Modeling tissue co-regulation to estimate tissue-specific contributions to disease

Tiffany Amariuta1,2, Katherine Siewert-Rocks1,2, Alkes L. Price1,2,3

1. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
2. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
3. Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

Abstract

Integrative analyses of genome-wide association studies (GWAS) and gene expression data across diverse tissues and cell types have enabled the identification of putative disease-critical tissues. However, co-regulation of genetic effects on gene expression across tissues makes it difficult to distinguish biologically causal tissues from tagging tissues. While previous work emphasized the potential of accounting for tissue co-regulation, tissue-specific disease effects have not previously been formally modeled. Here, we introduce a new method, tissue co-regulation score regression (TCSC), that disentangles causal tissues from tagging tissues and partitions disease heritability (or covariance) into tissue-specific components. TCSC leverages gene-disease association statistics across tissues from transcriptome-wide association studies (TWAS), which implicate both causal and tagging genes and tissues. TCSC regresses TWAS chi-square statistics (or products of z-scores) on tissue co-regulation scores reflecting correlations of predicted gene expression across genes and tissues. In simulations, TCSC powerfully distinguishes causal tissues from tagging tissues while controlling type I error. We applied TCSC to GWAS summary statistics for 78 diseases and complex traits (average N = 302K) and gene expression prediction models for 48 GTEx tissues. TCSC identified 27 causal tissue-trait pairs at 10% FDR, including well-established findings, biologically plausible novel findings (e.g. aorta artery and glaucoma), and increased specificity of known tissue-trait associations (e.g. subcutaneous adipose, but not visceral adipose, and HDL). TCSC also identified 30 causal tissue-trait covariance pairs at 10% FDR. For the positive genetic covariance between eosinophil count and white blood cell count, whole blood contributed positive covariance while LCLs contributed negative covariance; this suggests that genetic covariance may reflect distinct tissue-specific contributions. Overall, TCSC is a powerful method for distinguishing causal tissues from tagging tissues, improving our understanding of disease and complex trait biology.
Introduction

Most diseases are driven by tissue-specific or cell-type-specific mechanisms, thus the inference of causal disease tissues is an important goal. For many polygenic diseases and complex traits, disease-associated tissues have previously been identified via the integration of genome-wide association studies (GWAS) with tissue-level functional data characterizing expression quantitative trait loci (eQTLs)\(^2\text{-}^5\), gene expression\(^6\text{-}^8\), or epigenetic features\(^9\text{-}^{16}\). However, it is likely that most disease-associated tissues are not actually causal, due to the high correlation of eQTL effects (resp. gene expression or epigenetic features) across tissues; the correlation of eQTL effects across tissues, i.e. tissue co-regulation, can arise due to shared eQTLs or distinct eQTLs in linkage disequilibrium (LD)\(^2\text{-}^17,^18^,^19\). One approach to address this involves comparing eQTL-disease colocalizations across different tissues\(^2\); however, this approach relies on colocalizations with disease that are specific to a single tissue, and may implicate co-regulated tagging tissues that colocalize with disease. Another approach leverages multi-trait fine-mapping methods to simultaneously evaluate all tissues for colocalization with disease; however, this locus-based approach does not produce genome-wide estimates and it remains the case that many (causal or tagging) tissues may colocalize with disease under this framework. To our knowledge, no previous study has formally modeled genetic co-regulation across tissues to statistically disentangle causal from tagging tissues.

Here, we introduce a new method, tissue co-regulation score regression (TCSC), that disentangles causal tissues from tagging tissues and partitions disease heritability (or genetic covariance of two diseases/traits) into tissue-specific components. TCSC leverages gene-disease association statistics across tissues from transcriptome-wide association studies (TWAS)\(^19\text{-}^{20},^{17}\). A challenge is that TWAS association statistics include the effects of both co-regulated tissues (see above) and co-regulated genes\(^17\text{-}^{21}\). To address this, TCSC regresses TWAS chi-square statistics (or products of z-scores for two diseases/traits) on tissue co-regulation scores reflecting correlations of predicted gene expression across genes and tissues. (TCSC is conceptually related to gene co-regulation score regression (GCSC)\(^21\), a method for identifying disease-enriched gene sets that models gene co-regulation but does not model tissue co-regulation.) Distinct from previous methods that analyze each tissue marginally, TCSC jointly models contributions from each tissue to identify causal tissues (analogous to the distinction in GWAS between marginal association and fine-mapping\(^22\)). We validate TCSC using extensive simulations and apply TCSC to 78 diseases and complex traits (average \(N = 302K\)) and 48 GTEx tissues\(^18\), showing that TCSC recapitulates known biology and identifies biologically plausible novel tissue-trait pairs (or tissue-trait covariance pairs) while attaining increased specificity relative to previous methods.

Results

Overview of TCSC regression

TCSC estimates the disease heritability explained by cis-genetic components of gene expression in each tissue when jointly modeling contributions from each tissue. We refer to
tissues with nonzero contributions as “causal” tissues (with the caveat that joint-fit effects of
gene expression on disease may not reflect biological causality; see Discussion). TCSC leverages
the fact that TWAS $\chi^2$ statistics for each gene and tissue include both causal effects of that gene
and tissue on disease and tagging effects of co-regulated genes and tissues. We define co-
regulation based on squared correlations in cis-genetic expression, which can arise due to
shared causal eQTLs and/or LD between causal eQTLs. TCSC determines that a tissue is causal
for disease if genes and tissues with high co-regulation to that tissue have higher TWAS $\chi^2$
statistics than genes and tissues with low co-regulation to that tissue.

In detail, the expected TWAS $\chi^2$ statistic for gene $g$ and tagging tissue $t$ is

$$E[\chi^2_{g,t}] = N \sum_{t'} l(g, t; t') h^2_{ge(t')}/G_t + 1,$$

(1)

where $N$ is GWAS sample size, $t'$ indexes causal tissues, $l(g, t; t')$ are tissue co-regulation scores
(defined as $l(g, t; t') = \sum g' r^2(W_{g,t}, W_{g',t'})$, where $W$ denotes the cis-genetic component of
gene expression for a gene-tissue pair across individuals and the sum is over genes $g'$ within +/-
1 Mb to gene $g$), $h^2_{ge(t')}$ is the disease heritability explained by the cis-genetic component of
gene expression in tissue $t'$, and $G_t$ is the number of significantly cis-heritable genes in tissue $t'$
(see below). Equation (1) allows us to estimate $h^2_{ge(t')}$ via a multiple linear regression of TWAS
$\chi^2$ statistics (for each gene and tagging tissue) on tissue co-regulation scores (Figure 1). To
facilitate comparisons across diseases/traits, we primarily report the proportion of disease
heritability explained by the cis-genetic component of gene expression in tissue $t'$
($\pi_{t'} = h^2_{ge(t')}/h^2_g$), where $h^2_g$ is the common variant SNP-heritability estimated by S-
LDSC.$^{12,23,24}$

TCSC can also estimate the genetic covariance between two diseases explained by cis-
genetic components of gene expression in each tissue, using products of TWAS z-scores. In
detail, the expected product of TWAS z-scores in disease 1 and disease 2 for gene $g$ and tagging
tissue $t$ is

$$E[z^1_{gt} \times z^2_{gt}] = \sqrt{N_1 N_2} \sum_{t'} l(g, t; t') \omega_{ge(t')}/G_{t'} + \rho N_s/\sqrt{N_1 N_2},$$

(2)

where $N_1$ is GWAS sample size for disease 1, $N_2$ is GWAS sample size for disease 2, $t'$ indexes
causal tissues, $l(g, t; t')$ are tissue co-regulation scores (see above), $\omega_{ge(t')}$ is the genetic
covariance explained by the cis-genetic component of gene expression in tissue $t'$, $G_{t'}$ is the
number of significantly cis-heritable genes in tissue $t'$ (see below), $\rho$ is the phenotypic
correlation between disease 1 and disease 2, and $N_s$ is the number of overlapping GWAS
samples between disease 1 and disease 2. Equation (2) allows us to estimate $\omega_{ge(t')}$ via a
multiple linear regression of products of TWAS z-scores in disease 1 and disease 2 (for each
gene and tagging tissue) on tissue co-regulation scores. We note that the last term in Equation
(2) is not known a priori but is accounted for via the regression intercept, analogous to previous
work.$^{25}$ To facilitate comparisons across diseases/traits, we primarily report the proportion of
genetic covariance explained by the cis-genetic component of gene expression in tissue $t'$

\[ \zeta' = \omega_{ge(t')} / |\omega_g| \]

where $\omega_g$ is the common variant genetic covariance estimated by cross-trait LDSC$^{26}$.

We restrict gene expression prediction models and TWAS association statistics for each tissue to significantly cis-heritable genes in that tissue, defined as genes with significantly positive cis-heritability (2-sided $p < 0.01$; estimated using GCTA$^{27}$) and positive adjusted-$R^2$ in cross-validation prediction. We note that quantitative estimates of the disease heritability explained by the cis-genetic component of gene expression in tissue $t'$ ($h^2_{ge(t')} \equiv \omega_{ge(t')} |\omega_{g} |$) are impacted by the number of significantly cis-heritable genes in tissue $t'$ ($G_{t'}$), which may be sensitive to eQTL sample size. For each disease (or pair of diseases), we use a genomic block-jackknife with 200 blocks to estimate standard errors on the disease heritability (or covariance) explained by cis-genetic components of gene expression in each tissue, and compute 1-sided $P$-values for non-zero heritability (or 2-sided $P$-values for non-zero covariance) and false discovery rates (FDR) accordingly; we primarily report causal tissues with FDR < 10%. Further details, including correcting for bias in tissue co-regulation scores arising from differences between cis-genetic vs. cis-predicted expression (analogous to GCSC$^{21}$) and utilizing regression weights to improve power, are provided in the Methods section. We have publicly released open-source software implementing TCSC regression (see Code Availability), as well as all GWAS summary statistics, TWAS association statistics, tissue co-regulation scores, and TCSC output from this study (see Data Availability).

Simulations

We performed extensive simulations to evaluate the robustness and power of TCSC, using the TWAS simulator of Mancuso et al.$^{28}$ (see Code Availability). We used simulated genotypes to simulate gene expression values (for each gene and tissue) and complex trait phenotypes, and computed TWAS association statistics for each gene and tissue. In our default simulations, the number of tissues was set to 10. The gene expression sample size (in each tissue) varied from 100 to 1,000 (with the value of 300 corresponding most closely to the GTEx data$^{18}$ used in our analyses of real diseases/traits; see below). The number of genes was set to 1,245, distributed across 249 genomic blocks (representing cis regions of length 1 Mb on chromosome 1) with each block uniformly containing 50 SNPs and 5 genes. Each gene had 5 causal cis-eQTLs, consistent with the upper range of independent eQTLs per gene detected in GTEx$^{18}$ and others studies$^{29-32}$. The cis-eQTL effect sizes for each gene were drawn from a multivariate normal distribution across tissues to achieve a specified level of co-regulation (see below), and the cis-heritability of each gene was approximated as the sum of squared standardized cis-eQTL effect sizes. In each tissue, the average cis-heritability (across genes) was tuned to 0.08 (sd = 0.05, ranging from 0.01 to 0.40) in order to achieve an average estimated cis-heritability (across significantly cis-heritable genes) varying from 0.11 to 0.31 (across gene expression sample sizes), matching GTEx$^{18}$, the proportions of expressed genes that were significantly cis-heritable were also matched to GTEx data$^{18}$. The 10 tissues were split into three tissue categories to mimic biological tissue modules in GTEx$^{18}$ (tissues 1-3, tissues 4-6, and
tissues 7-10), and average cis-genetic correlations between tissues (averaged across genes) were set to 0.789 within the same tissue category, 0.737 between tissue categories, and 0.751 overall\textsuperscript{33} (Methods). The GWAS sample size was set to 100,000. The 10 tissues included one causal tissue explaining 5% of complex trait heritability and nine non-causal tissues. 10% of the genes (125 of the 1,245 genes) had nonzero (normally distributed) gene-trait effects in the causal tissue\textsuperscript{34}. The SNP-heritability of the trait was set to 5%, all of which was explained by gene expression in the causal tissue; we note that TCSC is not impacted by SNP-heritability that is not explained by gene expression. Other parameter values were also explored, including other values of the number of causal tissues and the number of tagging tissues. Further details of the simulation framework are provided in the Methods section. We primarily focused our comparisons between TCSC and other methods on analyses of real diseases/traits (see below), because the power of TCSC relative to other methods is likely to be highly sensitive to assumptions about the role of gene expression in disease architectures. However, we included comparisons to RTC Coloc\textsuperscript{2} (which also uses eQTL data) in our simulations.

We first evaluated the bias in TCSC estimates of the disease heritability explained by the cis-genetic component of gene expression in tissue \(t'\) (\(h^2\text{ge}(t)\)), for both causal and non-causal tissues. For causal tissues, TCSC produced conservative estimates of \(h^2\text{ge}(t)\), particularly at smaller eQTL sample sizes (Figure 2A, Supplementary Table 1). As noted above, estimates of \(h^2\text{ge}(t)\) are impacted by the number of significantly cis-heritable genes in tissue \(t'\) (\(G_{t'}\)), which may be sensitive to eQTL sample size. Estimates were more conservative when setting \(G_{t'}\) to the number of significantly cis-heritable genes, and less conservative when setting \(G_{t'}\) to the number of true cis-heritable genes. In the latter case, estimates were unbiased at the largest eQTL sample size, suggesting that the conservative bias is due to noise in gene expression prediction models. For non-causal tissues, TCSC produced very slightly negative estimates of \(h^2\text{ge}(t)\) (e.g. \(-7.2 \times 10^{-4}\), s.e. = \(2.8 \times 10^{-4}\)) at eQTL sample size of 100 when setting \(G_{t'}\) to the number of true cis-heritable genes). We did not include a comparison to RTC Coloc\textsuperscript{2} in our analyses of bias, because RTC Coloc does not provide quantitative estimates of \(h^2\text{ge}(t)\).

We next evaluated the type I error of TCSC for non-causal tissues. TCSC was conservative, with type I error less than 5% at a significance threshold of \(p = 0.05\) (Figure 2B, Supplementary Table 1). The conservative type I error was most pronounced at small eQTL sample sizes. We determined that the conservative type I error is primarily due to conservative jackknife standard errors (1.3x-1.6x; see Supplementary Table 2) rather than the very slight negative bias, which is generally two to three orders of magnitude smaller than the jackknife standard error for a given estimate (Supplementary Table 3); we hypothesize that the standard errors are conservative relative to the empirical standard error due to variation of causal signal across the genome\textsuperscript{35}. For comparison, we also evaluated the type I error of RTC Coloc\textsuperscript{2}. We determined that RTC Coloc was not well-calibrated at lower eQTL sample sizes, with type I error of 8.4% at eQTL sample size of 100 (Supplementary Figure 1, Supplementary Table 4).

We next evaluated the power of TCSC for causal tissues. We determined that TCSC was well-powered to detect causal tissues at realistic eQTL sample sizes. When analyzing 300 eQTL
samples, TCSC attained 72% power at a nominal significance threshold of $p < 0.05$ (Figure 2C) and 26% power at a stringent significance threshold of $p < 0.005$ (corresponding to 10% per-trait FDR across tissues in these simulations, analogous to our analyses of real diseases/traits) (Supplementary Table 1). As expected, power increased at larger eQTL sample sizes, due to lower standard errors on point estimates of $h_{g_e(t)}^2$ (Figure 2A). For comparison, we also evaluated the power of RTC Coloc². We determined that RTC Coloc attained substantially lower power, which did not vary with eQTL sample size (Supplementary Figure 1, Supplementary Table 4). We note that eQTL sample size is more likely to be a limiting factor when disentangling co-regulated tissues using TCSC, whereas GWAS sample size is more likely to be a limiting factor in other analyses such as the colocalization analyses used by RTC Coloc.

We similarly evaluated the robustness and power of TCSC when estimating tissue-specific contributions to the genetic covariance between two diseases/traits. We employed the same simulation framework described above and set the genetic correlation of the two simulated traits to 0.5. We first evaluated the bias in TCSC estimates of the genetic covariance explained by the cis-genetic component of gene expression in tissue $t'$ ($\omega_{g_e(tr)}$), for both causal and non-causal tissues. For causal tissues, TCSC produced conservative estimates of $\omega_{g_e(tr)}$, particularly at small eQTL sample sizes and when setting $G_{t'}$, to the number of significantly cis-heritable genes (rather than the number of true cis-heritable genes) (Figure 3A, Supplementary Table 5), analogous to single-trait simulations. For non-causal tissues, TCSC again produced very slightly negative estimates of $\omega_{g_e(tr)}$ (Figure 3A). We next evaluated the type I error of cross-trait TCSC for non-causal tissues. TCSC was conservative at lower eQTL sample sizes, with type I error less than 5% at $P = 0.05$ (Figure 3B). Finally, we evaluated the power of cross-trait TCSC for causal tissues. We determined that cross-trait TCSC was modestly powered at realistic eQTL sample sizes. When analyzing 300 gene expression samples, TCSC attained 29% power at $p < 0.05$ (Figure 3C) and 5.6% power at $p < 0.005$ (Supplementary Table 5).

We performed 6 secondary analyses. First, we set the eQTL sample size of the causal tissue to 100 individuals and the eQTL sample sizes of the non-causal tissues to range between 100-1,000 individuals. We observed inflated type I error for non-causal tissues (particularly for non-causal tissues with larger eQTL sample sizes), implying that large variations in eQTL sample sizes may compromise type I error (Supplementary Figure 2). Second, we varied the true values of $h_{g_e(tr)}^2$ (or $\omega_{g_e(tr)}$) for causal tissues. We determined that patterns of bias, type I error, and power were similar across different parameter values (Supplementary Figures 3-4). Third, we varied the number of causal tissues, considering 1, 2, or 3 causal tissues. We determined that the power of TCSC decreased with each additional causal tissue (Supplementary Figures 5-6). Fourth, we varied the number of non-causal tissues from 0 to 9. We determined that the conservative bias in estimates of $h_{g_e(tr)}^2$ (or $\omega_{g_e(tr)}$) for causal tissues became less pronounced as the number of non-causal tissues increased (Supplementary Figures 7-8). Fifth, we modified TCSC to not correct for bias in tissue co-regulation scores arising from differences between cis-genetic and cis-predicted expression. We determined that this increased the conservative bias in estimates of $h_{g_e(tr)}^2$ for causal tissues and also resulted in anti-conservative jackknife standard errors and type I error for non-causal tissues (Supplementary Figure 9), although the
impact on cross-trait simulations was limited (Supplementary Figure 10). Finally, we modified TCSC to use “cheating” tissue co-regulation scores constructed using true eQTL effects. We determined that this exacerbated the conservative bias in estimates of $h^2_{ge(t^r)}$ (or $\omega_{ge(t^r)}$) for causal tissues, while also producing anti-conservative jackknife standard errors and type I error for non-causal tissues (Supplementary Figures 11-12), perhaps because correlation patterns in TWAS $\chi^2$ statistics reflect in-sample correlations of cis-predicted expression. Further details of these secondary analyses are provided in the Supplementary Note.

Identifying tissue-specific contributions to 78 diseases and complex traits

We applied TCSC to publicly available GWAS summary statistics for 78 diseases and complex traits (average $N = 302K$; Supplementary Table 6) and gene expression data for 48 GTEx tissues (Table 1) (see Data Availability). The 78 diseases/traits (which include 33 diseases/traits from UK Biobank) were selected to have z-score > 6 for nonzero SNP-heritability (as in previous studies), with no pair of traits having squared genetic correlation > 0.16 and substantial sample overlap (Methods). The 48 GTEx tissues were aggregated into 39 meta-tissues (average $N = 266$, range: $N = 101-320$ individuals, 23 meta-tissues with $N = 320$) in order to reduce variation in eQTL sample size across tissues (Table 1 and Methods); below, we refer to these as “tissues” for simplicity. We constructed gene expression prediction models for an average of 3,993 significantly cis-heritable protein-coding genes (as defined above) in each tissue. We primarily report the proportion of disease heritability explained by the cis-genetic component of gene expression in tissue $t^r$ ($\pi_{t^r} = \frac{h^2_{ge(t^r)}}{h^2_g}$), as well as its statistical significance (using per-trait FDR).

TCSC identified 27 causal tissue-trait pairs with significantly positive contributions to disease/trait heritability at 10% FDR, spanning 9 distinct tissues and 23 distinct diseases/trait (Figure 4 and Supplementary Table 7). Many of the significant findings recapitulated known biology, including associations of whole blood with blood cell traits such as white blood cell count ($\pi_{t^r} = 0.21$, s.e. = 0.064, $P = 5.7 \times 10^{-4}$); liver with lipid traits such as LDL ($\pi_{t^r} = 0.20$, s.e. = 0.050, $P = 2.9 \times 10^{-5}$); and thyroid with hypothyroidism ($\pi_{t^r} = 0.31$, s.e. = 0.11, $P = 1.9 \times 10^{-3}$).

TCSC also identified several biologically plausible findings not previously reported in the genetics literature. First, aorta artery was associated with glaucoma ($\pi_{t^r} = 0.15$, s.e. = 0.051, $P = 1.3 \times 10^{-3}$). TCSC also identified aorta artery as a causal tissue for diastolic blood pressure (DBP) ($\pi_{t^r} = 0.078$, s.e. = 0.024, $P = 5.1 \times 10^{-4}$), which is consistent with DBP measuring the pressure exerted on the aorta when the heart is relaxed. High blood pressure is a known risk factor for glaucoma, explaining the role of aorta artery in genetic susceptibility to glaucoma. Second, esophagus muscularis was associated with the lung trait FEV1/FVC (\pi_{t^r} = 0.15, s.e. = 0.053, $P = 1.7 \times 10^{-3}$), consistent with previous work reporting that acidification of the esophagus leads to airway constriction and birth defects in the esophagus have consequences for pulmonary function likely stemming from shared causal mechanisms in Sonic hedgehog. FEV1/FVC is computed by dividing the amount of air an individual can expel from their lungs in one second (FEV1) by total lung capacity (FVC); therefore, the esophagus muscularis could...
impact either FEV1 and/or FVC. We sought to assess tissue-specific contributions to FEV1 and FVC separately. For FVC, no causal tissues were identified at 10% FDR, and the tissues with the most significant P-values were coronary artery and aorta artery (nominal $P = 0.02$ and 0.07, respectively; FDR > 10%), not esophagus muscularis. For FEV1, we analyzed summary statistics from ref.50 and determined that the tissue with the most significant P-value was esophagus muscularis (nominal $P = 8.3 \times 10^{-3}$; FDR > 10%), suggesting that esophagus muscularis may specifically impact FEV1 (we note that FEV1/FVC, FEV1, and FVC had similar z-scores for nonzero SNP heritability, implying similar power in this analysis). Third, TCSC identified heart left ventricle (in addition to whole blood) as a causal tissue for platelet count ($\pi_{e} = 0.091$, s.e. = 0.031, $P = 1.7 \times 10^{-3}$), consistent with the role of platelets in the formation of blood clots in cardiovascular disease51-54. In cardiovascular disease, platelets are recruited to damaged heart vessels after cholesterol plaques rupture, resulting in blood clots due to the secretion of coagulating molecules55; antiplatelet drugs have been successful at reducing adverse cardiovascular outcomes56. Moreover, the left ventricle serves as a muscle to pump blood throughout the body57, likely modulating platelet counts and other blood cell counts, creating detectable changes in serum from which platelet counts are measured. Other significant findings are discussed in the Supplementary Note, and numerical results for all tissues and diseases/traits analyzed are reported in Supplementary Table 8.

TCSC also increased the specificity of known tissue-trait associations. For high density lipoprotein (HDL), previous studies reported that deletion of a cholesterol transporter gene in adipose tissue reduces HDL levels, consistent with the fact that adipose tissues are storage sites of cholesterol and express genes involved in cholesterol transport and HDL lipidation58,59. While there are three adipose tissues represented in the GTEx data that we analyzed (subcutaneous, visceral, and breast tissue), TCSC specifically identified subcutaneous adipose ($\pi_{e} = 0.16$, s.e. = 0.054, $P = 1.5 \times 10^{-3}$; Figure 4), but not visceral adipose or breast tissue ($P > 0.05$; Supplementary Table 8), as a causal tissue for HDL. Previous studies have established that levels of adiponectin, a hormone released by adipose tissue to regulate insulin, are significantly positively correlated with HDL60-62 and more recently, a study has reported that adiponectin levels are associated specifically with subcutaneous adipose tissue and not visceral adipose tissue63; thus, the specific role of subcutaneous adipose tissue in HDL may be due to a causal mechanism related to adiponectin. We note that TCSC did not identify liver as a causal tissue for HDL (FDR > 10%), which may be due to limited power in liver due to smaller eQTL sample size. For waist-hip ratio adjusted for BMI (WHRadjBMI), previous studies reported colocalization of WHRadjBMI GWAS variants with cis-eQTLs in subcutaneous adipose, visceral adipose, liver, and whole blood64, consistent with WHRadjBMI measuring adiposity in the intraabdominal space which is likely regulated by metabolically active tissues65. TCSC specifically identified subcutaneous adipose ($\pi_{e} = 0.10$, s.e. = 0.037, $P = 2.4 \times 10^{-3}$; Figure 4), but not visceral adipose, breast, liver, or whole blood ($P > 0.05$; Supplementary Table 8), as a causal tissue for WHRadjBMI. The causal mechanism may involve adiponectin secreted from subcutaneous adipose tissue, which is negatively correlated with WHRadjBMI66. We note that the $P$ value distributions across traits are similar for subcutaneous adipose (median $P = 0.42$) and visceral adipose (median $P = 0.56$) and are comparable to the other 37 analyzed (median $P = 0.20 - 0.84$, Supplementary Table 9). For BMI, previous studies have broadly implicated the central...
nervous system, but did not reveal more precise contributions\textsuperscript{67,12,68,69,7,70}. TCSC specifically identified brain cerebellum (\(\pi_{t'} = 0.042, \text{s.e.} = 0.015, P = 2.6 \times 10^{-3}\)), but not brain cortex or brain limbic (\(P > 0.05\); Supplementary Table 8), as a causal tissue for BMI. Our finding is consistent with brain cerebellum accounting for the majority of the brain’s neurons and serving as a key area of neurogenesis in the developing brain\textsuperscript{71,72}.

We performed a secondary analysis in which we removed tissues with eQTL sample size less than 320 individuals, as these tissues may often be underpowered (Figure 2C). Results are reported in Supplementary Figure 13 and Supplementary Table 10. The number of causal tissue-trait pairs with significantly positive contributions to disease/trait heritability (at 10% FDR) increased from 27 to 33, likely due to a decrease in multiple hypothesis testing burden from removing underpowered tissues. The 33 significant tissue-trait pairs reflect a gain of 14 newly significant tissue-trait pairs (and a loss of 8 formerly significant tissue-trait pairs, of which 5 were lost because the tissue was removed), but estimates of \(\pi_{t'}\) for each significant tissue-trait pair were not statistically different from our primary analysis (Supplementary Table 11).

Notably, among the newly significant tissue-trait pairs, esophagus mucosa was associated with eczema (\(\pi_{t'} = 0.085, \text{s.e.} = 0.032, P = 4.0 \times 10^{-3}\)), consistent with the comorbidity of acid reflux and eczema and hypotheses of a causal mechanism in which acid reflux stimulates the immune cells in the mucosa, prompting an allergic response such as eczema\textsuperscript{73}. Other newly significant findings are discussed in the Supplementary Note, and numerical results for all tissues and diseases/traits are reported in Supplementary Table 10.

We also performed a brain-specific analysis in which we applied TCSC to 41 brain traits (average \(N = 226K\), Supplementary Table 12) while restricting to 13 individual GTEx brain tissues (Supplementary Table 13), analogous to previous work\textsuperscript{7}. The 41 brain traits reflect a less stringent squared genetic correlation threshold of 0.25. The 13 GTEx brain tissues were analyzed without merging tissues into meta-tissues, and irrespective of eQTL sample size (range: \(N = 101\text{--}189\) individuals); we expected power to be limited due to the eQTL small sample sizes and substantial co-regulation among individual brain tissues. TCSC identified 16 brain tissue-brain trait pairs at 10% FDR (Supplementary Figure 14, Supplementary Table 14)—substantially more than the 2 brain tissue-brain trait pairs in Figure 4, although most results are not directly comparable due to the more fine-grained individual brain tissues analyzed. For ADHD, TCSC identified brain hippocampus as a causal tissue (\(\pi_{t'} = 0.28, \text{s.e.} = 0.10, P = 2.5 \times 10^{-3}\)), consistent with the correlation between hippocampal volume and ADHD diagnosis in children\textsuperscript{74}. A recent ADHD GWAS identified a locus implicating the FOXP2 gene\textsuperscript{75}, which has been proposed to regulate dopamine secretion in mice\textsuperscript{76}; hippocampal activation results in the firing of dopamine neurons\textsuperscript{77}. For BMI, TCSC identified brain amygdala (in addition to brain cerebellum, a constituent tissue of the meta-tissue implicated in Figure 4) as a causal tissue (\(\pi_{t'} = 0.54, \text{s.e.} = 0.023, P = 8.3 \times 10^{-3}\), consistent with previous work linking the amygdala to obesity and dietary self-control\textsuperscript{78}, although no previous study has implicated the amygdala in genetic regulation of BMI. Other significant findings are discussed in the Supplementary Note, and numerical results for all brain tissues and brain traits analyzed are reported in Supplementary Table 14.
Comparisons of TCSC to other methods

We compared TCSC to two previous methods, RTC Coloc\(^2\) and LDSC-SEG\(^7\), that identify disease-critical tissues using gene expression data. RTC Coloc identifies disease-critical tissues based on tissue specificity of eQTL-GWAS colocalizations. LDSC-SEG identifies disease-critical tissues based on heritability enrichment of specifically expressed genes. We note that RTC Coloc and LDSC-SEG analyze each tissue marginally, whereas TCSC jointly models contributions from each tissue to identify causal tissues (analogous to the distinction in GWAS between marginal association and fine-mapping\(^22\)). Thus, we hypothesized that RTC Coloc and LDSC-SEG may output multiple highly statistically significant associated tissues for a given trait, whereas TCSC may output a single causal tissue with weaker statistical evidence of causality. To assess whether TCSC indeed attains higher specificity, we evaluated the results of each method both for causal tissues identified by TCSC and for the most strongly co-regulated tagging tissue (based on Spearman \(\rho\) for estimated eQTL effect sizes, averaged across genes, from ref.\(^18\)). Our primary analyses focused on 5 representative traits, defined as the 5 traits with highest \(z\)-score for nonzero SNP-heritability among the 10 diseases/traits with at least one tissue-trait association for each of the three methods (Methods).

Results for the 5 representative traits are reported in Figure 5 and Supplementary Table 15; results for all 23 diseases/traits with causal tissue-trait associations identified by TCSC (Figure 4) are reported in Supplementary Figure 15 and Supplementary Table 16, and complete results for all diseases/traits and tissues included in these comparisons are reported in Supplementary Table 17. We reached three main conclusions. First, for a given disease/trait, RTC Coloc typically implicates a broad set of tissues (not just strongly co-regulated tissues) (Figure 5A); for example, for BMI, RTC Coloc implicated 6 of 10 tissues in Figure 5 (and 9 of 17 tissues in Supplementary Figure 15). Second, for a given disease/trait, LDSC-SEG typically implicates a small set of strongly co-regulated tissues (Figure 5B); for BMI, LDSC-SEG implicated 2 of 10 tissues in Figure 5 (and 2 of 17 tissues in Supplementary Figure 15), consisting of brain cereb. and its most strongly co-regulated tagging tissue. Third, for a given disease/trait, TCSC typically implicates one causal tissue (Figure 5C); for BMI, TCSC implicated only brain cereb. as a causal tissue, with even the most strongly co-regulated tagging tissue reported as non-significant. For diastolic blood pressure (DBP), RTC Coloc implicated 9 of 10 tissues, LDSC-SEG implicated 2 of 10 tissues (consisting of aorta artery and its most strongly co-regulated tagging tissue), and TCSC implicated only aorta artery; as noted above, this result is consistent with DBP measuring the pressure exerted on the aorta while the heart is relaxed\(^38\). As expected, the higher specificity of TCSC in identifying unique causal tissues involves a trade-off of reduced sensitivity, with less significant (lower \(-\log_{10} P\)-value and lower \(-\log_{10} FDR\)) results for causal tissues in Figure 5C than in Figure 5A and Figure 5B (Supplementary Table 16). We observed similar patterns for HDL and WHRadjBMI, for which the increased specificity of TCSC is discussed above (Supplementary Figure 15, Supplementary Table 17, Supplementary Note).

We also observed similar patterns when comparing TCSC to RTC Coloc and LDSC-SEG in the brain-specific analysis of Supplementary Figure 14 (Supplementary Figure 16, Supplementary Table 18, Supplementary Note).
Identifying tissue-specific contributions to the genetic covariance between two diseases/traits

We applied cross-trait TCSC to 256 pairs of disease/traits (Supplementary Table 19) and gene expression data for 48 GTEx tissues (Table 1) [see Data Availability]. Of 3,003 pairs of the 78 disease/traits analyzed above, the 256 pairs of diseases/traits were selected based on significantly nonzero genetic correlation ($p < 0.05 / 3,003$; see Methods). The 48 GTEx tissues were aggregated into 39 meta-tissues, as before (Table 1 and Methods). We primarily report the proportion of genetic covariance explained by the cis-genetic component of gene expression in tissue $t'$ ($\zeta_{t'} = \omega_{ge(t')} / |\omega_{g}|$), as well as its statistical significance (using per-trait FDR).

TCSC identified 30 causal tissue-trait covariance pairs with significant contributions to trait covariance at 10% FDR, spanning 21 distinct tissues and 23 distinct trait pairs (Figure 6A and Supplementary Table 20). For 27 of the 30 causal tissue-trait covariance pairs, the causal tissue was non-significant for both constituent traits in the single-trait analysis of Supplementary Table 8. Findings that recapitulated known biology included both examples involving a tissue-trait pair that was significant in the single-trait analysis (Figure 4) and examples in which both tissue-trait pairs were non-significant in the single-trait analysis (Supplementary Table 8). Consistent with the significant contribution of liver to LDL heritability in the single-trait analysis, TCSC identified a positive contribution of liver to the genetic covariance of LDL and total cholesterol ($\zeta_{t'} = 0.090$, s.e. = 0.029, $P = 1.0 \times 10^{-3}$), and consistent with the positive contributions of whole blood to eosinophil count heritability and to white blood cell count heritability in the single-trait analysis, TCSC identified a positive contribution of whole blood to the genetic covariance of eosinophil count and white blood cell count ($\zeta_{t'} = 0.32$, s.e. = 0.12, $P = 2.4 \times 10^{-3}$). TCSC also identified a negative contribution of muscle skeletal to the genetic covariance of WHRadjBMI and total protein ($\zeta_{t'} = -0.27$, s.e. = 0.084, $P = 8.0 \times 10^{-4}$), consistent with reduced muscle mass in individuals with high waist-hip-ratio$^{79}$; this suggests that cross-trait TCSC can reveal tissue-trait biology distinct from what is identified in the single-trait analyses.

TCSC also identified several biologically plausible findings not previously reported in the genetics literature. First, LCLs had a significantly negative contribution to the genetic covariance of eosinophil count and white blood cell count ($\zeta_{t'} = -0.081$, s.e. = 0.028, $P = 1.8 \times 10^{-3}$, in contrast to the positive contribution of whole blood). This is plausible as previous studies have reported the suppression of proliferation of lymphocytes (the white blood cell hematopoietic lineage from which LCLs are derived) by molecules secreted from eosinophils$^{80-82}$. The contrasting results for whole blood and LCLs suggest that genetic covariance may reflect distinct tissue-specific contributions. Second, brain substantia nigra had a significantly positive contribution to the genetic covariance of BMI and red blood cell count (RBC count) ($\zeta_{t'} = 0.28$, s.e. = 0.084, $P = 4.6 \times 10^{-4}$), while pancreas had a significantly negative contribution ($\zeta_{t'} = -0.25$, s.e. = 0.079, $P = 8.7 \times 10^{-4}$). In the brain, energy metabolism is regulated by oxidation and previous work has shown that red blood cells play a large role in these metabolic processes as oxygen sensors$^{83}$; in addition, previous studies have reported differences in the level of oxidative enzymes in red blood cells between individuals with high BMI and low BMI$^{84,85}$.
suggesting that genes regulating oxidative processes might have pleiotropic effects on RBC count and BMI. In the pancreas, pancreatic inflammation (specifically acute pancreatitis) is associated with reduced levels of red blood cells, or anemia, while pancreatic fat is associated with metabolic disease and increased BMI. Once again, the contrasting results for brain substantia nigra and pancreas suggest that genetic covariance may reflect distinct tissue-specific contributions. Third, brain substantia nigra had a significantly negative contribution to the genetic covariance of age at first birth and height ($\zeta_{t'} = -0.11$, s.e. = 0.032, $P = 4.5 \times 10^{-4}$).

Previous work in C. elegans reported that fecundity is positively regulated by dopamine, which is produced in the substantia nigra. Dopamine also plays a role in regulating the levels of key growth hormones such as IGF-1 and IGF-BP3, and has been previously shown to be associated with height.

Other significant findings are discussed in the Supplementary Note. Numerical results for all tissues and disease/trait pairs analyzed are reported in Supplementary Table 21.

As noted above, for 27 of the 30 causal tissue-trait covariance pairs, the causal tissue was non-significant for both constituent traits. We sought to formally assess whether differences in tissue-specific contributions to genetic covariance vs. constituent trait heritability were statistically significant. Specifically, for each causal tissue-trait covariance pair, we estimated the differences between the tissue-specific contribution to covariance ($\zeta_{t'}$) and the tissue-specific contributions to heritability for each constituent trait ($\pi_{t'}$). We identified eight tissue-trait covariance pairs for which these differences were statistically significant at 10% FDR for both constituent traits and $\pi_{t'}$ was non-significant for both constituent traits (Figure 6A and Supplementary Table 22). For BMI and RBC count, the positive contribution of brain substantia nigra and the negative contribution of pancreas to genetic covariance were each significantly larger than the respective contributions of those tissues to BMI and RBC count heritability, which were non-significant (Figure 6B). Other examples are discussed in the Supplementary Note. Numerical results for all tissues and trait pairs are reported in Supplementary Table 22.

Discussion

We developed a new method, tissue co-regulation score regression (TCSC), that disentangles causal tissues from tagging tissues and partitions disease heritability (or genetic covariance of two diseases/trait) into tissue-specific components. We applied TCSC to 78 diseases and complex traits and 48 GTEx tissues, identifying 27 tissue-trait pairs (and 30 tissue-trait covariance pairs) with significant tissue-specific contributions. TCSC identified biologically plausible novel tissue-trait pairs, including associations of aorta artery with glaucoma, esophagus muscularis with FEV1/FVC, and heart left ventricle with platelet count. TCSC also identified biologically plausible novel tissue-trait covariance pairs, including a negative contribution of LCLs to the covariance of eosinophil count and white blood cell count (in contrast to the positive contribution of whole blood) and a positive contribution of brain substantia nigra and a negative contribution of pancreas to the covariance of BMI and red blood cell count; in particular, our findings suggest that genetic covariance may reflect distinct tissue-specific contributions.
TCSC differs from previous methods in jointly modeling contributions from each tissue to disentangle causal tissues from tagging tissues (analogous to the distinction in GWAS between marginal association and fine-mapping\(^2\)). We briefly discuss several other methods that use eQTL or gene expression data to identify disease-associated tissues. RTC Coloc identifies disease-associated tissues based on tissue specificity of eQTL-GWAS colocalizations\(^2\); this study made a valuable contribution in emphasizing the importance of tissue co-regulation, but did not model tissue-specific effects, such that RTC Coloc may implicate many tissues (Figure 5A). LDSC-SEG identifies disease-critical tissues based on heritability enrichment of specifically expressed genes\(^7\); this distinguishes a focal tissue from the set of all tissues analyzed, but does not distinguish closely co-regulated tissues (Figure 5B). MaxCPP models contributions to heritability enrichment of fine-mapped eQTL variants across tissues or meta-tissues\(^4\); although this approach proved powerful when analyzing eQTL effects that were meta-analyzed across all tissues, it has limited power to identify disease-critical tissues: fine-mapped eQTL annotations for blood (resp. brain) were significant conditional on annotations constructed using all tissues only when meta-analyzing results across a large set of blood (resp. brain) traits (Fig. 4 of ref.\(^4\)). eQTLenrich compares eQTL enrichments of disease-associated variants across tissues\(^3\); this approach produced compelling findings for eQTL that were aggregated across tissues, but tissue-specific analyses often implicated many tissues (Fig. 1d of ref.\(^3\)). MESC estimates the proportion of heritability causally mediated by gene expression in assayed tissues\(^93\); this study made a valuable contribution in its strict definition and estimation of mediated effects (see below), but did not jointly model distinct tissues and had limited power to distinguish disease-critical tissues (Fig. 3 of ref.\(^5\)). CAFEH leverages multi-trait fine-mapping methods to simultaneously evaluate all tissues for colocalization with disease\(^5\); however, this locus-based approach does not produce genome-wide estimates and it remains the case that many (causal or tagging) tissues may colocalize with disease under this framework. Likewise, methods for identifying tissues associated to disease/trait covariance do not distinguish causal tissues from tagging tissues\(^94,95\).

We note several limitations of our work. First, joint-fit effects of gene expression on disease may not reflect biological causality; if a causal tissue or cell type is not assayed\(^96\), TCSC may identify a co-regulated tissue (e.g. a tissue whose cell type composition favors a causal cell type) as causal. We anticipate that this limitation will become less severe as potentially causal tissues, cell types and contexts are more comprehensively assayed. Second, TCSC does not achieve a strict definition or estimation of mediated effects; this is conceptually appealing and can, in principle, be achieved my modeling non-mediated effects, but may result in limited power to distinguish disease-critical tissues\(^93\). Third, TCSC has low power at small eQTL sample sizes; in addition, TCSC estimates are impacted by the number of significantly cis-heritable genes in a focal tissue, which can lead to conservative bias at small eQTL sample sizes. We anticipate that these limitations will become less severe as eQTL sample sizes increase. Fourth, TCSC is susceptible to large variations in eQTL sample size, which may compromise type I error; therefore, there is a tradeoff between maximizing the number of tissues analyzed and limiting the variation in eQTL sample size. Fifth, TCSC assumes that causal gene expression-disease effects are independent across tissues; this assumption may become invalid for tissues and cell
types assayed at high resolution. However, we verified via simulations that in the case of two
causal tissues with identical gene expression-trait effects, TCSC correctly identified both tissues
as causal and split estimates of explained heritability across the tissues (Supplementary Figure 17). Sixth, TCSC does not formally model measurement error in tissue co-regulation scores, but
instead applies a heuristic bias correction. We determined that the bias correction generally
performs well in simulations. Seventh, TCSC does not produce locus-specific estimates or
identify causal tissues at specific loci. However, genome-wide results from TCSC may be used as
a prior for locus-based methods (analogous to GWAS fine-mapping with functional priors). Finally, we did not apply TCSC to single-cell RNA-seq (scRNA-seq) data, which represents a
promising new direction as scRNA-seq sample sizes increase; we caution that scRNA-seq
data may require new eQTL modeling approaches. Despite these limitations, TCSC is a
powerful and generalizable approach for modeling tissue co-regulation to estimate tissue-
specific contributions to disease.

Code Availability

FUSION software: http://gusevlab.org/projects/fusion/

Data Availability

We have made 78 GWAS summary statistics and 41 brain-specific summary statistics publicly
available at https://github.com/TiffanyAmariuta/TCSC/tree/main/sumstats, TWAS association
statistics publicly available at
scores publicly available at
https://github.com/TiffanyAmariuta/TCSC/tree/main/coregulation_scores, and TCSC output
will be publicly available at https://github.com/TiffanyAmariuta/TCSC/tree/main/results.

Acknowledgements

We thank Huwenbo Shi, Martin Zhang, and Benjamin Strober for helpful discussions. This work
was funded by NIH grants U01 HG009379, R01 MH101244, R37 MH107649, R01 HG006399, R01
MH115676 and U01 HG012009.
**Online Methods**

**TCSC regression**

TCSC leverages the fact that the TWAS $\chi^2$ statistic for a gene-tissue pair includes the direct effects of the gene on the disease as well as the tagging effects of co-regulated tissues and genes with shared eQTLs or eQTLs in LD. Thus, genes that are co-regulated across many tissues will tend to have higher $\chi^2$ statistics than genes regulated in a single tissue. TCSC determines that a tissue causally contributes to disease if genes with high co-regulation to the tissue have higher TWAS $\chi^2$ statistics than genes with low co-regulation to the tissue.

We model the genetic component of gene expression as a linear combination of SNP-level effects:

$$W_{jgt} = \sum_m X_{jm} \beta_{gtm},$$

where $W_{jgt}$ is the cis-genetic component of gene expression in individual $j$ for gene $g$ and tissue $t$, $X_{jm}$ is the standardized genotype of individual $j$ for SNP $m$, and $\beta_{gtm}$ is the effect of the $m^{th}$ SNP on gene expression of gene $g$ in tissue $t$.

TCSC assumes that true gene-disease effects are independent and additive across tissues. However, $cis$ eQTLs are highly correlated across tissues, leading to tagging from co-regulated tissues. We model phenotype as a linear combination of genetic components of gene expression across genes in different tissues:

$$Y_j = \sum_t \sum_g W_{jgt} \alpha_{gt} + \epsilon_j,$$

where $Y_j$ is the (binary or continuous-valued) phenotype of individual $j$, $\alpha_{gt}$ is the gene-trait effect size and $\epsilon_j$ is the error term.

We define the disease heritability explained by $cis$-predicted expression across all tissues as follows:

$$h^2_{ge} = \sum_t \sum_g \alpha^2_{gt},$$

analogous to the relationship between SNP effect sizes and SNP-heritability. It follows that the disease heritability explained by a particular tissue $t'$ is

$$h^2_{ge(t')} = \sum_g \alpha^2_{gt'},$$

Let $\alpha_{gt'}$ be a random variable drawn from a normal distribution with mean zero and tissue-specific variance $var(\alpha_{gt'}) = \tau_{t'}$. Then

$$h^2_{ge(t')} = \sum_g var(\alpha_{gt'}) = \sum_g \tau_{t'} G_{t'},$$

where $G_{t'}$ is the number of $cis$ heritable genes in $t'$. With this variance term, we can define a polygenic model that relates TWAS $\chi^2$ statistics to co-regulation scores, which explicitly model the covariance structure of the $\chi^2$ statistics. This strategy is analogous to modeling the dependence of GWAS $\chi^2$ statistics on LD scores. We model the expected value of the TWAS $\chi^2$ across genes and tissues as follows:
\[ E[\chi^2_{gt}] = N \sum_t \tau_t l(g, t; t') + 1, \]  

(1)

where \( g \) indexes genes, \( t \) indexes tissues, \( \chi^2_{gt} \) is a vector of TWAS \( \chi^2 \) statistics across all significantly cis-heritable genes and tissues, \( N \) is GWAS sample size, \( \tau_t \) is the per-gene heritability explained by predicted gene expression in tissue \( t' \), and \( l(g, t; t') \) are tissue and gene co-regulation scores (see below). From the derivation, the genome-wide tissue-specific contribution to disease heritability is estimated as

\[ h^2_{ge(t')} = G_t \tau_t. \]

(2)

For the analysis of tissue-specific contributions to the covariance between two diseases, we can extend TCSC by using products of TWAS z-scores. Following the polygenic model described above, the expected product of TWAS z-scores in disease 1 and disease 2 for gene \( g \) and tagging tissue \( t \) is

\[ E[z_{gt}^1 \times z_{gt}^2] = \sqrt{N_1 N_2} \sum_t l(g, t; t') \omega_{ge(t')} / G_{t'} + \rho N_s / \sqrt{N_1 N_2} \]

(2)

where \( N_1 \) is GWAS sample size for disease 1, \( N_2 \) is GWAS sample size for disease 2, \( t' \) indexes causal tissues, \( l(g, t; t') \) are tissue co-regulation scores (see below), \( \omega_{ge(t')} \) is the genetic covariance explained by the cis-genetic component of gene expression in tissue \( t' \), \( G_{t'} \) is the number of significantly cis-heritable genes in tissue \( t' \) (see below), \( \rho \) is the phenotypic correlation between disease 1 and disease 2, and \( N_s \) is the number of overlapping GWAS samples between disease 1 and disease 2. The last term represents the intercept\(^26\), and while we use a free intercept in the multivariate regression on co-regulation scores, the estimation of this term only plays a role in the estimation of regression weights (see below).

For estimates of \( h^2_{ge(t')} \) and \( \omega_{ge(t')} \), we use a free intercept; the estimation of \( \rho N_s / \sqrt{N_1 N_2} \) serves only to inform the heteroscedasticity weights (see below) and is not used in the multivariate TCSC regression to estimate \( \omega_{ge(t')} \). To estimate standard errors, we use a genomic block jackknife over 200 genomic blocks with an equal number of genes in each. The standard deviation is computed as the square root of the weighted variance across the jackknife estimates (where the weight of each block is equal to the sum of the regression weights for the genes in that block) multiplied by 200 blocks. We expect that the jackknife standard error will be conservative relative to the empirical standard error across estimates due to variation in causal signal across loci\(^35\).

**Estimating tissue co-regulation scores and correcting for bias**

We define the co-regulation score of gene \( g \) with tissues \( t \) and \( t' \) as

\[ l(g, t; t') = \sum_g r^2(W_{gt}, W_{g't'}), \]

(3)
where $W$ denotes the cis-genetic component of gene expression for a gene-tissue pair across individuals and genes $g'$ are within +/- 1 Mb of the focal gene $g$. TCSC corrects for bias in tissue co-regulation scores arising from differences between cis-genetic vs. cis-predicted expression (analogous to GCSC\textsuperscript{21}). We apply bias correction to co-regulation scores in the special case when $t = t'$. While co-regulation scores aim to estimate the squared correlation of cis-predicted gene expression of gene $g$ and tissue $t$ (corresponding to the TWAS $\chi^2_g,t$ statistic) with the true cis-genetic component of gene expression of cis-genes in tissue $t'$, when $g = g'$ and $t = t'$, the estimated value of $r^2(W_{g,t}, W_{g',t'})$ will always equals one. However, in practice, the prediction $R^2$ of a gene expression model will always be less than one, except in the case of a perfect model, causing the value of $r^2(W_{g,t}, W_{g',t'})$ to be systematically inflated. Therefore, when $g = g'$ and $t = t'$, we set

\begin{equation}
    r^2(W_{g,t}, W_{g',t'}) = R^2 / h^2_{GCTA}
\end{equation}

where $R^2$ is the cross-validation prediction statistic of the gene expression model for gene $g$ in tissue $t$ and $h^2_{GCTA}$ is the GCTA-estimated cis-heritability of gene expression for gene $g$ in tissue $t$. The quotient $R^2 / h^2_{GCTA}$ is the accuracy of the gene expression prediction model, which reflects the upper bound on how much the cis-predicted expression can be correlated with the true cis-genetic component of gene expression.

**TCSC regression weights**

TCSC uses three sets of regression weights to increase power (analogous to GCSC\textsuperscript{21}). The first regression weight is inversely proportional to $L(g, t)$, the total co-regulation score of each gene-tissue pair summed across tissues $t'$:

\begin{equation}
    L(g, t) = \sum_{t'} \sum_{g'} r^2(W_{g,t}, W_{g',t'})
\end{equation}

(without applying bias correction; see above), which allows TCSC to properly account for redundant contributions of co-regulated genes to TWAS $\chi^2$ statistics.

The second regression weight is inversely proportional to $T(g, t)$, the number of tissues in which a gene is significantly cis-heritable:

\begin{equation}
    T(g, t) = \sum_{t' \ni g \cdot t'} \text{significantly cis-heritable} 1,
\end{equation}

thereby up-weighting signal from genes that are regulated in a limited number of tissues and preventing TCSC from attributing more weight to genes that are co-regulated across many tissues.

The third regression weight is inversely proportional to $H_{h^2}(g, t)$, the heteroscedasticity of $\chi^2$ statistics, and is computed differently for estimates of $h^2_{ge}(t')$ than for estimates of $\omega_{ge}(t')$ (analogous to GCSC\textsuperscript{21} and cross-trait LDSC\textsuperscript{26}, respectively).
For estimates of $h^2_{ge(t)}$, we estimate $H_{h^2}(g, t)$ in two steps. First, we make a crude estimate of heritability explained by predicted expression ($h^2_{ge}$) as follows:

$$
\mu_X = N \mu_L h^2_{ge} + 1,
$$

where $\mu_X$ is the mean $\chi^2$ statistic:

$$
\mu_X = \frac{\sum t' \sum g' t^2_{g't'}}{\sum t' g_{tt'}},
$$

where $N$ is the GWAS sample size, $g'$ iterates over significantly cis-heritable genes and $t'$ iterates over tissues, and $\mu_L$ is the mean value of total co-regulation across tissues $t'$,

$$
\mu_L = \frac{\sum t' \sum g' t L(g', t)}{\sum t' g_{tt'}},
$$

Then, we compute the heteroscedasticity for each significantly cis-heritable gene-tissue pair as

$$
H_{h^2}(g, t) = (N L(g, t) h^2_{ge} + 1)^2.
$$

Finally, we combine the three regression weights as follows:

$$
Weight_{h^2}(g, t) = \frac{1}{L(g, t) T(g, t) H_{h^2}(g, t)}.
$$

For estimates of $\omega_{ge(t')}$, we estimate $H_{\omega}(g, t)$ in two steps. First, we regress the products of TWAS z-scores on total tissue co-regulation scores, $L(g, t)$, using regression weights, $Weight_{\omega}(g, t)$, computed as follows:

$$
Weight_{\omega}(g, t) = \frac{1}{L(g, t) T(g, t) H_{\omega}(g, t)}
$$

where $H_{\omega}(g, t)$ is first estimated as follows:

$$
H_{\omega}(g, t) = \left(N_1 L(g, t) h^2_{ge} (\text{trait 1}) + 1\right) \left(N_2 L(g, t) h^2_{ge} (\text{trait 2}) + 1\right) + \frac{\left(\frac{\sum t' g'_{tt'} \omega_{ge} L(g, t)}{\sum t' g_{tt'}}\right)^2}{\frac{\sum t' g'_{tt'}}{T'}} ^{\frac{1}{2}}
$$

where $h^2_{ge} (\text{trait 1})$ is the crude heritability estimate for trait 1 and $h^2_{ge} (\text{trait 2})$ is the crude heritability estimate for trait 2, $\omega_{ge}$ is estimated as $\frac{\sum t' g'_{tt'} \omega_{ge} L(g, t)}{\sum t' g_{tt'}}$, $N_1$ is the sample size of the first GWAS, $N_2$ is the sample size of the second GWAS, and $T'$ is the total number of tissues in the regression.

Second, we use the regression intercept to estimate the product $\rho N_S$:

$$
\rho N_S = \text{intercept} \ast \sqrt{N_1 N_2},
$$

where $\rho$ represents the phenotypic correlation between trait 1 and 2 and $N_S$ represents the number of shared samples between GWAS 1 and 2. We also use the coefficient of the
regression to update our estimate of \( \omega_{ge} \), such that we may update the heteroscedasticity

weight as follows:

\[
H_{\omega}(g, t) = \left( N_1 L(g, t) h_{ge}^2(\text{trait 1}) + 1 \right) \left( N_2 L(g, t) h_{ge}^2(\text{trait 2}) + 1 \right) \\
+ \left( \frac{\sqrt{N_1 N_2} \omega_{ge} L(g, t)}{\sum_{t'} T_{t'}} + \rho N_5 \sqrt{N_1 N_2} \right). 
\]

Finally, we combine the three regression weights as follows:

\[
\text{Weight}_{\omega}(g, t) = \frac{1}{L(g, t) T(g, t) H_{\omega}(g, t)}. 
\]

Simulating TCSC

We employed a widely used TWAS simulation framework (Mancuso Lab TWAS Simulator, see Code Availability) to assess the power, bias, and calibration of TCSC in the presence of co-regulation across genes and tissues. We simulated a genome in which there are 1,245 genes, of which 125 (10%) are causal. Each gene belongs to one of 249 independent genomic blocks, representing approximately 1Mb intervals of chromosome 1, which is 249 Mb in length. Each block contains 50 variants and 5 genes.

Each primary simulation consists of 10 tissues, of which at least one is causal, defined as having nonzero gene-trait effect sizes. We create a covariance structure among tissues mimicking empirical GTEx data. We use a previously published method to estimate the causal cross-tissue correlation of eQTL effect sizes which is 0.75. We observe that not all GTEx tissues are equally correlated to one another. We estimate three different cross-tissue eQTL correlation quantities: (1) average correlation across all pairs of tissues = 0.75, (2) average correlation across similar tissues = 0.80, e.g. brain (13 in GTEx) or adipose (2 in GTEx) tissues, and (3) average correlation across dissimilar tissues, e.g. pairs of brain and adipose tissues = 0.74. To represent these biological modules, we let simulated tissues 1-3 have higher correlation of true eQTL effects to one another than to other tissues; likewise for tissues 4-6 and 7-10. We set covariance parameters, described below, such that the similar tissues had an average eQTL correlation of 0.789 across genes, dissimilar tissues have an average eQTL correlation of 0.737, and the average eQTL correlation across any pair of tissues is 0.751. We do not simulate linkage disequilibrium, which is another source of co-regulation in real data.

Rather, we simulate co-regulation strictly using shared eQTLs across genes and tissues. We do not expect this to impact the interpretation of our empirical results. We simulate each gene has 5 true eQTLs, based on the upper bound of empirical data from GTEx and others, as well as the value used in other TWAS simulation methods. Between co-regulated tissues, each pair of genes shares 3 eQTLs. The minimum allowed \( cis \)-heritability of a gene is 0.01 in our simulations.

\( Cis \)-heritability is approximated as the sum of squared true \( cis \)-eQTL effect sizes, as done previously; this approximation of \( cis \)-heritability is more accurate than if we were to model linkage disequilibrium between \( cis \)-eQTLs. Effect sizes for the 3 shared eQTLs across tissues are sampled from a multivariate normal distribution with mean 0 and a variance-covariance matrix.

We define the variance and covariance terms of this matrix such that (1) the proportion of genes detected as significantly \( cis \)-heritable by GCTA at a given sample size and (2) the average
cis heritability of detected genes at a given sample size match empirical observations from GTEx data at sample sizes N = 100, 200, 300 and 500. As a result, the diagonal of the variance-covariance matrix, e.g. the variance term, is set to 0.015, and the off-diagonal elements are set to the product of the variance term and the desired correlation for each tissue pair, described above.

For each of 1,000 independent simulations per analysis, we simulate a GWAS (N = 100,000) by creating a complex trait which is the summation of the genetic components of causal gene expression (in the causal tissue). We use simulated genotypes drawn from a standard normal distribution. Gene-trait effect sizes are drawn from a normal distribution with mean 0 and variance 1. In cross-trait TCSC analysis, effect sizes across genes between the two traits are correlated with default $R_g = 0.5$. To simulate a GWAS trait, we first compute the genetic component of each gene, which is the product of GWAS cohort genotypes and eQTL effects, such that we have 125 gene-specific traits. We then add noise to each gene-specific trait such that the total variance of the phenotype explained by the five eQTLs from the causal tissue is equal to a specified value; the value of $h^2_{ge(t)}$ in primary simulations is 5%. Then, we multiply each gene-specific trait by the causal gene-trait effect size, consistent with the additive generative model of gene-level effects on trait (see above). Finally, we take the sum across all gene-specific traits to make one complex trait, where the total variance of the trait explained by gene effects from the causal tissue is $h^2_{ge(t)}$, e.g. 5%.

We simulate an eQTL cohort of various gene expression sample sizes (N = 100, 200, 300, 500, 1000) using simulated genotypes drawn from a standard normal distribution. We simulate total gene expression in the eQTL cohort by adding a desired amount of noise to the genetic component of gene expression, e.g. the product of individual genotypes and true eQTL effect sizes, with variance equal to one minus the gene expression heritability, which is the sum of squared eQTL effects. Next, we fit gene expression prediction models by regressing the total gene expression on eQTL cohort genotypes of cis variants using lasso regularization, a standard approach used in TWAS. We define significantly cis-heritable genes as genes with GCTA heritability $P$ value $< 0.01^{20}$ and heritability estimate $> 0$, and adjusted-$R^2 > 0$ in cross-validation prediction.

Then we estimate co-regulation scores at each different eQTL sample size by predicting gene expression into a cohort of 500 individuals, to approximate the size of the European sample of 1000 Genomes (N = 489). Using significantly cis-heritable genes from each tissue at a given sample size, we estimate gene and tissue co-regulation scores $l(g, t; t')$ as described above, including bias correction. In simulations, cis genes are defined as genes within the same 1 Mb block.

Then we apply TWAS to individual-level simulated GWAS data and gene expression prediction models. We predict gene expression into each of the 100,000 GWAS cohort individuals across all significantly cis-heritable genes for each tissue. We regress each complex trait on predicted gene expression to obtain TWAS $z$-scores. Finally, we run TCSC by regressing TWAS $\chi^2$ statistics, or products of TWAS $z$-scores, on bias-corrected gene and tissue co-regulation scores.

Simulating the RTC Coloc method
We simulated the RTC Coloc method\(^2\) by leveraging our existing TCSC simulation framework such that both methods could be compared via application to same simulated data. We precisely followed the steps of the RTC Coloc method. We first simulate individual level genotypes of a GWAS (\(N = 100,000\)) using a standard normal distribution, as described. We then regress the simulated complex traits on each of the GWAS genotypes in turn using a Bonferroni significance threshold to identify genome-wide significant variants. We then select one null SNP per GWAS variant for comparison. Next, we simulate an eQTL cohort consisting of total gene expression and genotypes, using the same underlying true eQTL effect sizes as for TCSC simulations. Then we perform colocalization analysis of GWAS variants with eQTLs, across 10 tissues at 4 different eQTL sample sizes, to obtain the regulatory trait concordance (RTC) score. This is repeated for the set of null variants. Next, we perform colocalization analysis of eQTL variants between pairs of tissues to obtain tissue-sharing RTC scores, and similarly repeat this for null variants. GWAS-eQTL RTC scores are divided by tissue-sharing RTC scores summed across variants. Tissue-specific enrichment is computed as the ratio of this quotient to the null quotient. The enrichment \(P\) value is obtained using a Wilcox test comparing the values of the quotient to the values of the null quotient.

**Gene expression prediction models and tissue co-regulation scores in GTEx data**

We downloaded GTEx v8 gene expression data for 49 tissues. We excluded tissues with fewer than 100 samples, e.g. kidney cortex (\(n = 69\)). We retained only European samples for each tissue, as labeled by GTEx via PCA of genotypes. We constructed gene expression models for two scenarios: (1) subsampling to 320 individuals including meta-analyzed tissues (Table 1) or (2) using all European samples per tissue. We recommend meta-analyzing gene expression prediction models across tissues in the case of tissues with low eQTL sample size (e.g. \(< 320\) samples) and high pairwise genetic correlation (e.g. \(> 0.93\)). We determined in simulations that TCSC is sensitive to eQTL sample size differences, such that a tagging tissue with larger sample size than a causal tissue can produce false positive results; the subsampling approach was designed to mitigate this issue. For the subsampling procedure, we first set aside tissues with more than 320 samples; we chose 320 based on the average GTEx tissue sample size (\(N = 271\)) and robustness of TCSC in simulations at \(N = 300\). Then, we grouped tissues with genetic correlation, e.g. marginal effect size correlation as reported by GTEx, with \(R_g > 0.93\), an arbitrary threshold that produced biologically plausible groups of related tissues, separating groups of brain tissues based on cranial compartment. We meta-analyzed gene expression prediction models for these grouped tissues in order to achieve a total sample size of 320 individuals where each tissue contributed an approximately equal number of samples, using an inverse-variance weighted meta-analysis across genes that were significantly \(cis\)-heritable in two or more constituent tissues. The prediction weights of genes that were significantly \(cis\)-heritable in a single constituent tissue were left unmodified.

To construct gene expression prediction models, we applied FUSION\(^20\) (Code Availability) to individual-level GTEx data by regressing measured gene expression on genotypes of common variants (\(MAF > 0.05\)) and covariates provided by GTEx\(^18\). FUSION uses several different regression models: single eQTL, elastic net, lasso, and BLUP and the following covariates: sex, 5 genotyping principal components, PEER factors\(^{101}\), and assay type. We
defined significantly cis-heritable genes as protein-coding genes with GCTA heritability $p < 0.01^{20}$, heritability estimate $> 0$, and adjusted-$R^2 > 0$ in cross-validation prediction.

We used gene expression prediction models of significantly cis-heritable genes to predict expression into 489 European individuals from 1000 Genomes$^{102}$. We then estimated tissue co-regulation scores using Equation (3) and Equation (4), where cis-predicted gene expression is used to estimate the cis-genetic component of gene expression.

GWAS summary statistics and TWAS association statistics

We collected GWAS summary statistics from 78 independent heritable complex diseases and traits (average $N = 302K$) with heritability $z$-score $> 6$. We estimated the heritability of all summary statistics and genetic correlation of all pairs of summary statistics. We excluded traits with heritability $z$-score $< 6$, using S-LDSC with the baseline-LD v2.2 model$^{12,23,24}$ and as done previously$^{23}$. We excluded one of each pair of traits that are both genetically correlated and have significantly overlapping samples. Specifically, for any pair of non-UK Biobank traits with an estimated sample overlap greater than the following threshold -- squared cross-trait LDSC intercept $/ (trait 1 S-LDSC intercept \times trait 2 S-LDSC intercept) > 0.1^{26}$. For any pair of UK Biobank traits with a squared genetic correlation $> 0.1$, the trait with the larger SNP heritability $z$-score was retained. For any pair of non-UK Biobank traits with a squared genetic correlation $> 0.1$, we additionally ran a brain related traits, e.g. psychiatric disorders and behavioral phenotypes, excluding multi case-control studies and case vs case studies. Then, we applied our standard filters as described above, but relaxing the threshold of squared genetic correlation to 0.25.

We used FUSION$^{20}$ (Code Availability) to compute TWAS association statistics for each pair of signed GWAS summary statistics and each significantly cis-heritable gene-tissue pair, across the two scenarios described above. We further removed genes within the MHC (chromosome 6, 29 Mb - 33 Mb) and TWAS $\chi^2 > 80$ or $\chi^2 > 0.001N$, where $N$ is the GWAS sample size, as previously used for quality control in the heritability analysis of GWAS summary statistics$^{12}$.

RTC Coloc and LDSC-SEG analysis of GWAS summary statistics and GTEx tissues

We downloaded supplementary tables for the RTC coloc method$^3$ and for LDSC-SEG$^7$. For traits in our set of 78 GWAS summary statistics that were not analyzed by the LDSC-SEG study and for traits that are inherently brain-related (as these traits require a different procedure for generating tissue-specific gene sets), we ran LDSC-SEG ourselves. To this end, we downloaded LD scores for GTEx tissues and specifically expressed gene set SNP-level annotations (https://alkesgroup.broadinstitute.org/LDSCORE/LDSC_SEG_lldscores/) and ran LDSC-SEG as previously described$^7$. For brain-related traits, we additionally ran a brain-specific analysis using LDSC-SEG, also as previously described$^7$. Briefly, specifically expressed genes were determined via a t-test of the sentinel brain tissue against all other brain tissues, rather than against all other non-brain GTEx tissues, as done in the primary analysis of the LDSC-SEG study. For traits in our set that were not analyzed by the RTC Coloc study, of which there were few, we did not apply their method, as it was too computationally intensive to apply to real trait data.
## Table 1. GTEx meta-tissues and constituent tissues analyzed

For each meta-tissue we list the constituent tissue(s) and total sample size. Daggers denote meta-tissues with more than one constituent tissue; for these meta-tissues, each constituent tissue has equal sample size up to rounding error (an exception is the transverse intestine meta-tissue, which includes 176 transverse colon samples and all 144 small intestine samples).

<table>
<thead>
<tr>
<th>Meta-tissue</th>
<th>Constituent Tissue(s)</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adipose Subcutaneous</td>
<td>Adipose Subcutaneous</td>
<td>320</td>
</tr>
<tr>
<td>Adipose Visceral Omentum</td>
<td>Adipose Visceral Omentum</td>
<td>320</td>
</tr>
<tr>
<td>Adrenal Gland</td>
<td>Adrenal Gland</td>
<td>200</td>
</tr>
<tr>
<td>Aorta Artery</td>
<td>Aorta Artery</td>
<td>320</td>
</tr>
<tr>
<td>† Brain Basal Ganglia</td>
<td>Putamen, Caudate, Nucleus Accumbens</td>
<td>320</td>
</tr>
<tr>
<td>† Brain Cereb.</td>
<td>Cerebellum, Cerebellar Hemisphere</td>
<td>320</td>
</tr>
<tr>
<td>† Brain Cortex</td>
<td>Frontal, Anterior, Cingulate</td>
<td>320</td>
</tr>
<tr>
<td>† Brain Limbic</td>
<td>Amygdala, Hippocampus, Hypothalamus</td>
<td>320</td>
</tr>
<tr>
<td>Brain Spinal Cord</td>
<td>Brain Spinal Cord</td>
<td>115</td>
</tr>
<tr>
<td>Brain Substantia Nigra</td>
<td>Brain Substantia Nigra</td>
<td>101</td>
</tr>
<tr>
<td>Breast Mammary Gland</td>
<td>Breast Mammary Gland</td>
<td>320</td>
</tr>
<tr>
<td>Coronary Artery</td>
<td>Coronary Artery</td>
<td>180</td>
</tr>
<tr>
<td>Cultured Fibroblasts</td>
<td>Cultured Fibroblasts</td>
<td>320</td>
</tr>
<tr>
<td>Esophagus Mucosa</td>
<td>Esophagus Mucosa</td>
<td>320</td>
</tr>
<tr>
<td>Esophagus Muscularis</td>
<td>Esophagus Muscularis</td>
<td>320</td>
</tr>
<tr>
<td>Heart Atrial Appendage</td>
<td>Heart Atrial Appendage</td>
<td>320</td>
</tr>
<tr>
<td>Heart Left Ventricle</td>
<td>Heart Left Ventricle</td>
<td>320</td>
</tr>
<tr>
<td>LCLs</td>
<td>LCLs</td>
<td>116</td>
</tr>
<tr>
<td>Liver</td>
<td>Liver</td>
<td>183</td>
</tr>
<tr>
<td>Lung</td>
<td>Lung</td>
<td>320</td>
</tr>
<tr>
<td>Minor Salivary Gland</td>
<td>Minor Salivary Gland</td>
<td>118</td>
</tr>
<tr>
<td>Muscle Skeletal</td>
<td>Muscle Skeletal</td>
<td>320</td>
</tr>
<tr>
<td>Ovary</td>
<td>Ovary</td>
<td>140</td>
</tr>
<tr>
<td>Pancreas</td>
<td>Pancreas</td>
<td>252</td>
</tr>
<tr>
<td>Pituitary</td>
<td>Pituitary</td>
<td>220</td>
</tr>
<tr>
<td>Prostate</td>
<td>Prostate</td>
<td>186</td>
</tr>
<tr>
<td>Skin (sun exposed)</td>
<td>Skin (sun exposed)</td>
<td>320</td>
</tr>
<tr>
<td>Skin (sun unexposed)</td>
<td>Skin (sun unexposed)</td>
<td>320</td>
</tr>
<tr>
<td>† Sigmoid Intestine</td>
<td>Sigmoid Colon, Gastroesophageal Junction</td>
<td>320</td>
</tr>
<tr>
<td>Spleen</td>
<td>Spleen</td>
<td>185</td>
</tr>
<tr>
<td>Stomach</td>
<td>Stomach</td>
<td>269</td>
</tr>
<tr>
<td>Tibial Artery</td>
<td>Tibial Artery</td>
<td>320</td>
</tr>
<tr>
<td>Tibial Nerve</td>
<td>Tibial Nerve</td>
<td>320</td>
</tr>
<tr>
<td>Testis</td>
<td>Testis</td>
<td>277</td>
</tr>
<tr>
<td>Thyroid</td>
<td>Thyroid</td>
<td>320</td>
</tr>
<tr>
<td>† Transverse Intestine</td>
<td>Transverse Colon, Small Intestine</td>
<td>320</td>
</tr>
<tr>
<td>Uterus</td>
<td>Uterus</td>
<td>108</td>
</tr>
<tr>
<td>Vagina</td>
<td>Vagina</td>
<td>122</td>
</tr>
<tr>
<td>Whole Blood</td>
<td>Whole Blood</td>
<td>320</td>
</tr>
</tbody>
</table>
Figure 1. Overview of TCSC regression. (A) Input data to TCSC includes (1) GWAS summary statistics for a disease and (2) gene expression prediction models for each tissue, which are used to produce (3) TWAS summary statistics for the disease for each tissue. (B) TCSC computes tissue co-regulation scores $L(g, t; t')$ for each gene-tissue pair $(g, t)$ with potentially causal tissues $t'$. (C) TCSC regresses TWAS chi-squares on tissue co-regulation scores to estimate tissue-specific contributions to disease.
Figure 2. Robustness and power of TCSC regression in simulations. (A) Bias in estimates of disease heritability explained by the cis-genetic component of gene expression in tissue $t'$ ($h_{ge(t')}$) for causal (light and dark purple) and non-causal (gray) tissues, across 1,000 simulations per eQTL sample size. The dashed line indicates the true value of $h_{ge(t')}$ for causal tissues. Light purple indicates that $G_{t'}$ was set to the number of true cis-heritable genes, dark purple indicates that $G_{t'}$ was set to the number of significantly cis-heritable genes. (B) Percentage of estimates of $h^2_{ge(t')}$ for non-causal tissues that were significantly positive at $p < 0.05$, across 1,000 simulations per eQTL sample size. (C) Percentage of estimates of $h^2_{ge(t')}$ for causal tissues that were significantly positive at $p < 0.05$, across 1,000 simulations per eQTL sample size. We note that (B) type I error and (C) power are not impacted by the value of $G_{t'}$. Error bars denote 95% confidence intervals. Numerical results are reported in Supplementary Table 1.
Figure 3. Robustness and power of cross-trait TCSC in simulations. (A) Bias in estimates of genetic covariance explained by the cis-genetic component of gene expression in tissue $t'$ ($\omega_{ge(t')}$) for causal (light and dark purple) and non-causal (gray) tissues, across 1,000 simulations per eQTL sample size. The dashed line indicates the true value of $\omega_{ge(t')}$ for causal tissues. Light purple indicates that $G_{t'}$ was set to the number of true cis-heritable genes, dark purple indicates that $G_{t'}$ was set to the number of significantly cis-heritable genes. (B) Percentage of estimates of $\omega_{ge(t')}$ for non-causal tissues that were significantly positive at $p < 0.05$, across 1,000 simulations per eQTL sample size. (C) Percentage of estimates of $\omega_{ge(t')}$ for causal tissues that were significantly positive at $p < 0.05$, across 1,000 simulations per eQTL sample size. We note that (B) type I error and (C) power are not impacted by the value of $G_{t'}$. Error bars denote 95% confidence intervals. Numerical results are reported in Supplementary Table 5.
**Figure 4. TCSC estimates tissue-specific contributions to disease and complex trait heritability.**

We report estimates of the proportion of disease heritability explained by the cis-genetic component of gene expression in tissue $t'$ ($\pi_{t'}$) for 27 causal tissue-trait pairs with significantly positive contributions to disease/trait heritability at 10% FDR, spanning 9 distinct tissues and 23 distinct diseases/traits. Tissues are ordered alphabetically. Daggers denote meta-tissues with more than one constituent tissue. Diseases/traits are ordered with respect to causal tissues. Asterisks denote significance at FDR < 5%. Dashed boxes denote results highlighted in the text. Numerical results are reported in **Supplementary Table 7.**

Figure 5. Comparison of disease-critical tissues identified by RTC Coloc, LDSC-SEG and TCSC.

We report -log_{10}FDR values (restricted to FDR \leq 10\%) for (A) RTC Coloc, (B) LDSC-SEG, (C) TCSC, across 5 representative traits (see text) and 10 tissues consisting of the causal tissues identified by TCSC and the most strongly co-regulated tagging tissues, ordered consecutively. In panel (B), results of brain-specific LDSC-SEG analyses are reported for BMI (a brain-related trait). Purple circles in panels (A) and (B) denote the causal tissue-trait pairs found by TCSC. Daggers denote meta-tissues involving more than one constituent tissue. Numerical results are reported in Supplementary Table 15.
Figure 6. Cross-trait TCSC estimates tissue-specific contributions to the genetic covariance of two diseases/traits. (A) We report estimates of the proportion of genetic covariance explained by the cis-genetic component of gene expression in tissue $t'$ ($\xi_{t'}$) for 30 causal tissue-trait covariance pairs with significant contributions to trait covariance at 10% FDR, spanning 21 distinct tissues and 23 distinct trait pairs. Tissues are ordered alphabetically. Daggers denote meta-tissues with more than one constituent tissue. Trait pairs are ordered by positive (+) or negative (-) genetic covariance, and further ordered with respect to causal tissues. Underlined traits are those for which TCSC identified a causal tissue in Figure 4: for eosinophil count, WBC count, and platelet count the causal tissue was whole blood, and for LDL the causal tissue was liver. Double asterisks denote trait pairs for which the differences between the tissue-specific contribution to covariance and the tissue-specific contributions to heritability were significant for both constituent traits and the tissue-specific contributions to heritability were non-significant for both constituent traits. Asterisks denote significance at FDR < 5%. Dashed boxes denote results highlighted in the text. Numerical results are reported in Supplementary Table.
References


Medic, N. *et al.* Increased body mass index is associated with specific regional alterations in brain structure. *Int J Obes (Lond)* **40**, 1177-82 (2016).


81. Peterson, C.G., Skoog, V. & Venge, P. Human eosinophil cationic proteins (ECP and EPX) and their suppressive effects on lymphocyte proliferation. Immunobiology 171, 1-13 (1986).


