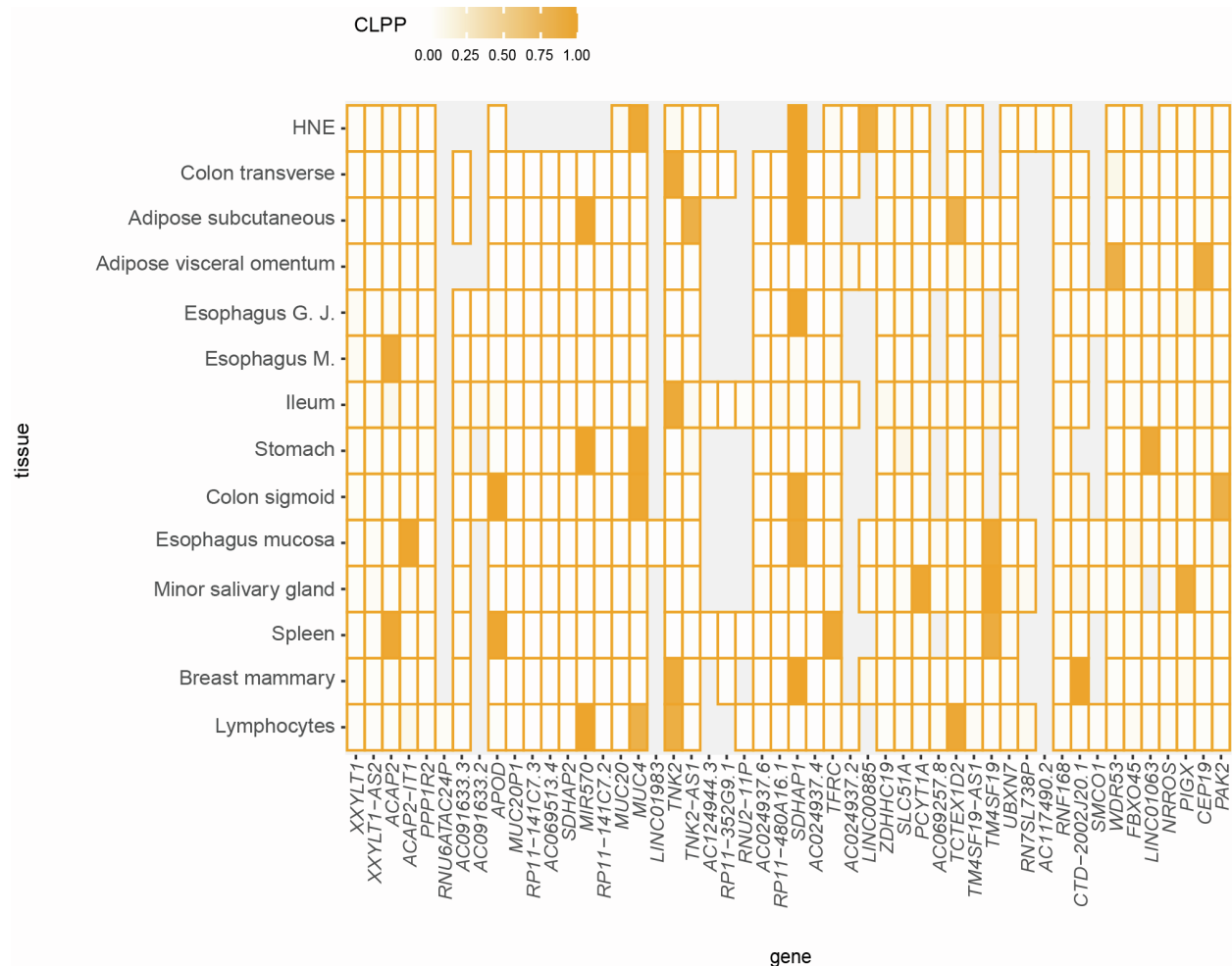
Figure S3. Heatmap of colocalization evidence across genes and tissues for a 1Mb region encompassing the peak lung-associated variants at chromosome 3 by COLOC2. COLOC2 analysis is conducted based on the likelihood from each single gene-by-tissue pair, calculated from SNPs within 0.1Mb of the peak GWAS variants. The color intensity corresponds to the COLOC2 colocalization evidence as measured by colocalization posterior probability (CLPP) ranging from 0 to 1. The genes on the x-axis are annotated by GENCODE version 26 for hg38 GTEX V8 and are ordered by their chromosomal positions. Grey indicates insufficient expression levels attained for the gene in the tissue under study. The eQTL data used for all gene/tissue pairs was sourced from GTEx version 8 release, with exception to the HNE eQTLs. eQTL analysis in HNE is conducted using FastQTL with RNA-sequencing of HNE from 94 CF

Canadians enrolled in the Canadian CF Gene Modifier study. Esophagus G.J represents Esophagus Gastroesophageal Junction; Esophagus M. represents Esophagus Muscularis; Ileum represents Small Intestine Terminal Ileum; Lymphocytes represents Cells EBV-transformed lymphocyte.

Supplemental Tables

Null Scenarios	H_{01}	H_{02}	$H_{03,(1)}$	$H_{03,(2)}$	$H_{04,(1)}$	$H_{04,(2)}$
Parameter Values	$\lambda_{Z_c} = 0;$ $\lambda_{T_c} = 0$	$\lambda_{Z_c} = 0;$ $\lambda_{T_c} = 5.48$	$\lambda_{Z_c} = 6.57;$ $\lambda_{T_c} = 0$	$\lambda_{Z_c} = 5.73$ $\lambda_{T_c} = 0$	$\lambda_{Z_c} = 6.57;$ $\lambda_{T_c} = 5.48$	$\lambda_{Z_c} = 5.73$ $\lambda_{T_c} = 5.48$

Table S1. Parameter settings under composite null hypothesis when conducting colocalization tests for a single gene with overlapping samples. Each column corresponds to a specific null scenario when there is no colocalization. The values of λ_{Z_c} and λ_{T_c} represent the standardized true effect size of a GWAS associated variant and an eQTL variant, respectively. For example, a GWAS association, λ_{Z_c} is set to be 0, 5.73, or 6.57 such that 0, 0.5 or 0.8 power is achieved to detect the signal at the significance level of 10^{-8}. If there is an eQTL, λ_{T_c} is set to be 5.48 such that 0.4 power is achieved to detect signal at the significance level of 10^{-8}. H_{01} represents the scenario when there are no SNP-phenotype associations and no eQTL; H_{02} represents the scenario when there are no SNP-phenotype associations but eQTLs are present; $H_{03,(1)}$ and $H_{03,(2)}$ represent two scenarios where SNP-phenotype associations are present but no eQTL; $H_{04,(1)}$ and $H_{04,(2)}$ represent two scenarios where both SNP-phenotype association and eQTL are present, but occur at two independent SNPs. See Section *Simulation Details for Overlapping Samples* for other simulation details.

Null Scenarios	Type I error of SS2				
	Correlation =0	Correlation =0.3	Correlation =0.5	Correlation =0.7	Correlation =0.9
H_{01}	0.0033	0.0031	0.0033	0.0031	0.0030
H_{02}	0.0430	0.0426	0.0425	0.0430	0.0425
$H_{03,(1)}$	0.0243	0.0271	0.0266	0.0269	0.0259
$H_{03,(2)}$	0.0178	0.0202	0.0199	0.0190	0.0202
$H_{04,(1)}$	0.0257	0.0281	0.0277	0.0277	0.0269
$H_{04,(2)}$	0.0187	0.0208	0.0206	0.0209	0.0217

Table S2. Empirical type I error rates of SS2 when 50% of the participants in the eQTL study overlap with the GWAS study with varying phenotypic correlations. The LD pattern at the simulated region follows that at the *MUC20/MUC4* locus. Simulations are performed under the null hypothesis of no colocalization and with varying phenotypic correlation (0.3, 0.5, 0.7 and 0.9). For comparison, the type I error rate when no samples are overlapping is also shown (correlation = 0). Table S1 defines the parameters used under each null scenario. The nominal type I error is set at alpha = 0.05. In total, 10^4 replications are simulated for each null scenario. See Section *Simulation Details for Overlapping Samples* for other simulation details.

Null Scenarios	Type I error of SS2			
	correlation=0.3	correlation=0.5	correlation=0.7	correlation=0.9
H_{01}	0.0015	0.0024	0.0038	0.0039
H_{02}	0.0373	0.0375	0.0385	0.0382
$H_{03,(1)}$	0.0229	0.0242	0.0241	0.0224
$H_{03,(2)}$	0.0241	0.0251	0.0254	0.0231
$H_{04,(1)}$	0.0183	0.0181	0.0175	0.0174
$H_{04,(2)}$	0.0188	0.0186	0.0176	0.0185

Table S3. Empirical type I error rates of SS2 when 100% of the participants in the eQTL study overlap with the GWAS study, with varying phenotypic correlations.

The LD pattern at the simulated region follows that at the *MUC20/MUC4 locus.* Simulations are performed under the null hypothesis of no colocalization and with varying phenotypic correlation (0.3, 0.5, 0.7 and 0.9). Table S1 defines the parameters used under each null scenario. The nominal type I error is set at alpha = 0.05. In total, 10^4 replications are simulated for each null scenario. See Section *Simulation Details for Overlapping Samples* for other simulation details.

Null Scenarios	Type I error of SS2				
	eQTL sample size =200	eQTL sample size =200	eQTL sample size =300	eQTL sample size =400	eQTL sample size =500
H_{01}	0.0037	0.0040	0.0053	0.0065	0.0086
H_{02}	0.0495	0.0478	0.0487	0.0504	0.0493
$H_{03,(1)}$	0.0266	0.0265	0.0227	0.0255	0.0243
$H_{03,(2)}$	0.0288	0.0274	0.0233	0.0263	0.0255
$H_{04,(1)}$	0.0199	0.0165	0.0222	0.0202	0.0261
$H_{04,(2)}$	0.0190	0.0166	0.0218	0.0210	0.0255

Table S4. Empirical type I error rates of SS2 when 100% of the participants in the eQTL study overlap with the GWAS study, with phenotypic correlation 0.5, and varying sample sizes of the eQTL study. The LD pattern at the simulated region follows that at the *MUC20/MUC4* locus. Simulations are performed under the null hypothesis of no colocalization and with varying eQTL sample size (100, 200, 300, 400 or 500). Table S1 defines the parameters used under each null scenario. The nominal type I error is set at alpha = 0.05. In total, 10^4 replications are simulated for each null scenario. See Section *Simulation Details for Overlapping Samples* for other simulation details.

Null Scenarios	Type I error of SS2				
	eQTL sample size =100	eQTL sample size =200	eQTL sample size =300	eQTL sample size =400	eQTL sample size =500
H_{01}	0.0037	0.009	0.0122	0.0177	0.0216
H_{02}	0.0495	0.0502	0.0122	0.0572	0.0576
$H_{03,(1)}$	0.0266	0.0258	0.0122	0.024	0.0252
$H_{03,(2)}$	0.0288	0.0276	0.0122	0.0262	0.0262
$H_{04,(1)}$	0.0199	0.0185	0.0122	0.0209	0.0296
$H_{04,(2)}$	0.019	0.019	0.0122	0.0211	0.0304

Table S5. Empirical type I error rates of SS2 when 100% of the participants in the eQTL study overlap with the GWAS study, with strong phenotypic correlation 0.9, and varying sample sizes of the eQTL study. The LD pattern at the simulated region follows that at the *MUC20/MUC4* locus. Simulations are performed under the null hypothesis of no colocalization and with varying eQTL sample size (100, 200, 300, 400 or 500). Table S1 defines the parameters used under each null scenario. The nominal type I error is set at alpha = 0.05. In total, 10^4 replications are simulated for each null scenario. See Section *Simulation Details for Overlapping Samples* for other simulation details.

Null Scenarios	Parameter Values
H_{01}	$\lambda_{Z_c} = 0;\ \lambda_{T_c} = 0$
H_{02}	$\lambda_{Z_c} = 0;\ \lambda_{T_c} = 7.01$
H_{03}	$\lambda_{Z_c} = 5.73;\ \lambda_{T_c} = 0$
H_{04}	$\lambda_{Z_c} = 5.73;\ \lambda_{T_c} = 7.01$

Table S6. Parameter settings under composite null hypothesis when conducting colocalization tests for a single gene with independent samples. Each row corresponds to a specific null scenario when there is no colocalization. H_{01} represents the scenario when there is no SNP-phenotype association and no eQTL; H_{02} represents the scenario when there is no SNP-phenotype association but eQTLs are present; H_{03} represents the scenario where SNP-phenotype associations are present but no eQTL; H_{04} represents the scenario where both SNP-phenotype association and eQTL are present, but occur at two independent SNPs. Values of λ_{Z_c} and λ_{T_c} represent the standardized true effect size of a GWAS associated variant and an eQTL variant, respectively. GWAS association, λ_{Z_c} is set to be 5.73 such that 0.5 power is achieved to detect the signal at the significance level of 10^{-8}. If there is an eQTL, λ_{T_c} is set to be 7.01 such that 0.9 power is achieved to detect the signal at the significance level of 10^{-8}. See Section *Simulation Details for a Single Gene-by-Tissue Pair* for other simulation details.

Scenarios	Parameter Values
Scenario 1: one shared SNP for GWAS and eQTL study (Figure 2A and Figure S1A)	$\lambda_{Z_{c1}} = 6.57;\ \lambda_{Z_{c2}} = 0$ $\lambda_{T_{c1}} = 3.40, 4.45, 5.21, 5.73, 6.25\ \text{or}\ 7.01;$ $\lambda_{T_{c2}} = 0$
Scenario 2: two independent GWAS SNPs, and one eQTL SNP is shared with the first GWAS SNP (Figure 2B and Figure S1B)	$\lambda_{Z_{c1}} = 6.57;\ \lambda_{Z_{c2}} = 5.73$ $\lambda_{T_{c1}} = 3.40, 4.45, 5.21, 5.73, 6.25\ \text{or}\ 7.01;$ $\lambda_{T_{c2}} = 0$
Scenario 3: two independent GWAS SNPs, and one eQTL SNP is shared with the second GWAS SNP (Figure 2C and Figure S1C)	$\lambda_{Z_{c1}} = 6.57;\ \lambda_{Z_{c2}} = 5.73$ $\lambda_{T_{c1}} = 0;$ $\lambda_{T_{c2}} = 3.40, 4.45, 5.21, 5.73, 6.25\ \text{or}\ 7.01$
Scenario 4: two independent eQTL SNPs, one GWAS SNP colocalizes with the eQTL SNP; the non-overlapping eQTL association is mild (Figure 2D and Figure S1D)	$\lambda_{Z_{c1}} = 6.57;\ \lambda_{Z_{c2}} = 0$ $\lambda_{T_{c1}} = 3.40, 4.45, 5.21, 5.73, 6.25\ \text{or}\ 7.01;$ $\lambda_{T_{c2}} = 5.73$
Scenario 5: two independent eQTL SNPs, one GWAS SNP co-localizes with the eQTL SNP; the non-overlapping eQTL association is strong (Figure 2E and Figure S1E)	$\lambda_{Z_{c1}} = 6.57;\ \lambda_{Z_{c2}} = 0$ $\lambda_{T_{c1}} = 3.40, 4.45, 5.21, 5.73, 6.25\ \text{or}\ 7.01;$ $\lambda_{T_{c2}} = 7.01$

Scenario 6: two independent GWAS SNPs, two independent eQTL SNPs and both SNPs are shared (Figure 2F and Figure S1F)	$\lambda_{Z_{c1}} = 6.57; \lambda_{Z_{c2}} = 5.73$ $\lambda_{T_{c1}} = 3.40, 4.45, 5.21, 5.73, 6.25 \text{ or } 7.01;$ $\lambda_{T_{c2}} = 7.01$

Table S7. Parameter settings of the six alternative scenarios simulated to access the power/true positive rate of different methods. $\lambda_{Z_{c1}}$ and $\lambda_{Z_{c2}}$ denote the standardized true effect sizes of two GWAS associated variants, while $\lambda_{T_{c1}}$ and $\lambda_{T_{c2}}$ denote the standardized true effect sizes of two eQTL variants. For the GWAS associated variants, $\lambda_{Z_{c1}}$ is set to be 6.57 such that 0.8 power is achieved to detect the GWAS signal at the significance level 10^{-8}; $\lambda_{Z_{c2}}$ is set to be 0 or 5.73 such that 0 or 0.5 power is achieved to detect the GWAS signal at the significance level 10^{-8}. For the eQTL variants, $\lambda_{T_{c1}}$ is set to be 0, 3.4, 4.09, 4.45, 5.21, 5.73 or 7.01 such that 0, 0.01, 0.05, 0.1, 0.3, 0.5 or 0.9 power is achieved to detect the eQTL signal at the significance level 10^{-8}; same for the value of $\lambda_{T_{c2}}$. See Section *Simulation Details for a Single Gene-by-Tissue Pair* for other simulation details.

Locus	Proportion of genes with eQTL association but do not colocalize	FWER of SS2				
		100 genes	200 genes	300 genes	400 genes	500 genes
MUC20/MUC4	0%	0.0352	0.0348	0.0364	0.0367	0.0370
	20%	0.0240	0.0272	0.0274	0.0283	0.0283
	40%	0.0247	0.0267	0.0274	0.0291	0.0292
	60%	0.0213	0.0227	0.0253	0.0279	0.0291
	80%	0.0194	0.0223	0.0258	0.0293	0.0309
	100%	0.0213	0.0270	0.0293	0.0311	0.0332
SLC6A14	0%	0.0058	0.0055	0.0055	0.0053	0.0054
	20%	0.0050	0.0031	0.0025	0.0020	0.0018
	40%	0.0032	0.0021	0.0016	0.0014	0.0012
	60%	0.0026	0.0017	0.0014	0.0012	0.0010
	80%	0.0021	0.0014	0.0011	0.0010	0.0009
	100%	0.0018	0.0012	0.0010	0.0009	0.0008

Table S8. Empirical family-wise error rates (FWERs) of SS2 with different number of genes. The LD pattern at the simulated region follows that at the *MUC20/MUC4* and *SLC6A14* loci. Simulations are conducted using a different number of genes (columns)

and a varying proportion of genes with eQTL association (rows). The eQTL peaks are randomly generated from 6 different intervals (50%-60%, 60%-70%, 70%-80%, 80%-90%, 90%-95%, 95%-100% power is achieved to detect the eQTL association at the significance level of 10^{-8}) with probabilities defined by the proportion of the -log10(maximum eQTL p-value) within each interval observed at the corresponding locus. The height of the GWAS peak is set at 5.06 on the -log10p scale such that 10% power is achieved to detect the GWAS association at the significance level of 10^{-8}. None of the eQTL peaks colocalize with the GWAS peak for FWER evaluation. In total, 10^5 replications are simulated to evaluate FWER of 0.05. The empirical FWER is calculated by counting the proportion of 10^5 replications where at least one gene has a false colocalization claim. See Section *Simulation Details for Multiple Genes-Tissue Pairs* for other simulation details.

Locus	Proportion of genes with eQTL association but do not colocalize	FWER of SMR				
		100 genes	200 genes	300 genes	400 genes	500 genes
MUC20/MUC4	0%	0.0002	0.0004	0.0006	0.0007	0.0008
	20%	0.0021	0.0011	0.0008	0.0006	0.0004
	40%	0.0011	0.0005	0.0003	0.0002	0.0002
	60%	0.0006	0.0003	0.0002	0.0002	0.0001
	80%	0.0005	0.0003	0.0002	0.0001	0.0001
	100%	0.0004	0.0002	0.0002	7.00×10^{-5}	7.00×10^{-5}
SLC6A14	0%	0.0002	0.0003	0.0005	0.0005	0.0007
	20%	0.0052	0.0023	0.0016	0.0012	0.0009
	40%	0.0024	0.0012	0.0008	0.0006	0.0005
	60%	0.0016	0.0008	0.0005	0.0004	0.0003
	80%	0.0012	0.0006	0.0004	0.0002	0.0002
	100%	0.0009	0.0004	0.0003	0.0002	0.0001

Table S9. Empirical family-wise error rates (FWERs) of SMR with different number of genes. The LD pattern at the simulated region follows that at the *MUC20/MUC4* and *SLC6A14* loci. SMR is conducted under the default setting such that

a SNP is picked if only if the eQTL p-value is less than 5×10^{-8}. Simulations are conducted using a different number of genes (columns) and a varying proportion of genes with eQTL association (rows). The eQTL peaks are randomly generated from 6 different intervals (50%-60%, 60%-70%, 70%-80%, 80%-90%, 90%-95%, 95%-100% power is achieved to detect the eQTL association at the significance level of 10^{-8}) with probabilities defined by the proportion of the -log10(maximum eQTL p-value) within each interval observed at the corresponding locus. The height of the GWAS peak is set at 5.06 on the -log10p scale such that 10% power is achieved to detect the GWAS association at the significance level of 10^{-8}. None of the eQTL peaks colocalize with the GWAS peak for FWER evaluation. In total, 10^5 replications are simulated to evaluate FWER of 0.05. The empirical FWER is calculated by counting the proportion of 10^5 replications where at least one gene has a false colocalization claim. See Section *Simulation Details for Multiple Genes-Tissue Pairs* for other simulation details.

Locus	Proportion of genes with eQTL association but do not colocalize	FWER of SMR-multi				
		100 genes	200 genes	300 genes	400 genes	500 genes
MUC20/MUC4	0%	0.0002	0.0004	0.0006	0.0007	0.0008
	20%	0.0015	0.0006	0.0004	0.0003	0.0002
	40%	0.0006	0.0003	0.0002	0.0001	0.0001
	60%	0.0003	0.0002	0.0001	9.00×10^{-5}	7.00×10^{-5}
	80%	0.0002	0.0002	8.00×10^{-5}	6.00×10^{-5}	6.00×10^{-5}
	100%	0.0003	0.0002	7.00×10^{-5}	6.00×10^{-5}	5.00×10^{-5}
SLC6A14	0%	0.0002	0.0004	0.0005	0.0005	0.0007
	20%	0.0035	0.0015	0.0009	0.0007	0.0005
	40%	0.0014	0.0007	0.0004	0.0003	0.0002
	60%	0.0009	0.0004	0.0002	0.0002	0.0002
	80%	0.0007	0.0003	0.0002	0.0002	0.0002
	100%	0.0004	0.0002	0.0002	0.0001	0.0001

Table S10. Empirical family-wise error rates (FWERs) of Multi-SNP-based SMR test (SMR-multi) with different number of genes. The LD pattern at the simulated region follows that at the *MUC20/MUC4* and *SLC6A14* loci. Multi-SNP-based SMR test

(SMR-multi) is conducted under the default setting such that a SNP is picked if only if the eQTL p-value is less than 5×10^{-8}. Simulations are conducted using a different number of genes (columns) and a varying proportion of genes with eQTL association (rows). The eQTL peaks are randomly generated from 6 different intervals (50%-60%, 60%-70%, 70%-80%, 80%-90%, 90%-95%, 95%-100% power is achieved to detect the eQTL association at the significance level of 10^{-8}) with probabilities defined by the proportion of the -log10(maximum eQTL p-value) within each interval observed at the corresponding locus. The height of the GWAS peak is set at 5.06 on the -log10p scale such that 10% power is achieved to detect the GWAS association at the significance level of 10^{-8}. None of the eQTL peaks colocalize with the GWAS peak for FWER evaluation. In total, 10^5 replications are simulated to evaluate FWER of 0.05. The empirical FWER is calculated by counting the proportion of 10^5 replications where at least one gene has a false colocalization claim. See Section *Simulation Details for Multiple Genes-Tissue Pairs* for other simulation details.

Locus	Proportion of genes with eQTL association but do not colocalize	False positive rate of COLOC				
		100 genes	200 genes	300 genes	400 genes	500 genes
MUC20/MUC4	0%	0.0265	0.0489	0.0704	0.0912	0.1118
	20%	0.0223	0.0406	0.0587	0.0758	0.0927
	40%	0.0180	0.0320	0.0460	0.0597	0.0729
	60%	0.0134	0.0232	0.0330	0.0424	0.0515
	80%	0.0088	0.0142	0.0197	0.0252	0.0300
	100%	0.0009	0.0017	0.0022	0.0030	0.0037
SLC6A14	0%	0.0350	0.0634	0.0902	0.1157	0.1400
	20%	0.0269	0.0509	0.0735	0.0952	0.1162
	40%	0.0218	0.0408	0.0592	0.0766	0.0938
	60%	0.0163	0.0304	0.0441	0.0574	0.0703
	80%	0.0111	0.0201	0.0289	0.0373	0.0458
	100%	0.0057	0.0093	0.0128	0.0164	0.0198

Table S11. Empirical false positive rates of COLOC with different number of genes.

The LD pattern at the simulated region follows that at the *MUC20/MUC4* and *SLC6A14*

loci. Simulations are conducted using a different number of genes (columns) and a

varying proportion of genes with eQTL association (rows). The eQTL peaks are randomly generated from 6 different intervals (50%-60%, 60%-70%, 70%-80%, 80%-90%, 90%-95%, 95%-100% power is achieved to detect the eQTL association at the significance level of 10^{-8}) with probabilities defined by the proportion of the -log10(maximum eQTL p-value) within each interval observed at the corresponding locus. The height of the GWAS peak is set at 5.06 on the -log10p scale such that 10% power is achieved to detect the GWAS association at the significance level of 10^{-8}. None of the eQTL peaks colocalize with the GWAS peak for FWER evaluation. In total, 10^5 replications are simulated to evaluate the false positive rates by applying the 0.8 threshold (as recommended by [4]) for the colocalization posterior probability. The empirical false positive rate for COLOC is calculated by counting the proportion of 10^5 replications where at least one gene has a false colocalization claim. See Section *Simulation Details for Multiple Genes-Tissue Pairs* for other simulation details.

Locus	Proportion of genes with eQTL association but do not colocalize	False positive rate of COLOC2				
		100 genes	200 genes	300 genes	400 genes	500 genes
MUC20/MUC4	0%	0.0020	0.0034	0.0048	0.0061	0.0073
	20%	0.0023	0.0023	0.0023	0.0023	0.0023
	40%	0.0025	0.0026	0.0027	0.0027	0.0027
	60%	0.0033	0.0033	0.0033	0.0034	0.0034
	80%	0.0038	0.0040	0.0040	0.0040	0.0040
	100%	0.0044	0.0044	0.0044	0.0044	0.0044
SLC6A14	0%	2.00×10^{-5}	5.00×10^{-5}	9.00×10^{-5}	0.0001	0.0002
	20%	0.0046	0.0047	0.0047	0.0048	0.0049
	40%	0.0054	0.0055	0.0055	0.0056	0.0056
	60%	0.0068	0.0069	0.0070	0.0070	0.0070
	80%	0.0079	0.0079	0.0080	0.0080	0.0080
	100%	0.0088	0.0088	0.0089	0.0089	0.0089

Table S12. Empirical false positive rates of COLOC2 with different number of genes. The LD pattern at the simulated region follows that at the _MUC20/MUC4_ and _SLC6A14_ loci. Simulations are conducted using a different number of genes (columns)

and a varying proportion of genes with eQTL association (rows). The eQTL peaks are randomly generated from 6 different intervals (50%-60%, 60%-70%, 70%-80%, 80%-90%, 90%-95%, 95%-100% power is achieved to detect the eQTL association at the significance level of 10^{-8}) with probabilities defined by the proportion of the $-\log 10$(maximum eQTL p-value) within each interval observed at the corresponding locus. The height of the GWAS peak is set at 5.06 on the $-\log 10 p$ scale such that 10% power is achieved to detect the GWAS association at the significance level of 10^{-8}. None of the eQTL peaks colocalize with the GWAS peak for FWER evaluation. In total, 10^5 replications are simulated to evaluate the false positive rates by applying the 0.8 threshold (as recommended by [4]) for the colocalization posterior probability. The empirical false positive rates for COLOC2 are calculated by counting the proportions of 10^5 replications where at least one gene has a false colocalization claim. See Section *Simulation Details for Multiple Genes-Tissue Pairs* for other simulation details.

Locus	eQTL height for the 5% of genes that colocalize	Power of SS2				
		100 genes	200 genes	300 genes	400 genes	500 genes
MUC20/MUC4	5.48-5.73	0.8880	0.8668	0.8508	0.8379	0.8268
	5.73-5.98	0.8891	0.8671	0.8512	0.8381	0.8269
	5.98-6.25	0.8893	0.8672	0.8508	0.8376	0.8261
	6.25-6.57	0.8891	0.8667	0.8499	0.8368	0.8255
	6.57-7.01	0.8885	0.8658	0.8488	0.8352	0.8235
SLC6A14	5.48-5.73	0.7126	0.7104	0.7018	0.6938	0.6855
	5.73-5.98	0.7450	0.7337	0.7200	0.7090	0.6992
	5.98-6.25	0.7642	0.7468	0.7306	0.7183	0.7076
	6.25-6.57	0.7752	0.7547	0.7376	0.7247	0.7138
	6.57-7.01	0.7819	0.7601	0.7424	0.7295	0.7184

Table S13. Power of SS2 with different number of genes. The LD pattern at the simulated region follows that at the *MUC20/MUC4* and *SLC6A14* loci. The height of the GWAS peak is set at 5.06 on the -log10p scale such that 10% power is achieved to detect the GWAS association at the significance level of 10^{-8}. Simulations are conducted using a different number of genes (columns) and a varying range of the eQTL height for the 5%

of genes that colocalize (rows). The eQTL peaks are set with 5 different intervals such that 40%-50%, 50%-60%, 60%-70%, 70%-80%, 80%-90% power is achieved to detect the eQTL association at the significance level of 10^{-8}. The 95% of genes that do not colocalize are simulated under a mixture of H_{03} and H_{04}, and details are demonstrated in the Supplemental Methods. The definitions of H_{03} and H_{04} are provided in Table S6. In total, 10^5 replications are simulated to evaluate power at 0.05 significance level. The power is calculated by counting the proportion of 10^5 replications where at least one gene is correctly identified with colocalization. See Section *Simulation Details for Multiple Genes-Tissue Pairs* for other simulation details.

Locus	eQTL height for the 5% of genes that colocalize	Power of SMR				
		100 genes	200 genes	300 genes	400 genes	500 genes
MUC20/MUC4	5.48-5.73	0.8299	0.7907	0.7540	0.7257	0.7015
	5.73-5.98	0.8467	0.7993	0.7633	0.7359	0.7126
	5.98-6.25	0.8561	0.8072	0.7717	0.7448	0.7225
	6.25-6.57	0.8633	0.8149	0.7808	0.7542	0.7322
	6.57-7.01	0.8704	0.8234	0.7906	0.7646	0.7435
SLC6A14	5.48-5.73	0.8005	0.7486	0.7067	0.6749	0.6519
	5.73-5.98	0.8184	0.7599	0.7186	0.6879	0.6652
	5.98-6.25	0.8292	0.7699	0.7293	0.6995	0.6767
	6.25-6.57	0.8379	0.7796	0.7398	0.7102	0.6886
	6.57-7.01	0.8449	0.7886	0.7504	0.7213	0.7005

Table S14. Power of SMR with different number of genes. The LD pattern at the

simulated region follows that at the *MUC20/MUC4* and *SLC6A14* loci. SMR is

conducted under the default setting such that a SNP is picked if only if the eQTL p-value

is less than 5×10^{-8}. The height of the GWAS peak is set at 5.06 on the -log10p scale such

that 10% power is achieved to detect the GWAS association at the significance level of

10^{-8}. Simulations are conducted using a different number of genes (columns) and a varying range of the eQTL height for the 5% of genes that colocalize (rows). The eQTL peaks are set with 5 different intervals such that 40%-50%, 50%-60%, 60%-70%, 70%-80%, 80%-90% power is achieved to detect the eQTL association at the significance level of 10^{-8}. The 95% of genes that do not colocalize are simulated under a mixture of H_{03} and H_{04}, and details are demonstrated in the Supplemental Methods. The definitions of H_{03} and H_{04} are provided in Table S6. In total, 10^5 replications are simulated to evaluate power at 0.05 significance level. The power is calculated by counting the proportion of 10^5 replications where at least one gene has correctly identified with colocalization. See Section *Simulation Details for Multiple Genes-Tissue Pairs* for other simulation details.

Locus	eQTL height for the 5% of genes that colocalize	Power of SMR-multi				
		100 genes	200 genes	300 genes	400 genes	500 genes
MUC20/MUC4	5.48-5.73	0.7955	0.7432	0.6964	0.6615	0.6316
	5.73-5.98	0.8077	0.7477	0.7009	0.6655	0.6363
	5.98-6.25	0.8133	0.7510	0.7050	0.6702	0.6410
	6.25-6.57	0.8162	0.7536	0.7074	0.6734	0.6453
	6.57-7.01	0.8161	0.7542	0.7094	0.6758	0.6481
SLC6A14	5.48-5.73	0.7659	0.7003	0.6483	0.6077	0.5785
	5.73-5.98	0.7789	0.7068	0.6548	0.6155	0.5859
	5.98-6.25	0.7850	0.7118	0.6604	0.6215	0.5923
	6.25-6.57	0.7883	0.7154	0.6652	0.6266	0.5983
	6.57-7.01	0.7884	0.7166	0.6686	0.6311	0.6036

Table S15. Power of Multi-SNP-based SMR test (SMR-multi) with different number of genes. The LD pattern at the simulated region follows that at the *MUC20/MUC4* and *SLC6A14* loci. SMR-multi is conducted under the default setting such that a SNP is picked if only if the eQTL p-value is less than 5×10^{-8}. The height of the GWAS peak is set at 5.06 on the -log10p scale such that 10% power is achieved to detect the GWAS

association at the significance level of 10^{-8}. Simulations are conducted using a different

number of genes (columns) and a varying range of the eQTL height for the 5% of genes

that colocalize (rows). The eQTL peaks are set with 5 different intervals such that 40%-

50%, 50%-60%, 60%-70%, 70%-80%, 80%-90% power is achieved to detect the eQTL

association at the significance level of 10^{-8}. The 95% of genes that do not colocalize are

simulated under a mixture of H_{03} and H_{04}, and details are demonstrated in the

Supplemental Methods. The definitions of H_{03} and H_{04} are provided in Table S6. In

total, 10^5 replications are simulated to evaluate power at 0.05 significance level. The

power is calculated by counting the proportion of 10^5 replications where at least one gene

is correctly identified with colocalization. See Section *Simulation Details for Multiple*

Genes-Tissue Pairs for other simulation details.

Locus	eQTL height for the 5% of genes that colocalize	True positive rate of COLOC2				
		100 genes	200 genes	300 genes	400 genes	500 genes
MUC20/MUC4	5.48-5.73	0.4976	0.5516	0.5783	0.5968	0.6105
	5.73-5.98	0.5193	0.5690	0.5942	0.6118	0.6249
	5.98-6.25	0.5353	0.5817	0.6051	0.6219	0.6343
	6.25-6.57	0.5477	0.5908	0.6129	0.6287	0.6405
	6.57-7.01	0.5581	0.5981	0.6184	0.6333	0.6439
SLC6A14	5.48-5.73	0.5801	0.6258	0.6407	0.6494	0.6555
	5.73-5.98	0.5979	0.6364	0.6497	0.6581	0.6638
	5.98-6.25	0.6118	0.6443	0.6565	0.6643	0.6697
	6.25-6.57	0.6232	0.6516	0.6625	0.6693	0.6745
	6.57-7.01	0.6330	0.6572	0.6669	0.6734	0.6781

Table S16. True positive rates of COLOC2 with different number of genes. The LD

pattern at the simulated region follows that at the *MUC20/MUC4* and *SLC6A14* loci. The

height of the GWAS peak is set at 5.06 on the -log10p scale such that 10% power is

achieved to detect the GWAS association at the significance level of 10^{-8}. Simulations are

conducted using a different number of genes (columns) and a varying range of the eQTL

height for the 5% of genes that colocalize (rows). The eQTL peaks are set with 5

different intervals such that 40%-50%, 50%-60%, 60%-70%, 70%-80%, 80%-90% power

is achieved to detect the eQTL association at the significance level of 10^{-8}. The 95% of

genes that do not colocalize are simulated under a mixture of H_{03} and H_{04}, and details

are demonstrated in the Supplemental Methods. The definitions of H_{03} and H_{04} are

provided in Table S6. In total, 10^5 replications are simulated to evaluate the true positive

rates by applying the 0.8 threshold (as recommended by [4]) for the colocalization posterior

probability. The true positive rates for COLOC2 are calculated by counting the

proportion of 10^5 replications where at least one gene is correctly identified with

colocalization. See Section *Simulation Details for Multiple Genes-Tissue Pairs* for other

simulation details.

Null Scenarios	False positive rate of COLOC2			
	eQTL sample size = 200	eQTL sample size = 500	eQTL sample size = 1000	eQTL sample size = 2000
H_{01}	0.0022	0.0025	0.0024	0.0021
H_{02}	0.0468	0.0468	0.0467	0.0467
H_{03}	0.0880	0.0692	0.0550	0.0417
H_{04}	0.0012	0.0012	0.0012	0.0012

Table S17. False positive rates of COLOC2 for a single hypothesis test with different sample sizes for the eQTL study. The LD pattern at the simulated region follows that at the *SLC6A14* locus. Sample size is fixed at 2000 participants in the GWAS study, and is varied for the eQTL study (columns). Each row corresponds to a specific null scenario when there is no colocalization. See Table S6 for the explanations of each null scenario with the corresponding parameter settings. The false positive rates are calculated by applying the 0.8 threshold (as recommended by [4]) for the colocalization posterior probability. In total, 10^4 replications are simulated for each null scenario. See Section *Simulation Details for a Single Gene-by-Tissue Pair* for other simulation details.

Null Scenarios	Type I error of SMR		Type I error of SMR-multi	
	(eQTL p<0.01)	(eQTL p<0.05)	(eQTL p<0.01)	(eQTL p<0.05)
H_{01}	0.0067	0.0071	0.003	6.00×10^{-4}
H_{02}	0.0392	0.0392	0.0162	0.006
H_{03}	0.0766	0.0934	0.0843	0.0807
H_{04}	0.0404	0.0404	0.0483	0.0606

Table S18. Empirical type I error rates of SMR and Multi-SNP-based SMR test (SMR-multi) for a single hypothesis test under eQTL p-value thresholds 0.01 and 0.05. SMR and Multi-SNP-based SMR test (SMR-multi) are conducted under the setting such that a SNP is picked if only if the eQTL p-value is less than 0.01(eQTL p<0.01) or 0.05 (eQTL p<0.05). The LD pattern at the simulated region follows that at the *MUC20/MUC4* locus. Each row corresponds to a specific null scenario when there is no colocalization. Table S6 defines the parameters used under each null scenario. The nominal type I error is set at alpha = 0.05. In total, 10^4 replications are simulated for each null scenario.

Null Scenarios	Type I error of SS2				
	Correlation =0	Correlation =0.3	Correlation =0.5	Correlation =0.7	Correlation =0.9
H_{01}	0.0033	0.0091	0.0249	0.0422	0.0489
H_{02}	0.0430	0.052	0.0643	0.0884	0.1404
$H_{03,(1)}$	0.0243	0.0222	0.0254	0.0282	0.0320
$H_{03,(2)}$	0.0178	0.0300	0.0338	0.0405	0.0501
$H_{04,(1)}$	0.0257	0.0226	0.0258	0.0288	0.0317
$H_{04,(2)}$	0.0187	0.0312	0.0336	0.0397	0.0478

Table S19: Empirical type I error rates of SS2 when all of 2000 participants in the eQTL study overlap with the GWAS study, with varying phenotypic correlations.

The LD pattern at the simulated region follows that at the *MUC20/MUC4* locus. Simulations are performed under the null hypothesis of no colocalization and with varying phenotypic correlation (0.3, 0.5, 0.7 and 0.9). For comparison, the type I error rate when no samples are overlapping is also shown (correlation = 0). Table S1 defines the parameters used under each null scenario. The nominal type I error is set at alpha = 0.05. In total, 10^4 replications are simulated for each null scenario. See Section *Simulation Details for Overlapping Samples* for other simulation details.

Supplemental Methods

Covariance of Summary Statistics from Meta-analyses with Related Individuals within Sub-studies

Suppose the GWAS meta-analysis consists of C sub-studies with sample sizes $n_c, c = 1, 2, \ldots C$. Let $\hat{\beta}_c = (\hat{\beta}_{c,1}, \ldots, \hat{\beta}_{c,m})^\top$ and $Z_c = (Z_{c,1}, Z_{c,2}, \cdots, Z_{c,m})^\top$ denote the marginal effect sizes and Z-scores, respectively, from study c. Let $Z_{\text{meta}} = (Z_{\text{meta},1}, Z_{\text{meta},2}, \ldots Z_{\text{meta},m})^\top$, where $Z_{\text{meta},j}$ represents the Z-score obtained from the meta-analysis for SNP j. Then

$$Z_{meta,j} = \frac{\sum_{c=1}^C w_{c,j} \hat{\beta}_{c,j}}{\sqrt{\sum_{c=1}^C w_{c,j}}} \tag{2}$$

$$= \frac{\sum_{c=1}^C \sqrt{w_{c,j}} Z_{c,j}}{\sqrt{\sum_{c=1}^C w_{c,j}}}, \tag{3}$$

where $w_{c,j}$ represents the weight for study c and can take different forms (i.e. the inverse variance of $\hat{\beta}_{c,j}$). In particular, for the sample size weighted meta-analysis, $w_{c,j} = n_c$ and $Z_{meta,j} = \frac{\sum_{c=1}^C \sqrt{n_c} Z_{c,j}}{\sqrt{n \lambda_c}}$, where λ_c is the genomic control.[7] For SNP j and SNP l,

$$\text{cov}(Z_{meta,j}, Z_{meta,l}) = \text{cov}\left(\frac{\sum_{c=1}^C \sqrt{w_{c,j}} Z_{c,j}}{\sqrt{\sum_{c=1}^C w_{c,j}}}, \frac{\sum_{c=1}^C \sqrt{w_{c,l}} Z_{c,l}}{\sqrt{\sum_{c=1}^C w_{c,l}}}\right)$$

$$= \frac{\sum_{c=1}^C \sqrt{w_{c,j}} \sqrt{w_{c,l}}}{\sqrt{\sum_{c=1}^C w_{c,j}} \sqrt{\sum_{c=1}^C w_{c,l}}} \text{cov}(Z_{c,j}, Z_{c,l}). \tag{4}$$

Therefore, the covariance of meta-Z scores between two SNPs is a weighted sum of covariance measures for the sub-studies.

If all participants in the sub-study c are independent, we consider fitting the following single-marker model with additional covariates to obtain the marginal summary statistics:

$$Y = G_j\beta + X\alpha + \epsilon \text{ with } \epsilon \sim N(0, \sigma^2 I), \tag{5}$$

for SNP j. $Y = (y_1, \dots, y_{n_c})^{\mathsf{T}}$ denote a $n_c \times 1$ vector of phenotypic values, where each component y_i is the phenotype for the ith individual. Assume there are m SNPs at the locus, $G_j = (G_{j1}, \dots, G_{jn_c})^{\mathsf{T}}$ is a $n_c \times 1$ vector of genotype for SNP j, where each component G_{ij} is the genotype for the ith individual. X is a $n_c \times q$ matrix of covariates (i.e. sex and age) and α is a $q \times 1$ vector of fixed effects for the covariates, including the intercept.

We define $U_j = (G_j, X)$, which is a $n_c \times (q + 1)$ matrix that contains the genotype for SNP j and additional covariates. Let $B = (1, 0, 0, \dots, 0)$, a $1 \times (1 + q)$ vector with the first component to be 1 and other components to be 0. We obtain $\hat{\beta}_{c,j}$ and $Z_{c,j}$ by the least square method:

$$\hat{\beta}_{c,j} = B(U_j^{\mathsf{T}} U_j)^{-1} U_j^{\mathsf{T}} Y \ ; \tag{6}$$

$$Z_{c,j} = \frac{B(U_j^{\mathsf{T}} U_j)^{-1} U_j^{\mathsf{T}} Y}{\sqrt{\hat{\sigma}_j^2 B(U_j^{\mathsf{T}} U_j)^{-1} B^{\mathsf{T}}}}, \tag{7}$$

where $\hat{\sigma}_j^2$ is the maximum likelihood estimator (MLE) of σ_j in model (5). For SNP j

and SNP l,

$$\mathrm{cov}\left(Z_{c,j}, Z_{c,l}\right) = \mathrm{cov}\left(\frac{B\left(U_j^\mathsf{T} U_j\right)^{-1} U_j^\mathsf{T} Y}{\sqrt{\hat{\sigma}_j^2 B\left(U_j^\mathsf{T} U_j\right)^{-1} B^\mathsf{T}}}, \frac{B(U_l^\mathsf{T} U_l)^{-1} U_l^\mathsf{T} Y}{\sqrt{\hat{\sigma}_l^2 B(U_l^\mathsf{T} U_l)^{-1} B^\mathsf{T}}}\right)$$

$$= \frac{B\left(U_j^\mathsf{T} U_j\right)^{-1} U_j^\mathsf{T}}{\sqrt{\hat{\sigma}_j^2 B\left(U_j^\mathsf{T} U_j\right)^{-1} B^\mathsf{T}}} \sigma^2 I \frac{U_l(U_l^\mathsf{T} U_l)^{-1} B^\mathsf{T}}{\sqrt{\hat{\sigma}_l^2 B(U_l^\mathsf{T} U_l)^{-1} B^\mathsf{T}}},$$

where $\hat{\sigma}_i^2$ and $\hat{\sigma}_l^2$ approaches σ^2 in probability. By Slutsky's theorem, the

asymptotically covariance is

$$\mathrm{cov}\left(Z_{c,j}, Z_{c,l}\right) = \frac{B\left(U_j^\mathsf{T} U_j\right)^{-1} U_j^\mathsf{T} U_l(U_l^\mathsf{T} U_l)^{-1} B^\mathsf{T}}{\sqrt{B\left(U_j^\mathsf{T} U_j\right)^{-1} B^\mathsf{T} B(U_l^\mathsf{T} U_l)^{-1} B^\mathsf{T}}}. \tag{8}$$

Zheng et al[8] showed equation (8) can be simplified such that

$$\mathrm{cov}\left(Z_{c,j}, Z_{c,l}\right) = \frac{G_j^\mathsf{T}(I - P_X)G_l}{\sqrt{G_j^t(I - P_X)G_j}\sqrt{G_l^t(I - P_X)G_l}} \tag{9}$$

$$= \mathrm{cor}\left(\hat{\epsilon}_{G_j|X}, \hat{\epsilon}_{G_l|X}\right),$$

where $P_X = X(X^\mathsf{T} X)^{-1} X^\mathsf{T}$. $\hat{\epsilon}_{G_j|X}$ denotes the residuals obtained by regressing G_j on X,

for $j = 1, .. m$. In particular, if both phenotypes and genotypes are standardized, and

there are no additional covariates expect the genotypes $(B = 1$ and $U_j = G_j)$, equation

(9) can be simplified as

$$\mathrm{cov}\left(Z_{c,j}, Z_{c,l}\right) = \frac{G_j^\mathsf{T} G_l}{\sqrt{G_j^\mathsf{T} G_j G_l^\mathsf{T} G_l}}.$$

The asymptotic covariance of Z scores between two SNPs is their pairwise Pearson correlation coefficient.[8,9]

If the sub-study c contains related samples, the first approach to obtain the summary statistics (for SNP j) is considering a linear fixed-effect model:

$$Y = G_j\beta + X\alpha + \epsilon, \text{ with } \epsilon \sim N(0, \Sigma_\epsilon), \tag{10}$$

where Σ_ϵ denote the matrix that contains information about the sample relatedness (i.e. the kinship coefficient matrix). If Σ_ϵ is known, let H denote the Cholesky decomposition of Σ_ϵ^{-1} such that $H^t H = \Sigma_\epsilon^{-1}$. We define $Y^* = HY, X^* = HX, G_j^* = HG_j, U_j^* = HU_j$ and $\epsilon^* = H\epsilon$. By the generalized least square (GLS) method, the effect size and Z-score for SNP j is

$$
\begin{aligned}
\hat{\beta}_{c,j} &= B\left(U_j^{*\top}U_j^*\right)^{-1}U_j^{*\top}Y^* \\
&= B\left(U_j^\top \Sigma_\epsilon^{-1} U_j\right)^{-1}U_j^\top \Sigma_\epsilon^{-1}Y; \tag{11}
\end{aligned}
$$

$$
\begin{aligned}
Z_{c,j} &= \frac{B\left(U_j^{*\top}U_j^*\right)^{-1}U_j^* Y^*}{\sqrt{B\left(U_j^{*\top}U_j^*\right)^{-1}B^\top}} \\
&= \frac{B\left(U_j^\top \Sigma_\epsilon^{-1} U_j\right)^{-1}U_j^\top \Sigma_\epsilon^{-1}Y}{\sqrt{B\left(U_j^\top \Sigma_\epsilon^{-1} U_j\right)^{-1}B^\top}}. \tag{12}
\end{aligned}
$$

Similar to equation (9),

$$
\begin{aligned}
\text{cov}\left(Z_{c,j}, Z_{c,l}\right) &= \frac{G_j^{*\top}(I - P_{X^*})G_l^*}{\sqrt{G_j^*(I - P_{X^*})G_j^*}\sqrt{G_l^{*\top}(I - P_{X^*})G_l^*}}, \text{ with } P_{X^*} = X^*(X^{*\top}X^*)^{-1}X^{*\top}; \\
&= \frac{G_j^\top P^* G_l}{\sqrt{G_j^\top P^* G_j}\sqrt{G_l^\top P^* G_l}}, \text{ with } P^* = \Sigma_\epsilon^{-1} - \Sigma_\epsilon^{-1}X(X^\top \Sigma_\epsilon^{-1}X)^{-1}X^\top \Sigma_\epsilon^{-1}. \tag{13}
\end{aligned}
$$

If Σ_ϵ is unknown, we can get a consistent estimator of Σ_ϵ first,[10] say $\hat{\Sigma}_\epsilon$, and then

obtain the effect size and Z-score by using $\hat{\Sigma}_\epsilon$:

$$\hat{\beta}_{c,j}(\hat{\Sigma}_\epsilon) = B\left(U_j^\mathsf{T}\hat{\Sigma}_\epsilon^{-1}U_j\right)^{-1}U_j^\mathsf{T}\hat{\Sigma}_\epsilon^{-1}Y; \tag{14}$$

$$Z_{c,j}(\hat{\Sigma}_\epsilon) = \frac{B\left(U_j^\mathsf{T}\hat{\Sigma}_\epsilon^{-1}U_j\right)^{-1}U_j^\mathsf{T}\hat{\Sigma}_\epsilon^{-1}Y}{\sqrt{B\left(U_j^\mathsf{T}\hat{\Sigma}_\epsilon^{-1}U_j\right)^{-1}B^\mathsf{T}}}. \tag{15}$$

By Slutsky theorem, equation (13) is the asymptotic covariance of Z-scores. In practice,

we estimate the covariance by $\hat{\Sigma}_\epsilon$:

$$\text{cov}(Z_{c,j}, Z_{c,l}) = \frac{G_j^\mathsf{T}\hat{P}^*G_l}{\sqrt{G_j^\mathsf{T}\hat{P}^*G_j}\sqrt{G_l^\mathsf{T}\hat{P}^*G_l}}, \text{ with } \hat{P}^* = \hat{\Sigma}_\epsilon^{-1} - \hat{\Sigma}_\epsilon^{-1}X\left(X^\mathsf{T}\hat{\Sigma}_\epsilon^{-1}X\right)^{-1}X^\mathsf{T}\hat{\Sigma}_\epsilon^{-1}. \tag{16}$$

Secondly, the linear mixed-effect model (LMM) with a random intercept is another

commonly used approach for related samples:[11,12]

$$Y = G_j\beta + X\alpha + Q + \epsilon, \text{ with } Q \sim N\left(0, \sum_{k=1}^{K}\tau_k\Phi_k\right) \text{ and } \epsilon \sim N(0, \sigma^2 I). \tag{17}$$

Q is the random intercept; K is the number of correlation matrices (Φ_k) included for a

given individual and τ_k is the corresponding variance component parameter, as defined

in [11]. For example, Φ_1 can be the standard genetic correlation matrix (GRM). Then Φ_2

could be the environmental relationship matrix, which captures the variance associated

with shared environmental effects between family members.[13] Q and ϵ are assumed to

be independent. α no longer includes the intercept in this case since the equation is

supposed to have a random intercept. Excluding the intercept, X becomes a $n_c \times (q -$

1) matrix of covariates and $U_j = (G_j, X)$ which becomes a $n_c \times q$ matrix that contains the genotype for SNP j and covariates. Given the observed covariates X and G_j ($j = 1, \dots, m$), the marginal distribution of Y is $N(G_j\beta + X\alpha, \Sigma_\epsilon)$, where $\Sigma_\epsilon = \sum_{k=1}^{K} \tau_k \Phi_k + \sigma^2 I$.

When Σ_ϵ is known, the MLE of β based on the marginal distribution of Y can be written in the same form as the estimate in (11), which is also the minimum variance unbiased estimate.[14] Therefore, we can estimate the covariance of Z-scores according to equation (13) under this case, which is the same as the covariance derived under linear fixed-effect model in (10) except excluding the column corresponding to the intercept in X.

When Σ_ϵ is unknown, the variance parameters contained in Σ_ϵ can be estimated by EM algorithm for maximum likelihood (ML) estimates or restricted maximum likelihood (REML) estimates.[14,15] Then the effect size is calculated by equation (14) (without the intercept in X),[14] and this procedure has been implemented in the R package nlme.[16] Therefore, the Z-scores and its covariance can be calculated by equation (15) and equation (16) (without the intercept in X), respectively, using variance component estimates from MLE or REML. Other types of summary statistics such as the score-based statistics have been implemented in R packages GMMAT based on the LMM assumption, with the covariance estimated by AI-REML algorithm.[17]

Independence of the stage 1 and the stage 2 test

We assume the vector of summary statistics $(Z_1, Z_2, \ldots, Z_m, T_1, T_2, \ldots T_m)$ has multivariate normal distribution. When there is no overlapping or related samples between two studies, the GWAS summary statistics and eQTL summary statistics are pairwise independent (i.e. $\text{cov}(Z_j, T_l) = 0$ for $j = 1,2, \ldots m$ and $l = 1,2, \ldots m$), and thus

$$\text{cov}\left[\begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_m \end{pmatrix}, \begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ T_m \end{pmatrix}\right] = \begin{pmatrix} \Sigma & 0_{m \times m} \\ 0_{m \times m} & \Sigma \end{pmatrix}.$$

Because of the multivariate normality of $(Z_1, Z_2, \ldots, Z_m, T_1, T_2, \ldots T_m)$, the vector of GWAS summary statistics $(Z_1, Z_2, \ldots, Z_m)$ is independent of the vector of eQTL summary statistics $(T_1, T_2, \ldots, T_m)$ when there is no overlapping or related samples between two studies.

Recall that the stage 2 test statistic is defined as $\sum_{j=1}^m a_j Z_j^2$, where $a_j = \frac{t_j - \bar{t}}{\sum_j t_j^2 - m\bar{t}^2}$ as a function of eQTL evidence. The distribution of $\sum_{j=1}^m a_j Z_j^2$ is accessed by treating $a_j's$ as the constants, where the variation of the eQTL summary statistics is only considered in the stage 1 test. Therefore, the stage 2 test statistic $\sum_{j=1}^m a_j Z_j^2$ and the stage 1 test statistic $\sum_{j=1}^m T_j^2$ are measurable functions of $(Z_1, Z_2, \ldots, Z_m)$ and $(T_1, T_2, \ldots, T_m)$, respectively, and they are independent due to the independence of $(Z_1, Z_2, \ldots, Z_m)$ and $(T_1, T_2, \ldots, T_m)$.

Type I Error Rate Control for a Single Hypothesis Testing

Let H_0 denote the null hypothesis that there is no colocalization given a specific gene-by-tissue pair, including multiple different scenarios: H_{01}, H_{02}, H_{03}, and H_{04}. The definitions of H_{01}, H_{02}, H_{03}, and H_{04} are provided in Table S6. For a specific gene-by-tissue pair, let R_1 and R_2 denote the event that the stage 1 eQTL test and the stage 2 SS test, respectively, provide significant p-values. Let R_{SS2} denote the event where the SS2 test rejects the null hypothesis of no colocalization. Thus, $R_{SS2} = R_1 \cap R_2$. The type I error rate for testing a single gene-by-tissue pair is

$$
\begin{aligned}
Pr(R_{SS2} \mid H_0) \quad &= Pr(R_2 \mid R_1, H_0) \times Pr(R_1 \mid H_0) \\
&= \sum_{i=1}^{4} Pr(R_{SS2} \mid H_{0i}) \times Pr(H_{0i} \mid H_0) \\
&= \sum_{i=1}^{4} Pr(R_2 \mid R_1, H_{0i}) \times Pr(R_1 \mid H_{0i}) \times Pr(H_{0i} \mid H_0) \qquad (18) \\
&= \sum_{i=1}^{4} Pr(R_2 \mid H_{0i}) \times Pr(R_1 \mid H_{0i}) \times Pr(H_{0i} \mid H_0) \qquad (19) \\
&\leq \sum_{i=1}^{4} \alpha \times Pr(H_{0i} \mid H_0) \qquad (20) \\
&= \alpha
\end{aligned}
$$

From equation (18) to equation (19), $Pr(R_2 \mid H_{0i}) = Pr(R_2 \mid R_1, H_{0i})$ because the stage 2 test is independent of the stage 1 test. We access the distribution of the stage 2 test statistics by conditioning on the observed eQTL summary statistics, where the variation of the eQTL summary statistics is only considered in the stage 1 test. In equation (20), for $i = 1$ or 3 when there is no eQTL, the stage 1 test controls the type I error rate within α

and thus $Pr(R_1 \mid H_{0i}) \leq \alpha$; for $i = 2$ or 4, the stage 2 test controls the type I error rate

within α and thus $Pr(R_2 \mid H_{0i}) \leq \alpha$. Therefore, $Pr(R_{SS2} \mid H_0) \leq \alpha$.

Upper Bound of Family-wise Error Rate for Multiple Hypothesis Testing

To understand the challenge of Family-wise error rate (FWER) control in this mixture

context, we first demonstrate a simple example of testing two genes at a locus, and then

formally derive a crude upper bound that works for any number of tests at a locus, which

would not change as the number of tests increases.

Consider one simple scenario where the locus of interest was identified through

GWAS as in the CF example, and there are two genes, A under H_{03} (no eQTL) and B

under H_{04} (eQTL but not overlapping with GWAS). Let α denote the nominal

significance level for the SS2 test and thus the stage 1 eQTL test is conducted at the $\frac{\alpha}{2}$

significance level for each gene. While testing the colocalization for gene A under H_{03},

let α^* and α_2^* denote the empirical false positive rate of the SS (stage 2) test at

significance level α and $\frac{\alpha}{2}$, respectively. For gene B, let 1-β denote the power of the

eQTL test in stage 1.

$$FWER = \text{Pr(SS2 rejects at least one gene)}$$

$$= \text{Pr(stage 1 rejects gene } A \text{ but not gene } B)$$

$\times$ Pr(stage 2 rejects gene A |stage 1 rejects gene A but not gene B)

$+$ Pr(stage 1 rejects gene B but not gene A)

$\times$ Pr(stage 2 rejects gene B |stage 1 rejects gene B but not gene A)

$+$ Pr(stage 1 rejects both genes)

$\times$ Pr(stage 2 rejects at least one gene |stage 1 rejects both genes)

By definition of α^* and the independence of the stage 1 and the stage 2 tests, Pr(stage 2 rejects gene A | stage 1 rejects gene A but not gene B) = Pr(stage 2 rejects gene A) = α^*, which could be greater or equal to α. Similarly, Pr(stage 2 rejects gene A | stage 1 rejects both genes)= $\alpha_2{}^*$, which could be greater or equal to $\frac{\alpha}{2}$. Thus, we can obtain

$$\text{FWER} \le \frac{\alpha}{2} \times \beta \times \alpha^* + (1 - \beta) \times \left(1 - \frac{\alpha}{2}\right) \times \alpha + \frac{\alpha}{2} \times (1 - \beta) \times \left(1 - \left(1 - \frac{\alpha}{2}\right)(1 - \alpha_2{}^*)\right).$$

If $\alpha^* = \alpha$ and $\alpha_2{}^* = \frac{\alpha}{2}, \left(1 - \left(1 - \frac{\alpha}{2}\right)(1 - \alpha_2{}^*)\right) \le \alpha$, then

$$\text{FWER} \le \frac{\alpha}{2} \times \beta \times \alpha + (1 - \beta) \times \left(1 - \frac{\alpha}{2}\right) \times \alpha + \frac{\alpha}{2} \times (1 - \beta) \times \alpha$$

$$= \alpha \times \left(\frac{\alpha}{2} \times \beta + (1 - \beta) \times \left(1 - \frac{\alpha}{2}\right) + \frac{\alpha}{2} \times (1 - \beta)\right)$$

$$\le \alpha.$$

When the magnitude of α^* and $\alpha_2{}^*$ are unknown, a crude upper bound for this FWER can be specified by assuming the maximum value of α^* and $\alpha_2{}^*$ to be 1, leading to

FWER $\le \frac{\alpha}{2} + \alpha - \frac{\alpha^2}{2}$; this is not necessarily less than or equal to α but provides a benchmark for the worst case scenario.

Now we consider deriving the upper bound of the FWERs when we are interested in testing colocalization for K gene-by-tissue pairs at a locus. Let α denote the nominal significance level for the SS2 test. Let $H_{0,K}$ denote the null hypothesis that there are no colocalizations for the K gene-by-tissue pairs at the locus. If there is no SNP-phenotype association (GWAS) at the locus (denoted by $H_{0,K,nsp}$), the K tests are under H_{01} or H_{02}; if there is SNP-phenotype association (GWAS) at the locus (denoted by $H_{0,K,sp}$), the K tests are under H_{03} or H_{04}. Let R^*_{SS2} denote the event that the SS2 test rejects at least one tests among the K tests. Therefore. for a locus with K tests, the FWER of the SS2 test is

$$
\begin{aligned}
&Pr\big(R^*_{SS2} \mid H_{0,K}\big) \\
&= Pr\big(R^*_{SS2} \mid H_{0,K,nsp}\big) \times Pr\big(H_{0,K,nsp} \mid H_{0,K}\big) \qquad\qquad (21)\\
&+ Pr\big(R^*_{SS2} \mid H_{0,K,sp}\big) \times Pr\big(H_{0,K,sp} \mid H_{0,K}\big). \qquad\qquad (22)
\end{aligned}
$$

$Pr\big(H_{0,K,sp} \mid H_{0,K}\big)$ represents the probability that the locus has phenotype-SNP association given there are no colocalizations for the K gene-by-tissue pairs, which is unknown in practice. Therefore, we need to find the maximum values for both $Pr\big(R^*_{SS2} \mid H_{0,K,nsp}\big)$ in (21) and $Pr\big(R^*_{SS2} \mid H_{0,K,sp}\big)$ in (22) to obtain the upper bound for $Pr\big(R^*_{SS2} \mid H_{0,K}\big)$. Let $R_{1,P}$ represent the event that the stage 1 test has P significant gene-by-tissue pairs. Let $R^*_{2,P}$ represents the event that the stage 2 test rejects at least one gene-by-tissue pairs among the P significant gene-by-tissue pairs from the stage 1 test.

For a locus with no phenotype-SNP association (under $H_{0,K,nsp}$), the K tests consist of mixture of H_{01} or H_{02}. If there are P significant gene-by-tissue pairs from the stage 1 test, the stage 2 test is implemented at significant level $\frac{\alpha}{P}$ per test by Bonferroni correction, and thus $Pr\left(R_{2,P}^* \mid R_{1,P}, H_{0,K,nsp}\right) \leq \frac{\alpha}{P} \times P = \alpha$. Therefore,

$$
\begin{aligned}
Pr\left(R_{SS2}^* \mid H_{0,K,nsp}\right) &= \sum_{P=1}^{K} Pr\left(R_{2,P}^* \mid R_{1,P}, H_{0,K,nsp}\right) \times Pr\left(R_{1,P} \mid H_{0,K,nsp}\right) \\
&\leq \sum_{P=1}^{K} \left(\frac{\alpha}{P} \times P\right) \times Pr\left(R_{1,P} \mid H_{0,K,nsp}\right) \\
&= \alpha \times \sum_{P=1}^{K} Pr\left(R_{1,P} \mid H_{0,K,nsp}\right) \\
&\leq \alpha
\end{aligned}
\tag{23}
$$

For a locus with phenotype-SNP association (under $H_{0,K,sp}$), we need to consider three cases separately: the K tests are all under H_{04}; the K tests are all under H_{03}; the K tests consist of a mixture of H_{03} and H_{04}. Let $H_{0,K,h4}, H_{0,K,h3}$ and $H_{0,K,h34}$ denote the three cases. respectively. Therefore,

$$
\begin{aligned}
Pr\left(R_{SS2}^* \mid H_{0,K,sp}\right) \quad &= Pr\left(R_{SS2}^* \mid H_{0,K,h4}\right) \times Pr\left(H_{0,K,h4} \mid H_{0,K,sp}\right) \\
&+ Pr\left(R_{SS2}^* \mid H_{0,K,h3}\right) \times Pr\left(H_{0,K,h3} \mid H_{0,K,sp}\right) \\
&+ Pr\left(R_{SS2}^* \mid H_{0,K,h34}\right) \times Pr\left(H_{0,K,h34} \mid H_{0,K,sp}\right).
\end{aligned}
\tag{24}
$$

Under $H_{0,K,h4}$, when all the K tests are under H_{04}, $Pr\left(R_{2,P}^* \mid R_{1,P}, H_{0,K,nsp}\right) \leq \frac{\alpha}{P} \times P = \alpha$. Thus,

$$Pr\left(R_{SS2}^* \mid H_{0,K,h4}\right) = \sum_{P=1}^{K} Pr\left(R_{2,P}^* \mid R_{1,P}, H_{0,K,h4}\right) \times Pr\left(R_{1,P} \mid H_{0,K,h4}\right)$$

$$\leq \sum_{P=1}^{K} \left(\frac{\alpha}{P} \times P\right) \times Pr\left(R_{1,P} \mid H_{0,K,h4}\right)$$

$$= \alpha \times \sum_{P=1}^{K} Pr\left(R_{1,P} \mid H_{0,K,h4}\right)$$

$$\leq \alpha. \tag{25}$$

Under $H_{0,K,h3}$ when all the K tests are under H_{03}, the stage 1 test for each gene-by-tissue pair is implemented under the null that there is no eQTL, and adjust the α for the total number of tests by Bonferroni correction, $\frac{\alpha}{K}$. Thus,

$$Pr\left(R_{SS2}^* \mid H_{0,K,h3}\right) \leq Pr\left(\text{ the stage 1 test rejects at least one gene-by-tissue pairs } \mid H_{0,K,h3}\right)$$

$$\leq \frac{\alpha}{K} \times K$$

$$= \alpha. \tag{26}$$

So far we have proved that when all the K tests are under H_{03} or H_{04}, the upper bound on the FWER of the two-stage SS2 test is α.

Under $H_{0,K,h34}$ when K tests consist of H_{03} and H_{04}, let H_{0,K_1,K_2} denote the event that there are K_1 tests under H_{04} and K_2 tests under H_{03}. We define R_{1,P_1,P_2} to be the event that there are P_1 significant stage 1 eQTL tests among those K_1 tests and P_2 significant stage 1 eQTL tests among those K_2 tests. In this case,

$$Pr\left(R_{SS2}^* \mid H_{0,K,h_{34}}\right) = \sum_{K_1 \geq 1, K_2 \geq 1} Pr\left(R_{SS2}^* \mid H_{0,K_1,K_2}\right) \times Pr\left(H_{0,K_1,K_2} \mid H_{0,K,h_{34}}\right),$$

$$Pr\left(R_{SS2}^* \mid H_{0,K_1,K_2}\right) = \sum_{P_1 \geq 1, P_2 = 0} Pr\left(R_{2,p}^* \mid R_{1,P_1,P_2}, H_{0,K_1,K_2}\right) \times Pr\left(R_{1,P_1,P_2} \mid H_{0,K_1,K_2}\right)$$

$$+ \sum_{P_1 = 0, P_2 \geq 1} Pr\left(R_{2,p}^* \mid R_{1,P_1,P_2}, H_{0,K_1,K_2}\right) \times Pr\left(R_{1,P_1,P_2} \mid H_{0,K_1,K_2}\right)$$

$$+ \sum_{P_1 \geq 1, P_2 \geq 1} Pr\left(R_{2,p}^* \mid R_{1,P_1,P_2}, H_{0,K_1,K_2}\right) \times Pr\left(R_{1,P_1,P_2} \mid H_{0,K_1,K_2}\right)$$

$$= \sum_{P_1=1}^{K_1} Pr\left(R_{2,p_1}^* \mid R_{1,P_1,0}, H_{0,K_1,K_2}\right) \times Pr\left(R_{1,P_1,0} \mid H_{0,K_1,K_2}\right) \tag{27}$$

$$+ \sum_{P_2=1}^{K_2} Pr\left(R_{2,p_2}^* \mid R_{1,0,P_2}, H_{0,K_1,K_2}\right) \times Pr\left(R_{1,0,P_2} \mid H_{0,K_1,K_2}\right) \tag{28}$$

$$+ \sum_{P_1 \geq 1, P_2 \geq 1} Pr\left(R_{2,p}^* \mid R_{1,P_1,P_2}, H_{0,K_1,K_2}\right) \times Pr\left(R_{1,P_1,P_2} \mid H_{0,K_1,K_2}\right). \tag{29}$$

The term (27) demonstrates the scenario when all the significant stage 1 eQTL tests are within the K_1 gene-by-tissue pairs under H_{04}. In this scenario, $P = P_1$ and the stage 2 test is implemented at significance level $\frac{\alpha}{P_1}$ per test. Therefore, term (27)

$$\sum_{P_1=1}^{K_1} Pr\left(R_{2,p_1}^* \mid R_{1,P_1,0}, H_{0,K_1,K_2}\right) \times Pr\left(R_{1,P_1,0} \mid H_{0,K_1,K_2}\right)$$

$$\leq \left(\frac{\alpha}{P_1} \times P_1\right) \times \sum_{P_1=1}^{K_1} Pr\left(R_{1,P_1,0} \mid H_{0,K_1,K_2}\right)$$

$$= \alpha \times \sum_{P_1=1}^{K_1} Pr\left(R_{1,P_1,0} \mid H_{0,K_1,K_2}\right)$$

$$= \alpha \times S_{(1,0)}, \tag{30}$$

Where $S_{(1,0)} = \sum_{P_1=1}^{K_1} Pr\left(R_{1,P_1,0} \mid H_{0,K_1,K_2}\right)$ which represents the probability that there is at least one significant stage 1 eQTL test among the K_1 gene-by-tissue pairs but no significant stage 1 eQTL test among the K_2 gene-by-tissue pairs.

Let α_p^* denote the empirical false positive rate of the stage 2 test at the significance

level $\frac{\alpha}{P}$ under H_{03} at the locus of interest. We define $\tilde{\alpha}$ to be $\max\limits_{p=1,...K}(\alpha_p^* \times P)$, which can

be interpreted as an upper bound of the empirical probability that the stage 2 test rejects

at least one gene-by-tissue pairs among those P significant stage 1 tests at the locus. In

particular when $\alpha_p^* = \frac{\alpha}{P}$ (for $P = 1, ... K$), $\tilde{\alpha} = \alpha$, and $\tilde{\alpha}$ could be bigger than α if

α_p^* is inflated. The term (28) demonstrates the scenario when all the significant stage 1

eQTL tests are within the K_2 gene-by-tissue pairs under H_{03}. In this scenario, $P = P_2$;

the stage 2 test is implemented at significance level $\frac{\alpha}{P_2}$ per test.

$$
\sum_{P_2=1}^{K_2} Pr\big(R_{2,p_2}^* \mid R_{1,0,P_2}, H_{0,K_1,K_2}\big) \times Pr\big(R_{1,0,P_2} \mid H_{0,K_1,K_2}\big)
$$
$$
\leq \sum_{P_2=1}^{K_2} (\alpha_{P_2}^* \times P_2) \times \sum_{P_2=1}^{K_2} Pr\big(R_{1,0,P_2} \mid H_{0,K_1,K_2}\big)
$$
$$
\leq \tilde{\alpha} \times S_{(0,1)} \tag{31}
$$

, where $S_{(0,1)} = \sum_{P_2=1}^{K_2} Pr\big(R_{1,0,P_2} \mid H_{0,K_1,K_2}\big)$ which represents the probability that there

are at least one significant stage 1 eQTL tests among the K_2 tests but no significant stage

1 eQTL tests among the K_1 tests.

The term (29) demonstrates the scenario when those significant stage 1 eQTL tests

present both in the K_1 and K_2 gene-by-tissue pairs. In this scenario,

$$
\sum_{P_1 \geq 1, P_2 \geq 1} Pr\big(R_{2,P}^* \mid R_{1,P_1,P_2}, H_{0,K_1,K_2}\big) \times Pr\big(R_{1,P_1,P_2} \mid H_{0,K_1,K_2}\big) \leq \sum_{P_1 \geq 1, P_2 \geq 1} (\alpha_P^* \times P) \times
$$
$$
\leq \tilde{\alpha}\big(1 - S_{(0,0)} - S_{(0,1)} - S_{(1,0)}\big), \tag{32}
$$

where $S_{(0,0)}$ represent the probability that under H_{0,K_1,K_2}, there are no significant stage 1 eQTL tests among the K gene-by-tissue pairs. Note that

$$S_{(0,1)} = Pr(\text{ there are at least one significant stage 1 tests among the } K \text{ tests } | H_{0,K_1,K_2})$$

$$-Pr(\text{there are at least one significant stage 1 tests among the } K_1 \text{ tests } | H_{0,K_1,K_2})$$

$$\geq 1 - S_{(0,0)} - \alpha,$$

and therefore

$$\alpha \geq 1 - S_{(0,0)} - S_{(0,1)} . \tag{33}$$

We define δ to be $\tilde{\alpha} - \alpha$ and combine the results in (30), (31), (32) and (33) to obtain

$$
\begin{aligned}
Pr(R^*_{SS2} | H_{0,K_1,K_2}) \quad &\leq \alpha S_{(0,1)} + \tilde{\alpha} S_{(1,0)} + \tilde{\alpha}(1 - S_{(0,0)} - S_{(1,0)} - S_{(0,1)}) \\
&= \alpha S_{(0,1)} + \tilde{\alpha}(1 - S_{(0,0)} - S_{(0,1)}) \\
&= \alpha S_{(0,1)} + (\alpha + \delta)(1 - S_{(0,0)} - S_{(0,1)}) \\
&= \alpha(1 - S_{(0,0)}) + \delta(1 - S_{(0,0)} - S_{(0,1)}) \\
&\leq \alpha(1 - S_{(0,0)}) + \delta \times \alpha \\
&\leq \alpha + \delta \times \alpha \\
&= \alpha \times (1 + \delta).
\end{aligned}
\tag{34}
$$

Therefore under the $H_{0,K,h_{34}}$,

$$
\begin{aligned}
Pr(R^*_{SS2} | H_{0,K,h_{34}}) \quad &= \sum_{K_1 \geq 1, K_2 \geq 1} Pr(R^*_{SS2} | H_{0,K_1,K_2}) \times Pr(H_{0,K_1,K_2} | H_{0,K,h_{34}}) \\
&\leq \sum_{K_1 \geq 1, K_2 \geq 1} \alpha(1 + \delta) \times Pr(H_{0,K_1,K_2} | H_{0,K,h_{34}}) \\
&\leq \alpha \times (1 + \delta).
\end{aligned}
\tag{35}
$$

So far, we have shown that when the K tests consist of a mixture of H_{03} and H_{04}, the upper bound is $\alpha \times (1 + \delta)$; by assuming the maximum value of $\tilde{\alpha}$ to be 1,

$\alpha \times (1 + \delta) = 2\alpha - \alpha^2$, which is bigger than α. This upper bound works for any value of K, which would not change as the number of tests increases.

Taking into consideration all of the scenarios, the FWER of the SS2 in (21) and (22)

can be calculated by

$$
\begin{aligned}
Pr(R^*_{SS2} \mid H_{0,K}) \quad &= Pr(R^*_{SS2} \mid H_{0,K,nsp}) \times Pr(H_{0,K,nsp} \mid H_{0,K}) \\
&+ \{Pr(R^*_{SS2} \mid H_{0,K,h_4}) \times Pr(H_{0,K,h_4} \mid H_{0,K,sp}) \\
&+ Pr(R^*_{SS2} \mid H_{0,K,h_3}) \times Pr(H_{0,K,h_3} \mid H_{0,K,sp}) \\
&+ Pr(R^*_{SS2} \mid H_{0,K,h_{34}}) \times Pr(H_{0,K,h34} \mid H_{0,K,sp})\} \times Pr(H_{0,K,sp} \mid H_{0,K}) \\
&\le \alpha \times \{Pr(H_{0,K,nsp} \mid H_{0,K}) + Pr(H_{0,K,h_4} \mid H_{0,K}) + Pr(H_{0,K,h_3} \mid H_{0,K})\} \\
&+ \alpha(1 + \delta) \times Pr(H_{0,K,h_{34}} \mid H_{0,K}) \\
&\le \alpha \times (1 + \delta).
\end{aligned}
$$

If we want to strictly control the upper bound of the FWER at a certain level such as

0.05, $\alpha \times (1 + \delta) = 0.05$, where $\delta = \tilde{\alpha} - \alpha$. This crude upper bound can be used to

find the nominal level for performing the test in practice. By assuming a maximum value

of $\tilde{\alpha}$ to be 1, we can solve $\alpha = 0.0253$. However, this procedure is found to be

unnecessary from our empirical studies, because $\tilde{\alpha}$ never approaches 1 and we did not

observe a single iteration with the FWER greater than the specified significance level

without this procedure.

Simulation Details for a Single Gene-by-Tissue Pair

We generate the GWAS summary statistics Z and the eQTL summary statistics T from

$N(\Sigma\Lambda_Z, \Sigma)$ and $N(\Sigma\Lambda_T, \Sigma)$, respectively, where Σ is the LD matrix from the locus of

interest. $\Lambda_Z = (\lambda_{Z_1}, \lambda_{Z_2}, \dots \lambda_{Z_c}, \dots \lambda_{Z_m})^\top$ and $\Lambda_T = (\lambda_{T_1}, \lambda_{T_2}, \dots \lambda_{T_c}, \dots \lambda_{T_m})^\top$ are vectors

with each component being the true standardized effect size of the corresponding SNP,

respectively for GWAS and eQTL analyses. In particular, λ_{Z_c} and λ_{T_c} represent the value

for the associated SNP from the GWAS and eQTL studies, respectively. When there is

SNP-phenotype association, we use the lead GWAS SNP from our CF study[18,19] as the

associated SNP with value λ_{Z_c}, and we define a SNP ($r < 0.002$ and > -0.002 with the

associated) SNP as an independent eQTL SNP with value λ_{T_c}.

For type I error evaluation, we consider all four scenarios of the composite null

hypothesis. For example, under H_{01} when there is no SNP-phenotype association and no

eQTL, $\lambda_{Z_c} = 0$ and $\lambda_{T_c} = 0$. In contrast, under H_{04} when both SNP-phenotype

association and eQTL are present but occurring at independent SNPs, $\lambda_{Z_c} \neq 0$ and $\lambda_{T_c} \neq$

0, and we set $\lambda_{Z_c} = 5.73$ and $\lambda_{T_c} = 7.01$ to be consistent with the previous simulation

study in [18]. Table S6 provides details for the parameter settings under the four null

scenarios. Under each scenario, 10^4 replications are simulated.

To study power, we simulate six scenarios. For scenarios with two association

peaks, we used the lead GWAS SNP from our CF study at the locus to be the first

associated SNP, while the second associated SNP is defined as the next adjacent SNP

with $r < 0.002$ and > -0.002 with the first associated SNP. We vary the magnitude of

eQTL evidence to measure the relationship between power and colocalization strength.

Table S7 provides details for the parameter settings under the six alternative scenarios.

Under each scenario, 10^4 replications are simulated.

Simulation Details for Multiple Genes-Tissue Pairs

The GWAS summary statistics are generated from $N(\Sigma\Lambda_Z, \Sigma)$ with $\lambda_{Z_c} = 4.5$, and Σ is simulated based on the LD within the *MUC20/MUC4* and *SLC6A14* loci. The two loci are defined by including SNPs within 0.1Mb of either side of the top associated GWAS SNP from the CF study. For FWER evaluation with one GWAS association at the simulated locus, we generate 600 sets of independent eQTL summary statistics corresponding to eQTL analyses of 600 genes of interest under H_{03} or H_{04}, where different proportions of genes with eQTL evidence under H_{04} are considered (0%, 20%, 40%, 60%, 80% and 100%). For a given gene with eQTL signals under H_{04}, the eQTL summary statistics are generated from $N(\Sigma\Lambda_T, \Sigma)$ where the value of λ_{T_c} is randomly generated from 6 different intervals (50%-60%, 60%-70%, 70%-80%, 80%-90%, 90%-95%, 95%-100% power is achieved to detect the eQTL association at the significance level of 10^{-8}) with probabilities according to the proportion of the -log10 (maximum eQTL p-value) within each interval observed at the locus. For other genes with eQTL under H_{03}, the eQTL summary statistics are generated from $N(0, \Sigma)$.

For power evaluation, with the same GWAS peak, we set 5% of genes with eQTL signals as colocalizing with the GWAS signal (Scenario 1 in Figure 1). For the *SLC6A14* locus, we set 47.5% of genes as having no eQTL signal (under H_{03}), and 47.5% of genes

as having eQTL signals that are independent from the GWAS signal (under H_{04}); while modeling the *MUC20/MUC4* locus, we set 19% of genes under H_{03}, and 76% of genes under H_{04}. The GWAS associated SNP and the eQTL associated SNPs are selected in the same way in section *Simulation Details for a Single Gene-by-Tissue Pair*. For genes under the alternative, the eQTL summary statistics are generated from $N(\Sigma\Lambda_T, \Sigma)$, where the value of λ_{T_c} is randomly selected from an interval (i.e. [5.48, 5.73]). For genes with the eQTL under H_{04}, λ_{T_c} is selected from 4.45 to 9.48 such that 10% to 100% power is achieved at genome-wide significance level 10^{-8}. We evaluate power as the average increase in strength of eQTL among the 5% of genes under the alternative, and consider 5 different intervals: [5.48, 5.73], [5.73, 5.98], [5.98, 6.25], [6.25, 6.57] and [6.57, 7.01]. Under each scenario, 10^5 replications are simulated.

Simulation Details for Overlapping Samples

Given the sample size n_{GWAS} and n_{eQTL} for a GWAS and eQTL study, respectively, we randomly select n_{GWAS} individuals from the genotype data of the CF lung GWAS study.[19] A region is defined which included the SNPs 0.1Mb around the top GWAS SNP at the *MUC4/MUC20* locus. All or half of the participants in the eQTL study are randomly selected and simulated such that they are overlapping with the participants in the GWAS.

For individual i, let g_{1i} and g_{2i} denote the genotype of the associated SNP for the GWAS and eQTL study respectively. We simulate a pair of phenotypes for each individual $\begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} = \begin{pmatrix} g_{1i}\beta_1 + \epsilon_{1i} \\ g_{2i}\beta_2 + \epsilon_{2i} \end{pmatrix}$, where β_1 and β_2 denote the effect sizes for the GWAS and eQTL studies, respectively. Hormozdiari et al[20] has shown that

$$\lambda_{Z_c} = \frac{\beta_1}{\sqrt{\text{var}(\epsilon_{1i})}} \times \sqrt{n_{GWAS}};$$
$$\lambda_{T_c} = \frac{\beta_2}{\sqrt{\text{var}(\epsilon_{2i})}} \times \sqrt{n_{eQTL}}. \tag{1}$$

We set different values for λ_{Z_c} and λ_{T_c} under the composite null hypothesis, and then calculate the corresponding values for β_1 and β_2 according to equation (1). For example, when the GWAS signal and eQTL signal are distinct (H_{04}), we set $\Lambda_{T_c} = 5.48$ to mimic the observed eQTL peak for *MUC4* in HNE. Given 100 participants in the eQTL study $(n_{eQTL} = 100)$ and $var(\epsilon_{2i}) = 1$, we obtain $\beta_1 = 0.123$. We consider two different values of λ_{Z_c} (5.73 or 6.57) such that 0.5 or 0.8 power is achieved to detect the GWAS association at the genome-wide significance level of 10^{-8}. Given 2000 participants in the GWAS study and $var(\epsilon_{2i}) = 1$, we obtain $\beta_1 = 0.128$ and 0.147, respectively. Parameter settings for λ_{Z_c} and λ_{T_c} under the composite null hypothesis is provided in Table S1. $\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \sim \text{MVN} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_\epsilon \right)$, where Σ_ϵ is the covariance matrix with the diagonal elements set to be 1. If an individual is included in both studies, the off-diagonal terms are set to C which represents the phenotypic correlation; otherwise it is set to be 0.

Simulation studies are conducted for different values of phenotypic correlation, C= 0.3, 0.5, 0.7 or 0.9. We perform univariate linear regression to obtain the marginal summary statistics and then apply SS2. Under each scenario, 10^4 replications are simulated.

Supplemental References

1. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., and Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nature Genetics *48*, 481-487. 10.1038/ng.3538.
2. Wu, Y., Zeng, J., Zhang, F., Zhu, Z., Qi, T., Zheng, Z., Lloyd-Jones, L.R., Marioni, R.E., Martin, N.G., Montgomery, G.W., et al. (2018). Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. Nature Communications *9*, 918. 10.1038/s41467-018-03371-0.
3. 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 Genetics *10*, e1004383. 10.1371/journal.pgen.1004383.
4. Dobbyn, A., Huckins, L.M., Boocock, J., Sloofman, L.G., Glicksberg, B.S., Giambartolomei, C., Hoffman, G.E., Perumal, T.M., Girdhar, K., Jiang, Y., et al. (2018). Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-localization with Schizophrenia GWAS. Am J Hum Genet *102*, 1169-1184. 10.1016/j.ajhg.2018.04.011.
5. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al. (2013). The Genotype-Tissue Expression (GTEx) project. Nature Genetics *45*, 580-585. 10.1038/ng.2653.
6. Ongen, H., Buil, A., Brown, A.A., Dermitzakis, E.T., and Delaneau, O. (2016). Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics *32*, 1479-1485. 10.1093/bioinformatics/btv722.
7. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature genetics *47*, 291-295.
8. Xu, Z., Duan, Q., Yan, S., Chen, W., Li, M., Lange, E., and Li, Y. (2015). DISSCO: direct imputation of summary statistics allowing covariates. Bioinformatics *31*, 2434-2442.
9. 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. 10.1016/j.ajhg.2016.10.003.
10. Baltagi, B.H. (2008). Econometrics, 4th ed. 2008. Edition (Springer Berlin Heidelberg). 10.1007/978-3-540-76516-5.

11.   Park, J.Y., Wu, C., Basu, S., McGue, M., and Pan, W. (2018). Adaptive SNP-set association testing in generalized linear mixed models with application to family studies. Behavior genetics *48*, 55-66.

12.   Park, J.Y., Wu, C., and Pan, W. (2018). An adaptive gene-level association test for pedigree data. BMC genetics *19*, 39-43.

13.   Hill, W.D., Arslan, R.C., Xia, C., Luciano, M., Amador, C., Navarro, P., Hayward, C., Nagy, R., Porteous, D.J., and McIntosh, A.M. (2018). Genomic analysis of family data reveals additional genetic effects on intelligence and personality. Molecular psychiatry *23*, 2347-2362.

14.   Laird, N.M., and Ware, J.H. (1982). Random-effects models for longitudinal data. Biometrics, 963-974.

15.   Lindstrom, M.J., and Bates, D.M. (1988). Newton—Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. Journal of the American Statistical Association *83*, 1014-1022.

16.   Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., Van Willigen, B., and Maintainer, R. (2017). Package 'nlme'. Linear and nonlinear mixed effects models, version *3*.

17.   Chen, H., Conomos, M.P., and Chen, M.H. (2019). Package 'GMMAT'.

18.   Gong, J., Wang, F., Xiao, B., Panjwani, N., Lin, F., Keenan, K., Avolio, J., Esmaeili, M., Zhang, L., He, G., et al. (2019). Genetic association and transcriptome integration identify contributing genes and tissues at cystic fibrosis modifier loci. PLOS Genetics *15*, e1008007. 10.1371/journal.pgen.1008007.

19.   Corvol, H., Blackman, S.M., Boëlle, P.-Y., Gallins, P.J., Pace, R.G., Stonebraker, J.R., Accurso, F.J., Clement, A., Collaco, J.M., Dang, H., et al. (2015). Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. Nature Communications *6*, 8382. 10.1038/ncomms9382.

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.