Integrating gene expression with summary association statistics to identify susceptibility genes for 30 complex traits

Nicholas Mancuso[1], Huwenbo Shi[2], Pagé Goddard[3], Gleb Kichaev[2], Alexander Gusev[4,5,6,*], Bogdan Pasaniuc[1,2,7,*]

1. Department of Pathology & Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA
2. Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA, USA
3. Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, Los Angeles, CA, USA
4. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.
5. Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA.
6. Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA.
7. Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA.

* indicates equal contribution

Abstract

Although genome-wide association studies (GWASs) have identified thousands of risk loci for many complex traits and diseases, the causal variants and genes at these loci remain largely unknown. We leverage recently introduced methods to integrate gene expression measurements from 45 expression panels with summary GWAS data to perform 30 transcriptome-wide association studies (TWASs). We identify 1,196 susceptibility genes whose expression is associated with these traits; of these, 168 reside more than 0.5Mb away from any previously reported GWAS significant variant, thus providing new risk loci. Second, we find 43 pairs of traits with significant genetic correlation at the level of predicted expression; of these, 8 are not found through genetic correlation at the SNP level. Third, we use bi-directional regression to find evidence for BMI causally influencing triglyceride levels, and triglyceride levels causally influencing LDL. Taken together, our results provide insights into the role of expression to susceptibility of complex traits and diseases.

Introduction

Although genome-wide association studies (GWASs) have identified tens of thousands of common genetic variants associated with many complex traits[1], with some notable exceptions[2; 3], the causal variants and genes at these loci remain unknown. Multiple lines of evidence show that GWAS risk variants co-localize with genetic variants that regulate expression—i.e. expression quantitative trait loci (eQTL)[4]. This suggests that a substantial proportion of GWAS risk variants influence complex trait by regulating gene expression levels of their target genes[4-7]. Analyses of genotype, phenotype, and gene expression measurements from multiple tissues in the same set of individuals can directly investigate this plausible chain of causality. However, doing so is challenging due to cost and tissue availability; therefore, GWAS and eQTL data sets remain largely independent (i.e. no overlapping subjects)[8; 9]. Recent work demonstrated that using eQTL data to predict expression into the much larger GWAS followed by association testing can identify new susceptibility genes[10-12]. This approach, referred to as transcriptome-wide association study (TWAS), provides testable hypotheses under the molecular cascade of genetic variation impacting expression which in turn impacts complex trait.

In this work we connect TWAS to a test for non-zero genetic covariance between expression and trait, and extend it to estimate the genetic correlation between expression and trait. This interpretation enables us to develop new methods that characterize the relationship between complex traits using gene effects instead of single nucleotide polymorphism (SNP) effects. In particular, we estimate the genetic correlation between pairs of traits at the level of predicted expression; this is analogous to computing genome-wide genetic correlation between traits[13], with correlations being determined over gene effects rather than SNP effects. Finally, we use a bi-directional regression approach[14] to investigate putative causal direction for pairs of traits. This approach compares models that regress over estimated effects for identified susceptibility genes and is conceptually similar to recent work[15] which uses effects of GWAS risk SNPs.

We analyze 30 GWASs spanning over 2.3 million phenotype measurements[16-29] jointly with 45 expression panels sampled from more than 35 tissues and perform 30 TWASs to gain insights into the role of expression in complex trait etiology. First, we identify 1,196 genes associated with these complex traits and diseases resulting in 1,789 distinct gene-trait pairs. Of these pairs,

168 did not overlap (0.5Mb from TSS) a genome-wide significant SNP for that respective trait, which we consider to be novel risk loci. We also find 219 cases where association signal is stronger in TWAS suggesting that allelic heterogeneity plays a role in regulating expression. Consistent with previous reports[11; 12], the vast majority of susceptibility genes were not proximal to the GWAS index SNP. Second, we estimate genetic correlation between these traits at the level of predicted expression and identify 43 pairs with significantly non-zero estimates; of these, 35 can be identified through genetic correlation analyses at the SNP level with 8 being identified only by analyzing predicted expression. These results suggest that a significant component of genetic correlation between complex traits can be explained by predicted expression. Lastly, we perform bi-directional analyses to provide evidence for putative causal effects between pairs of traits. Using this approach, we find evidence consistent with a causal model where body mass index (BMI) influences triglyceride levels, in line with earlier work[15]. We also report a novel result suggesting that triglyceride levels influence low-density lipoprotein (LDL) levels. Overall, our results shed light on shared biological mechanisms responsible for susceptibility to disease and complex trait, as well as potential downstream effects between traits.

Methods

Data Sets

We used summary association statistics from 30 large-scale (N>20,000 subjects) GWAS including various anthropometric[16; 28; 29] (BMI, femoral neck bone mineral density (BMD), forearm BMD, height, lumbar spine), hematopoietic[24; 27] (hemoglobin, HBA1C, mean cell hemoglobin (MCH), MCH concentration, mean cell volume, number of platelets, packed cell volume, red blood cell count), immune-related[18; 20] (Crohn's disease, inflammatory bowel disease, rheumatoid arthritis, and ulcerative colitis), metabolic[17; 23] (age of menarche, fasting glucose, fasting insulin, high-density lipoprotein, HOMA-B, HOMA-IR, low-density lipoprotein, triglycerides, type 2 diabetes, and total cholesterol levels), neurological[19] (schizophrenia), and social phenotypes[22] (college and educational attainment; see Supplementary Table 1). We removed SNPs that were strand-ambiguous or those with minor allele frequency $\leq$ 1% (see Supplementary Table 1).

Gene expression data from RNA-Seq data were obtained from the CommonMind Consortium[30] (CMC; brain; N=613), the Genotype-Tissue Expression project[8] (GTEx; 41 tissues; see Supplementary Table 2 for sample size per tissue), Metabolic Syndrome in Men (METSIM; adipose; N=563)[31; 32]. Expression microarray data were obtained from the Netherlands Twins Registry[33] (NTR; blood; N=1,247), and the Young Finns Study[34; 35] (YFS; blood; N=1,264).

Performing TWAS using GWAS summary statistics

We estimated SNP heritability for observed expression levels partitioned into cis-h_g^2 (1 Mb region surrounding the gene) and trans-h_g^2 (rest of genome) components. We used the AI-REML algorithm implemented in GCTA[36], which allows estimates to fall outside of the (0, 1) boundaries to maintain unbiasedness. To control for confounding, we included batch variables and the top 20 principal components estimated from genome-wide SNPs. Genes with significant cis-h_g^2 ($p < 0.05$ in a likelihood ratio test between the cis-only and joint model) in expression data were used for prediction. We performed a prediction-based transcriptome wide association study (TWAS) for each of the 30 GWAS using the summary approach described in ref[11]. In brief, we estimated the strength of association between predicted expression of gene and complex trait (z_{TWAS}), as function of the vector of GWAS association summary Z-scores at a given cis locus $\boldsymbol{z}_T$ and the LD-adjusted weights vector learned from the gene expression data $\boldsymbol{w}_{GE}$ as

$$z_{TWAS} = \frac{\boldsymbol{w}_{GE}'\boldsymbol{z}_T}{\sqrt{var(\boldsymbol{w}_{GE}'\boldsymbol{z}_T)}} = \frac{\boldsymbol{w}_{GE}'\boldsymbol{z}_T}{\sqrt{\boldsymbol{w}_{GE}'\boldsymbol{V}\boldsymbol{w}_{GE}}}.$$

where $\boldsymbol{V}$ is a covariance matrix across SNPs at the locus (i.e. LD). We estimated $\boldsymbol{w}_{GE}$ using GBLUP[37] from eQTL data and computed z_{TWAS} using GWAS summary data for all 30 traits and the ~36k gene expression measurements across all studies. We removed all loci in the human leukocyte antigen (HLA) region due to complex LD patterns. We conservatively account for multiple tests using trait-specific Bonferroni correction factors (see Supplementary Table 2).

Estimating the proportion of trait variance explained by predicted expression

We use an LD-Score[38; 39] approach[11] to quantify the heritability for a complex trait explained by predicted expression (denoted here as h_{GE}^2). The expected χ^2 statistic under a polygenic trait is $E[\chi^2] = 1 + \frac{N_T \ell}{M} h_{GE}^2 + N_T a$ where N_T is the number of individuals in the GWAS, M is the

number of genes, ℓ is the LD-score, and a is the effect of population structure. We estimate ℓ for each gene by predicting expression for 503 European samples in 1000Genomes[40] using the BLUP weights (see above) and then computing sample correlation. For each complex trait we perform LD-Score regression using z^2_{TWAS} (which is asymptotically equivalent to χ^2) to infer h^2_{GE}. We estimate heritability for each expression study separately, to account for varying sample sizes and repeated gene measurements.

Estimating genetic correlation of expression and complex trait from summary data

Let expression and trait be modeled as a linear function of the genotypes in a ~1Mb local region flanking the gene: $\boldsymbol{y}_{GE} = \boldsymbol{X}\boldsymbol{\beta}_{GE} + \boldsymbol{\epsilon}_{GE}$ and $\boldsymbol{y}_T = \boldsymbol{X}\boldsymbol{\beta}_T + \boldsymbol{\epsilon}_T$ where $\boldsymbol{X}$ is the standardized genotype matrix, $\boldsymbol{\beta}_{GE}$ ($\boldsymbol{\beta}_T$) are the standardized effects, and $\boldsymbol{\epsilon}_{GE}$ ($\boldsymbol{\epsilon}_T$) is environmental noise for expression (trait). The local covariance between expression and complex trait is

$$cov(\boldsymbol{y}_{GE}, \boldsymbol{y}_T) = cov(\boldsymbol{X}\boldsymbol{\beta}_{GE} + \boldsymbol{\epsilon}_{GE}, \boldsymbol{X}\boldsymbol{\beta}_T + \boldsymbol{\epsilon}_T) = \boldsymbol{\beta}'_{GE} cov(\boldsymbol{X}, \boldsymbol{X})\boldsymbol{\beta}_T + cov(\boldsymbol{\epsilon}_{GE}, \boldsymbol{\epsilon}_T)$$
$$= \boldsymbol{\beta}'_{GE}\boldsymbol{V}\boldsymbol{\beta}_T + cov(\boldsymbol{\epsilon}_{GE}, \boldsymbol{\epsilon}_T),$$

where V is the LD matrix. If no individuals are shared between studies then $cov(\boldsymbol{\epsilon}_{GE}, \boldsymbol{\epsilon}_T) = 0$, (as in eQTL and GWAS studies). The local genetic correlation can be computed as

$$\rho_{g,local} = \frac{\boldsymbol{\beta}'_{GE}\boldsymbol{V}\boldsymbol{\beta}_T}{\sqrt{h^2_{g,local}(GE)}\sqrt{h^2_{g,local}(T)}}$$

where $h^2_{g,local}(GE)$ is the local SNP-heritability[41] for expression (trait) estimated at the locus captured by $\boldsymbol{X}$; however, this requires knowing the true effect sizes. Previous work[41] describes a method to obtain unbiased estimates for $\boldsymbol{\beta}_i$ using genome-wide association summary statistics (i.e. Z-scores) and reference LD. Given association statistics $\boldsymbol{z}_T$, an LD-adjusted effect size estimate is computed as $\boldsymbol{\beta}_T = \frac{1}{\sqrt{N_T}}\boldsymbol{V}^{-1}\boldsymbol{z}_T$. Hence, an estimate of the local genetic covariance[42] is given by

$$\boldsymbol{\beta}'_{GE}\boldsymbol{V}\boldsymbol{\beta}_T = \frac{1}{\sqrt{N_{GE}}\sqrt{N_T}}(\boldsymbol{z}'_{GE}\boldsymbol{V}^{-1})\boldsymbol{V}(\boldsymbol{V}^{-1}\boldsymbol{z}_T) = \boldsymbol{b}'_{GE}\boldsymbol{V}^{-1}\boldsymbol{b}_T$$

where $\boldsymbol{b}_{GE}$ ($\boldsymbol{b}_T$) are the marginal (i.e. LD-unadjusted) effect sizes[41; 43]. It follows that

$$\frac{1}{\sqrt{N_T}} z_{TWAS} = \frac{1}{\sqrt{N_T}} \frac{\boldsymbol{\beta}'_{GE} \boldsymbol{z}_T}{\sqrt{var(\boldsymbol{\beta}'_{GE} \boldsymbol{z}_T)}} = \frac{\boldsymbol{b}'_{GE} \boldsymbol{V}^{-1} \boldsymbol{b}_T}{\sqrt{h^2_{g,local}(GE)}} = \rho_{g,local} \sqrt{h^2_{g,local}(T)}.$$

We standardize this estimate to obtain our final local genetic correlation estimate as

$$\rho_{g,local} = \frac{z_{TWAS}}{\sqrt{N_T} \sqrt{h^2_{g,local}(T)}}.$$

In practice we use the variance explained by the local index (i.e. smallest p-value) SNP as proxy for $h^2_{g,local}(T)$.

Local components of genetic correlation characterize the shared SNP effect between complex trait and expression; however, we can interpret $\rho_{g,local}$ as the standardized effect of predicted expression on trait. Using this definition, we estimate the genetic correlation between two complex traits as the Pearson correlation across the vector of $\boldsymbol{\rho}_{g,local}$ across all genes; we term this estimate as ρ_{GE}. We test for significance assuming that $\rho_{GE} \sqrt{\frac{M-2}{1-\rho^2_{GE}}} \sim T(n-2)$ where M is the number of genes. This procedure is unbiased in principle provided that effects of genes within single trait are not correlated. This assumption may be violated; hence, we computed trait correlation using one gene per 1Mb locus. To determine if estimates of ρ_{GE} were sensitive to changes in scale, we recomputed ρ_{GE} using non-standardized estimates of genetic covariance. We found our estimates to be highly correlated ($r = 0.94$; $p < 2.2 \times 10^{-16}$), indicating little importance in using correlation versus covariance. We report results using standardized effects for consistency across figures and tables.

Estimating putative casual relationships between pairs of traits

To glean insight into the underlying causal relationship between pairs of traits, we perform a bi-directional regression[14] and estimate two different values of ρ_{GE} by varying gene sets. Before describing the approach, we first review several causal models that explain non-zero ρ_{GE} between two traits (see Figure 1). Models A and B depict causal relationships in which the effects of a gene set are mediated by one trait on the other. We can formally state model A (without loss of generality for B). Let T_1 be defined as $\boldsymbol{y}_{T_1} = \boldsymbol{G}_{T_1} \boldsymbol{\beta}_{T_1} + \boldsymbol{\epsilon}_{T_1}$ where $\boldsymbol{G}_{T_1}$ denotes the matrix of predicted expression at the causal genes, $\boldsymbol{\beta}_{T_1}$ are the effect sizes, and $\boldsymbol{\epsilon}_{T_1}$ is environmental noise. We define T_2 as,

$$\boldsymbol{y}_{T_2} = \boldsymbol{y}_{T_1}\boldsymbol{\gamma}_{T_1} + \boldsymbol{G}_{T_2}\boldsymbol{\beta}_{T_2} + \boldsymbol{\epsilon}_{T_2} = \boldsymbol{G}_{T_1}\boldsymbol{\beta}_{T_1}\boldsymbol{\gamma}_{T_1} + \boldsymbol{G}_{T_2}\boldsymbol{\beta}_{T_2} + \boldsymbol{\epsilon}_{T_2'}$$

where $\boldsymbol{\gamma}_{T_1}$ is the causal effect of T_1 on T_2, $\boldsymbol{G}_{T_2}\boldsymbol{\beta}_{T_2}$ are the remaining causal genes and their effects for T_2, and $\epsilon_{T_2'}$ is the combined environment component. Under model A, the causal gene set for T_1 will have a non-zero effect on T_2 (i.e. $\boldsymbol{\gamma}_{T_1} \neq 0$); however, if T_1 does not cause T_2, this effect will be zero since unrelated genes have no downstream effect. Bi-directional regression provides a test to distinguish between models A and B by regressing estimated effect sizes for gene sets under model A (i.e. $\boldsymbol{\beta}_{T_1} \sim \boldsymbol{\beta}_{T_1}\boldsymbol{\gamma}_{T_1}$) and comparing to estimates under model B (i.e. $\boldsymbol{\beta}_{T_2} \sim \boldsymbol{\beta}_{T_2}\boldsymbol{\gamma}_{T_2}$). Since the causal gene sets for each trait are unknown, we use their identified susceptibility genes as proxy. We estimate ρ_{GE} conditional on the gene set for trait i and denote its value as $\rho_{j\,|\,i}$. This procedure is repeated by ascertaining the gene set for trait j to obtain $\rho_{i\,|\,j}$. We perform a Welch's t-test[44] to determine if estimates of $\rho_{i\,|\,j}$ and $\rho_{j\,|\,i}$ are significantly different, thus providing evidence consistent with a causal direction. This approach is conceptually similar to bi-directional regression analyses of estimated SNP effects on two complex traits[15; 45]. We stress that while a bi-directional approach is capable of rejecting model A in favor of model B (or vice-versa), it cannot rule out model C, in which a shared pathway (or set of pathways) drive both traits independently.

Results

TWAS identifies 1,196 susceptibility genes for 30 complex traits and diseases

We integrated the 30 GWAS summary data with gene expression to identify 1,196 susceptibility genes (i.e. gene with at least one significant trait association) comprising 5,490 total associations (after Bonferroni correction; see Methods). Of these associations, we observed 1,789 distinct gene-trait pairs with 783 found in anthropometric traits, 423 in metabolic traits, 215 in immune-related traits, 213 in hematopoietic traits, 137 in neurological traits (i.e. schizophrenia), and 18 in social traits (see Table 1; see Supplementary Tables 3-4). For example, the 137 susceptibility genes found for schizophrenia included *SNX19* (cerebellum; $p=2.2 \times 10^{-8}$) and *NMRAL1* (muscle; $p=9.7 \times 10^{-7}$); this is consistent with a previously reported study[12] that used different methods and expression data (see Supplementary Table 5). We did not find susceptibility genes for forearm

bone mineral density (BMD), HOMA-B, and mean cell hemoglobin concentration, which is consistent with low GWAS signal for these traits (see Table 1). Indeed, the number of GWAS risk loci strongly correlated with the number of identified susceptibility genes ($r=0.99$; $p < 2.2 \times 10^{-16}$) which reflects the underlying polygenicity of these traits. We explored putative molecular function and pathways enriched with identified susceptibility genes using the PANTHER database[46], but were underpowered to detect molecular function for most individual traits (see Supplementary Note).

Next, we quantified the overlap of susceptibility genes and GWAS signals. Of the 1,789 identified gene-trait pairs, 168 (9%) were not proximal (more than 0.5Mb from TSS) to any genome-wide significant SNP for that respective trait thus yielding new risk loci. Conversely, of the 1,526 GWAS risk loci, 1,405 (92%) overlapped with at least one eGene (i.e. gene with heritable expression levels in at least one of the considered expression panels) and 551 (36%) overlapping at least one susceptibility gene (see Table 1). Focusing on the 1,621 associations that overlapped a genome-wide significant SNP, we observed 1,488 (83%) genes that were not nearest, suggesting that the traditional heuristic of prioritizing genes closest to GWAS SNPs is typically not supported by evidence from eQTL data (see Supplementary Figure 1). While GWAS SNPs provide the majority of the power in this approach, the flexibility of TWAS to leverage allelic heterogeneity provides a significant gain[11]. We found 219 instances across 19 traits where association signal was stronger in TWAS compared to GWAS, with an average 1.2× increase in χ^2 statistics. For example, predicted expression in *CCDC88B* (a gene involved in T-cell maturation and inflammation[47]) exhibited strong association with Crohn's disease ($p_{TWAS}=6.32 \times 10^{-8}$) whereas the index SNP (i.e. top overlapping GWAS SNP) at site rs11231774 was only suggestive ($p_{GWAS}=2.47 \times 10^{-6}$). This effect was most dramatic for height, with 108 susceptibility genes having stronger signal than GWAS index SNPs. We observed a 2.6× increase in χ^2 statistics for predicted expression in *CRELD1* ($p_{TWAS}=1.55 \times 10^{-10}$) compared to the index SNP rs1473183 ($p_{GWAS}=6.33 \times 10^{-5}$).

Recent work[48] applied a similar approach[12] using summary eQTL from blood and GWAS data to identify 71 genes for 28 complex traits[48]. Of the investigated traits, 12 overlapped our study. Surprisingly, despite using independent methods and expression data we were able to validate 40 out of 51 associations (see Supplementary Table 6). Overall, we identified 564 genes for these

traits in contrast to 63 genes reported in that study. This increase in power can be attributed to two reasons. First, we integrate multiple expression panels sampled from many tissues, which assays many more genes. Second, we use a method that jointly tests the entire locus, rather than index SNPs. We have shown that many identified susceptibility genes contain signals of allelic heterogeneity; therefore, using individual SNPs will decrease power.

Genes associated to multiple traits

We investigated the degree of pleiotropic susceptibility genes (i.e. gene associated with more than one trait) in our data and found 380 (32%) identified genes associated with multiple traits (see Supplementary Figure 2). For example, the gene *IKZF3* displayed strong associations in Crohn's disease (blood; $p=1.6 \times 10^{-9}$), HDL levels (blood; $p=6.6 \times 10^{-15}$), IBD (blood; $p=7.9 \times 10^{-16}$), rheumatoid arthritis (blood; $p=6.0 \times 10^{-8}$), and ulcerative colitis (blood; $p=9.2 \times 10^{-10}$). Indeed, *IKZF3* has been shown to influence lymphocyte development and differentiation[49; 50]. These traits are known to have a strong autoimmune component[51]; hence, association with predicted *IKZF3* expression levels is consistent with a model where cis-regulated variation in *IKZF3* product levels contributes to risk. Similarly, we observed three susceptibility genes shared between education years and height (see Figure 2): *ABCB9* (heart; $p_{height}=1.38 \times 10^{-15}$, $p_{ey}=1.28 \times 10^{-6}$), *BTN2A3P* (adipose; $p_{height}=3.82 \times 10^{-12}$, $p_{ey}=1.90 \times 10^{-7}$), and *MPHOSPH9* (thyroid; $p_{height}=5.84 \times 10^{-18}$, $p_{ey}=1.30 \times 10^{-6}$). This is consistent with a recent study[13] that reported a non-zero genetic correlation between height and education years ($\rho_g = 0.13$, $p=3.82 \times 10^{-6}$).

Effect of cis expression on trait is consistent across tissues

Having established the importance of individual predicted gene expression levels for these traits, we next estimated the amount of trait variance explained by predicted expression using all examined genes, including those not significantly associated, using an LD-Score regression approach (see Methods). We found 108 tissue-trait pairs across 17 traits and 33 tissues where the cumulative effect of all measured genes on trait was significantly greater ($p < 0.05 / 45$) than the significant-only set. For example, in height we estimated $h^2_{GE} = 0.068$ (Jack-knife SE=0.02; $p=5.6 \times 10^{-4}$; see Supplementary Table 7) using all 3,733 measured genes in YFS and $h^2_{GE} = 0.015$ (Jack-knife SE=6.9×10^{-3}; p=0.026) using the 169 YFS susceptibility genes ($p_{ALL>SIG}=5.6 \times 10^{-3}$). This suggests that there exist additional susceptibility genes for height, which we are underpowered to detect. However, for most trait-tissue pairs we did not observe a significant

difference at our given sample sizes. Indeed, we measured a significant association between expression study sample size and number of eGenes (r=0.2; SE=0.05; p=6.4×10^{-8}), which indicates that smaller studies lack power to find eGenes, thus underestimating the total h^2_{GE}.

We next asked whether any tissues are burdened with increased levels of risk for a given trait. To test this hypothesis, we examined the difference between estimated trait variance explained per gene with the average. Our results did not suggest tissue-specific enrichment at current sample sizes (see Supplementary Table 8). Given no observable difference in tissue-specific risk, we expect local estimates of genetic correlation to be highly similar across tissues. When estimating $\rho_{g,local}$, we observed consistent effect size estimates in both sign and magnitude estimates across tissues (mean tissue-tissue r=0.82; see Figure 3). These results are compatible with earlier work that found cis effects on expression is largely consistent across tissues[52]. To obtain a meta estimate of local genetic correlation for gene-trait pairs with measurements in multiple tissues, we use the mean genetic correlation across all expression panels in all following analyses.

Genetic correlation between traits using predicted expression

To evaluate the shared contribution of predicted expression on pairs of traits, we computed expression correlation (ρ_{GE}; see Methods) using nominally significant ($p_{TWAS} < 0.05$) genes. This approach is similar to estimating genetic correlation (ρ_g) between two complex traits[13]; however, it differs in that correlation is computed through predicted components of gene expression rather than SNP effects. For 435 distinct pairs, we discovered 43 significant expression correlations, 22 of which had previously reported non-zero genetic correlations[13] (see Figure 4; see Supplementary Table 9). For example, age of menarche and BMI had an estimated $\rho_{GE} = -0.32$ (95% CI [-0.32, -0.21]; p=7.97×10^{-8}). This negative correlation is consistent with estimates published in epidemiological studies[53] in addition to studies probing genetic correlation across complex traits[13]. Using estimates of ρ_{GE}, we clustered traits and observed groups forming naturally in the trait-trait matrix (see Figure 4). Interestingly, BMI clustered with insulin-related traits (HOMA-B, HOMA-IR, and fasting insulin). Our estimates were highly consistent with LD-Score regression results (see Figure 4; Supplementary Table 9). Out of 435 pairs of traits, 35 demonstrated significance for ρ_{GE} and ρ_g, whereas 8 and 27 were exclusive for ρ_{GE} and ρ_g, respectively. Given the high degree of concordance between estimates of ρ_{GE} and ρ_g, we tested if any were significantly different and found four insulin-related pairs of traits and three blood-

related pairs with more extreme values for ρ_{GE} (see Supplementary Table 9). Differences for these pairs of traits can be partially explained by overconfident standard errors in ρ_{GE} (see Supplementary Table 10). Overall, we found ρ_{GE} to explain the majority of variation in ρ_g (r^2 = 0.72).

Bi-directional regression suggests putative causal relationships

Given pairs of traits with significant estimates of ρ_{GE}, we aimed to distinguish among possible causal explanations by performing bi-directional regression analyses (see Methods). To empirically validate our approach, we regressed HDL, LDL, and triglycerides with total cholesterol. Total cholesterol (TC) is the direct consequence of summing over triglyceride, HDL, and LDL levels, thus we expect to observe increased signal for $\rho_{TC \mid Lipid}$ compared to $\rho_{Lipid \mid TC}$. Of these three, we found evidence for triglycerides influencing total cholesterol ($p=2.34 \times 10^{-3}$). We observed consistent, but not significant, evidence for the effect of LDL on TC ($p=6.79 \times 10^{-2}$) and HDL on TC ($p=5.56\times 10^{-1}$; see Figure 5). These results suggest that point-estimates from the bi-directional approach favor the correct model, but may not have adequate power required for significance.

We tested the 43 pairs of traits identified above (see Table 3) while ascertaining on susceptibility genes and observed asymmetric effects at $p < 0.05$ for BMI-triglycerides and LDL-triglycerides (see Figure 6). For example, in the bi-directional analysis on BMI and triglycerides, we observed a significant effect for $\rho_{TG \mid BMI} = 0.62$ (95% CI [0.27, 0.83]; $p=2.06 \times 10^{-3}$). By contrast, the reverse analysis estimate overlapped with zero at $\rho_{BMI \mid TG} = -0.04$ (95% CI [-0.49, 0.42]; p=0.86). Individual estimates for $\rho_{TG \mid BMI}$ and $\rho_{BMI \mid TG}$ were significantly different (p=0.01; Welch's t-test), which is consistent with a model where BMI directly influences triglyceride levels. In practice, we used susceptibility genes found through TWAS ($p \sim 1 \times 10^{-6}$), but this may be too strict an inclusion threshold for genes which we lack power to detect. We report analyses using weaker thresholds and observe similar results (see Supplementary Tables 11, 12). Our result reinforces previous estimates of putative causal effect where BMI influences triglyceride levels[15; 54].

Discussion

In this work we used GWAS summary statistics from 30 complex traits and diseases jointly with expression data sampled across 45 expression panels to identify susceptibility genes for complex traits. We identified 1,196 susceptibility genes for 27 of the 30 complex traits. We use estimates of local genetic correlation between gene expression and trait to compute ρ_{GE}, which quantifies the shared effect of predicted expression levels between two complex traits. Using this definition, we found 43 pairs of traits to be significantly correlated, of which 8 were novel. To provide evidence of possible causal direction, we adapted a recently proposed causality test[15] to operate at the gene level. Our results support triglycerides (TG) influencing LDL, and BMI influencing triglycerides. As more GWAS and eQTL summary results become publicly available, we expect additional studies to integrate cross-trait information to make inferences about mechanistic bases for complex trait.

Assuming gene expression mediates the effect of genetics on complex trait, testing for association between the predicted component of expression and trait is equivalent with a two-sample Mendelian randomization test for a causal effect of expression on trait[55; 56]. This test for causality is valid provided SNPs do not exhibit pleiotropic effects; therefore, the TWAS associations are not proof of causal relationships between expression and complex trait. This set of assumptions extends to our bi-directional approach to infer causal direction. A bi-directional regression is capable of distinguishing between direction of effect, but cannot rule out pleiotropy.

We conclude with several caveats. First, we note that using estimates of genetic correlation to find susceptibility genes may still be biased due to confounding. The expression weights used for TWAS may tag variants that are causal through other genes or non-genic mechanisms. In principle, this can be partially remedied by jointly testing multiple genes with trait; however, a correctly specified model would require covariance estimates between observed, not predicted, expression levels—which is not available in summary data. In this work we combined estimates across tissues by taking the mean effect to compute the genetic correlation between trait and expression. This approach is unbiased, but may be inefficient. Recent work[57] describes a random-effects model to combine estimates across tissues to increase power. Finally, our method to estimate correlation between traits using the genetically predicted component of gene expression makes several simplifying assumptions. We remedied the non-independence of genes by

sampling single genes within a 1Mb region, an approach which has been used previously[45]. However, a more powerful approach may take correlations across genes into account.

Acknowledgements

We would like to thank Valerie Arboleda, Robert Brown, Kathy Burch, and Malika Kumar for helpful discussions and feedback. We also thank Dr. Nicole Soranzo for sharing summary data for the platelet traits.

CMC: Data were generated as part of the CommonMind Consortium supported by funding from Takeda Pharmaceuticals Company Limited, F. Hoffman-La Roche Ltd and NIH grants R01MH085542, R01MH093725, P50MH066392, P50MH080405, R01MH097276, RO1-MH-075916, P50M096891, P50MH084053S1, R37MH057881 and R37MH057881S1, HHSN271201300031C, AG02219, AG05138 and MH06692. Brain tissue for the study was obtained from the following brain bank collections: the Mount Sinai NIH Brain and Tissue Repository, the University of Pennsylvania Alzheimer's Disease Core Center, the University of Pittsburgh NeuroBioBank and Brain and Tissue Repositories and the NIMH Human Brain Collection Core. CMC Leadership: Pamela Sklar, Joseph Buxbaum (Icahn School of Medicine at Mount Sinai), Bernie Devlin, David Lewis (University of Pittsburgh), Raquel Gur, Chang-Gyu Hahn (University of Pennsylvania), Keisuke Hirai, Hiroyoshi Toyoshiba (Takeda Pharmaceuticals Company Limited), Enrico Domenici, Laurent Essioux (F. Hoffman-La Roche Ltd), Lara Mangravite, Mette Peters (Sage Bionetworks), Thomas Lehner, Barbara Lipska (NIMH)

Web Resources

TWAS: http://bogdan.bioinformatics.ucla.edu/software/twas/

CMC: https://www.synapse.org/cmc/

GTEx: http://www.gtexportal.org/home/

GCTA: http://cnsgenomics.com/software/gcta/

Figures

Figure 1. Illustration of several causal models that explain expression correlation for traits T_1 and T_2 given their causal gene sets. Model A) trait 1 directly influences trait 2. In this case, the effect of genes $G_1^1, \ldots, G_p^1$ on trait 2 is mediated by trait 1 which implies $\{G_i^1\}_{i=1}^p \subsetneq \{G_i^2\}_{i=1}^q$. Model B) trait 2 directly influences trait 1. Similarly, the effect of genes $G_1^2, \ldots, G_q^2$ on trait 1 is mediated by trait 2 which implies $\{G_i^2\}_{i=1}^q \subsetneq \{G_i^1\}_{i=1}^p$. Model C) traits 1 and 2 are influenced independently through unobserved trait or traits.

Figure 2. Susceptibility genes shared for education years and height. We indicate $-\log_{10}$ p-values for eQTLs in green and trait-specific GWAS in black using separate axes to simplify illustration. Their respective TWAS p-values are *ABCB9* (heart; $p_{height}=1.38 \times 10^{-15}$, $p_{ey}=1.28 \times 10^{-6}$), *BTN2A3P* (adipose; $p_{height}=3.82 \times 10^{-12}$, $p_{ey}=1.90 \times 10^{-7}$), and *MPHOSPH9* (thyroid; $p_{height}=5.84 \times 10^{-18}$, $p_{ey}=1.30 \times 10^{-6}$).

Figure 3. Histogram and density estimate for correlation of $\rho_{g,local}$ across tissues. We computed the correlation across pairs of different tissues using local estimates of genetic correlation between expression on trait. The majority of tissues exhibited high correlation over the underlying gene effects on trait with an estimated mean $r = 0.82$

Figure 4. Estimates of genetic correlation ρ_g obtained from LD-Score vs estimates of expression correlation ρ_{GE} using nominally significant TWAS results. A) Correlation matrix for 30 traits. The lower triangle contains ρ_{GE} and the upper triangle contains ρ_g estimates. Estimates of correlation that are significantly non-zero ($p < 0.05 / 435$) are marked with a star (*). Strength and direction of correlation is indicated by size and color. We found 43 significantly correlated traits using cis expression and 62 using genome-wide SNPs. B) Linear relationship between estimates of ρ_{GE} and ρ_g. We indicate whether individual estimates were significant in either approach by color. Non-significant trait pairs are reduced in size for visibility.

Figure 5. Estimates of expression correlation ρ_{GE} for HDL, LDL, and TG with total cholesterol. Column A) Estimates of ρ_{GE} using nominally significant genes ($p < 0.05$). Column B) We repeated the analysis using only susceptibility genes found in the x-axis trait but not found in the y-axis trait. Column C) Same analysis as Column B, but using the other trait's susceptibility genes. All three analyses resulted in stronger point estimates for $\rho_{TC \mid Lipid}$ when conditioning on

HDL/LDL/TG genes compared to $\rho_{Lipid \mid TC}$; however, significance was only observed for $\rho_{TC \mid TG}$ ($p=2.34 \times 10^{-3}$).

Figure 6. Estimates of expression correlation ρ_{GE} for triglycerides with BMI and triglycerides with LDL. We present results for pairs of traits that displayed a significant difference ($p < 0.05$; Welch's t-test) in their conditional estimates. These results are consistent with a causal model where BMI influences TG and TG influences LDL.

Tables

Table 1. Summary of GWAS and TWAS results. The majority (92%) of GWAS risk loci overlap with at least one eGene, of which 40% contain at least one susceptibility gene. We report 168 (9%) of identified gene-trait pairs do not overlap a GWAS variant, which provide novel risk loci for follow up.

		Number of GWAS				Number of Susceptibility	
Trait	Short Name	Loci	Loci with eGene	Loci with Single Susceptibility Gene	Loci with at least one Susceptibility Gene	Genes that overlap GWAS	Genes that do not overlap GWAS
Age at Menarche	AM	70	60	14	19	34	9
Body Mass Index	BMI	76	60	10	18	44	11
College	COL	5	5	2	2	1	4
Crohn's Disease	CD	50	48	4	17	65	5
Education Years	EY	7	4	2	2	2	11
Fasting Glucose	FG	12	11	2	5	8	1
Fasting Insulin	FI	0	0	0	0	0	1
emoral Neck BMD	FN	20	20	2	2	2	1
Forearm BMD	FA	3	3	0	0	0	0
Hemoglobin	HB	22	21	2	5	22	3
HBA1C	HBA1C	10	10	0	1	4	0
Height	HEIGHT	482	454	94	225	669	52
High Density Lipoprotein	HDL	100	95	11	29	98	4
HOMA-B	HOMA-B	4	3	0	0	0	0
HOMA-IR	HOMA-IR	0	0	0	0	0	1
Inflammatory Bowel Disease	IBD	63	59	12	23	70	11
Low Density Lipoprotein	LDL	75	72	8	25	84	3
Lumbar Spine	LS	24	23	2	3	4	0
ICH Concentration	MCHC	5	3	0	0	0	0
Mean Cell Hemoglobin	MCH	35	31	5	17	46	7

Mean Cell Volume	MCV	43	40	8	20	49	1
Number of Platelets	PLT	35	34	6	13	30	8
Packed Cell Volume	PCV	14	13	1	3	5	1
Red Blood Cell Count	RBC	25	21	3	10	35	2
Rheumatoid Arthritis	RA	44	41	7	13	30	5
Schizophrenia	SCZ	95	74	15	31	113	24
Total Cholesterol	TC	88	85	13	40	117	0
Triglycerides	TG	70	67	4	18	59	1
Type 2 Diabetes	T2D	12	12	0	1	3	0
Ulcerative Colitis	UC	37	36	5	9	27	2
Total		1526	1405	232	551	1621	168

Table 2. Novel risk loci. Identified susceptibility genes that do not overlap a genome-wide significant SNP ($p < 5 \times 10^{-8}$) within 0.5Mb for the tested trait.

Trait	Gene	Chr	Tx Start	Tx End	Index SNP	Index SNP P	Expression Training Set	TWAS P
AM	*NUCKS1*	chr1	205681946	205719372	rs1832775	7.10E-08	GTEX_Skin_Not_Sun_Exposed_Suprapubic	1.70E-08
AM	*RAB7L1*	chr1	205737113	205744574	rs1832775	7.10E-08	GTEX_Whole_Blood	7.96E-07
AM	*SLC26A9*	chr1	205882176	205912588	rs1832775	7.10E-08	CMC	3.07E-07
AM							GTEX_Brain_Frontal_Cortex_BA9	5.78E-10
AM	*PMS2P5*	chr7	72476618	72520245	rs13238203	1.10E-02	CMC	1.27E-06
AM							GTEX_Brain_Cerebellum	8.41E-08
AM							GTEX_Muscle_Skeletal	9.19E-07
AM							GTEX_Lung	6.17E-07
AM							GTEX_Esophagus_Mucosa	4.38E-08
AM	*STAG3L2*	chr7	74298091	74306731	rs4717903	1.50E-07	GTEX_Thyroid	1.91E-07
AM							GTEX_Artery_Tibial	5.99E-09
AM							GTEX_Nerve_Tibial	2.84E-10
AM							GTEX_Esophagus_Muscularis	5.47E-07
AM	*TMEM180*	chr10	104221148	104236005	rs2274351	1.80E-05	GTEX_Heart_Atrial_Appendage	6.32E-07
AM	*CCDC65*	chr12	49297892	49315359	rs11168850	3.70E-06	GTEX_Artery_Aorta	2.92E-07
AM	*COG6*	chr13	40229763	40326765	rs9548873	2.90E-07	GTEX_Nerve_Tibial	4.78E-09
AM							GTEX_Artery_Tibial	2.55E-08
AM	*INO80E*	chr16	30007529	30017111	rs747973	1.10E-07	GTEX_Skin_Sun_Exposed_Lower_leg	6.92E-07
BMI							GTEX_Artery_Tibial	9.44E-07
BMI							CMC	6.09E-07
BMI	*STAG3L1*	chr7	72467817	72476448	rs7801936	1.01E-03	GTEX_Skin_Sun_Exposed_Lower_leg	5.34E-08
BMI							GTEX_Thyroid	6.48E-07
BMI							GTEX_Muscle_Skeletal	1.58E-08
BMI	*SLC27A4*	chr9	131102838	131123749	rs2270204	3.25E-07	GTEX_Esophagus_Mucosa	1.15E-06
BMI							GTEX_Lung	6.40E-08
BMI							GTEX_Artery_Tibial	2.15E-07
BMI	*URM1*	chr9	131133597	131154295	rs2270204	3.25E-07	GTEX_Skin_Sun_Exposed_Lower_leg	6.32E-08
BMI							GTEX_Nerve_Tibial	2.60E-07
BMI							GTEX_Adipose_Subcutaneous	1.77E-08
BMI							GTEX_Esophagus_Muscularis	5.19E-07
BMI	*CERCAM*	chr9	131181438	131199630	rs2270204	3.25E-07	GTEX_Thyroid	5.46E-07
BMI	*TUBA1C*	chr12	49621708	49667121	rs12821008	3.21E-05	YFS	2.25E-07
BMI							GTEX_Adipose_Subcutaneous	6.00E-07
BMI							GTEX_Artery_Tibial	5.10E-07
BMI							GTEX_Heart_Atrial_Appendage	1.05E-06
BMI							GTEX_Esophagus_Mucosa	5.45E-07
BMI							GTEX_Pituitary	2.88E-07
BMI	*INO80E*	chr16	30007529	30017111	rs4787491	2.92E-06	GTEX_Breast_Mammary_Tissue	2.23E-07
BMI							GTEX_Adipose_Visceral_Omentum	6.60E-07
BMI							GTEX_Nerve_Tibial	4.34E-07
BMI							GTEX_Colon_Transverse	1.23E-06
BMI							GTEX_Thyroid	3.41E-07
BMI							GTEX_Testis	6.65E-07

BMI							GTEX_Skin_Sun_Exposed_Lower_leg	1.29E-06
BMI	*GGNBP2*	chr17	34900736	34946276	rs12150665	2.06E-07	GTEX_Esophagus_Mucosa	4.40E-08
BMI							GTEX_Artery_Tibial	4.12E-07
BMI	*DHRS11*	chr17	34948225	34957233	rs12150665	2.06E-07	GTEX_Muscle_Skeletal	3.93E-07
BMI	*RP11-6N17.9*	chr17	46022789	46051804	rs6504108	3.09E-07	GTEX_Thyroid	1.90E-07
BMI							GTEX_Nerve_Tibial	1.09E-06
BMI	*CDK5RAP3*	chr17	46047893	46059152	rs6504108	3.09E-07	GTEX_Thyroid	4.34E-09
BMI	*RP11-6N17.10*	chr17	46057763	46073562	rs6504108	3.09E-07	GTEX_Thyroid	2.38E-08
COL	*AFF3*	chr2	100163715	100722045	rs13033732	4.03E-07	CMC	5.89E-08
COL	*RNF123*	chr3	49726931	49758962	rs3796386	1.04E-06	GTEX_Thyroid	1.07E-06
COL	*AC091729.9*	chr7	1200009	1204903	rs7793318	5.74E-06	GTEX_Spleen	7.54E-07
COL	*ABCB9*	chr12	123405497	123451056	rs12316131	7.40E-07	GTEX_Skin_Not_Sun_Exposed_Suprapubic	6.57E-07
CD	*RIT1*	chr1	155867598	155880706	rs821551	7.01E-08	GTEX_Thyroid	3.37E-08
CD	*SMIM19*	chr8	42396297	42408140	rs2923396	1.20E-06	GTEX_Brain_Cortex	6.42E-07
CD	*CISD1*	chr10	60028861	60049019	rs1624017	2.95E-07	GTEX_Whole_Blood	1.28E-07
CD							GTEX_Testis	1.56E-07
CD	*PPP1R14B*	chr11	64011950	64014413	rs11231774	2.47E-06	GTEX_Artery_Tibial	8.56E-09
CD	*CCDC88B*	chr11	64107689	64125006	rs11231774	2.47E-06	GTEX_Brain_Cerebellum	6.32E-08
CD							CMC	1.11E-06
EY	*SDCCAG8*	chr1	243419306	243663393	rs12080886	5.73E-07	GTEX_Whole_Blood	7.77E-07
EY							GTEX_Skin_Sun_Exposed_Lower_leg	1.01E-06
EY	*ABCB9*	chr12	123405497	123451056	rs7980687	1.59E-06	GTEX_Heart_Left_Ventricle	1.28E-06
EY							GTEX_Skin_Not_Sun_Exposed_Suprapubic	5.06E-07
EY	*MPHOSPH9*	chr12	123640942	123717785	rs7980687	1.59E-06	GTEX_Thyroid	1.30E-06
EY	*STK24*	chr13	99102452	99174379	rs17574378	1.52E-07	NTR	1.96E-09
EY	*EIF3CL*	chr16	28390902	28415206	rs8049439	1.52E-07	GTEX_Adipose_Subcutaneous	5.27E-07
EY							GTEX_Lung	4.50E-09
EY							GTEX_Skin_Sun_Exposed_Lower_leg	4.15E-07
EY							GTEX_Muscle_Skeletal	1.07E-08
EY	*SULT1A1*	chr16	28616907	28620649	rs8049439	1.52E-07	MET	1.22E-06
EY							GTEX_Pancreas	3.18E-08
EY							GTEX_Breast_Mammary_Tissue	7.53E-07
EY							GTEX_Artery_Tibial	1.29E-06
EY							GTEX_Artery_Aorta	1.76E-07
EY	*RP11-1348G14.4*	chr16	28814096	28829149	rs8049439	1.52E-07	GTEX_Esophagus_Muscularis	2.85E-07
EY							GTEX_Lung	9.86E-07
EY							GTEX_Nerve_Tibial	7.77E-07
EY							MET	3.05E-08
EY	*TUFM*	chr16	28853731	28857729	rs8049439	1.52E-07	GTEX_Nerve_Tibial	5.25E-07
EY							GTEX_Artery_Aorta	3.64E-07

EY							GTEX_Artery_Tibial	2.42E-07
EY							GTEX_Spleen	4.83E-07
EY							GTEX_Lung	4.90E-07
EY							GTEX_Skin_Not_Sun_Exposed_Suprapubic	2.65E-08
EY							GTEX_Whole_Blood	5.88E-07
EY							GTEX_Colon_Sigmoid	4.45E-07
EY	*MIR4721*	chr16	28855239	28855328	rs8049439	1.52E-07	GTEX_Whole_Blood	6.25E-07
EY							GTEX_Artery_Tibial	1.18E-06
EY	*SH2B1*	chr16	28857920	28885534	rs8049439	1.52E-07	GTEX_Adipose_Subcutaneous	6.52E-07
EY							MET	6.60E-07
EY	*NFATC2IP*	chr16	28962317	28977767	rs8049439	1.52E-07	GTEX_Skin_Not_Sun_Exposed_Suprapubic	5.01E-07
FG	*MAPRE3*	chr2	27193238	27250087	rs7586601	2.82E-07	GTEX_Skin_Not_Sun_Exposed_Suprapubic	1.21E-06
FI							GTEX_Thyroid	1.11E-06
FI	*KNOP1*	chr16	19717673	19729492	rs1858973	1.61E-05	GTEX_Lung	7.85E-07
FI							GTEX_Adrenal_Gland	5.27E-07
FN	*FGFRL1*	chr4	1005609	1020686	rs35654957	5.10E-08	GTEX_Thyroid	2.86E-07
HB	*UBE2Q2*	chr15	76135626	76193388	rs1976748	2.07E-07	CMC	7.61E-07
HB	*WNT3*	chr17	44839871	44896126	rs916888	8.24E-07	GTEX_Artery_Aorta	1.24E-06
HB	*CCDC117*	chr22	29168661	29185289	rs13056243	2.44E-07	YFS	9.31E-07
[EIGHT	*ECHDC2*	chr1	53361581	53387446	rs1769316	1.21E-05	YFS	1.24E-06
[EIGHT	*RP4-612B15.3*	chr1	87169184	87170145	rs4656136	1.47E-06	GTEX_Adipose_Subcutaneous	1.08E-06
[EIGHT							GTEX_Esophagus_Mucosa	1.15E-06
[EIGHT							GTEX_Artery_Tibial	4.37E-07
[EIGHT							GTEX_Skin_Not_Sun_Exposed_Suprapubic	6.43E-07
[EIGHT							GTEX_Muscle_Skeletal	1.04E-06
[EIGHT	*CNIH4*	chr1	224544512	224564586	rs6693287	5.68E-08	GTEX_Skin_Sun_Exposed_Lower_leg	5.03E-07
[EIGHT							GTEX_Thyroid	1.52E-07
[EIGHT							GTEX_Esophagus_Muscularis	4.86E-07
[EIGHT							GTEX_Nerve_Tibial	3.92E-07
[EIGHT							GTEX_Adipose_Subcutaneous	2.45E-07
[EIGHT							GTEX_Breast_Mammary_Tissue	2.77E-07
[EIGHT							GTEX_Muscle_Skeletal	3.76E-08
[EIGHT							GTEX_Pancreas	2.70E-07
[EIGHT	*SH3YL1*	chr2	218135	256340	rs2290911	5.73E-07	GTEX_Esophagus_Mucosa	7.97E-08
[EIGHT							GTEX_Heart_Atrial_Appendage	2.31E-08
[EIGHT							GTEX_Nerve_Tibial	1.63E-07
[EIGHT							GTEX_Esophagus_Muscularis	2.29E-09
[EIGHT	*LBX2-AS1*	chr2	74729743	74732192	rs2268424	2.42E-05	GTEX_Lung	2.98E-07
[EIGHT	*MAT2A*	chr2	85766100	85772403	rs1465821	2.31E-07	GTEX_Brain_Cerebellum	1.23E-06
[EIGHT	*CRELD1*	chr3	9975523	9987097	rs1473183	6.33E-05	GTEX_Skin_Not_Sun_Exposed_Suprapubic	6.32E-10

[EIGHT							GTEX_Skin_Sun_Exposed_Lower_leg	2.61E-09
[EIGHT							GTEX_Artery_Tibial	9.45E-07
[EIGHT							GTEX_Adipose_Subcutaneous	1.55E-10
[EIGHT	*P4HTM*	chr3	49027340	49044581	rs990211	4.34E-06	GTEX_Nerve_Tibial	3.46E-07
[EIGHT	*TMEM128*	chr4	4237268	4249969	rs11729131	6.33E-05	GTEX_Lung	9.70E-07
[EIGHT	*FUCA2*	chr6	143815948	143833020	rs2161972	5.73E-07	MET	5.14E-07
[EIGHT	*HIBADH*	chr7	27565058	27702620	rs10951187	3.68E-07	CMC	6.19E-08
[EIGHT	*ATP5J2*	chr7	99055783	99063824	rs17277546	6.07E-08	GTEX_Testis	2.70E-07
[EIGHT							GTEX_Thyroid	3.77E-10
[EIGHT	*WDR60*	chr7	158649268	158738883	rs4909278	4.84E-06	GTEX_Nerve_Tibial	9.18E-07
[EIGHT	*C9orf156*	chr9	100666770	100684852	rs953199	5.68E-08	GTEX_Brain_Cortex	1.34E-08
[EIGHT							GTEX_Adipose_Subcutaneous	1.54E-09
[EIGHT	*RP11-4O1.2*	chr9	114794834	114800010	rs10441737	1.31E-06	GTEX_Esophagus_Muscularis	7.93E-08
[EIGHT							GTEX_Adipose_Visceral_Omentum	6.60E-07
[EIGHT	*SUSD1*	chr9	114803060	114937577	rs4310281	1.31E-06	GTEX_Esophagus_Muscularis	1.03E-07
[EIGHT	*MEGF9*	chr9	123363195	123476765	rs10739570	1.08E-07	GTEX_Adipose_Subcutaneous	1.35E-07
[EIGHT	*PSMD5*	chr9	123578331	123605299	rs10739570	1.08E-07	GTEX_Liver	6.56E-09
[EIGHT							GTEX_Brain_Hippocampus	2.48E-08
[EIGHT	*PSMD5-AS1*	chr9	123605319	123616651	rs10739570	1.08E-07	GTEX_Ovary	2.61E-07
[EIGHT							GTEX_Uterus	4.40E-09
[EIGHT	*PHF19*	chr9	123617928	123631189	rs10739570	1.08E-07	GTEX_Esophagus_Muscularis	5.10E-09
[EIGHT	*DAB2IP*	chr9	124329380	124547809	rs12683062	1.21E-05	GTEX_Thyroid	1.92E-07
[EIGHT	*MSRB2*	chr10	23384426	23410942	rs10828323	1.24E-06	YFS	1.75E-07
[EIGHT							GTEX_Adipose_Subcutaneous	1.32E-06
[EIGHT	*RP13-39P12.3*	chr10	79542623	79552934	rs1268956	1.08E-07	GTEX_Whole_Blood	2.47E-07
[EIGHT							GTEX_Nerve_Tibial	1.17E-08
[EIGHT							GTEX_Lung	1.50E-07
[EIGHT							GTEX_Lung	9.34E-07
[EIGHT							MET	5.40E-07
[EIGHT	*DLG5*	chr10	79550548	79686348	rs1268956	1.08E-07	GTEX_Adipose_Visceral_Omentum	1.31E-06
[EIGHT							GTEX_Brain_Cerebellum	2.79E-07
[EIGHT							GTEX_Esophagus_Mucosa	5.85E-08
[EIGHT	*SFTPD*	chr10	81697495	81708861	rs1304463	2.45E-07	GTEX_Esophagus_Gastroesophageal_Junction	2.19E-07
[EIGHT	*FAM35A*	chr10	88854952	88951222	rs6586076	5.73E-07	GTEX_Skin_Not_Sun_Exposed_Suprapubic	1.75E-07
[EIGHT	*PLEKHA1*	chr10	124134093	124191871	rs4612730	9.64E-08	NTR	2.29E-07
[EIGHT	*MGMT*	chr10	131265447	131565884	rs11016853	3.06E-06	GTEX_Testis	2.71E-07
[EIGHT	*H2AFJ*	chr12	14927269	14930936	rs10772776	2.31E-07	YFS	7.11E-07
[EIGHT	*ATF1*	chr12	51157788	51214943	rs10747592	1.31E-06	YFS	1.35E-07
[EIGHT	*MED4*	chr13	48649863	48669277	rs12875433	3.84E-07	GTEX_Skin_Sun_Exposed_Lower_leg	3.96E-08
[EIGHT	*IQGAP1*	chr15	90931472	91045475	rs6496620	2.58E-07	MET	1.28E-06

EIGHT							GTEX_Stomach	7.09E-07
EIGHT	*RRN3*	chr16	15153878	15188192	rs3751877	7.13E-06	GTEX_Heart_Left_Ventricle	2.76E-08
EIGHT							GTEX_Adipose_Subcutaneous	8.06E-07
EIGHT							GTEX_Testis	2.92E-07
EIGHT							GTEX_Pituitary	4.10E-09
EIGHT							GTEX_Heart_Atrial_Appendage	2.29E-07
EIGHT	*INO80E*	chr16	30007529	30017111	rs11642612	9.64E-08	GTEX_Skin_Sun_Exposed_Lower_leg	8.62E-08
EIGHT							YFS	5.68E-08
EIGHT							GTEX_Esophagus_Muscularis	4.10E-07
EIGHT							GTEX_Artery_Tibial	2.16E-07
EIGHT							GTEX_Colon_Transverse	1.10E-06
EIGHT	*YPEL3*	chr16	30103634	30107521	rs11642612	9.64E-08	GTEX_Breast_Mammary_Tissue	3.99E-07
EIGHT	*RP11-455F5.3*	chr16	30107750	30115437	rs11642612	9.64E-08	GTEX_Skin_Sun_Exposed_Lower_leg	6.88E-07
EIGHT	*RP11-67A1.2*	chr16	68111242	68156174	rs6499141	2.65E-07	GTEX_Testis	1.30E-06
EIGHT	*RP11-173M1.8*	chr17	25641668	25642952	rs7207347	1.38E-06	GTEX_Esophagus_Gastroesophageal_Junction	1.33E-06
EIGHT							GTEX_Muscle_Skeletal	1.46E-07
EIGHT	*CRHR1*	chr17	43861645	43913194	rs7222389	5.73E-07	GTEX_Colon_Transverse	3.94E-07
EIGHT							GTEX_Heart_Left_Ventricle	9.53E-10
EIGHT	*MAPT*	chr17	43971747	44105699	rs7222389	5.73E-07	GTEX_Brain_Cerebellar_Hemisphere	1.76E-07
EIGHT	*KANSL1*	chr17	44107281	44270166	rs7225082	5.73E-07	GTEX_Skin_Sun_Exposed_Lower_leg	7.91E-08
EIGHT	*ARL17A*	chr17	44376499	44439163	rs12947764	5.48E-06	GTEX_Esophagus_Muscularis	2.16E-07
EIGHT	*LRRC37A2*	chr17	44590075	44633014	rs8077487	8.81E-06	GTEX_Stomach	2.17E-07
EIGHT	*UTP18*	chr17	49337896	49375292	rs9896627	5.73E-07	GTEX_Muscle_Skeletal	1.27E-06
EIGHT	*DUS3L*	chr19	5785152	5791249	rs11672480	6.33E-05	GTEX_Adipose_Subcutaneous	9.21E-07
EIGHT	*COX6B1*	chr19	36139124	36149686	rs2280743	5.73E-07	GTEX_Nerve_Tibial	4.02E-07
EIGHT	*C20orf194*	chr20	3229947	3388309	rs2207994	2.33E-06	GTEX_Nerve_Tibial	1.12E-06
EIGHT							NTR	1.11E-07
EIGHT	*YWHAB*	chr20	43514239	43537175	rs2239535	2.93E-07	GTEX_Artery_Tibial	1.26E-06
EIGHT							GTEX_Esophagus_Mucosa	7.65E-07
EIGHT	*UBE2L3*	chr22	21903735	21978323	rs5754102	1.31E-07	CMC	3.41E-08
EIGHT							GTEX_Nerve_Tibial	2.64E-09
EIGHT	*CCDC116*	chr22	21987085	21991616	rs5754102	1.31E-07	GTEX_Artery_Aorta	8.21E-08
EIGHT							GTEX_Esophagus_Muscularis	4.29E-07
EIGHT	*MORC2-AS1*	chr22	31318294	31322640	rs1076301	2.53E-06	GTEX_Thyroid	6.20E-07
EIGHT	*DESI1*	chr22	41994031	42017061	rs5758341	7.37E-06	GTEX_Breast_Mammary_Tissue	6.05E-07
HDL	*RETSAT*	chr2	85569077	85581821	rs10460587	3.32E-07	GTEX_Whole_Blood	1.61E-07
HDL	*HRAS*	chr11	532241	535567	rs7117022	3.29E-05	GTEX_Artery_Tibial	3.08E-07
HDL	*TYRO3*	chr15	41851219	41871536	rs721772	2.26E-07	GTEX_Lung	7.33E-07
HDL	*KNOP1*	chr16	19717673	19729492	rs11865578	1.72E-06	GTEX_Skin_Not_Sun_Exposed_Suprapubic	6.84E-07
HDL							GTEX_Lung	1.00E-06

OMA-IR							GTEX_Lung	8.13E-07
OMA-IR	*KNOP1*	chr16	19717673	19729492	rs1858973	2.05E-05	GTEX_Artery_Coronary	1.13E-06
OMA-IR							GTEX_Thyroid	8.21E-07
OMA-IR							GTEX_Adrenal_Gland	1.32E-06
IBD							GTEX_Artery_Tibial	1.48E-07
IBD							GTEX_Nerve_Tibial	2.56E-07
IBD							GTEX_Lung	1.78E-08
IBD	*GBAP1*	chr1	155183615	155197325	rs734073	6.71E-07	GTEX_Adipose_Subcutaneous	6.75E-08
IBD							GTEX_Esophagus_Mucosa	1.44E-07
IBD							GTEX_Heart_Left_Ventricle	4.12E-07
IBD							GTEX_Thyroid	5.47E-07
IBD	*GBA*	chr1	155204238	155211066	rs734073	6.71E-07	GTEX_Esophagus_Mucosa	1.16E-06
IBD	*FAM189B*	chr1	155216995	155225274	rs734073	6.71E-07	GTEX_Thyroid	2.46E-07
IBD	*HCN3*	chr1	155247217	155258498	rs734073	6.71E-07	GTEX_Nerve_Tibial	5.23E-07
IBD	*ADCY3*	chr2	25042038	25142886	rs7583409	1.77E-06	YFS	1.17E-06
IBD	*SATB2*	chr2	200134222	200322819	rs62180151	3.77E-07	GTEX_Skin_Sun_Exposed_Lower_leg	1.05E-06
IBD	*TMEM180*	chr10	104221148	104236005	rs7078511	1.73E-06	GTEX_Breast_Mammary_Tissue	6.03E-07
IBD	*PPP1R14B*	chr11	64011950	64014413	rs11231774	1.35E-05	GTEX_Artery_Tibial	5.85E-08
IBD	*CCDC88B*	chr11	64107689	64125006	rs11231774	1.35E-05	GTEX_Brain_Cerebellum	5.24E-07
IBD							CMC	5.09E-07
IBD	*RMI2*	chr16	11439294	11445620	rs12598132	2.54E-07	GTEX_Muscle_Skeletal	1.06E-06
IBD							GTEX_Breast_Mammary_Tissue	7.26E-07
IBD	*ZFP90*	chr16	68573115	68601039	rs1094281	4.79E-07	YFS	1.05E-06
LDL	*WDR25*	chr14	100842754	100996640	rs999045	2.81E-05	GTEX_Testis	5.44E-07
LDL	*ERAL1*	chr17	27181974	27188085	rs9892942	1.59E-06	YFS	8.33E-07
LDL	*DHRS13*	chr17	27224798	27230089	rs9892942	1.59E-06	YFS	6.00E-07
MCH	*RP11-69E11.4*	chr1	39987951	40011859	rs3916164	2.25E-07	GTEX_Skin_Not_Sun_Exposed_Suprapubic	5.99E-07
MCH							GTEX_Esophagus_Muscularis	1.25E-06
MCH	*PABPC4*	chr1	40026484	40042521	rs3916164	2.25E-07	GTEX_Nerve_Tibial	2.95E-07
MCH	*RP1-18D14.7*	chr1	47691468	47696422	rs741959	8.45E-08	GTEX_Whole_Blood	2.48E-07
MCH	*RPS6KB2*	chr11	67195934	67202879	rs596603	3.41E-07	GTEX_Skin_Sun_Exposed_Lower_leg	2.94E-07
MCH							GTEX_Testis	1.26E-07
MCH	*AP003419.16*	chr11	67198837	67202870	rs596603	3.41E-07	GTEX_Muscle_Skeletal	1.19E-06
MCH							GTEX_Skin_Sun_Exposed_Lower_leg	5.15E-07
MCH							GTEX_Testis	1.57E-07
MCH							GTEX_Esophagus_Muscularis	6.93E-07
MCH							GTEX_Skin_Not_Sun_Exposed_Suprapubic	1.47E-07
MCH							GTEX_Artery_Tibial	9.85E-07
MCH	*PTPRCAP*	chr11	67202980	67205153	rs596603	3.41E-07	GTEX_Skin_Sun_Exposed_Lower_leg	4.40E-08
MCH							GTEX_Adipose_Subcutaneous	1.23E-07

MCH							GTEX_Brain_Nucleus_accumbens_basal_ganglia	1.85E-08
MCH							GTEX_Testis	6.29E-07
MCH							GTEX_Esophagus_Muscularis	5.36E-08
MCH							GTEX_Brain_Cortex	1.07E-07
MCH							GTEX_Brain_Cerebellum	1.19E-06
MCH							GTEX_Muscle_Skeletal	4.86E-07
MCH							GTEX_Brain_Cerebellar_Hemisphere	9.09E-08
MCH							GTEX_Nerve_Tibial	5.98E-08
MCH							GTEX_Ovary	2.42E-07
MCH							GTEX_Breast_Mammary_Tissue	5.17E-07
MCH							GTEX_Thyroid	3.44E-07
MCH	*GSTP1*	chr11	67351065	67354124	rs596603	3.41E-07	GTEX_Artery_Tibial	1.12E-06
MCV	*COX4I2*	chr20	30225690	30232800	rs6060399	1.96E-06	GTEX_Muscle_Skeletal	7.16E-07
PLT	*MUTYH*	chr1	45794913	45805629	rs4660853	1.10E-07	MET	2.32E-07
PLT	*TESK2*	chr1	45809554	45956840	rs4660853	1.10E-07	YFS	5.92E-08
PLT	*CCDC17*	chr1	46085715	46089731	rs4660853	1.10E-07	YFS	2.25E-07
PLT	*IPP*	chr1	46159997	46216485	rs4660853	1.10E-07	YFS	1.04E-06
PLT	*TMEM180*	chr10	104221148	104236005	rs2281880	7.66E-07	GTEX_Brain_Cerebellar_Hemisphere	1.12E-06
PLT							GTEX_Liver	7.61E-07
PLT	*ACTR1A*	chr10	104238985	104262512	rs2281880	7.66E-07	NTR	2.91E-07
PLT	*BAZ2A*	chr12	56989379	57024115	rs2958139	6.56E-08	YFS	1.17E-06
PLT	*PRIM1*	chr12	57125363	57146146	rs2958139	6.56E-08	NTR	7.76E-07
PCV	*PLEKHH2*	chr2	43864438	43995126	rs1368087	1.28E-04	GTEX_Thyroid	1.20E-06
RBC	*FBXL20*	chr17	37408896	37557909	rs8182252	5.09E-08	GTEX_Small_Intestine_Terminal_Ileum	3.69E-07
RBC	*COX4I2*	chr20	30225690	30232800	rs6060359	3.34E-07	GTEX_Muscle_Skeletal	5.36E-07
RA							GTEX_Pancreas	8.16E-09
RA							GTEX_Skin_Sun_Exposed_Lower_leg	2.73E-07
RA							GTEX_Nerve_Tibial	1.22E-08
RA	*SUOX*	chr12	56391042	56399309	rs773125	8.50E-08	GTEX_Colon_Transverse	1.21E-09
RA							GTEX_Artery_Aorta	2.04E-07
RA							GTEX_Esophagus_Muscularis	8.40E-07
RA							GTEX_Lung	3.43E-07
RA							CMC	5.04E-07
RA	*RPS26*	chr12	56435685	56438007	rs773125	8.50E-08	GTEX_Breast_Mammary_Tissue	6.41E-07
RA							GTEX_Ovary	3.40E-07
RA	*SLC26A10*	chr12	58013692	58019934	rs1633360	9.10E-08	GTEX_Heart_Left_Ventricle	1.06E-09
RA	*METTL21B*	chr12	58166382	58176324	rs1633360	9.10E-08	GTEX_Esophagus_Muscularis	8.02E-07
RA	*RNF40*	chr16	30772932	30787628	rs56656810	2.80E-06	GTEX_Nerve_Tibial	9.58E-08
RA							GTEX_Adipose_Subcutaneous	1.00E-06
SCZ	*CAD*	chr2	27440257	27466660	rs4665386	1.87E-05	GTEX_Pancreas	3.32E-07

SCZ	*NRBP1*	chr2	27650656	27665124	rs12474906	9.11E-08	GTEX_Skin_Not_Sun_Exposed_Suprapubic	8.83E-07
SCZ	*CEBPZ*	chr2	37428774	37458740	rs3770752	2.69E-06	GTEX_Adipose_Subcutaneous	4.62E-07
SCZ	*ALMS1P*	chr2	73872045	73912694	rs56145559	7.83E-08	CMC	9.22E-07
SCZ	*10-Sep*	chr2	110300373	110371234	rs9330316	7.70E-08	CMC	9.09E-07
SCZ	*EMB*	chr5	49692030	49737234	rs76329678	5.75E-07	CMC	5.60E-07
SCZ	*PFDN1*	chr5	139624634	139682689	rs2563297	8.04E-07	GTEX_Skin_Sun_Exposed_Lower_leg	4.66E-07
SCZ	*SRA1*	chr5	139929651	139937041	rs13168670	4.71E-07	NTR	1.02E-06
SCZ	*TMCO6*	chr5	140019011	140024993	rs13168670	4.71E-07	GTEX_Brain_Cerebellar_Hemisphere	5.80E-07
SCZ	*IK*	chr5	140027383	140042065	rs13168670	4.71E-07	MET	1.32E-06
SCZ	*DND1*	chr5	140050380	140053171	rs13168670	4.71E-07	MET	8.37E-07
SCZ	*ENDOG*	chr9	131580778	131584955	rs2805099	1.91E-06	GTEX_Thyroid	1.99E-08
SCZ	*ARL14EP*	chr11	30344645	30359770	rs1765142	2.42E-06	GTEX_Lung	9.54E-07
SCZ							GTEX_Esophagus_Muscularis	8.78E-08
SCZ							GTEX_Skin_Sun_Exposed_Lower_leg	1.48E-07
SCZ	*PCNX*	chr14	71374121	71582099	rs67981189	1.65E-07	GTEX_Nerve_Tibial	8.99E-07
SCZ							GTEX_Adipose_Subcutaneous	1.48E-07
SCZ	*CORO7*	chr16	4404542	4466962	rs6500602	2.65E-07	CMC	1.73E-07
SCZ							GTEX_Muscle_Skeletal	9.68E-07
SCZ	*NMRAL1*	chr16	4511677	4524651	rs6500602	2.65E-07	GTEX_Heart_Left_Ventricle	1.90E-07
SCZ							GTEX_Thyroid	1.10E-06
SCZ	*CPNE7*	chr16	89642175	89663654	rs34753377	1.07E-07	CMC	4.18E-09
SCZ	*GRAP*	chr17	18923989	18950336	rs4273100	7.84E-07	NTR	1.69E-07
SCZ	*EPN2*	chr17	19140689	19240028	rs4273100	7.84E-07	GTEX_Artery_Tibial	1.08E-06
SCZ	*RP11-135L13.4*	chr17	19237620	19239440	rs4273100	7.84E-07	GTEX_Artery_Tibial	1.90E-07
SCZ	*RNF112*	chr17	19314490	19320589	rs4273100	7.84E-07	GTEX_Testis	1.83E-07
SCZ	*PRRG2*	chr19	50084586	50094265	rs56873913	2.06E-07	GTEX_Skin_Sun_Exposed_Lower_leg	5.45E-08
SCZ	*PRR12*	chr19	50094911	50129696	rs56873913	2.06E-07	GTEX_Artery_Aorta	1.08E-06
SCZ							GTEX_Nerve_Tibial	4.71E-08
SCZ	*CBR3*	chr21	37507262	37518860	rs6517329	6.37E-06	GTEX_Artery_Tibial	9.14E-07
SCZ							GTEX_Skin_Not_Sun_Exposed_Suprapubic	7.37E-07
TG	*L3MBTL3*	chr6	130339727	130462594	rs12530176	1.02E-07	GTEX_Breast_Mammary_Tissue	5.50E-07
UC	*SATB2*	chr2	200134222	200322819	rs2344700	1.53E-07	GTEX_Skin_Sun_Exposed_Lower_leg	7.70E-08
UC	*TNPO3*	chr7	128594233	128695227	rs4728142	2.94E-07	YFS	1.30E-06

Table 3. Significant estimates of ρ_{GE} for 43 pairs of traits. We performed bi-directional regression and obtain conditional estimates of ρ_{GE}, which provides evidence for a putative causal direction. We observed three pairs of traits with a significant difference between their directional estimates; namely, BMI influencing TG, TG influencing LDL, and TG influencing TC. We mark entries with $M < 3$ with "-".

[a] determined by Welch–Satterthwaite equation

Trait 1	Trait 2	All nominally significant genes				Ascertain for trait 1				Ascertain for trait 2				Test for difference		
		Rho GE	SE	P	M	Rho GE	SE	P	M	Rho GE	SE	P	M	t	P	Approx M[a]
AM	BMI	-0.33	0.06	7.97E-08	257	-0.74	0.13	6.45E-05	23	-0.89	0.07	5.76E-08	21	1.05	2.99E-01	33
BMI	COL	-0.31	0.07	1.02E-05	190	-0.58	0.17	3.02E-03	24	0.28	0.25	6.50E-01	5	-2.81	2.30E-02	8
BMI	EY	-0.31	0.06	5.00E-06	210	-0.43	0.19	3.51E-02	24	-0.59	0.47	2.93E-01	5	0.32	7.64E-01	5
BMI	FI	0.39	0.06	3.06E-07	164	0.60	0.10	1.77E-03	24	-	-	-	0	-	-	-
BMI	HDL	-0.34	0.06	1.66E-08	256	-0.76	0.12	2.75E-05	23	-0.39	0.16	2.35E-02	33	-1.77	8.20E-02	53
BMI	HOMA-B	0.31	0.07	4.34E-05	168	0.73	0.07	5.50E-05	24	-	-	-	0	-	-	-
BMI	HOMA-IR	0.36	0.06	2.14E-06	162	0.55	0.11	5.40E-03	24	-	-	-	0	-	-	-
BMI	TG	0.29	0.06	5.16E-06	233	0.62	0.10	2.06E-03	22	-0.04	0.22	8.62E-01	19	2.74	1.12E-02	25
CD	IBD	0.93	0.01	< 2.00E-16	366	0.99	0.00	1.47E-06	8	0.97	0.01	1.39E-09	16	2.34	3.17E-02	17
CD	UC	0.51	0.05	6.66E-16	218	0.96	0.01	1.02E-10	19	0.74	0.09	5.54E-02	7	2.34	5.81E-02	6
COL	EY	0.95	0.00	< 2.00E-16	363	0.99	0.00	1.21E-02	4	0.99	0.00	6.38E-05	6	-1.14	3.18E-01	4
FA	FN	0.57	0.05	6.04E-14	149	-	-	-	0	0.98	-	1.29E-01	3	-	-	-
FA	LS	0.60	0.04	< 2.00E-16	170	-	-	-	0	0.76	-	4.49E-01	3	-	-	-
FG	FI	0.65	0.05	< 2.00E-16	133	-0.74	0.45	1.55E-01	5	-	-	-	1	-	-	-
FG	HOMA-B	-0.60	0.06	1.72E-13	125	-0.98	0.08	2.23E-03	5	-	-	-	0	-	-	-
FG	HOMA-IR	0.92	0.01	< 2.00E-16	136	0.46	0.19	4.41E-01	5	-	-	-	1	-	-	-
FI	HDL	-0.31	0.07	3.83E-05	168	-	-	-	0	-0.14	0.17	4.51E-01	33	-	-	-
FI	HOMA-B	0.97	0.00	< 2.00E-16	243	-	-	-	1	-	-	-	0	-	-	-
FI	HOMA-IR	0.99	0.00	< 2.00E-16	383	-	-	-	0	-	-	-	0	-	-	-
FI	TG	0.57	0.05	3.38E-14	152	-	-	-	1	0.53	0.13	2.08E-02	19	-	-	-
FN	LS	0.86	0.01	< 2.00E-16	264	1.00	-	-	2	-1.00	-	-	2	-	-	-
HB	MCH	0.37	0.06	1.58E-06	156	0.66	0.12	7.58E-02	8	0.58	0.11	9.29E-03	19	0.48	6.37E-01	19
HB	MCHC	0.40	0.08	2.35E-05	105	0.27	0.24	5.10E-01	8	-	-	-	0	-	-	-
HB	PCV	0.97	0.00	< 2.00E-16	338	0.99	0.00	4.83E-06	8	1.00	-	-	2	-	-	-
HB	PLT	-0.36	0.08	1.54E-05	141	0.29	0.23	4.86E-01	8	-0.49	0.24	5.57E-02	16	2.35	2.97E-02	19
HB	RBC	0.95	0.01	< 2.00E-16	260	0.94	0.02	4.61E-03	6	0.71	0.09	1.01E-02	12	2.47	2.95E-02	12
HBA1C	T2D	0.46	0.07	5.08E-07	110	-	-	-	1	-	-	-	1	-	-	-
HBA1C	TG	0.37	0.07	9.54E-06	137	-	-	-	1	-0.19	0.23	4.29E-01	19	-	-	-

HDL	HOMA-IR	-0.32	0.07	3.43E-05	159	-0.08	0.17	6.42E-01	33	-	-	-	0	-	-	-
HDL	T2D	-0.32	0.07	6.23E-06	186	-0.33	0.17	5.38E-02	34	-	-	-	1	-	-	-
HDL	TG	-0.74	0.03	< 2.00E-16	274	-0.64	0.14	1.74E-04	29	-0.68	0.23	1.48E-02	12	0.14	8.88E-01	19
OMA-B	HOMA-IR	0.97	0.00	< 2.00E-16	227	-	-	-	0	-	-	-	1	-	-	-
OMA-B	TG	0.43	0.07	5.22E-07	127	-	-	-	0	0.47	0.14	4.45E-02	19	-	-	-
ↄMA-IR	TG	0.48	0.06	3.14E-09	138	-	-	-	1	0.55	0.12	1.46E-02	19	-	-	-
IBD	UC	0.96	0.00	< 2.00E-16	415	0.98	0.01	0.00E+00	24	-	-	-	1	-	-	-
LDL	TC	0.97	0.00	< 2.00E-16	452	0.94	0.02	4.36E-05	10	0.86	0.04	1.23E-07	23	1.89	6.79E-02	30
LDL	TG	0.54	0.04	< 2.00E-16	231	0.07	0.19	7.25E-01	25	0.56	0.13	3.55E-02	14	-2.17	3.69E-02	36
MCH	MCHC	0.63	0.05	3.55E-15	127	0.67	0.09	1.53E-03	19	-	-	-	0	-	-	-
MCH	MCV	0.96	0.00	< 2.00E-16	320	1.00	0.00	1.30E-07	7	1.00	0.00	3.88E-08	7	-0.90	3.92E-01	9
MCH	RBC	-0.81	0.03	< 2.00E-16	207	-0.95	0.04	6.49E-09	17	-0.68	0.32	6.42E-02	8	-0.85	4.24E-01	7
MCV	RBC	-0.80	0.03	< 2.00E-16	208	-0.94	0.05	7.82E-09	18	-0.67	0.35	9.76E-02	7	-0.75	4.83E-01	6
PCV	RBC	0.96	0.00	< 2.00E-16	278	-	-	-	1	0.89	0.04	9.61E-05	12	-	-	-
TC	TG	0.61	0.04	< 2.00E-16	248	0.24	0.14	1.63E-01	36	0.76	0.08	1.79E-03	14	-3.22	2.34E-03	47

References

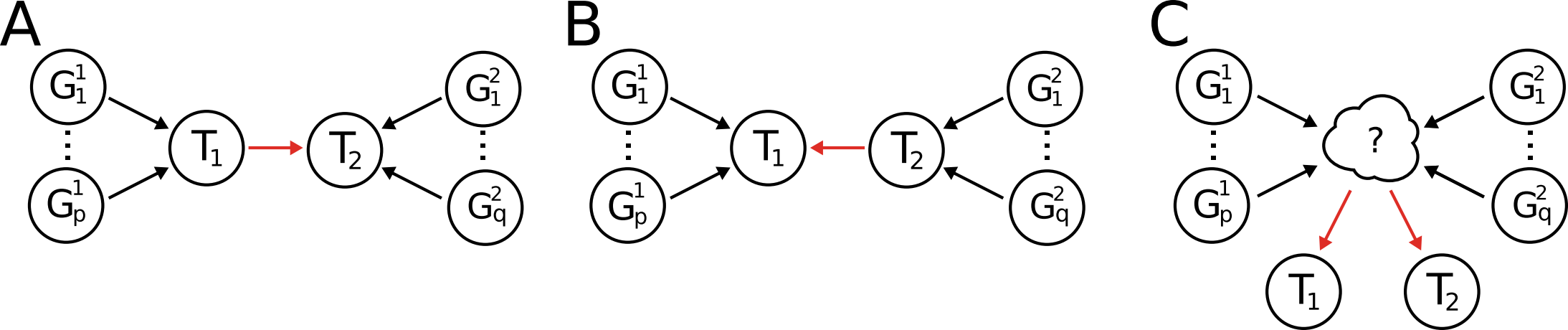
1. Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., et al. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Research 42, D1001-D1006.
2. Claussnitzer, M., Dankel, S.N., Kim, K.-H., Quon, G., Meuleman, W., Haugen, C., Glunk, V., Sousa, I.S., Beaudry, J.L., Puviindran, V., et al. (2015). FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. New England Journal of Medicine 373, 895-907.
3. Consortium, T.I.M.S.G.C.T.W.T.C.C. (2011). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 476, 214-219.
4. Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., and Cox, N.J. (2010). Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. PLoS Genet 6, e1000888.
5. Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S., et al. (2008). Genetics of gene expression and its effect on disease. Nature 452, 423-428.
6. Nica, A.C., Montgomery, S.B., Dimas, A.S., Stranger, B.E., Beazley, C., Barroso, I., and Dermitzakis, E.T. (2010). Candidate Causal Regulatory Effects by Integration of Expression QTLs with Complex Trait Genetic Associations. PLoS Genet 6, e1000895.
7. Albert, F.W., and Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. Nat Rev Genet 16, 197-212.
8. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al. (2013). The Genotype-Tissue Expression (GTEx) project. Nat Genet 45, 580-585.
9. Lappalainen, T., Sammeth, M., Friedlander, M.R., t Hoen, P.A.C., Monlong, J., Rivas, M.A., Gonzalez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature 501, 506-511.
10. Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., Consortium, G.T., Nicolae, D.L., et al. (2015). A gene-based association method for mapping traits using reference transcriptome data. Nat Genet 47, 1091-1098.
11. Gusev A, K.A., Shi H, Bhatia G, Chung W, Penninx B, Jansen R, de Geus E, Boomsma DI, Wright FA, Sullivan PF, Nikkola E, Alvarez M, Civelek M, Lusis AJ, Lehtimäki T, Raitoharju E, Kähönen M, Seppälä I, Raitakari OT, Kuusisto J, Laakso M, Price AL, Pajukanta P, Pasaniuc B. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nature Genetics.
12. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., et al. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet advance online publication.
13. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R., ReproGen, C., Psychiatric Genomics, C., Genetic Consortium for Anorexia Nervosa of the Wellcome

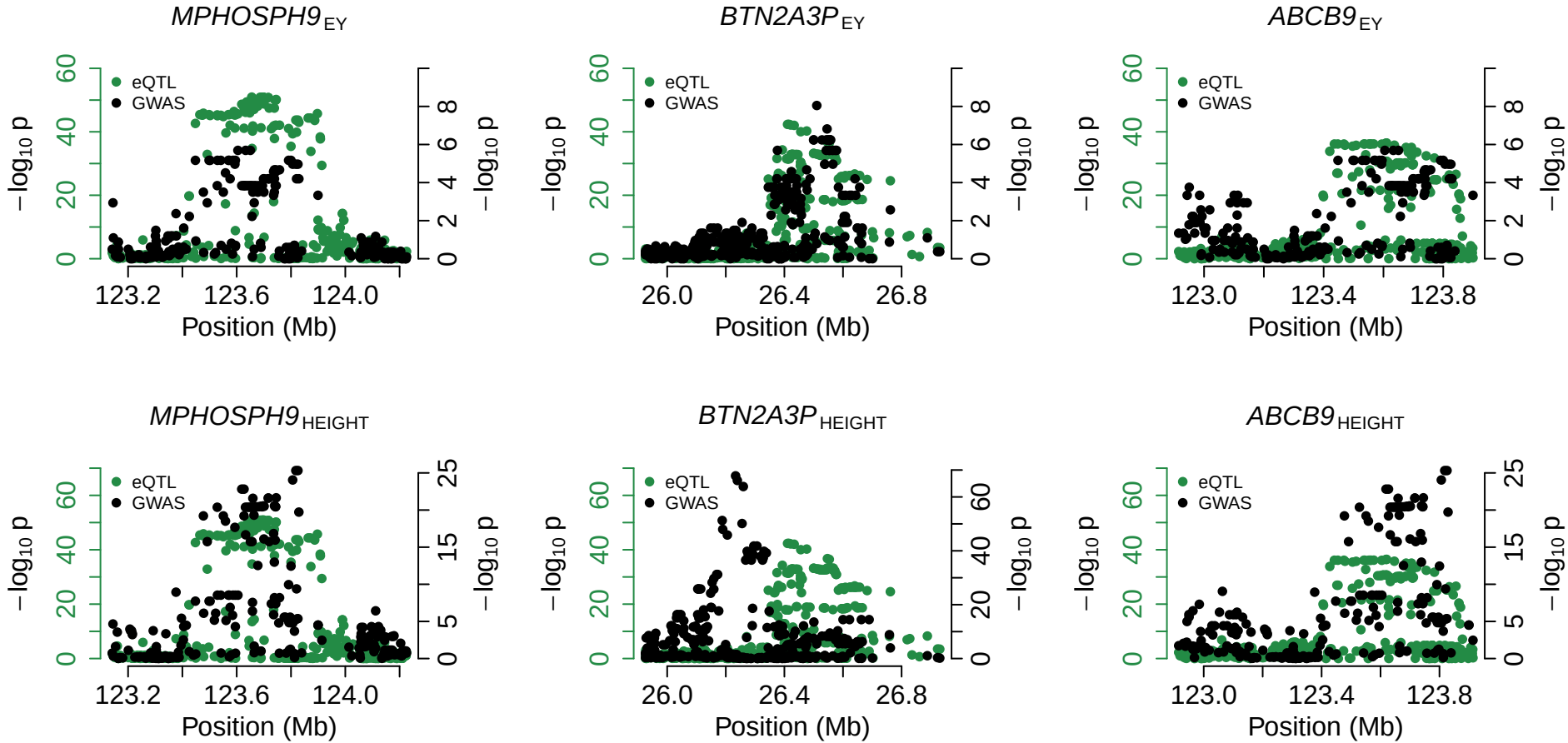
Trust Case Control, C., Duncan, L., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat Genet 47, 1236-1241.
14. Davey Smith, G., and Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Human Molecular Genetics 23, R89-R98.
15. Pickrell, J.K., Berisa, T., Liu, J.Z., Segurel, L., Tung, J.Y., and Hinds, D.A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. Nat Genet advance online publication.
16. Zheng, H., Forgetta, V., Hsu, Y., Estrada, K., RoselloDiez, A., Leo, P.J., Dahia, C.L., ParkMin, K.H., Tobias, J.H., Kooperberg, C., et al. (2015). Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. Nature 526, 112-117.
17. Morris AP, V.B., Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, Strawbridge RJ, Khan H, Grallert H, Mahajan A. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 44, 981-990.
18. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet 47, 979-986.
19. Schizophrenia Working Group of the Psychiatric Genomics, C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421-427.
20. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., et al. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 506, 376-381.
21. Perry, J.R.B., Day, F., Elks, C.E., Sulem, P., Thompson, D.J., Ferreira, T., He, C., Chasman, D.I., Esko, T., Thorleifsson, G., et al. (2014). Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. Nature 514, 92-97.
22. Rietveld, C.A., Medland, S.E., Derringer, J., Yang, J., Esko, T., Martin, N.W., Westra, H.-J., Shakhbazov, K., Abdellaoui, A., Agrawal, A., et al. (2013). GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment. Science 340, 1467-1471.
23. Global Lipids Genetics, C. (2013). Discovery and refinement of loci associated with lipid levels. Nat Genet 45, 1274-1283.
24. Soranzo, N., Sanna, S., Wheeler, E., Gieger, C., Radke, D., Dupuis, J., Bouatia-Naji, N., Langenberg, C., Prokopenko, I., Stolerman, E., et al. (2010). Common Variants at 10 Genomic Loci Influence Hemoglobin A(1C) Levels via Glycemic and Nonglycemic Pathways. Diabetes 59, 3229-3239.
25. Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A.U., Wheeler, E., Glazer, N.L., Bouatia-Naji, N., Gloyn, A.L., et al. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 42, 105-116.
26. Gieger, C., Radhakrishnan, A., Cvejic, A., Tang, W., Porcu, E., Pistis, G., Serbanovic-Canic, J., Elling, U., Goodall, A.H., Labrune, Y., et al. (2011). New gene functions in megakaryopoiesis and platelet formation. Nature 480, 201-208.
27. van der Harst, P., Zhang, W., Mateo Leach, I., Rendon, A., Verweij, N., Sehmi, J., Paul, D.S., Elling, U., Allayee, H., Li, X., et al. (2012). Seventy-five genetic loci influencing the human red blood cell. Nature 492, 369-375.

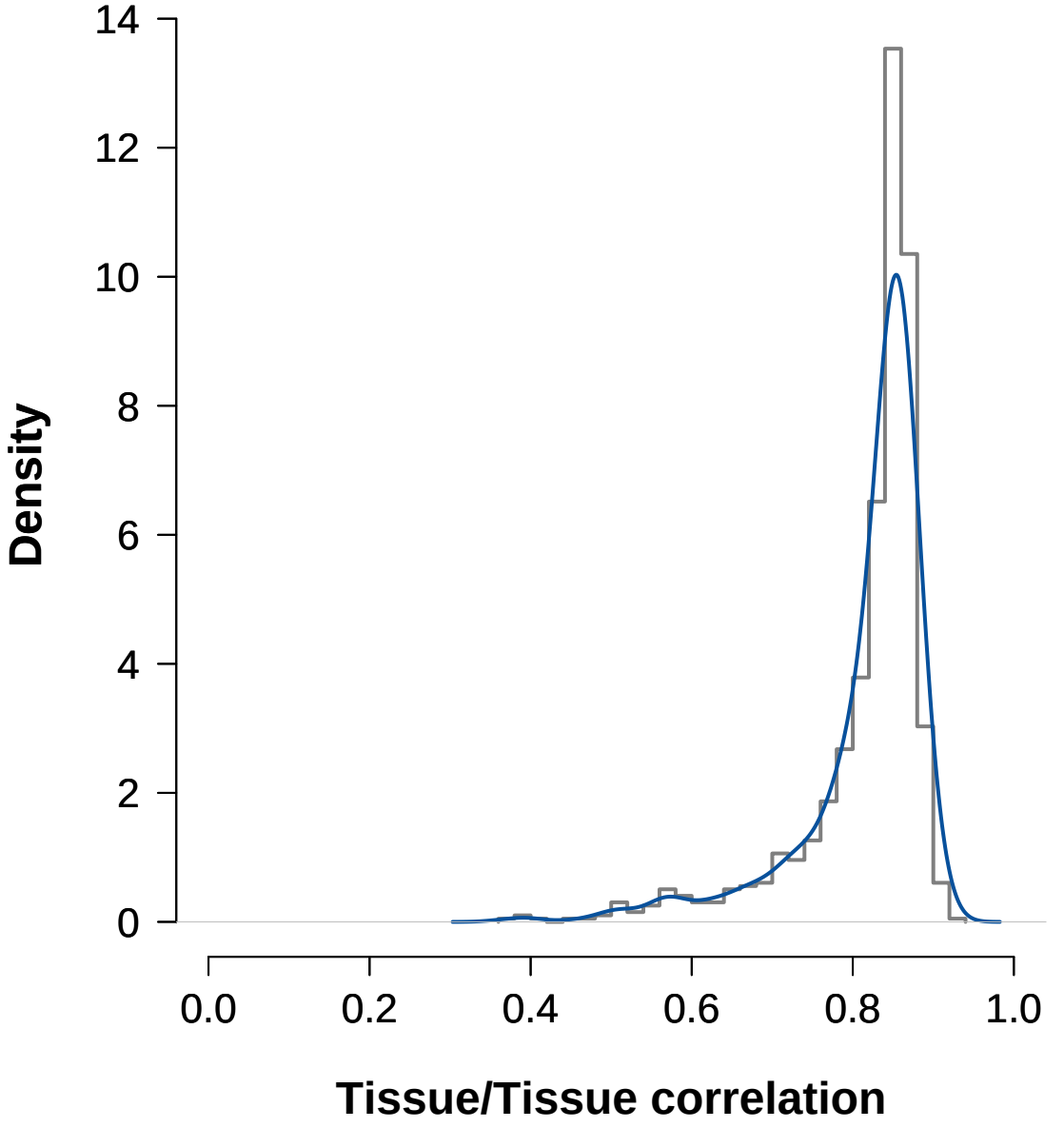
28. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. Nature 518, 197-206.
29. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J.a., Kutalik, Z., et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet 46, 1173-1186.
30. Fromer, M., Roussos, P., Sieberts, S.K., Johnson, J.S., Kavanagh, D.H., Perumal, T.M., Ruderfer, D.M., Oh, E.C., Topol, A., Shah, H.R., et al. (2016). Gene Expression Elucidates Functional Impact of Polygenic Risk for Schizophrenia. bioRxiv.
31. Stančáková, A., Civelek, M., Saleem, N.K., Soininen, P., Kangas, A.J., Cederberg, H., Paananen, J., Pihlajamäki, J., Bonnycastle, L.L., Morken, M.A., et al. (2012). Hyperglycemia and a Common Variant of GCKR Are Associated With the Levels of Eight Amino Acids in 9,369 Finnish Men. Diabetes 61, 1895-1902.
32. Stančáková, A., Javorský, M., Kuulasmaa, T., Haffner, S.M., Kuusisto, J., and Laakso, M. (2009). Changes in Insulin Sensitivity and Insulin Release in Relation to Glycemia and Glucose Tolerance in 6,414 Finnish Men. Diabetes 58, 1212-1221.
33. Wright, F.A., Sullivan, P.F., Brooks, A.I., Zou, F., Sun, W., Xia, K., Madar, V., Jansen, R., Chung, W., Zhou, Y.-H., et al. (2014). Heritability and genomics of gene expression in peripheral blood. Nat Genet 46, 430-437.
34. Nuotio, J., Oikonen, M., Magnussen, C.G., Jokinen, E., Laitinen, T., Hutri-Kähönen, N., Kähönen, M., Lehtimäki, T., Taittonen, L., Tossavainen, P., et al. (2014). Cardiovascular risk factors in 2011 and secular trends since 2007: The Cardiovascular Risk in Young Finns Study. Scandinavian Journal of Public Health 42, 563-571.
35. Raitakari, O.T., Juonala, M., Rönnemaa, T., Keltikangas-Järvinen, L., Räsänen, L., Pietikäinen, M., Hutri-Kähönen, N., Taittonen, L., Jokinen, E., Marniemi, J., et al. (2008). Cohort Profile: The Cardiovascular Risk in Young Finns Study. International Journal of Epidemiology 37, 1220-1226.
36. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. GCTA: A Tool for Genome-wide Complex Trait Analysis. The American Journal of Human Genetics 88, 76-82.
37. de los Campos, G., Vazquez, A.I., Fernando, R., Klimentidis, Y.C., and Sorensen, D. (2013). Prediction of Complex Human Traits Using the Genomic Best Linear Unbiased Predictor. PLoS Genet 9, e1003608.
38. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics, C., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47, 291-295.
39. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet 47, 1228-1235.
40. The Genomes Project, C. (2015). A global reference for human genetic variation. Nature 526, 68-74.

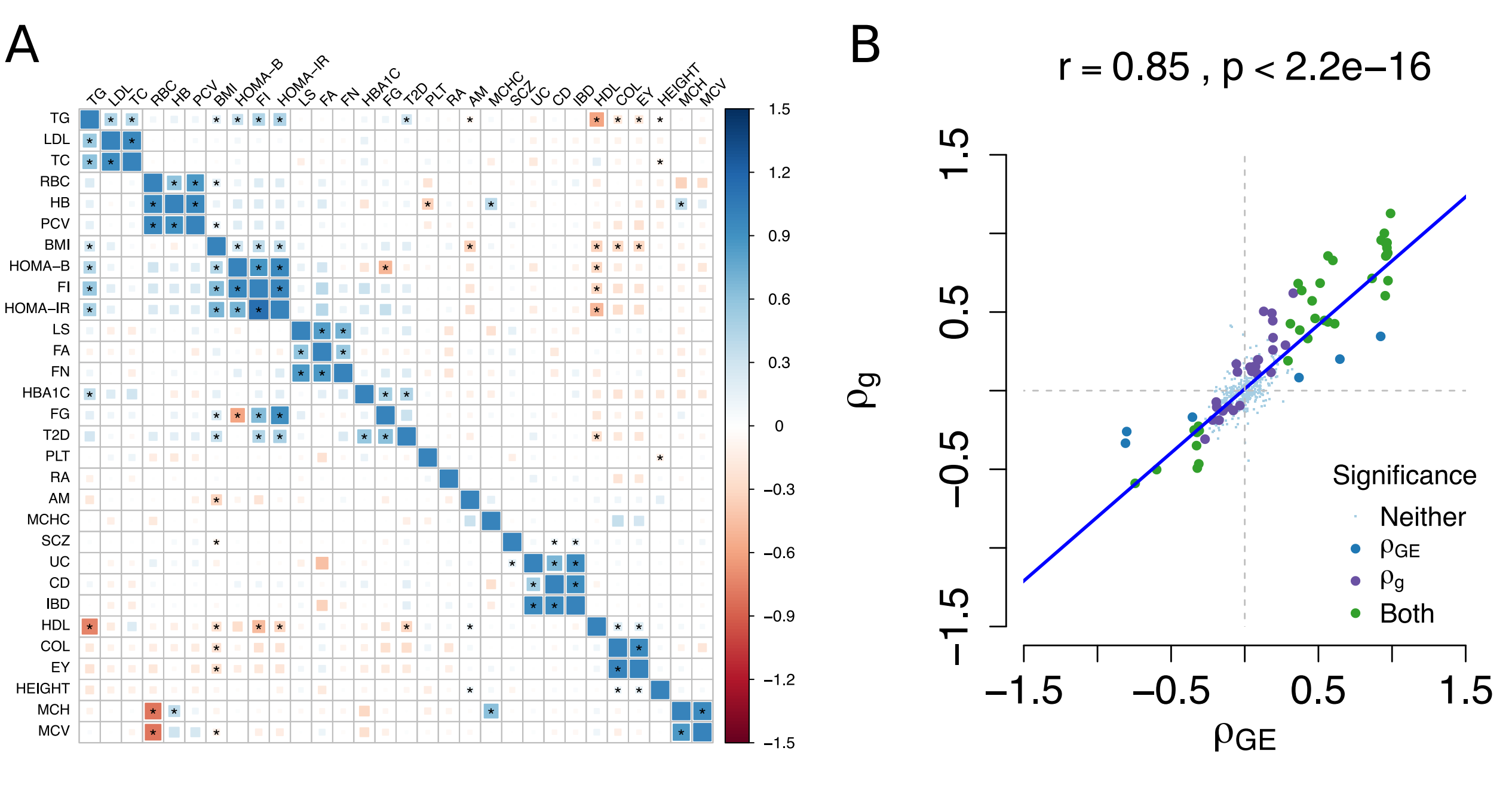
41. Shi, H., Kichaev, G., and Pasaniuc, B. (2016). Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. The American Journal of Human Genetics.
42. Shi, H.M., Nicholas; Pasaniuc, Bogdan;. (2016). Identifying genetic overlap among 30 complex traits from GWAS summary data. (In preperation).
43. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Madden, P.A.F., Heath, A.C., Martin, N.G., Montgomery, G.W., Weedon, M.N., Loos, R.J., et al. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet 44, 369-375.
44. Welch, B.L. (1947). The Generalization of 'Student's' Problem when Several Different Population Variances are Involved. Biometrika 34, 28-35.
45. Do, R., Willer, C.J., Schmidt, E.M., Sengupta, S., Gao, C., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., et al. (2013). Common variants associated with plasma triglycerides and risk for coronary artery disease. Nat Genet 45, 1345-1352.
46. Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. Nat Protocols 8, 1551-1566.
47. Kennedy, J.M., Fodil, N., Torre, S., Bongfen, S.E., Olivier, J.-F., Leung, V., Langlais, D., Meunier, C., Berghout, J., Langat, P., et al. (2014). CCDC88B is a novel regulator of maturation and effector functions of T cells during pathological inflammation. The Journal of Experimental Medicine 211, 2519-2535.
48. Pavlides, J.M.W., Zhu, Z., Gratten, J., McRae, A.F., Wray, N.R., and Yang, J. (2016). Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. Genome Medicine 8, 1-6.
49. Hosokawa, Y., Maeda, Y., Takahashi, E.-i., Suzuki, M., and Seto, M. (1999). Human Aiolos, an Ikaros-Related Zinc Finger DNA Binding Protein: cDNA Cloning, Tissue Expression Pattern, and Chromosomal Mapping. Genomics 61, 326-329.
50. Quintana, F.J., Jin, H., Burns, E.J., Nadeau, M., Yeste, A., Kumar, D., Rangachari, M., Zhu, C., Xiao, S., Seavitt, J., et al. (2012). Aiolos promotes TH17 differentiation by directly silencing Il2 expression. Nat Immunol 13, 770-777.
51. Farh, K.K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J.H., Shishkin, A.A., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature 518, 337-343.
52. Gutierrez-Arcelus, M., Ongen, H., Lappalainen, T., Montgomery, S.B., Buil, A., Yurovsky, A., Bryois, J., Padioleau, I., Romano, L., Planchon, A., et al. (2015). Tissue-Specific Effects of Genetic and Epigenetic Variation on Gene Regulation and Splicing. PLoS Genet 11, e1004958.
53. Parsons, T.J., Power, C., Logan, S., and Summerbelt, C.D. (1999). Childhood predictors of adult obesity: a systematic review. International Journal of Obesity 23.
54. Fall, T., Hägg, S., Mägi, R., Ploner, A., Fischer, K., Horikoshi, M., Sarin, A.-P., Thorleifsson, G., Ladenvall, C., Kals, M., et al. (2013). The Role of Adiposity in Cardiometabolic Traits: A Mendelian Randomization Analysis. PLoS Medicine 10, e1001474.
55. Pickrell, J. (2015). Fulfilling the promise of Mendelian randomization. bioRxiv.
56. Davey Smith, G., and Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? International Journal of Epidemiology 32, 1-22.

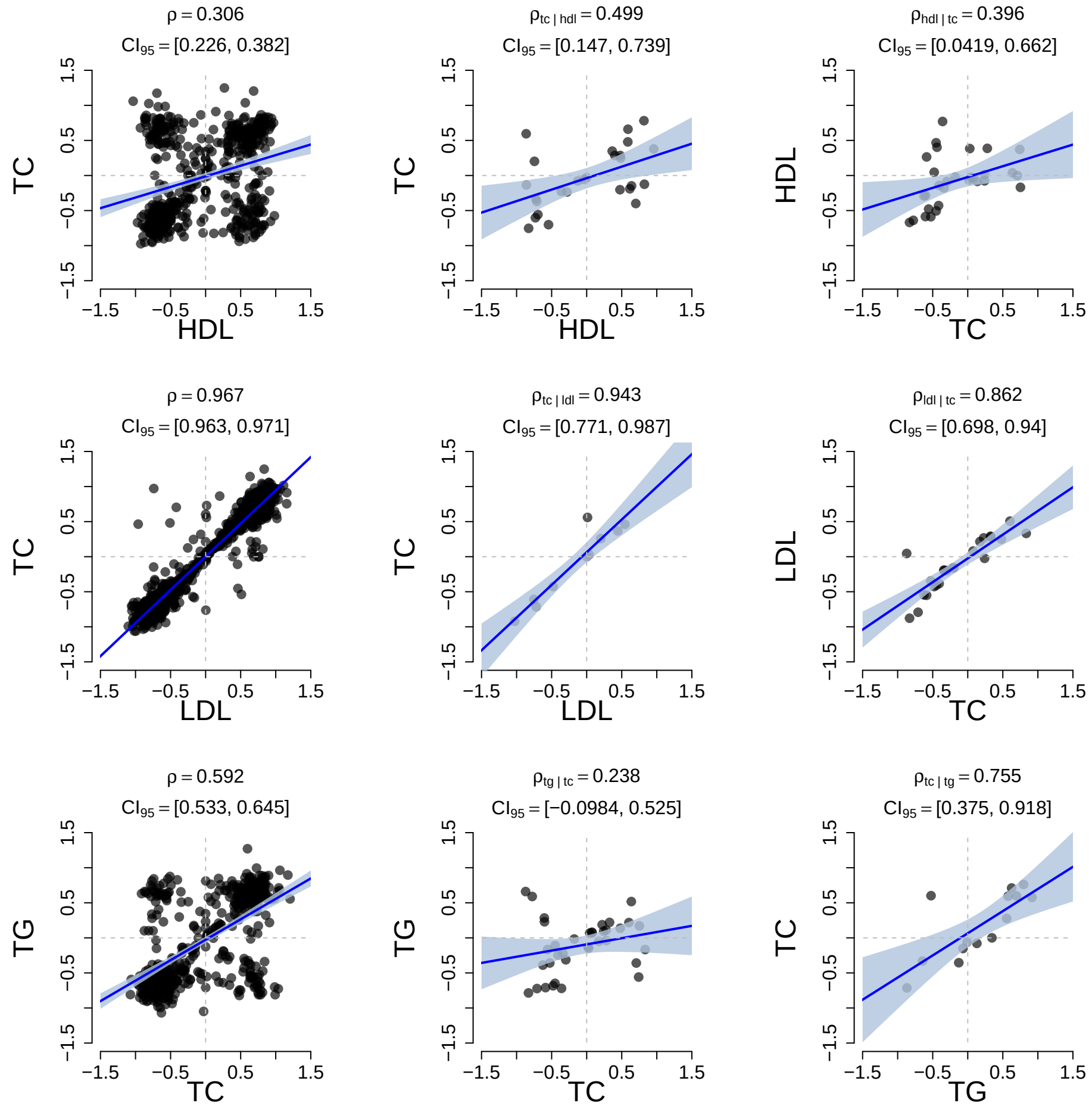
57. Wang, J., Gamazon, Eric R., Pierce, Brandon L., Stranger, Barbara E., Im, Hae K., Gibbons, Robert D., Cox, Nancy J., Nicolae, Dan L., and Chen, Lin S. (2016). Imputing Gene Expression in Uncollected Tissues Within and Beyond GTEx. The American Journal of Human Genetics 98, 697-708.

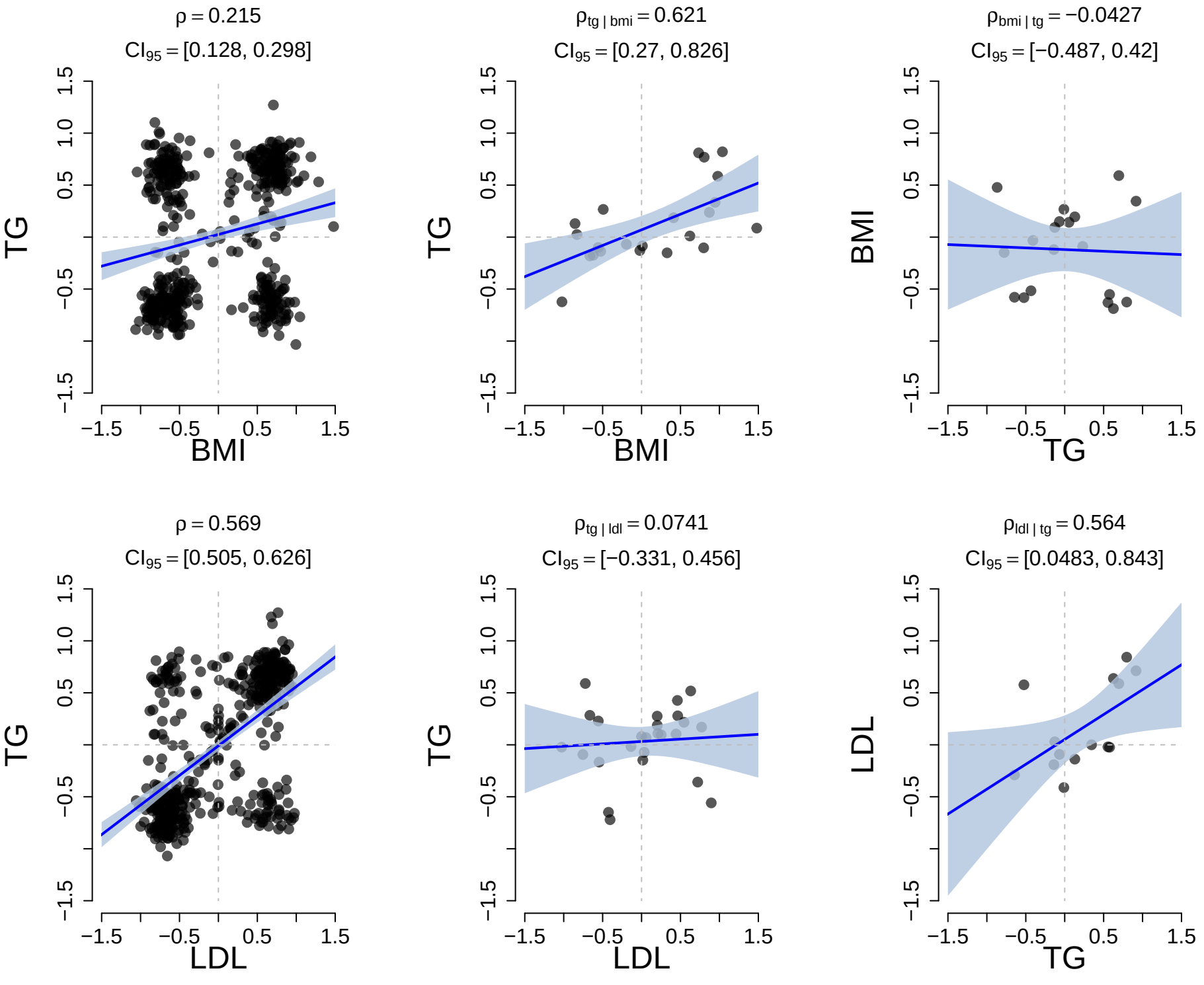

A
G_1^1
G_p^1
T_1
T_2
G_1^2
G_q^2
B
G_1^1
G_p^1
T_1
T_2
G_1^2
G_q^2
C
G_1^1
G_p^1
?
G_1^2
G_q^2
T_1
T_2


MPHOSPH9 EY
BTN2A3P EY
ABCB9 EY
MPHOSPH9 HEIGHT
BTN2A3P HEIGHT
ABCB9 HEIGHT
eQTL
GWAS
−log10 p
Position (Mb)
123.2
123.6
124.0
26.0
26.4
26.8
123.0
123.4
123.8


Density
14
12
10
8
6
4
2
0
0.0
0.2
0.4
0.6
0.8
1.0
Tissue/Tissue correlation


A
TG
LDL
TC
RBC
HB
PCV
BMI
HOMA–B
FI
HOMA–IR
LS
FA
FN
HBA1C
FG
T2D
PLT
RA
AM
MCHC
SCZ
UC
CD
IBD
HDL
COL
EY
HEIGHT
MCH
MCV
1.5
1.2
0.9
0.6
0.3
0
−0.3
−0.6
−0.9
−1.2
−1.5
B
r = 0.85 , p < 2.2e−16
ρg
ρGE
1.5
0.5
−0.5
−1.5
Significance
Neither
ρGE
ρg
Both


ρ = 0.306
CI95 = [0.226, 0.382]
TC
HDL
ρtc | hdl = 0.499
CI95 = [0.147, 0.739]
TC
HDL
ρhdl | tc = 0.396
CI95 = [0.0419, 0.662]
HDL
TC
ρ = 0.967
CI95 = [0.963, 0.971]
TC
LDL
ρtc | ldl = 0.943
CI95 = [0.771, 0.987]
TC
LDL
ρldl | tc = 0.862
CI95 = [0.698, 0.94]
LDL
TC
ρ = 0.592
CI95 = [0.533, 0.645]
TG
TC
ρtg | tc = 0.238
CI95 = [−0.0984, 0.525]
TG
TC
ρtc | tg = 0.755
CI95 = [0.375, 0.918]
TC
TG
−1.5
−0.5
0.5
1.5


ρ = 0.215
CI95 = [0.128, 0.298]
TG
BMI
ρtg | bmi = 0.621
CI95 = [0.27, 0.826]
TG
BMI
ρbmi | tg = −0.0427
CI95 = [−0.487, 0.42]
BMI
TG
ρ = 0.569
CI95 = [0.505, 0.626]
TG
LDL
ρtg | ldl = 0.0741
CI95 = [−0.331, 0.456]
TG
LDL
ρldl | tg = 0.564
CI95 = [0.0483, 0.843]
LDL
TG
−1.5
−0.5
0.5
1.5
1.0