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

Tiffany Amariuta[1,2], Katherine Siewert-Rocks[1,2], Alkes L. Price[1,2,3]

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[1]. 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)[2-5], gene expression[6-8], or epigenetic features[9-16]. 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)[2,17,18,5]. 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[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. 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)[19,20,17]. A challenge is that TWAS association statistics include the effects of both co-regulated tissues (see above) and co-regulated genes[17,21]. 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 χ^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[17]. TCSC determines that a tissue is causal for disease if genes and tissues with high co-regulation to that tissue have higher TWAS χ^2 statistics than genes and tissues with low co-regulation to that tissue.

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

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

where N is GWAS sample size, t' indexes causal tissues, $l(g,t;t')$ are tissue co-regulation scores (defined as $l(g,t;t') = \sum_{g'} r^2\left(W_{g,t}, W_{g',t'}\right)$, 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 χ^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

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

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), ρ 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' ($\zeta_{t'} = \omega_{ge(t')} / |\omega_g|$), where ω_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' ($h^2_{ge(t')}$) 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 nonzero heritability (or 2-sided P-values for nonzero 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

175 tissues 7-10), and average *cis*-genetic correlations between tissues (averaged across genes)
176 were set to 0.789 within the same tissue category, 0.737 between tissue categories, and 0.751
177 overall[33] (**Methods**). The GWAS sample size was set to 100,000. The 10 tissues included one
178 causal tissue explaining 5% of complex trait heritability and nine non-causal tissues. 10% of the
179 genes (125 of the 1,245 genes) had nonzero (normally distributed) gene-trait effects in the
180 causal tissue[34]. The SNP-heritability of the trait was set to 5%, all of which was explained by
181 gene expression in the causal tissue; we note that TCSC is not impacted by SNP-heritability that
182 is not explained by gene expression. Other parameter values were also explored, including
183 other values of the number of causal tissues and the number of tagging tissues. Further details
184 of the simulation framework are provided in the **Methods** section. We primarily focused our
185 comparisons between TCSC and other methods on analyses of real diseases/traits (see below),
186 because the power of TCSC relative to other methods is likely to be highly sensitive to
187 assumptions about the role of gene expression in disease architectures. However, we included
188 comparisons to RTC Coloc[2] (which also uses eQTL data) in our simulations.
189
190 We first evaluated the bias in TCSC estimates of the disease heritability explained by the
191 *cis*-genetic component of gene expression in tissue t' ($h^2_{ge(t')}$), for both causal and non-causal
192 tissues. For causal tissues, TCSC produced conservative estimates of $h^2_{ge(t')}$, particularly at
193 smaller eQTL sample sizes (**Figure 2A**, **Supplementary Table 1**). As noted above, estimates of
194 $h^2_{ge(t')}$ are impacted by the number of significantly *cis*-heritable genes in tissue t' ($G_{t'}$), which
195 may be sensitive to eQTL sample size. Estimates were more conservative when setting $G_{t'}$ to
196 the number of significantly *cis*-heritable genes, and less conservative when setting $G_{t'}$ to the
197 number of true *cis*-heritable genes. In the latter case, estimates were unbiased at the largest
198 eQTL sample size, suggesting that the conservative bias is due to noise in gene expression
199 prediction models. For non-causal tissues, TCSC produced very slightly negative estimates of
200 $h^2_{ge(t')}$ (e.g. -7.2×10^{-4}, s.e. = 2.8×10^{-4} at eQTL sample size of 100 when setting $G_{t'}$ to
201 the number of true *cis*-heritable genes). We did not include a comparison to RTC Coloc[2] in our
202 analyses of bias, because RTC Coloc does not provide quantitative estimates of $h^2_{ge(t')}$.
203
204 We next evaluated the type I error of TCSC for non-causal tissues. TCSC was
205 conservative, with type I error less than 5% at a significance threshold of p = 0.05 (**Figure 2B**,
206 **Supplementary Table 1**). The conservative type I error was most pronounced at small eQTL
207 sample sizes. We determined that the conservative type I error is primarily due to conservative
208 jackknife standard errors (1.3x-1.6x; see **Supplementary Table 2**) rather than the very slight
209 negative bias, which is generally two to three orders of magnitude smaller than the jackknife
210 standard error for a given estimate (**Supplementary Table 3**); we hypothesize that the standard
211 errors are conservative relative to the empirical standard error due to variation of causal signal
212 across the genome[35]. For comparison, we also evaluated the type I error of RTC Coloc[2]. We
213 determined that RTC Coloc was not well-calibrated at lower eQTL sample sizes, with type I error
214 of 8.4% at eQTL sample size of 100 (**Supplementary Figure 1**, **Supplementary Table 4**).
215
216 We next evaluated the power of TCSC for causal tissues. We determined that TCSC was
217 well-powered to detect causal tissues at realistic eQTL sample sizes. When analyzing 300 eQTL

218 samples, TCSC attained 72% power at a nominal significance threshold of $p < 0.05$ (**Figure 2C**)
219 and 26% power at a stringent significance threshold of $p < 0.005$ (corresponding to 10% per-
220 trait FDR across tissues in these simulations, analogous to our analyses of real diseases/traits)
221 (**Supplementary Table 1**). As expected, power increased at larger eQTL sample sizes, due to
222 lower standard errors on point estimates of $h^2_{ge(t\prime)}$ (**Figure 2A**). For comparison, we also
223 evaluated the power of RTC Coloc[2]. We determined that RTC Coloc attained substantially lower
224 power, which did not vary with eQTL sample size (**Supplementary Figure 1**, **Supplementary**
225 **Table 4**). We note that eQTL sample size is more likely to be a limiting factor when
226 disentangling co-regulated tissues using TCSC, whereas GWAS sample size is more likely to be a
227 limiting factor in other analyses such as the colocalization analyses used by RTC Coloc.
228
229 We similarly evaluated the robustness and power of TCSC when estimating tissue-
230 specific contributions to the genetic covariance between two diseases/traits. We employed the
231 same simulation framework described above and set the genetic correlation of the two
232 simulated traits to 0.5. We first evaluated the bias in TCSC estimates of the genetic covariance
233 explained by the *cis*-genetic component of gene expression in tissue t' ($\omega_{ge(t\prime)}$), for both causal
234 and non-causal tissues. For causal tissues, TCSC produced conservative estimates of $\omega_{ge(t\prime)}$,
235 particularly at small eQTL sample sizes and when setting $G_{t'}$ to the number of significantly *cis*-
236 heritable genes (rather than the number of true *cis*-heritable genes) (**Figure 3A**, **Supplementary**
237 **Table 5**), analogous to single-trait simulations. For non-causal tissues, TCSC again produced very
238 slightly negative estimates of $\omega_{ge(t\prime)}$ (**Figure 3A**). We next evaluated the type I error of cross-
239 trait TCSC for non-causal tissues. TCSC was conservative at lower eQTL sample sizes, with type I
240 error less than 5% at $P = 0.05$ (**Figure 3B**). Finally, we evaluated the power of cross-trait TCSC
241 for causal tissues. We determined that cross-trait TCSC was modestly powered at realistic eQTL
242 sample sizes. When analyzing 300 gene expression samples, TCSC attained 29% power at $p <$
243 0.05 (**Figure 3C**) and 5.6% power at $p < 0.005$ (**Supplementary Table 5**).
244
245 We performed 6 secondary analyses. First, we set the eQTL sample size of the causal
246 tissue to 100 individuals and the eQTL sample sizes of the non-causal tissues to range between
247 100-1,000 individuals. We observed inflated type I error for non-causal tissues (particularly for
248 non-causal tissues with larger eQTL sample sizes), implying that large variations in eQTL sample
249 sizes may compromise type I error (**Supplementary Figure 2**). Second, we varied the true values
250 of $h^2_{ge(t\prime)}$ (or $\omega_{ge(t\prime)}$) for causal tissues. We determined that patterns of bias, type I error, and
251 power were similar across different parameter values (**Supplementary Figures 3-4**). Third, we
252 varied the number of causal tissues, considering 1, 2, or 3 causal tissues. We determined that
253 the power of TCSC decreased with each additional causal tissue (**Supplementary Figures 5-6**).
254 Fourth, we varied the number of non-causal tissues from 0 to 9. We determined that the
255 conservative bias in estimates of $h^2_{ge(t\prime)}$ (or $\omega_{ge(t\prime)}$) for causal tissues became less pronounced
256 as the number of non-causal tissues increased (**Supplementary Figures 7-8**). Fifth, we modified
257 TCSC to not correct for bias in tissue co-regulation scores arising from differences between *cis*-
258 genetic and *cis*-predicted expression. We determined that this increased the conservative bias
259 in estimates of $h^2_{ge(t\prime)}$ for causal tissues and also resulted in anti-conservative jackknife
260 standard errors and type I error for non-causal tissues (**Supplementary Figure 9**), although the

261 impact on cross-trait simulations was limited (**Supplementary Figure 10**). Finally, we modified
262 TCSC to use "cheating" tissue co-regulation scores constructed using true eQTL effects. We
263 determined that this exacerbated the conservative bias in estimates of $h^2_{ge(t')}$ (or $\omega_{ge(t')}$) for
264 causal tissues, while also producing anti-conservative jackknife standard errors and type I error
265 for non-causal tissues (**Supplementary Figures 11-12**), perhaps because correlation patterns in
266 TWAS χ^2 statistics reflect in-sample correlations of *cis*-predicted expression. Further details of
267 these secondary analyses are provided in the **Supplementary Note**.
268
269 *Identifying tissue-specific contributions to 78 diseases and complex traits*
270
271 We applied TCSC to publicly available GWAS summary statistics for 78 diseases and
272 complex traits (average *N* = 302K; **Supplementary Table 6**) and gene expression data for 48
273 GTEx tissues[18] (**Table 1**) (see **Data Availability**). The 78 diseases/traits (which include 33
274 diseases/traits from UK Biobank[36]) were selected to have z-score > 6 for nonzero SNP-
275 heritability (as in previous studies[12,23,37]), with no pair of traits having squared genetic
276 correlation > 0.1[26] and substantial sample overlap (**Methods**). The 48 GTEx tissues were
277 aggregated into 39 *meta-tissues* (average N = 266, range: *N* = 101-320 individuals, 23 meta-
278 tissues with *N* = 320) in order to reduce variation in eQTL sample size across tissues (**Table 1**
279 and **Methods**); below, we refer to these as "tissues" for simplicity. We constructed gene
280 expression prediction models for an average of 3,993 significantly *cis*-heritable protein-coding
281 genes (as defined above) in each tissue. We primarily report the proportion of disease
282 heritability explained by the *cis*-genetic component of gene expression in tissue t'
283 ($\pi_{t'} = h^2_{ge(t')}/h^2_g$), as well as its statistical significance (using per-trait FDR).
284
285 TCSC identified 27 causal tissue-trait pairs with significantly positive contributions to
286 disease/trait heritability at 10% FDR, spanning 9 distinct tissues and 23 distinct diseases/traits
287 (**Figure 4** and **Supplementary Table 7**). Many of the significant findings recapitulated known
288 biology, including associations of whole blood with blood cell traits such as white blood cell
289 count ($\pi_{t'}$ = 0.21, s.e. = 0.064, *P* = 5.7×10^{-4}); liver with lipid traits such as LDL ($\pi_{t'}$ = 0.20, s.e.
290 = 0.050, *P* = 2.9×10^{-5}); and thyroid with hypothyroidism ($\pi_{t'}$ = 0.31, s.e. = 0.11, *P* =
291 1.9×10^{-3}).
292
293 TCSC also identified several biologically plausible findings not previously reported in the
294 genetics literature. First, aorta artery was associated with glaucoma ($\pi_{t'}$ = 0.15, s.e. = 0.051, *P* =
295 1.3×10^{-3}). TCSC also identified aorta artery as a causal tissue for diastolic blood pressure
296 (DBP) ($\pi_{t'}$ = 0.078, s.e. = 0.024, *P* = 5.1×10^{-4}), which is consistent with DBP measuring the
297 pressure exerted on the aorta when the heart is relaxed[38]. High blood pressure is a known risk
298 factor for glaucoma[39-43], explaining the role of aorta artery in genetic susceptibility to glaucoma.
299 Second, esophagus muscularis was associated with the lung trait FEV1/FVC[44] ($\pi_{t'}$ = 0.15, s.e. =
300 0.053, *P* = 1.7×10^{-3}), consistent with previous work reporting that acidification of the
301 esophagus leads to airway constriction[45] and birth defects in the esophagus have consequences
302 for pulmonary function likely stemming from shared causal mechanisms in Sonic hedgehog[46-49].
303 FEV1/FVC is computed by dividing the amount of air an individual can expel from their lungs in
304 one second (FEV1) by total lung capacity (FVC); therefore, the esophagus muscularis could

305 impact either FEV1 and/or FVC. We sought to assess tissue-specific contributions to FEV1 and
306 FVC separately. For FVC, no causal tissues were identified at 10% FDR, and the tissues with the
307 most significant P-values were coronary artery and aorta artery (nominal P = 0.02 and
308 0.07, respectively; FDR > 10%), not esophagus muscularis. For FEV1, we analyzed summary
309 statistics from ref.[50] and determined that the tissue with the most significant P-value was
310 esophagus muscularis (nominal $P = 8.3 \times 10^{-3}$; FDR > 10%), suggesting that esophagus
311 muscularis may specifically impact FEV1 (we note that FEV1/FVC, FEV1, and FVC had similar z-
312 scores for nonzero SNP heritability, implying similar power in this analysis). Third, TCSC
313 identified heart left ventricle (in addition to whole blood) as a causal tissue for platelet count
314 ($\pi_{t'}$ = 0.091, s.e. = 0.031, $P = 1.7 \times 10^{-3}$), consistent with the role of platelets in the formation
315 of blood clots in cardiovascular disease[51-54]. In cardiovascular disease, platelets are recruited to
316 damaged heart vessels after cholesterol plaques rupture, resulting in blood clots due to the
317 secretion of coagulating molecules[55]; antiplatelet drugs have been successful at reducing
318 adverse cardiovascular outcomes[56]. Moreover, the left ventricle serves as a muscle to pump
319 blood throughout the body[57], likely modulating platelet counts and other blood cell counts,
320 creating detectable changes in serum from which platelet counts are measured. Other
321 significant findings are discussed in the **Supplementary Note**, and numerical results for all
322 tissues and diseases/traits analyzed are reported in **Supplementary Table 8**.
323
324 TCSC also increased the specificity of known tissue-trait associations. For high density
325 lipoprotein (HDL), previous studies reported that deletion of a cholesterol transporter gene in
326 adipose tissue reduces HDL levels, consistent with the fact that adipose tissues are storage sites
327 of cholesterol and express genes involved in cholesterol transport and HDL lipidation[58,59]. While
328 there are three adipose tissues represented in the GTEx data that we analyzed (subcutaneous,
329 visceral, and breast tissue), TCSC specifically identified subcutaneous adipose ($\pi_{t'}$ = 0.16, s.e. =
330 0.054, $P = 1.5 \times 10^{-3}$; **Figure 4**), but not visceral adipose or breast tissue (P > 0.05;
331 **Supplementary Table 8**), as a causal tissue for HDL. Previous studies have established that
332 levels of adiponectin, a hormone released by adipose tissue to regulate insulin, are significantly
333 positively correlated with HDL[60-62] and more recently, a study has reported that adiponectin
334 levels are associated specifically with subcutaneous adipose tissue and not visceral adipose
335 tissue[63]; thus, the specific role of subcutaneous adipose tissue in HDL may be due to a causal
336 mechanism related to adiponectin. We note that TCSC did not identify liver as a causal tissue
337 for HDL (FDR > 10%), which may be due to limited power in liver due to smaller eQTL sample
338 size. For waist-hip ratio adjusted for BMI (WHRadjBMI), previous studies reported colocalization
339 of WHRadjBMI GWAS variants with *cis*-eQTLs in subcutaneous adipose, visceral adipose, liver,
340 and whole blood[64], consistent with WHRadjBMI measuring adiposity in the intraabdominal
341 space which is likely regulated by metabolically active tissues[65]. TCSC specifically identified
342 subcutaneous adipose ($\pi_{t'}$ = 0.10, s.e. = 0.037, $P = 2.4 \times 10^{-3}$; **Figure 4**), but not visceral
343 adipose, breast, liver, or whole blood (P > 0.05; **Supplementary Table 8**), as a causal tissue for
344 WHRadjBMI. The causal mechanism may involve adiponectin secreted from subcutaneous
345 adipose tissue, which is negatively correlated with WHRadjBMI[66]. We note that the P value
346 distributions across traits are similar for subcutaneous adipose (median P = 0.42) and visceral
347 adipose (median P = 0.56) and are comparable to the other 37 analyzed (median P = 0.20 –
348 0.84, **Supplementary Table 9**). For BMI, previous studies have broadly implicated the central

349 nervous system, but did not reveal more precise contributions[67,12,68,69,7,70]. TCSC specifically
350 identified brain cereb. ($\pi_{t'}$ = 0.042, s.e. = 0.015, $P = 2.6 \times 10^{-3}$), but not brain cortex or brain
351 limbic (P > 0.05; **Supplementary Table 8**), as a causal tissue for BMI. Our finding is consistent
352 with brain cerebellum accounting for the majority of the brain's neurons and serving as a key
353 area of neurogenesis in the developing brain[71,72].
354
355 We performed a secondary analysis in which we removed tissues with eQTL sample size
356 less than 320 individuals, as these tissues may often be underpowered (**Figure 2C**). Results are
357 reported in **Supplementary Figure 13** and **Supplementary Table 10.** The number of causal
358 tissue-trait pairs with significantly positive contributions to disease/trait heritability (at 10%
359 FDR) increased from 27 to 33, likely due to a decrease in multiple hypothesis testing burden
360 from removing underpowered tissues. The 33 significant tissue-trait pairs reflect a gain of 14
361 newly significant tissue-trait pairs (and a loss of 8 formerly significant tissue-trait pairs, of which
362 5 were lost because the tissue was removed), but estimates of $\pi_{t'}$ for each significant tissue-
363 trait pair were not statistically different from our primary analysis (**Supplementary Table 11**).
364 Notably, among the newly significant tissue-trait pairs, esophagus mucosa was associated with
365 eczema ($\pi_{t'}$ = 0.085, s.e. = 0.032, $P = 4.0 \times 10^{-3}$), consistent with the comorbidity of acid reflux
366 and eczema and hypotheses of a causal mechanism in which acid reflux stimulates the immune
367 cells in the mucosa, prompting an allergic response such as eczema[73]. Other newly significant
368 findings are discussed in the **Supplementary Note**, and numerical results for all tissues and
369 diseases/traits are reported in **Supplementary Table 10**.
370
371 We also performed a brain-specific analysis in which we applied TCSC to 41 brain traits
372 (average *N* = 226K, **Supplementary Table 12**) while restricting to 13 individual GTEx brain
373 tissues (**Supplementary Table 13**), analogous to previous work[7]. The 41 brain traits reflect a less
374 stringent squared genetic correlation threshold of 0.25. The 13 GTEx brain tissues were
375 analyzed without merging tissues into meta-tissues, and irrespective of eQTL sample size
376 (range: *N* = 101-189 individuals); we expected power to be limited due to the eQTL small
377 sample sizes and substantial co-regulation among individual brain tissues. TCSC identified 16
378 brain tissue-brain trait pairs at 10% FDR (**Supplementary Figure 14, Supplementary Table 14**)—
379 substantially more than the 2 brain tissue-brain trait pairs in **Figure 4**, although most results are
380 not directly comparable due to the more fine-grained individual brain tissues analyzed. For
381 ADHD, TCSC identified brain hippocampus as a causal tissue ($\pi_{t'}$ = 0.28, s.e. = 0.10, P =
382 2.5×10^{-3}), consistent with the correlation between hippocampal volume and ADHD
383 diagnosis in children[74]. A recent ADHD GWAS identified a locus implicating the *FOXP2* gene[75],
384 which has been reported to regulate dopamine secretion in mice[76]; hippocampal activation
385 results in the firing of dopamine neurons[77]. For BMI, TCSC identified brain amygdala (in addition
386 to brain cerebellum, a constituent tissue of the meta-tissue implicated in **Figure 4**) as a causal
387 tissue ($\pi_{t'}$ = 0.54, s.e. = 0.023, $P = 8.3 \times 10^{-3}$), consistent with previous work linking the
388 amygdala to obesity and dietary self-control[78], although no previous study has implicated the
389 amygdala in genetic regulation of BMI. Other significant findings are discussed in the
390 **Supplementary Note**, and numerical results for all brain tissues and brain traits analyzed are
391 reported in **Supplementary Table 14**.
392

393 *Comparisons of TCSC to other methods*
394
395 We compared TCSC to two previous methods, RTC Coloc[2] and LDSC-SEG[7], that identify
396 disease-critical tissues using gene expression data. RTC Coloc identifies disease-critical tissues
397 based on tissue specificity of eQTL-GWAS colocalizations. LDSC-SEG identifies disease-critical
398 tissues based on heritability enrichment of specifically expressed genes. We note that RTC
399 Coloc and LDSC-SEG analyze each tissue marginally, whereas TCSC jointly models contributions
400 from each tissue to identify causal tissues (analogous to the distinction in GWAS between
401 marginal association and fine-mapping[22]). Thus, we hypothesized that RTC Coloc and LDSC-SEG
402 may output multiple highly statistically significant associated tissues for a given trait, whereas
403 TCSC may output a single causal tissue with weaker statistical evidence of causality. To assess
404 whether TCSC indeed attains higher specificity, we evaluated the results of each method both
405 for causal tissues identified by TCSC and for the most strongly co-regulated tagging tissue
406 (based on Spearman ρ for estimated eQTL effect sizes, averaged across genes, from ref.[18]). Our
407 primary analyses focused on 5 representative traits, defined as the 5 traits with highest z-score
408 for nonzero SNP-heritability among the 10 diseases/traits with at least one tissue-trait
409 association for each of the three methods (**Methods**).
410
411 Results for the 5 representative traits are reported in **Figure 5** and **Supplementary Table**
412 **15**; results for all 23 diseases/traits with causal tissue-trait associations identified by TCSC
413 (**Figure 4**) are reported in **Supplementary Figure 15** and **Supplementary Table 16**, and
414 complete results for all diseases/traits and tissues included in these comparisons are reported
415 in **Supplementary Table 17**. We reached three main conclusions. First, for a given disease/trait,
416 RTC Coloc typically implicates a broad set of tissues (not just strongly co-regulated tissues)
417 (**Figure 5A**); for example, for BMI, RTC Coloc implicated 6 of 10 tissues in **Figure 5** (and 9 of 17
418 tissues in **Supplementary Figure 15**). Second, for a given disease/trait, LDSC-SEG typically
419 implicates a small set of strongly co-regulated tissues (**Figure 5B**); for BMI, LDSC-SEG implicated
420 2 of 10 tissues in **Figure 5** (and 2 of 17 tissues in **Supplementary Figure 15**), consisting of brain
421 cereb. and its most strongly co-regulated tagging tissue. Third, for a given disease/trait, TCSC
422 typically implicates one causal tissue (**Figure 5C**); for BMI, TCSC implicated only brain cereb. as a
423 causal tissue, with even the most strongly co-regulated tagging tissue reported as non-
424 significant. For diastolic blood pressure (DBP), RTC Coloc implicated 9 of 10 tissues, LDSC-SEG
425 implicated 2 of 10 tissues (consisting of aorta artery and its most strongly co-regulated tagging
426 tissue), and TCSC implicated only aorta artery; as noted above, this result is consistent with DBP
427 measuring the pressure exerted on the aorta while the heart is relaxed[38]. As expected, the
428 higher specificity of TCSC in identifying unique causal tissues involves a trade-off of reduced
429 sensitivity, with less significant (lower $-\log_{10}$P-value and lower $-\log_{10}$FDR) results for causal
430 tissues in **Figure 5C** than in **Figure 5A** and **Figure 5B**) (**Supplementary Table 16**). We observed
431 similar patterns for HDL and WHRadjBMI, for which the increased specificity of TCSC is
432 discussed above (**Supplementary Figure 15, Supplementary Table 17, Supplementary Note**).
433 We also observed similar patterns when comparing TCSC to RTC Coloc and LDSC-SEG in the
434 brain-specific analysis of **Supplementary Figure 14** (**Supplementary Figure 16**, **Supplementary**
435 **Table 18, Supplementary Note**).
436

437 *Identifying tissue-specific contributions to the genetic covariance between two diseases/traits*
438
439 We applied cross-trait TCSC to 256 pairs of disease/traits (**Supplementary Table 19**) and
440 gene expression data for 48 GTEx tissues[18] (**Table 1**) (see **Data Availability**). Of 3,003 pairs of
441 the 78 disease/traits analyzed above, the 256 pairs of diseases/traits were selected based on
442 significantly nonzero genetic correlation ($p < 0.05 / 3{,}003$; see **Methods**). The 48 GTEx tissues
443 were aggregated into 39 meta-tissues, as before (**Table 1** and **Methods**). We primarily report
444 the proportion of genetic covariance explained by the *cis*-genetic component of gene
445 expression in tissue t' ($\zeta_{t'} = \omega_{ge(t')} / |\omega_g|$), as well as its statistical significance (using per-trait
446 FDR).
447
448 TCSC identified 30 causal tissue-trait covariance pairs with significant contributions to
449 trait covariance at 10% FDR, spanning 21 distinct tissues and 23 distinct trait pairs (**Figure 6A**
450 and **Supplementary Table 20**). For 27 of the 30 causal tissue-trait covariance pairs, the causal
451 tissue was non-significant for both constituent traits in the single-trait analysis of
452 **Supplementary Table 8**. Findings that recapitulated known biology included both examples
453 involving a tissue-trait pair that was significant in the single-trait analysis (**Figure 4**) and
454 examples in which both tissue-trait pairs were non-significant in the single-trait analysis
455 (**Supplementary Table 8**). Consistent with the significant contribution of liver to LDL heritability
456 in the single-trait analysis, TCSC identified a positive contribution of liver to the genetic
457 covariance of LDL and total cholesterol ($\zeta_{t'}$ = 0.090, s.e. = 0.029, $P = 1.0 \times 10^{-3}$), and
458 consistent with the positive contributions of whole blood to eosinophil count heritability and to
459 white blood cell count heritability in the single-trait analysis, TCSC identified a positive
460 contribution of whole blood to the genetic covariance of eosinophil count and white blood cell
461 count ($\zeta_{t'}$ = 0.32, s.e. = 0.12, $P = 2.4 \times 10^{-3}$). TCSC also identified a negative contribution of
462 muscle skeletal to the genetic covariance of WHRadjBMI and total protein ($\zeta_{t'}$ = -0.27, s.e. =
463 0.084, $P = 8.0 \times 10^{-4}$), consistent with reduced muscle mass in individuals with high waist-hip-
464 ratio[79]; this suggests that cross-trait TCSC can reveal tissue-trait biology distinct from what is
465 identified in the single-trait analyses.
466
467 TCSC also identified several biologically plausible findings not previously reported in the
468 genetics literature. First, LCLs had a significantly negative contribution to the genetic covariance
469 of eosinophil count and white blood cell count ($\zeta_{t'}$ = -0.081, s.e. = 0.028, $P = 1.8 \times 10^{-3}$, in
470 contrast to the positive contribution of whole blood). This is plausible as previous studies have
471 reported the suppression of proliferation of lymphocytes (the white blood cell hematopoietic
472 lineage from which LCLs are derived) by molecules secreted from eosinophils[80-82]. The
473 contrasting results for whole blood and LCLs suggest that genetic covariance may reflect
474 distinct tissue-specific contributions. Second, brain substantia nigra had a significantly positive
475 contribution to the genetic covariance of BMI and red blood cell count (RBC count) ($\zeta_{t'}$ = 0.28,
476 s.e. = 0.084, $P = 4.6 \times 10^{-4}$), while pancreas had a significantly negative contribution ($\zeta_{t'}$ = -
477 0.25, s.e. = 0.079, $P = 8.7 \times 10^{-4}$). In the brain, energy metabolism is regulated by oxidation
478 and previous work has shown that red blood cells play a large role in these metabolic processes
479 as oxygen sensors[83]; in addition, previous studies have reported differences in the level of
480 oxidative enzymes in red blood cells between individuals with high BMI and low BMI[84,85],

481 suggesting that genes regulating oxidative processes might have pleiotropic effects on RBC
482 count and BMI. In the pancreas, pancreatic inflammation (specifically acute pancreatitis) is
483 associated with reduced levels of red blood cells, or anemia[86], while pancreatic fat is associated
484 with metabolic disease and increased BMI[87]. Once again, the contrasting results for brain
485 substantia nigra and pancreas suggest that genetic covariance may reflect distinct tissue-
486 specific contributions. Third, brain substantia nigra had a significantly negative contribution to
487 the genetic covariance of age at first birth and height ($\zeta_{t'}$ = -0.11, s.e. = 0.032, $P = 4.5 \times 10^{-4}$).
488 Previous work in *C. elegans* reported that fecundity is positively regulated by dopamine[88,89],
489 which is produced in the substantia nigra[90]. Therefore, it is plausible that reproductive
490 outcomes related to fecundity, such as age at first birth, are also regulated by dopamine via the
491 substantia nigra. Dopamine also plays a role in regulating the levels of key growth hormones
492 such as IGF-1 and IGF-BP3[91] and has been previously shown to be associated with height[92].
493 Other significant findings are discussed in the **Supplementary Note**. Numerical results for all
494 tissues and disease/trait pairs analyzed are reported in **Supplementary Table 21**.
495
496 As noted above, for 27 of the 30 causal tissue-trait covariance pairs, the causal tissue
497 was non-significant for both constituent traits. We sought to formally assess whether
498 differences in tissue-specific contributions to genetic covariance vs. constituent trait heritability
499 were statistically significant. Specifically, for each causal tissue-trait covariance pair, we
500 estimated the differences between the tissue-specific contribution to covariance ($\zeta_{t'}$) and the
501 tissue-specific contributions to heritability for each constituent trait ($\pi_{t'}$). We identified eight
502 tissue-trait covariance pairs for which these differences were statistically significant at 10% FDR
503 for both constituent traits and $\pi_{t'}$ was non-significant for both constituent traits (**Figure 6A** and
504 **Supplementary Table 22**). For BMI and RBC count, the positive contribution of brain substantia
505 nigra and the negative contribution of pancreas to genetic covariance were each significantly
506 larger than the respective contributions of those tissues to BMI and RBC count heritability,
507 which were non-significant (**Figure 6B**). Other examples are discussed in the **Supplementary**
508 **Note**. Numerical results for all tissues and trait pairs are reported in **Supplementary Table 22**.
509
510 **Discussion**
511
512 We developed a new method, tissue co-regulation score regression (TCSC), that
513 disentangles causal tissues from tagging tissues and partitions disease heritability (or genetic
514 covariance of two diseases/traits) into tissue-specific components. We applied TCSC to 78
515 diseases and complex traits and 48 GTEx tissues, identifying 27 tissue-trait pairs (and 30 tissue-
516 trait covariance pairs) with significant tissue-specific contributions. TCSC identified biologically
517 plausible novel tissue-trait pairs, including associations of aorta artery with glaucoma,
518 esophagus muscularis with FEV1/FVC, and heart left ventricle with platelet count. TCSC also
519 identified biologically plausible novel tissue-trait covariance pairs, including a negative
520 contribution of LCLs to the covariance of eosinophil count and white blood cell count (in
521 contrast to the positive contribution of whole blood) and a positive contribution of brain
522 substantia nigra and a negative contribution of pancreas to the covariance of BMI and red
523 blood cell count; in particular, our findings suggest that genetic covariance may reflect distinct
524 tissue-specific contributions.

525
526 TCSC differs from previous methods in jointly modeling contributions from each tissue
527 to disentangle causal tissues from tagging tissues (analogous to the distinction in GWAS
528 between marginal association and fine-mapping[22]). We briefly discuss several other methods
529 that use eQTL or gene expression data to identify disease-associated tissues. RTC Coloc
530 identifies disease-associated tissues based on tissue specificity of eQTL-GWAS colocalizations[2];
531 this study made a valuable contribution in emphasizing the importance of tissue co-regulation,
532 but did not model tissue-specific effects, such that RTC Coloc may implicate many tissues
533 (**Figure 5A**). LDSC-SEG identifies disease-critical tissues based on heritability enrichment of
534 specifically expressed genes[7]; this distinguishes a focal tissue from the set of all tissues
535 analyzed, but does not distinguish closely co-regulated tissues (**Figure 5B**). MaxCPP models
536 contributions to heritability enrichment of fine-mapped eQTL variants across tissues or meta-
537 tissues[4]; although this approach proved powerful when analyzing eQTL effects that were meta-
538 analyzed across all tissues, it has limited power to identify disease-critical tissues: fine-mapped
539 eQTL annotations for blood (resp. brain) were significant conditional on annotations
540 constructed using all tissues only when meta-analyzing results across a large set of blood (resp.
541 brain) traits (Fig. 4 of ref.[4]). eQTLenrich compares eQTL enrichments of disease-associated
542 variants across tissues[3]; this approach produced compelling findings for eQTL that were
543 aggregated across tissues, but tissue-specific analyses often implicated many tissues (Fig. 1d of
544 ref.[3]). MESC estimates the proportion of heritability causally mediated by gene expression in
545 assayed tissues[93]; this study made a valuable contribution in its strict definition and estimation
546 of mediated effects (see below), but did not jointly model distinct tissues and had limited
547 power to distinguish disease-critical tissues (Fig. 3 of ref.[93]). CAFEH leverages multi-trait fine-
548 mapping methods to simultaneously evaluate all tissues for colocalization with disease[5];
549 however, this locus-based approach does not produce genome-wide estimates and it remains
550 the case that many (causal or tagging) tissues may colocalize with disease under this
551 framework. Likewise, methods for identifying tissues associated to disease/trait covariance do
552 not distinguish causal tissues from tagging tissues[94,95].
553
554 We note several limitations of our work. First, joint-fit effects of gene expression on
555 disease may not reflect biological causality; if a causal tissue or cell type is not assayed[96], TCSC
556 may identify a co-regulated tissue (e.g. a tissue whose cell type composition favors a causal cell
557 type) as causal. We anticipate that this limitation will become less severe as potentially causal
558 tissues, cell types and contexts are more comprehensively assayed. Second, TCSC does not
559 achieve a strict definition or estimation of mediated effects; this is conceptually appealing and
560 can, in principle, be achieved my modeling non-mediated effects, but may result in limited
561 power to distinguish disease-critical tissues[93]. Third, TCSC has low power at small eQTL sample
562 sizes; in addition, TCSC estimates are impacted by the number of significantly *cis*-heritable
563 genes in a focal tissue, which can lead to conservative bias at small eQTL sample sizes. We
564 anticipate that these limitations will become less severe as eQTL sample sizes increase. Fourth,
565 TCSC is susceptible to large variations in eQTL sample size, which may compromise type I error;
566 therefore, there is a tradeoff between maximizing the number of tissues analyzed and limiting
567 the variation in eQTL sample size. Fifth, TCSC assumes that causal gene expression-disease
568 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[97]). 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[98-100,32]; we caution that scRNA-seq data may require new eQTL modeling approaches[98]. Despite these limitations, TCSC is a powerful and generalizable approach for modeling tissue co-regulation to estimate tissue-specific contributions to disease.

Code Availability

TCSC software: https://github.com/TiffanyAmariuta/TCSC/tree/main/analysis.
Mancuso Lab TWAS Simulator: https://github.com/mancusolab/twas_sim.
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 https://github.com/TiffanyAmariuta/TCSC/tree/main/twas_statistics, tissue co-regulation 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.

606 **Online Methods**

607

608 *TCSC regression*

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

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

617 $$W_{jgt} = \sum_m X_{jm}\beta_{gtm},$$

618 where W_{jgt} is the *cis*-genetic component of gene expression in individual *j* for gene *g* and tissue
619 *t*, X_{jm} is the standardized genotype of individual *j* for SNP *m*, and β_{gtm} is the effect of the mth
620 SNP on gene expression of gene *g* in tissue *t*.

621 TCSC assumes that true gene-disease effects are independent and additive across
622 tissues[1]. However, *cis* eQTLs are highly correlated across tissues, leading to tagging from co-
623 regulated tissues[2]. We model phenotype as a linear combination of genetic components of
624 gene expression across genes in different tissues:

625 $$Y_j = \sum_t \sum_g W_{jgt}\,\alpha_{gt} + \epsilon_j,$$

626 where Y_j is the (binary or continuous-valued) phenotype of individual *j*, α_{gt} is the gene-trait
627 effect size and ϵ_j is the error term.

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

630 $$h^2_{ge} = \sum_t \sum_g \alpha^2_{gt}\,,$$

631 analogous to the relationship between SNP effect sizes and SNP-heritability[25]. It follows that the
632 disease heritability explained by a particular tissue t' is

633 $$h^2_{ge(t')} = \sum_g \alpha^2_{gt'}.$$

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

636 $$h^2_{ge(t')} = \sum_g var(\alpha_{gt'}) = \sum_g \tau_{t'}\,G_{t'},$$

637 where $G_{t'}$ is the number of *cis* heritable genes in t'. With this variance term, we can define a
638 polygenic model that relates TWAS χ^2 statistics to co-regulation scores, which explicitly model
639 the covariance structure of the χ^2 statistics. This strategy is analogous to modeling the
640 dependence of GWAS χ^2 statistics on LD scores[25]. We model the expected value of the TWAS χ^2
641 across genes and tissues as follows:

$$E\left[\chi^2_{g,t}\right] = N \sum_{t'} \tau_{t'} l(g,t;t') + 1, \quad (1)$$

where g indexes genes, t indexes tissues, $\chi^2_{g,t}$ is a vector of TWAS χ^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

$$h^2_{ge(t')} = G_{t'} \tau_{t'}.$$

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

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

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), ρ 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

$$l(g,t;t') = \sum_{g'} r^2\left(W_{g,t}, W_{g',t'}\right), \quad (3)$$

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[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 $r^2\left(W_{g,t}, W_{g',t'}\right)$ 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 $r^2\left(W_{g,t}, W_{g',t'}\right)$ to be systematically inflated. Therefore, when $g = g'$ and $t = t'$, we set

$$r^2\left(W_{g,t}, W_{g',t'}\right) = R^2 / h^2_{GCTA}, \qquad (4)$$

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[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':

$$L(g,t) = \sum_{t'} \sum_{g'} r^2\left(W_{g,t}, W_{g',t'}\right)$$

(without applying bias correction; see above), which allows TCSC to properly account for redundant contributions of co-regulated genes to TWAS χ^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:

$$T(g,t) = \sum_{t' \ni g,t' significantly\ cis-heritable} 1,$$

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 χ^2 statistics, and is computed differently for estimates of $h^2_{ge(t')}$ than for estimates of $\omega_{ge(t')}$ (analogous to GCSC[21] and cross-trait LDSC[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:

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

where μ_χ is the mean χ^2 statistic:

$$\mu_\chi = \frac{\sum_{t'}\sum_{g'}\chi^2_{g',t'}}{\sum_{t'}G_{t'}},$$

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

$$\mu_L = \frac{\sum_{t'}\sum_{g'}L(g',t')}{\sum_{t'}G_{t'}}.$$

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

$$H_{h^2}(g,t) = \left(NL(g,t)h^2_{ge} + 1\right)^2.$$

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:

$$H_\omega(g,t) = \left(N_1L(g,t)h^2_{ge}(trait\ 1) + 1\right)\left(N_2L(g,t)h^2_{ge}(trait\ 2) + 1\right) + \left(\frac{\sqrt{N_1N_2}\,\omega_{ge}L(g,t)}{\sum_{t'}\frac{G_{t'}}{T'}}\right)^2,$$

where $h^2_{ge}(trait\ 1)$ is the crude heritability estimate for trait 1 and $h^2_{ge}(trait\ 2)$ is the crude heritability estimate for trait 2, ω_{ge} is estimated as $\frac{\sum_{t'}\sum_{g'}z^{(1)}_{g't'}z^{(2)}_{g't'}}{\sqrt{N_1N_2}}$, 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 ρN_s:

$$\rho N_s = intercept * \sqrt{N_1N_2},$$

where ρ 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

757 regression to update our estimate of ω_{ge}, such that we may update the heteroscedasticity
758 weight as follows:

759 $$H_{\omega}(g,t) = \left(N_1 L(g,t) h_{ge}^2(trait\ 1) + 1\right)\left(N_2 L(g,t) h_{ge}^2(trait\ 2) + 1\right)$$

760 $$+ \left(\frac{\sqrt{N_1 N_2}\,\omega_{ge} L(g,t)}{\sum_{t'} \frac{G_{t'}}{T'}} + \rho N_s \sqrt{N_1 N_2}\right)^2.$$

761

762 Finally, we combine the three regression weights as follows:

763 $$Weight_{\omega}(g,t) = \frac{1}{L(g,t)T(g,t)H_{\omega}(g,t)}.$$

764

765 *Simulating TCSC*

766 We employed a widely used TWAS simulation framework (Mancuso Lab TWAS
767 Simulator, see **Code Availability**) to assess the power, bias, and calibration of TCSC in the
768 presence of co-regulation across genes and tissues. We simulated a genome in which there are
769 1,245 genes, of which 125 (10%) are causal[34]. Each gene belongs to one of 249 independent
770 genomic blocks, representing approximately 1Mb intervals of chromosome 1, which is 249 Mb
771 in length. Each block contains 50 variants and 5 genes.

772 Each primary simulation consists of 10 tissues, of which at least one is causal, defined as
773 having nonzero gene-trait effect sizes. We create a covariance structure among tissues
774 mimicking empirical GTEx data. We use a previously published method to estimate the causal
775 cross-tissue correlation of eQTL effect sizes which is 0.75[33]. We observe that not all GTEx tissues
776 are equally correlated to one another. We estimate three different cross-tissue eQTL
777 correlation quantities: (1) average correlation across all pairs of tissues = 0.75, (2) average
778 correlation across similar tissues = 0.80, e.g. brain (13 in GTEx) or adipose (2 in GTEx) tissues,
779 and (3) average correlation across dissimilar tissues, e.g. pairs of brain and adipose tissues =
780 0.74. To represent these biological modules, we let simulated tissues 1-3 have higher
781 correlation of true eQTL effects to one another than to other tissues; likewise for tissues 4-6
782 and 7-10. We set covariance parameters, described below, such that the similar tissues had an
783 average eQTL correlation of 0.789 across genes, dissimilar tissues have an average eQTL
784 correlation of 0.737, and the average eQTL correlation across any pair of tissues is 0.751. We do
785 not simulate linkage disequilibrium, which is another source of co-regulation in real data.
786 Rather, we simulate co-regulation strictly using shared eQTLs across genes and tissues. We do
787 not expect this to impact the interpretation of our empirical results. We simulate each gene has
788 5 true eQTLs, based on the upper bound of empirical data from GTEx[18] and others[32], as well as
789 the value used in other TWAS simulation methods[31]. Between co-regulated tissues, each pair of
790 genes shares 3 eQTLs. The minimum allowed *cis*-heritability of a gene is 0.01 in our simulations.
791 *Cis*-heritability is approximated as the sum of squared true *cis*-eQTL effect sizes, as done
792 previously[21]; this approximation of *cis*-heritability is more accurate than if we were to model
793 linkage disequilibrium between *cis*-eQTLs. Effect sizes for the 3 shared eQTLs across tissues are
794 sampled from a multivariate normal distribution with mean 0 and a variance-covariance matrix.
795 We define the variance and covariance terms of this matrix such that (1) the proportion of
796 genes detected as significantly *cis*-heritable by GCTA at a given sample size and (2) the average

797 *cis* heritability of detected genes at a given sample size match empirical observations from GTEx
798 data at sample sizes N = 100, 200, 300 and 500. As a result, the diagonal of the variance-
799 covariance matrix, e.g. the variance term, is set to 0.015, and the off-diagonal elements are set
800 to the product of the variance term and the desired correlation for each tissue pair, described
801 above.
802 For each of 1,000 independent simulations per analysis, we simulate a GWAS (N =
803 100,000) by creating a complex trait which is the summation of the genetic components of
804 causal gene expression (in the causal tissue). We use simulated genotypes drawn from a
805 standard normal distribution. Gene-trait effect sizes are drawn from a normal distribution with
806 mean 0 and variance 1. In cross-trait TCSC analysis, effect sizes across genes between the two
807 traits are correlated with default R_g = 0.5. To simulate a GWAS trait, we first compute the
808 genetic component of each gene, which is the product of GWAS cohort genotypes and eQTL
809 effects, such that we have 125 gene-specific traits. We then add noise to each gene-specific
810 trait such that the total variance of the phenotype explained by the five eQTLs from the causal
811 tissue is equal to a specified value; the value of $h^2_{ge(t)}$ in primary simulations is 5%. Then, we
812 multiply each gene-specific trait by the causal gene-trait effect size, consistent with the additive
813 generative model of gene-level effects on trait (see above). Finally, we take the sum across all
814 gene-specific traits to make one complex trait, where the total variance of the trait explained
815 by gene effects from the causal tissue is $h^2_{ge(t)}$, e.g. 5%.
816 We simulate an eQTL cohort of various gene expression sample sizes (N = 100, 200, 300,
817 500, 1000) using simulated genotypes drawn from a standard normal distribution. We simulate
818 total gene expression in the eQTL cohort by adding a desired amount of noise to the genetic
819 component of gene expression, e.g. the product of individual genotypes and true eQTL effect
820 sizes, with variance equal to one minus the gene expression heritability, which is the sum of
821 squared eQTL effects. Next, we fit gene expression prediction models by regressing the total
822 gene expression on eQTL cohort genotypes of *cis* variants using lasso regularization, a standard
823 approach used in TWAS. We define significantly *cis*-heritable genes as genes with GCTA
824 heritability *P* value < 0.01[20] and heritability estimate > 0, and adjusted-R^2 > 0 in cross-validation
825 prediction.
826 Then we estimate co-regulation scores at each different eQTL sample size by predicting
827 gene expression into a cohort of 500 individuals, to approximate the size of the European
828 sample of 1000 Genomes (*N* = 489). Using significantly *cis*-heritable genes from each tissue at a
829 given sample size, we estimate gene and tissue co-regulation scores $l(g, t; t')$ as described
830 above, including bias correction. In simulations, *cis* genes are defined as genes within the same
831 1 Mb block.
832 Then we apply TWAS to individual-level simulated GWAS data and gene expression
833 prediction models. We predict gene expression into each of the 100,000 GWAS cohort
834 individuals across all significantly *cis*-heritable genes for each tissue. We regress each complex
835 trait on predicted gene expression to obtain TWAS z-scores. Finally, we run TCSC by regressing
836 TWAS χ^2 statistics, or products of TWAS z-scores, on bias-corrected gene and tissue co-
837 regulation scores.
838
839 *Simulating the RTC Coloc method*

840 We simulated the RTC Coloc method[2] by leveraging our existing TCSC simulation
841 framework such that both methods could be compared via application to same simulated data.
842 We precisely followed the steps of the RTC Coloc method. We first simulated individual level
843 genotypes of a GWAS (N = 100,000) using a standard normal distribution, as described. We then
844 regress the simulated complex traits on each of the GWAS genotypes in turn using a Bonferroni
845 significance threshold to identify genome-wide significant variants. We then select one null SNP
846 per GWAS variant for comparison. Next, we simulate an eQTL cohort consisting of total gene
847 expression and genotypes, using the same underlying true eQTL effect sizes as for TCSC
848 simulations. Then we perform colocalization analysis of GWAS variants with eQTLs, across 10
849 tissues at 4 different eQTL sample sizes, to obtain the regulatory trait concordance (RTC) score.
850 This is repeated for the set of null variants. Next, we perform colocalization analysis of eQTL
851 variants between pairs of tissues to obtain tissue-sharing RTC scores, and similarly repeat this
852 for null variants. GWAS-eQTL RTC scores are divided by tissue-sharing RTC scores summed
853 across variants. Tissue-specific enrichment is computed as the ratio of this quotient to the null
854 quotient. The enrichment *P* value is obtained using a Wilcox test comparing the values of the
855 quotient to the values of the null quotient.
856
857 *Gene expression prediction models and tissue co-regulation scores in GTEx data*
858 We downloaded GTEx v8 gene expression data for 49 tissues. We excluded tissues with
859 fewer than 100 samples, e.g. kidney cortex (n = 69). We retained only European samples for
860 each tissue, as labeled by GTEx via PCA of genotypes. We constructed gene expression models
861 for two scenarios: (1) subsampling to 320 individuals including meta-analyzed tissues (**Table 1**)
862 or (2) using all European samples per tissue. We recommend meta-analyzing gene expression
863 prediction models across tissues in the case of tissues with low eQTL sample size (e.g. < 320
864 samples) and high pairwise genetic correlation (e.g. > 0.93). We determined in simulations that
865 TCSC is sensitive to eQTL sample size differences, such that a tagging tissue with larger sample
866 size than a causal tissue can produce false positive results; the subsampling approach was
867 designed to mitigate this issue. For the subsampling procedure, we first set aside tissues with
868 more than 320 samples; we chose 320 based on the average GTEx tissue sample size (N = 271)
869 and robustness of TCSC in simulations at N = 300. Then, we grouped tissues with genetic
870 correlation, e.g. marginal effect size correlation as reported by GTEx, with R_g > 0.93, an arbitrary
871 threshold that produced biologically plausible groups of related tissues, separating groups of
872 brain tissues based on cranial compartment. We meta-analyzed gene expression prediction
873 models for these grouped tissues in order to achieve a total sample size of 320 individuals
874 where each tissue contributed an approximately equal number of samples, using an inverse-
875 variance weighted meta-analysis across genes that were significantly *cis*-heritable in two or
876 more constituent tissues. The prediction weights of genes that were significantly *cis*-heritable in
877 a single constituent tissue were left unmodified.
878 To construct gene expression prediction models, we applied FUSION[20] (**Code**
879 **Availability**) to individual-level GTEx data by regressing measured gene expression on
880 genotypes of common variants (MAF > 0.05) and covariates provided by GTEx[18]. FUSION uses
881 several different regression models: single eQTL, elastic net, lasso, and BLUP and the following
882 covariates: sex, 5 genotyping principal components, PEER factors[101], and assay type. We

883 defined significantly *cis*-heritable genes as protein-coding genes with GCTA heritability $p <$
884 0.01[20], heritability estimate > 0, and adjusted-$R^2 > 0$ in cross-validation prediction.
885 We used gene expression prediction models of significantly *cis*-heritable genes to
886 predict expression into 489 European individuals from 1000 Genomes[102]. We then estimated
887 tissue co-regulation scores using Equation (3) and Equation (4), where *cis*-predicted gene
888 expression is used to estimate the *cis*-genetic component of gene expression.
889
890 *GWAS summary statistics and TWAS association statistics*
891 We collected GWAS summary statistics from 78 independent heritable complex diseases
892 and traits (average N = 302K) with heritability z-score > 6. We estimated the heritability of all
893 summary statistics and genetic correlation of all pairs of summary statistics. We excluded traits
894 with heritability z-score < 6, using S-LDSC with the baseline-LD v2.2 model[12,23,24] and as done
895 previously[23]. We excluded one of each pair of traits that are *both* genetically correlated and
896 have significantly overlapping samples. Specifically, for any pair of non-UK Biobank traits with
897 an estimated sample overlap greater than the following threshold -- squared cross-trait LDSC
898 intercept / (trait 1 S-LDSC intercept * trait 2 S-LDSC intercept) > 0.1[26] -- the trait with the larger
899 SNP heritability z-score was retained. For any pair of UK Biobank traits with a squared genetic
900 correlation > 0.1, the trait with the larger SNP heritability z-score was retained[37]. In total, this
901 procedure resulted in 78 sets of GWAS summary statistics. For the brain-specific analysis, we
902 first selected brain-related diseases and complex traits, e.g. psychiatric disorders and behavioral
903 phenotypes, excluding multi case-control studies and case vs case studies. Then, we applied our
904 standard filters as described above, but relaxing the threshold of squared genetic correlation to
905 0.25.
906 We used FUSION[20] (**Code Availability**) to compute TWAS association statistics for each
907 pair of signed GWAS summary statistics and each significantly *cis*-heritable gene-tissue pair,
908 across the two scenarios described above. We further removed genes within the MHC
909 (chromosome 6, 29 Mb - 33 Mb) and TWAS $\chi^2 > 80$ or $\chi^2 > 0.001N$, where N is the GWAS sample
910 size, as previously used for quality control in the heritability analysis of GWAS summary
911 statistics[12].
912
913 *RTC Coloc and LDSC-SEG analysis of GWAS summary statistics and GTEx tissues*
914 We downloaded supplementary tables for the RTC coloc method[2] and for LDSC-SEG[7].
915 For traits in our set of 78 GWAS summary statistics that were not analyzed by the LDSC-SEG
916 study and for traits that are inherently brain-related (as these traits require a different
917 procedure for generating tissue-specific gene sets), we ran LDSC-SEG ourselves. To this end, we
918 downloaded LD scores for GTEx tissues and specifically expressed gene set SNP-level
919 annotations (https://alkesgroup.broadinstitute.org/LDSCORE/LDSC_SEG_ldscores/) and ran
920 LDSC-SEG as previously described[7]. For brain-related traits, we additionally ran a brain-specific
921 analysis using LDSC-SEG, also as previously described[7]. Briefly, specifically expressed genes were
922 determined via a t-test of the sentinel brain tissue against all other brain tissues, rather than
923 against all other non-brain GTEx tissues, as done in the primary analysis of the LDSC-SEG study.
924 For traits in our set that were not analyzed by the RTC Coloc study, of which there were few, we
925 did not apply their method, as it was too computationally intensive to apply to real trait data.

927 **Tables**
928

Meta-tissue	Constituent Tissue(s)	Sample Size
Adipose Subcutaneous	Adipose Subcutaneous	320
Adipose Visceral Omentum	Adipose Visceral Omentum	320
Adrenal Gland	Adrenal Gland	200
Aorta Artery	Aorta Artery	320
† Brain Basal Ganglia	Putamen, Caudate, Nucleus Accumbens	320
† Brain Cereb.	Cerebellum, Cerebellar Hemisphere	320
† Brain Cortex	Frontal, Anterior, Cingulate	320
† Brain Limbic	Amygdala, Hippocampus, Hypothalamus	320
Brain Spinal Cord	Brain Spinal Cord	115
Brain Substantia Nigra	Brain Substantia Nigra	101
Breast Mammary Gland	Breast Mammary Gland	320
Coronary Artery	Coronary Artery	180
Cultured Fibroblasts	Cultured Fibroblasts	320
Esophagus Mucosa	Esophagus Mucosa	320
Esophagus Muscularis	Esophagus Muscularis	320
Heart Atrial Appendage	Heart Atrial Appendage	320
Heart Left Ventricle	Heart Left Ventricle	320
LCLs	LCLs	116
Liver	Liver	183
Lung	Lung	320
Minor Salivary Gland	Minor Salivary Gland	118
Muscle Skeletal	Muscle Skeletal	320
Ovary	Ovary	140
Pancreas	Pancreas	252
Pituitary	Pituitary	220
Prostate	Prostate	186
Skin (sun exposed)	Skin (sun exposed)	320
Skin (sun unexposed)	Skin (sun unexposed)	320
† Sigmoid Intestine	Sigmoid Colon, Gastroesophageal Junction	320
Spleen	Spleen	185
Stomach	Stomach	269
Tibial Artery	Tibial Artery	320
Tibial Nerve	Tibial Nerve	320
Testis	Testis	277
Thyroid	Thyroid	320
† Transverse Intestine	Transverse Colon, Small Intestine	320
Uterus	Uterus	108
Vagina	Vagina	122
Whole Blood	Whole Blood	320

929
930 **Table 1. GTEx meta-tissues and constituent tissues analyzed.** For each meta-tissue we list the
931 constituent tissue(s) and total sample size. Daggers denote meta-tissues with more than one
932 constituent tissue; for these meta-tissues, each constituent tissue has equal sample size up to
933 rounding error (an exception is the transverse intestine meta-tissue, which includes 176
934 transverse colon samples and all 144 small intestine samples).
935

Figures

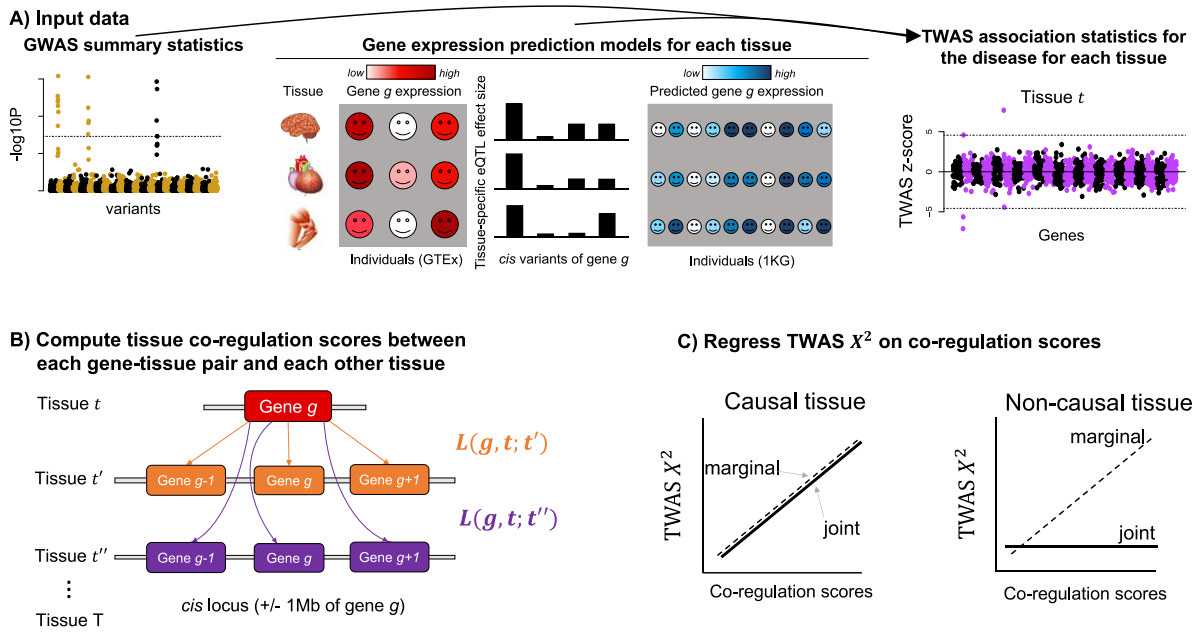

Figure 1. Overview of TCSC regression. (A) Input data to TCSC includes (1) GWAS summary statistics for a disease and (2) gene expression prediction models for each tissue, which are used to produce (3) TWAS summary statistics for the disease for each tissue. (B) TCSC computes tissue co-regulation scores $L(g, t; t')$ 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.

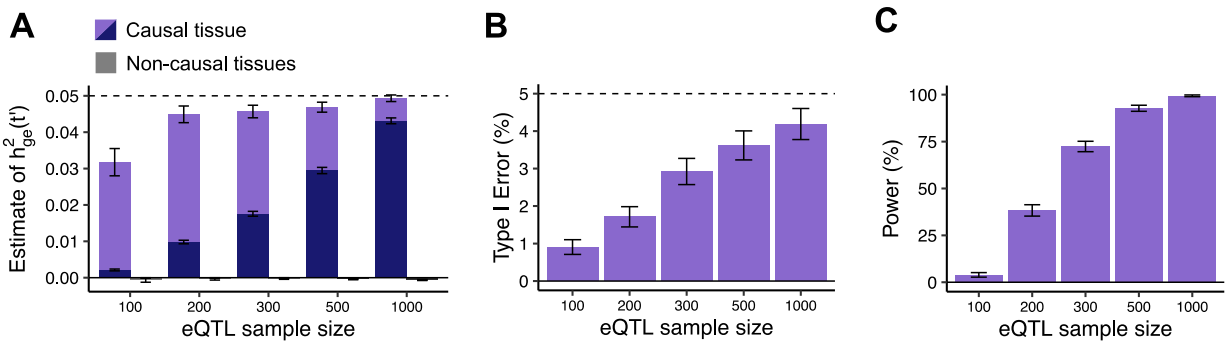

Figure 2. Robustness and power of TCSC regression in simulations. (A) Bias in estimates of disease heritability explained by the *cis*-genetic component of gene expression in tissue t' ($h^2_{ge(t')}$) for causal (light and dark purple) and non-causal (gray) tissues, across 1,000 simulations per eQTL sample size. The dashed line indicates the true value of $h^2_{ge(t')}$ 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**.

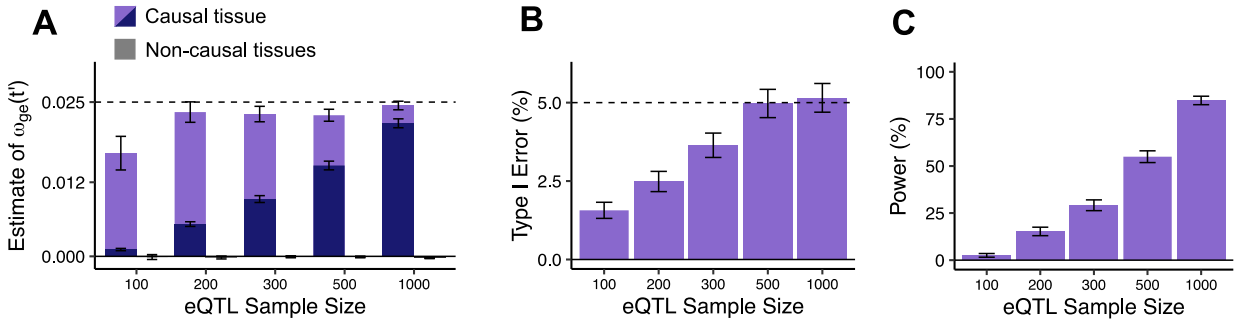

Figure 3. Robustness and power of cross-trait TCSC in simulations. (A) Bias in estimates of genetic covariance explained by the *cis*-genetic component of gene expression in tissue t' ($\omega_{ge(t')}$) for causal (light and dark purple) and non-causal (gray) tissues, across 1,000 simulations per eQTL 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**.

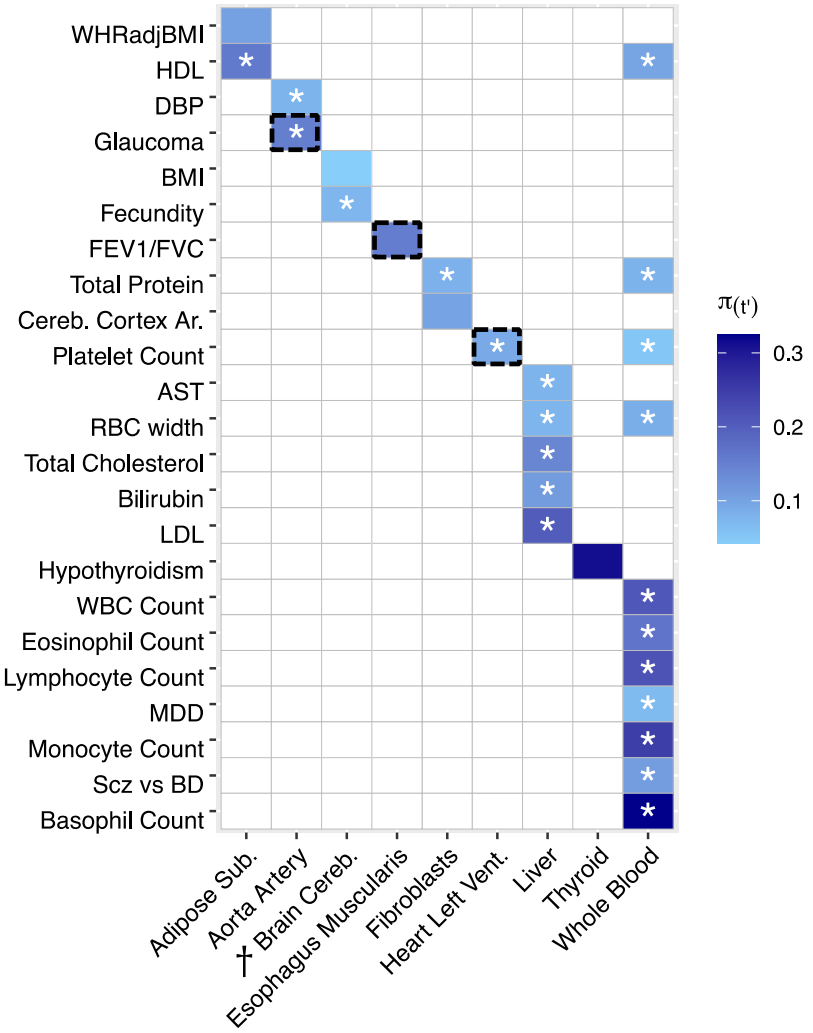

Figure 4. TCSC estimates tissue-specific contributions to disease and complex trait heritability. We report estimates of the proportion of disease heritability explained by the cis-genetic component of gene expression in tissue t' ($\pi_{t'}$) for 27 causal tissue-trait pairs with significantly positive 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**. WHRadjBMI: waist-hip-ratio adjusted for body mass index. HDL: high-density lipoprotein. DBP: diastolic blood pressure. BMI: body mass index. FEV1/FVC: forced expiratory volume in one second divided by forced vital capacity. Cereb. Cortex Ar.: cerebral cortex surface area. AST: aspartate aminotransferase. LDL: low-density lipoprotein. WBC Count: white blood cell count. MDD: major depressive disorder. Scz vs BD: schizophrenia versus bipolar disorder.

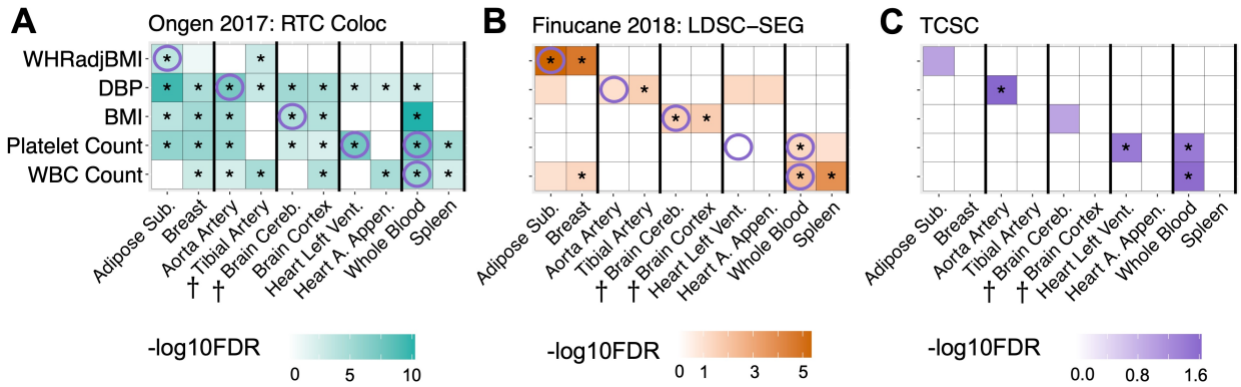

Figure 5. Comparison of disease-critical tissues identified by RTC Coloc, LDSC-SEG and TCSC. We report -$\log_{10}$FDR values (restricted to FDR $\leq$ 10%) for (A) RTC Coloc, (B) LDSC-SEG, (C) TCSC, across 5 representative traits (see text) and 10 tissues consisting of the causal tissues identified 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**.

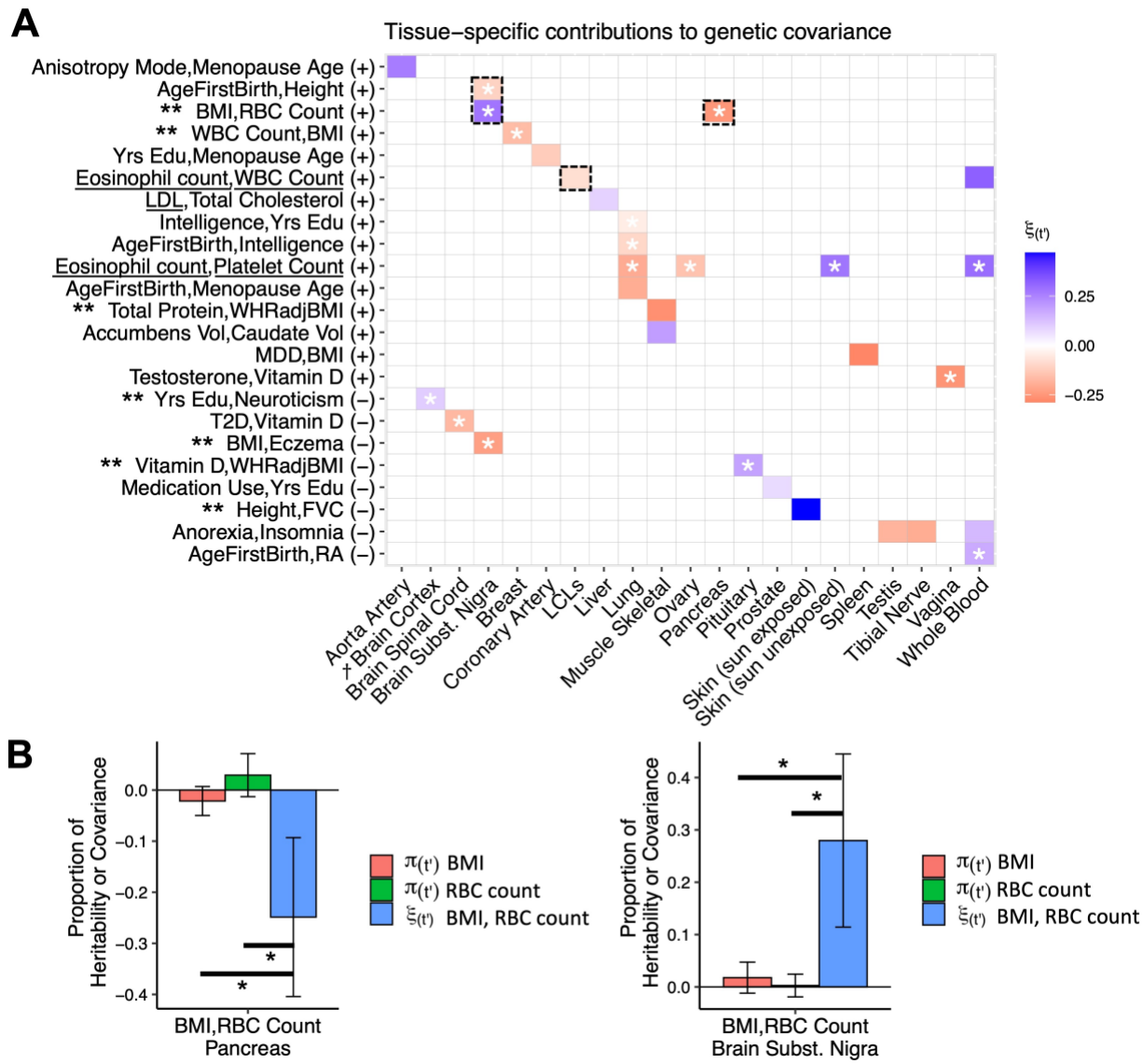

Figure 6. Cross-trait TCSC estimates tissue-specific contributions to the genetic covariance of two diseases/traits. (A) We report estimates of the proportion of genetic covariance explained by the *cis*-genetic component of gene expression in tissue t' ($\zeta_{t'}$) for 30 causal tissue-trait covariance 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**

1014 **20**. BMI: body mass index. RBC Count: red blood cell count. WBC Count: white blood cell count.
1015 LDL: low-density lipoprotein. Yrs Edu: years of education. WHRadjBMI: waist-hip-ratio adjusted
1016 for body mass index. Accumbens Vol: brain accumbens volume. Caudate Vol: brain caudate
1017 volume. MDD: major depressive disorder. T2D: type 2 diabetes. FVC: forced vital capacity. RA:
1018 rheumatoid arthritis. (B) For BMI and red blood cell count (RBC Count), we report estimates of
1019 the proportion of trait heritability for each trait and proportion of genetic covariance explained
1020 by the *cis*-genetic component of gene expression in pancreas (left panel) and brain substantia
1021 nigra (right panel). Lines with asterisks denote significant differences at 10% FDR between
1022 respective estimates, assessed by jackknifing the differences. Numerical results are reported in
1023 **Supplementary Table 22**.
1024

1025 **References**
1026

1027 1. Hekselman, I. & Yeger-Lotem, E. Mechanisms of tissue and cell-type specificity in
1028 heritable traits and diseases. *Nat Rev Genet* **21**, 137-150 (2020).
1029 2. Ongen, H. *et al.* Estimating the causal tissues for complex traits and diseases. *Nat Genet*
1030 **49**, 1676-1683 (2017).
1031 3. Gamazon, E.R. *et al.* Using an atlas of gene regulation across 44 human tissues to inform
1032 complex disease- and trait-associated variation. *Nat Genet* **50**, 956-967 (2018).
1033 4. Hormozdiari, F. *et al.* Leveraging molecular quantitative trait loci to understand the
1034 genetic architecture of diseases and complex traits. *Nat Genet* **50**, 1041-1047 (2018).
1035 5. Arvanitis, M., Tayeb, K., Strober, B.J. & Battle, A. Redefining tissue specificity of genetic
1036 regulation of gene expression in the presence of allelic heterogeneity. *Am J Hum Genet*
1037 **109**, 223-239 (2022).
1038 6. Calderon, D. *et al.* Inferring Relevant Cell Types for Complex Traits by Using Single-Cell
1039 Gene Expression. *Am J Hum Genet* **101**, 686-699 (2017).
1040 7. Finucane, H.K. *et al.* Heritability enrichment of specifically expressed genes identifies
1041 disease-relevant tissues and cell types. *Nat Genet* **50**, 621-629 (2018).
1042 8. Bryois, J. *et al.* Genetic identification of cell types underlying brain complex traits yields
1043 insights into the etiology of Parkinson's disease. *Nat Genet* **52**, 482-493 (2020).
1044 9. Maurano, M.T. *et al.* Systematic localization of common disease-associated variation in
1045 regulatory DNA. *Science* **337**, 1190-5 (2012).
1046 10. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex
1047 trait variants. *Nat Genet* **45**, 124-30 (2013).
1048 11. Pickrell, J.K. Joint analysis of functional genomic data and genome-wide association
1049 studies of 18 human traits. *Am J Hum Genet* **94**, 559-73 (2014).
1050 12. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-
1051 wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
1052 13. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human
1053 epigenomes. *Nature* **518**, 317-30 (2015).
1054 14. Backenroth, D. *et al.* FUN-LDA: A Latent Dirichlet Allocation Model for Predicting Tissue-
1055 Specific Functional Effects of Noncoding Variation: Methods and Applications. *Am J Hum*
1056 *Genet* **102**, 920-942 (2018).
1057 15. Amariuta, T. *et al.* IMPACT: Genomic Annotation of Cell-State-Specific Regulatory
1058 Elements Inferred from the Epigenome of Bound Transcription Factors. *Am J Hum Genet*
1059 **104**, 879-895 (2019).
1060 16. Boix, C.A., James, B.T., Park, Y.P., Meuleman, W. & Kellis, M. Regulatory genomic
1061 circuitry of human disease loci by integrative epigenomics. *Nature* **590**, 300-307 (2021).
1062 17. Wainberg, M. *et al.* Opportunities and challenges for transcriptome-wide association
1063 studies. *Nat Genet* **51**, 592-599 (2019).
1064 18. Consortium, G.T. The GTEx Consortium atlas of genetic regulatory effects across human
1065 tissues. *Science* **369**, 1318-1330 (2020).
1066 19. Gamazon, E.R. *et al.* A gene-based association method for mapping traits using
1067 reference transcriptome data. *Nat Genet* **47**, 1091-8 (2015).

1068 20. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association
1069 studies. *Nat Genet* **48**, 245-52 (2016).
1070 21. Siewert-Rocks, K.M., Kim, S.S., Yao, D.W., Shi, H. & Price, A.L. Leveraging gene co-
1071 regulation to identify gene sets enriched for disease heritability. *Am J Hum Genet* **109**,
1072 393-404 (2022).
1073 22. Schaid, D.J., Chen, W. & Larson, N.B. From genome-wide associations to candidate
1074 causal variants by statistical fine-mapping. *Nat Rev Genet* **19**, 491-504 (2018).
1075 23. Gazal, S. *et al.* Linkage disequilibrium-dependent architecture of human complex traits
1076 shows action of negative selection. *Nat Genet* **49**, 1421-1427 (2017).
1077 24. Gazal, S., Marquez-Luna, C., Finucane, H.K. & Price, A.L. Reconciling S-LDSC and LDAK
1078 functional enrichment estimates. *Nat Genet* **51**, 1202-1204 (2019).
1079 25. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from
1080 polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
1081 26. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits.
1082 *Nat Genet* **47**, 1236-41 (2015).
1083 27. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide
1084 complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
1085 28. Mancuso, N. *et al.* Probabilistic fine-mapping of transcriptome-wide association studies.
1086 *Nat Genet* **51**, 675-682 (2019).
1087 29. Hormozdiari, F. *et al.* Widespread Allelic Heterogeneity in Complex Traits. *Am J Hum*
1088 *Genet* **100**, 789-802 (2017).
1089 30. Abell, N.S. *et al.* Multiple causal variants underlie genetic associations in humans.
1090 *Science* **375**, 1247-1254 (2022).
1091 31. Li, Z. *et al.* METRO: Multi-ancestry transcriptome-wide association studies for powerful
1092 gene-trait association detection. *Am J Hum Genet* **109**, 783-801 (2022).
1093 32. Yazar, S. *et al.* Single-cell eQTL mapping identifies cell type-specific genetic control of
1094 autoimmune disease. *Science* **376**, eabf3041 (2022).
1095 33. Liu, X. *et al.* Functional Architectures of Local and Distal Regulation of Gene Expression
1096 in Multiple Human Tissues. *Am J Hum Genet* **100**, 605-616 (2017).
1097 34. Gazal, S. *et al.* Combining SNP-to-gene linking strategies to identify disease genes and
1098 assess disease omnigenicity. *Nat Genet* **54**, 827-836 (2022).
1099 35. Weissbrod, O. *et al.* Leveraging fine-mapping and multipopulation training data to
1100 improve cross-population polygenic risk scores. *Nat Genet* **54**, 450-458 (2022).
1101 36. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data.
1102 *Nature* **562**, 203-209 (2018).
1103 37. Gazal, S. *et al.* Functional architecture of low-frequency variants highlights strength of
1104 negative selection across coding and non-coding annotations. *Nat Genet* **50**, 1600-1607
1105 (2018).
1106 38. Homan, T.D., Bordes, S. & Cichowski, E. Physiology, Pulse Pressure. in *StatPearls*
1107 (Treasure Island (FL), 2022).
1108 39. Kass, M.A. *et al.* The Ocular Hypertension Treatment Study: a randomized trial
1109 determines that topical ocular hypotensive medication delays or prevents the onset of
1110 primary open-angle glaucoma. *Arch Ophthalmol* **120**, 701-13; discussion 829-30 (2002).

1111 40. Zhao, D., Cho, J., Kim, M.H. & Guallar, E. The association of blood pressure and primary
1112 open-angle glaucoma: a meta-analysis. *Am J Ophthalmol* **158**, 615-27 e9 (2014).
1113 41. Levine, R.M., Yang, A., Brahma, V. & Martone, J.F. Management of Blood Pressure in
1114 Patients with Glaucoma. *Curr Cardiol Rep* **19**, 109 (2017).
1115 42. De Moraes, C.G., Cioffi, G.A., Weinreb, R.N. & Liebmann, J.M. New Recommendations
1116 for the Treatment of Systemic Hypertension and their Potential Implications for
1117 Glaucoma Management. *J Glaucoma* **27**, 567-571 (2018).
1118 43. Leeman, M. & Kestelyn, P. Glaucoma and Blood Pressure. *Hypertension* **73**, 944-950
1119 (2019).
1120 44. Soler Artigas, M. *et al.* Sixteen new lung function signals identified through 1000
1121 Genomes Project reference panel imputation. *Nat Commun* **6**, 8658 (2015).
1122 45. Lang, I.M., Haworth, S.T., Medda, B.K., Forster, H. & Shaker, R. Mechanisms of airway
1123 responses to esophageal acidification in cats. *J Appl Physiol (1985)* **120**, 774-83 (2016).
1124 46. Cardoso, W.V. & Lu, J. Regulation of early lung morphogenesis: questions, facts and
1125 controversies. *Development* **133**, 1611-24 (2006).
1126 47. Shu, W. *et al.* Foxp2 and Foxp1 cooperatively regulate lung and esophagus
1127 development. *Development* **134**, 1991-2000 (2007).
1128 48. Jacobs, I.J., Ku, W.Y. & Que, J. Genetic and cellular mechanisms regulating anterior
1129 foregut and esophageal development. *Dev Biol* **369**, 54-64 (2012).
1130 49. Jia, X., Min, L., Zhu, S., Zhang, S. & Huang, X. Loss of sonic hedgehog gene leads to
1131 muscle development disorder and megaesophagus in mice. *FASEB J* **32**, 5703-5715
1132 (2018).
1133 50. Shrine, N. *et al.* New genetic signals for lung function highlight pathways and chronic
1134 obstructive pulmonary disease associations across multiple ancestries. *Nat Genet* **51**,
1135 481-493 (2019).
1136 51. Gregg, D. & Goldschmidt-Clermont, P.J. Cardiology patient page. Platelets and
1137 cardiovascular disease. *Circulation* **108**, e88-90 (2003).
1138 52. Coppinger, J.A. *et al.* Characterization of the proteins released from activated platelets
1139 leads to localization of novel platelet proteins in human atherosclerotic lesions. *Blood*
1140 **103**, 2096-104 (2004).
1141 53. Gawaz, M., Langer, H. & May, A.E. Platelets in inflammation and atherogenesis. *J Clin*
1142 *Invest* **115**, 3378-84 (2005).
1143 54. Davi, G. & Patrono, C. Platelet activation and atherothrombosis. *N Engl J Med* **357**, 2482-
1144 94 (2007).
1145 55. Badimon, L., Padro, T. & Vilahur, G. Atherosclerosis, platelets and thrombosis in acute
1146 ischaemic heart disease. *Eur Heart J Acute Cardiovasc Care* **1**, 60-74 (2012).
1147 56. Meadows, T.A. & Bhatt, D.L. Clinical aspects of platelet inhibitors and thrombus
1148 formation. *Circ Res* **100**, 1261-75 (2007).
1149 57. Berman, M.N., Tupper, C. & Bhardwaj, A. Physiology, Left Ventricular Function. in
1150 *StatPearls* (Treasure Island (FL), 2022).
1151 58. Chung, S., Sawyer, J.K., Gebre, A.K., Maeda, N. & Parks, J.S. Adipose tissue ATP binding
1152 cassette transporter A1 contributes to high-density lipoprotein biogenesis in vivo.
1153 *Circulation* **124**, 1663-72 (2011).

1154 59. McGillicuddy, F.C., Reilly, M.P. & Rader, D.J. Adipose modulation of high-density
1155 lipoprotein cholesterol: implications for obesity, high-density lipoprotein metabolism,
1156 and cardiovascular disease. *Circulation* **124**, 1602-5 (2011).
1157 60. Zoccali, C. *et al.* Adiponectin, metabolic risk factors, and cardiovascular events among
1158 patients with end-stage renal disease. *J Am Soc Nephrol* **13**, 134-141 (2002).
1159 61. Ryo, M. *et al.* Adiponectin as a biomarker of the metabolic syndrome. *Circ J* **68**, 975-81
1160 (2004).
1161 62. Toth, P.P. Adiponectin and high-density lipoprotein: a metabolic association through
1162 thick and thin. *Eur Heart J* **26**, 1579-81 (2005).
1163 63. Van Linthout, S. *et al.* Impact of HDL on adipose tissue metabolism and adiponectin
1164 expression. *Atherosclerosis* **210**, 438-44 (2010).
1165 64. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat
1166 distribution. *Nature* **518**, 187-196 (2015).
1167 65. Emdin, C.A. *et al.* Genetic Association of Waist-to-Hip Ratio With Cardiometabolic Traits,
1168 Type 2 Diabetes, and Coronary Heart Disease. *JAMA* **317**, 626-634 (2017).
1169 66. Smith, J., Al-Amri, M., Sniderman, A. & Cianflone, K. Leptin and adiponectin in relation to
1170 body fat percentage, waist to hip ratio and the apoB/apoA1 ratio in Asian Indian and
1171 Caucasian men and women. *Nutr Metab (Lond)* **3**, 18 (2006).
1172 67. Farooqi, I.S. Defining the neural basis of appetite and obesity: from genes to behaviour.
1173 *Clin Med (Lond)* **14**, 286-9 (2014).
1174 68. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity
1175 biology. *Nature* **518**, 197-206 (2015).
1176 69. Medic, N. *et al.* Increased body mass index is associated with specific regional
1177 alterations in brain structure. *Int J Obes (Lond)* **40**, 1177-82 (2016).
1178 70. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body
1179 mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet*
1180 **27**, 3641-3649 (2018).
1181 71. Wojcinski, A. *et al.* Cerebellar granule cell replenishment postinjury by adaptive
1182 reprogramming of Nestin(+) progenitors. *Nat Neurosci* **20**, 1361-1370 (2017).
1183 72. Andreotti, J.P. *et al.* Neurogenesis in the postnatal cerebellum after injury. *Int J Dev*
1184 *Neurosci* **67**, 33-36 (2018).
1185 73. Hait, E.J. & McDonald, D.R. Impact of Gastroesophageal Reflux Disease on Mucosal
1186 Immunity and Atopic Disorders. *Clin Rev Allergy Immunol* **57**, 213-225 (2019).
1187 74. Plessen, K.J. *et al.* Hippocampus and amygdala morphology in attention-
1188 deficit/hyperactivity disorder. *Arch Gen Psychiatry* **63**, 795-807 (2006).
1189 75. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention
1190 deficit/hyperactivity disorder. *Nat Genet* **51**, 63-75 (2019).
1191 76. Enard, W. *et al.* A humanized version of Foxp2 affects cortico-basal ganglia circuits in
1192 mice. *Cell* **137**, 961-71 (2009).
1193 77. Floresco, S.B., Todd, C.L. & Grace, A.A. Glutamatergic afferents from the hippocampus
1194 to the nucleus accumbens regulate activity of ventral tegmental area dopamine
1195 neurons. *J Neurosci* **21**, 4915-22 (2001).

1196 78. Kim, M.S. *et al.* Prefrontal Cortex and Amygdala Subregion Morphology Are Associated
1197 With Obesity and Dietary Self-control in Children and Adolescents. *Front Hum Neurosci*
1198 **14**, 563415 (2020).
1199 79. Elsayed, E.F. *et al.* Waist-to-hip ratio and body mass index as risk factors for
1200 cardiovascular events in CKD. *Am J Kidney Dis* **52**, 49-57 (2008).
1201 80. Peterson, C.G., Skoog, V. & Venge, P. Human eosinophil cationic proteins (ECP and EPX)
1202 and their suppressive effects on lymphocyte proliferation. *Immunobiology* **171**, 1-13
1203 (1986).
1204 81. Nakagome, K. *et al.* IL-5-induced hypereosinophilia suppresses the antigen-induced
1205 immune response via a TGF-beta-dependent mechanism. *J Immunol* **179**, 284-94 (2007).
1206 82. Onyema, O.O. *et al.* Eosinophils downregulate lung alloimmunity by decreasing TCR
1207 signal transduction. *JCI Insight* **4**(2019).
1208 83. Wei, H.S. *et al.* Erythrocytes Are Oxygen-Sensing Regulators of the Cerebral
1209 Microcirculation. *Neuron* **91**, 851-862 (2016).
1210 84. Olusi, S.O. Obesity is an independent risk factor for plasma lipid peroxidation and
1211 depletion of erythrocyte cytoprotectic enzymes in humans. *Int J Obes Relat Metab*
1212 *Disord* **26**, 1159-64 (2002).
1213 85. Ozata, M. *et al.* Increased oxidative stress and hypozincemia in male obesity. *Clin*
1214 *Biochem* **35**, 627-31 (2002).
1215 86. Druml, W., Laggner, A.N., Lenz, K., Grimm, G. & Schneeweiss, B. Pancreatitis in acute
1216 hemolysis. *Ann Hematol* **63**, 39-41 (1991).
1217 87. Sakai, N.S., Taylor, S.A. & Chouhan, M.D. Obesity, metabolic disease and the pancreas-
1218 Quantitative imaging of pancreatic fat. *Br J Radiol* **91**, 20180267 (2018).
1219 88. Schafer, W.R. & Kenyon, C.J. A calcium-channel homologue required for adaptation to
1220 dopamine and serotonin in Caenorhabditis elegans. *Nature* **375**, 73-8 (1995).
1221 89. Weinshenker, D., Garriga, G. & Thomas, J.H. Genetic and pharmacological analysis of
1222 neurotransmitters controlling egg laying in C. elegans. *J Neurosci* **15**, 6975-85 (1995).
1223 90. Triarhou, L.C. Introduction. Dopamine and Parkinson's disease. *Adv Exp Med Biol* **517**, 1-
1224 14 (2002).
1225 91. Zielonka, M. *et al.* Dopamine-Responsive Growth-Hormone Deficiency and Central
1226 Hypothyroidism in Sepiapterin Reductase Deficiency. *JIMD Rep* **24**, 109-13 (2015).
1227 92. Comings, D.E. *et al.* The dopamine D2 receptor (DRD2) as a major gene in obesity and
1228 height. *Biochem Med Metab Biol* **50**, 176-85 (1993).
1229 93. Yao, D.W., O'Connor, L.J., Price, A.L. & Gusev, A. Quantifying genetic effects on disease
1230 mediated by assayed gene expression levels. *Nat Genet* **52**, 626-633 (2020).
1231 94. Leyden, G.M. *et al.* Harnessing tissue-specific genetic variation to dissect putative causal
1232 pathways between body mass index and cardiometabolic phenotypes. *Am J Hum Genet*
1233 **109**, 240-252 (2022).
1234 95. Thom, C.S., Wilken, M.B., Chou, S.T. & Voight, B.F. Body mass index and adipose
1235 distribution have opposing genetic impacts on human blood traits. *Elife* **11**(2022).
1236 96. Umans, B.D., Battle, A. & Gilad, Y. Where Are the Disease-Associated eQTLs? *Trends*
1237 *Genet* **37**, 109-124 (2021).
1238 97. Weissbrod, O. *et al.* Functionally informed fine-mapping and polygenic localization of
1239 complex trait heritability. *Nat Genet* **52**, 1355-1363 (2020).

1240 98. Nathan, A. *et al.* Single-cell eQTL models reveal dynamic T cell state dependence of
1241 disease loci. *Nature* **606**, 120-128 (2022).
1242 99. Perez, R.K. *et al.* Single-cell RNA-seq reveals cell type-specific molecular and genetic
1243 associations to lupus. *Science* **376**, eabf1970 (2022).
1244 100. Soskic, B. *et al.* Immune disease risk variants regulate gene expression dynamics during
1245 CD4(+) T cell activation. *Nat Genet* **54**, 817-826 (2022).
1246 101. Stegle, O., Parts, L., Durbin, R. & Winn, J. A Bayesian framework to account for complex
1247 non-genetic factors in gene expression levels greatly increases power in eQTL studies.
1248 *PLoS Comput Biol* **6**, e1000770 (2010).
1249 102. Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human
1250 genomes. *Nature* **491**, 56-65 (2012).
1251