Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression

Vosa, U.; Claringbould, A.; Westra, H.J.; Bonder, M.J.; Deelen, P.; Zeng, B.; ... ; i2QTL Consortium

ARTICLES
https://doi.org/10.1038/s41588-021-00913-z

nature genetics

Large-scale *cis*- and *trans*-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression

Trait-associated genetic variants affect complex phenotypes primarily via regulatory mechanisms on the transcriptome. To investigate the genetics of gene expression, we performed *cis*- and *trans*-expression quantitative trait locus (eQTL) analyses using blood-derived expression from 31,684 individuals through the eQTLGen Consortium. We detected *cis*-eQTL for 88% of genes, and these were replicable in numerous tissues. Distal *trans*-eQTL (detected for 37% of 10,317 trait-associated variants tested) showed lower replication rates, partially due to low replication power and confounding by cell type composition. However, replication analyses in single-cell RNA-seq data prioritized intracellular *trans*-eQTL. *Trans*-eQTL exerted their effects via several mechanisms, primarily through regulation by transcription factors. Expression of 13% of the genes correlated with polygenic scores for 1,263 phenotypes, pinpointing potential drivers for those traits. In summary, this work represents a large eQTL resource, and its results serve as a starting point for in-depth interpretation of complex phenotypes.

eQTL have become a common tool to interpret regulatory mechanisms of variants identified by genome-wide association studies (GWASs). In particular, *cis*-eQTL, for which gene expression levels are affected by a gene-proximal (<1 Mb) SNP, have been widely used for this purpose. However, *cis*-eQTL generally explain only a modest proportion of disease heritability[1], suggesting additional routes to disease.

Trans-eQTL, for which the SNP is located distal to the gene (>5 Mb) or on another chromosome, usually have smaller effect sizes than *cis*-eQTL and thus require larger sample sizes for detection. However, *trans*-eQTL could be relevant for complex traits because, compared to stronger *cis*-eQTL effects, each individual *trans* effect is less likely to be dampened by compensatory post-transcriptional buffering or removed from the population by negative selection[2,3]. Indeed, genes regulated by weak eQTL effects are estimated to have more impact on the phenotype as compared to those regulated by strong eQTL effects[1]. Individual *trans*-eQTL SNPs can affect many genes and have a widespread impact on regulatory networks. Consequently, weak *trans*-eQTL have the potential to identify trait-relevant genes[4–10] and have already been used to prioritize genes that are likely to contribute to disease[4].

While *trans*-eQTL are useful for the identification of distal effects of a single variant, a different approach is required to determine the combined consequences of all variants associated with a polygenic trait. Polygenic scores (PGSs) summarize genome-wide combined risk for a complex disease into a single metric that may be used to stratify individuals into groups[11,12]. The recently proposed omnigenic model[13,14] postulates that the heritability of most complex traits is dominated by numerous weak *trans* effects and hypothesizes that those effects converge on a smaller set of trait-relevant 'core' genes. This suggests that associations between PGSs and gene expression (expression quantitative trait scores, eQTSs) could help to prioritize putative trait-relevant genes (Supplementary Note and Liu et al.[14]). While it remains unclear what fraction of the genome affects complex traits, we here systematically investigated *trans*-eQTL and eQTS to determine how genetic effects regulate genes and pathways and whether these effects could be informative about the biology of the respective traits.

To maximize statistical power to detect eQTL and eQTSs, we performed a large-scale meta-analysis on up to 31,684 blood samples from 37 eQTLGen Consortium cohorts. This allowed us to identify *cis*-eQTL for 16,987 genes, *trans*-eQTL for 6,298 genes and eQTS effects for 2,568 genes (false discovery rate (FDR) < 0.05, determined by permutations; Methods and Fig. 1). We replicated these eQTL across gene expression platforms, in other tissues and in single-cell (sc)RNA-seq data. While the overall concordance was good, formal replication remained limited, possibly due to the effects of genetics on blood cell composition, the limited sample size of the available replication datasets and the cell type-specific nature of distal effects. To demonstrate the utility of our resource, we combined the associations with additional data layers to gain biological insights into the mechanisms of blood eQTL and complex traits.

Results

Meta-analyses on local and distal gene expression. We performed *cis*-eQTL, *trans*-eQTL and eQTS meta-analyses using eQTLGen Consortium data from 31,684 individuals (Fig. 1a, Supplementary Table 1 and Supplementary Note). Because the consortium contained both array- and sequencing-based expression datasets, we integrated different platforms using co-regulation patterns between genes (Methods). Inter-platform *cis*-eQTL, *trans*-eQTL and eQTS replication analyses indicated good concordance between platforms (on average 93.2%, 99.2% and 99.4% for significant *cis*, *trans* and eQTS effects), enabling integrated gene-level meta-analyses. These analyses also demonstrated that effects identified by our approach replicate between different blood datasets (Methods, Supplementary Note and Supplementary Fig. 1a–c). We adopted a permutation-based strategy[4,15,16] to account for multiple testing in discovery meta-analyses (Methods and Supplementary Note), which was more conservative than a Benjamini–Hochberg FDR[17] and less stringent than the Bonferroni method (Supplementary Fig. 2).

In all analyses, we accounted for unknown technical confounders (such as batch effects) and biological confounders (such as inter-individual differences in cell type composition) by correcting expression data for up to 25 expression principal components (PCs) that were not associated with genetics (Methods). This correction

A full list of authors and their affiliations appears at the end of the paper.

adjusted for the majority of cell type-composition effects in a subset of samples (Biobank-based integrative omics study (BIOS) cohort, up to $N=3{,}831$; Supplementary Note and Supplementary Fig. 3). Nevertheless, we acknowledge that our dataset may include residual cell type-composition effects.

Local genetic effects on blood gene expression. We identified *cis*-eQTL (SNP–gene distance <1 Mb, FDR < 0.05; Methods) for 16,987 genes (88.2% of the 19,250 autosomal genes expressed in blood and tested in the *cis*-eQTL analysis; Fig. 1b).

After we observed that *cis*-eQTL replicated between whole-blood datasets (Supplementary Fig. 1a), we investigated the replicability of *cis*-eQTL in other tissues. In 47 post-mortem tissues (Genotype–Tissue Expression (GTEx) data)[18], we observed an average replication rate of 14.8% (discovery analysis without GTEx, Benjamini–Hochberg FDR < 0.05 in GTEx data; median across tissues of 15.0%, range of 3.6–29.6% when excluding whole-blood data) and, on average, a 94.9% concordance in allelic directions (median, 95.2%; range, 86.7–99.2% when excluding whole-blood data) among the *cis*-eQTL for which the lead SNP effect replicated in GTEx data (Supplementary Note, Extended Data Fig. 1 and Supplementary Data 1).

Genes highly expressed in blood but without detectable *cis*-eQTL were more likely to be intolerant to loss-of-function mutations in their coding region[19] (two-sided Wilcoxon rank-sum test, $P=6.5\times10^{-6}$; Fig. 2a and Supplementary Table 2), suggesting that eQTL on such genes are selectively constrained, as has been recently proposed[20].

Ninety-two percent of lead *cis*-eQTL SNPs were located within 100 kb of the gene (Fig. 2d), and stronger *cis*-eQTL typically had a smaller distance between the SNP and gene (within 20 kb for 84.1% of the top 20% strongest eQTL).

Lead *cis*-eQTL SNPs that were located >100 kb from the transcription start site (TSS) or transcription end site (TES) of the gene were likely to overlap with capture Hi-C (CHi-C) contacts (2.0-fold enrichment compared to when the location of the Hi-C target was flipped relative to the TSS; $P<3.3\times10^{-12}$; two-tailed two-sample test of equal proportions; Methods, Fig. 2e and Supplementary Note). This suggests that some long-range *cis*-eQTL are caused by physical interactions between genomic regions of the SNP and the gene. For example, a CHi-C contact for *IRS1* overlapped the lead eQTL SNP, mapping 630 kb downstream from *IRS1* (Fig. 2f). Similarly, we observed an enriched overlap with Hi-C contacts for short-range *cis*-eQTL effects (<100 kb, 1.3-fold; $P<9.1\times10^{-16}$; two-tailed two-sample test of equal proportions; Fig. 2e and Supplementary Note).

When comparing our results to the 5,440 protein-coding *cis*-eQTL genes previously identified in 5,311 samples[4], lead SNPs typically mapped closer to the *cis*-eQTL gene (Supplementary Fig. 4). In GWASs studies, larger sample sizes and denser imputation panels increase the resolution of signals in associated loci, especially for weaker effects. Additionally, GWAS simulations have indicated that lead GWAS signals map near the causal variant (within 33.5 kb in 80% of cases)[21]. Because the majority of lead *cis*-eQTL variants from our study map within 100 kb of the TSS and TES, we consider it likely that causal variants for gene expression are also within these regions.

One-third of trait-associated variants have distal effects. We focused on 10,317 trait-associated SNPs (GWAS $P\leq5\times10^{-8}$; Methods and Supplementary Table 3) and identified 59,786 *trans*-eQTL (SNP–gene distance >5 Mb; $P<8.3\times10^{-6}$, corresponding to FDR < 0.05; Supplementary Data 2 and Extended Data Fig. 2), representing 3,853 SNPs (37% of tested GWAS SNPs) and 6,298 genes (32% of tested genes; Fig. 1c). The largest previous *trans*-eQTL meta-analysis in blood[4] ($N=5{,}311$) identified *trans*-eQTL for 8% of tested SNPs, indicating that a larger sample size is beneficial for identification of distal effects. Similar to *cis*-eQTL, highly expressed genes without detectable *trans*-eQTL effects were more likely to be intolerant to loss-of-function variants (two-sided Wilcoxon rank-sum test, $P=6.4\times10^{-7}$; Fig. 2b), suggesting constrained expression of these genes.

While blood cell-composition SNPs[22] comprised 21% of tested SNPs, they represented the majority (64%) of *trans*-eQTL SNPs. Many of the identified *trans*-eQTL SNPs may regulate the abundance of a specific blood cell type and could result in *trans*-eQTL effects on genes specifically expressed in that cell type. Although we corrected the individual expression datasets for cell type-composition effects using PCs (Methods and Supplementary Note), the fact that numerous *trans*-eQTL emanate from known blood cell-composition SNPs suggested a residual effect of cell composition.

To prioritize *trans*-eQTL caused by intracellular mechanisms (that is, gene regulation within cells), we applied several analytic strategies (Supplementary Note). Replication in purified cell types and cell lines identified 4,018 (6.7%) *trans*-eQTL that replicated in at least one dataset or were supported by DNA methylation quantitative trait locus (mQTL) data (Benjamini–Hochberg FDR < 0.05; 93.3% average allelic concordance; Supplementary Note, Supplementary Fig. 5 and Supplementary Data 2). The replication rate in tissues from the GTEx project[23] was very low (discovery without GTEx, 0.07% of *trans*-eQTL replicated in any non-blood tissue, 0.09% in blood, Benjamini–Hochberg FDR < 0.05), but the allelic concordance of significant effects was, on average, 66% in non-blood tissues and 100% in blood (Supplementary Data 3). Despite these low replication rates, *trans*-eQTL showed an inflation of replication signal in the majority of tissues (Supplementary Fig. 6a), most notably in whole blood, the esophagus muscularis, liver, the heart atrial appendage and skin unexposed to the sun.

Compared to bulk datasets, scRNA-seq eQTL datasets are less impacted by cell composition and are therefore ideal for *trans*-eQTL replication. We performed replication meta-analyses in B cells, $CD4^+$ T cells, $CD8^+$ T cells, classical monocytes, non-classical monocytes, dendritic cells, natural killer cells and plasma cells from up to 1,139 individuals (cohort profiling 1,000 cells from 1,000 individuals (OneK1K), final $N=982$ and cohort profiling 1 million single blood cells from the Netherlands (1M-scBloodNL), $N=157$; Supplementary Note). Depending on the cell type, we could reliably test between 1,917 and 27,582 discovery *trans*-eQTL (Fig. 3a) and replicated 35 *trans*-eQTL at FDR < 0.05 (Supplementary Table 4), with two effects replicating in more than one cell type. For those *trans*-eQTL, the allelic concordance between the discovery and the replication analysis was 97%. For seven of the eight cell types, we observed inflation of replication signal (Supplementary Table 5 and Supplementary Fig. 6a) and greater than expected allelic concordance with the discovery analysis (Fig. 3a and Supplementary Table 5; two-sided binomial test, $P<0.05$). Similarly, *trans*-eQTL effect sizes correlated significantly with replication effects in the scRNA-seq data (r_b metric[24]; Methods and Fig. 3a; two-sided $P<0.05$) for four cell types (classical monocytes ($P=3.36\times10^{-8}$, $r_b=0.514$, standard error (SE) = 0.093), natural killer cells ($P=3.24\times10^{-4}$, $r_b=0.185$, SE = 0.051), $CD8^+$ lymphocytes ($P=3.41\times10^{-3}$, $r_b=0.454$, SE = 0.155) and B cells ($P=5.98\times10^{-3}$, $r_b=0.049$, SE = 0.018)). Data from more abundant cell types showed higher correlations with those from whole blood (Fig. 3a, Pearson $R^2=0.53$, two-sided $P=0.04$), and we observed similar correlations for replication datasets from several purified cell types (Supplementary Fig. 7). When confining the analysis to 729 *trans*-eQTL with an absolute average $Z>1.96$ over all types (corresponding to a nominal $P<0.05$, Supplementary Table 4), we observed a relatively high effect direction concordance of 84% (Fig. 3a and Supplementary Table 5, two-sided binomial test; $P=1.25\times10^{-84}$).

These results suggest that some of the *trans*-eQTL identified in blood are present within individual cells, although it remains challenging to prioritize individual effects. *Trans*-eQTL effect sizes

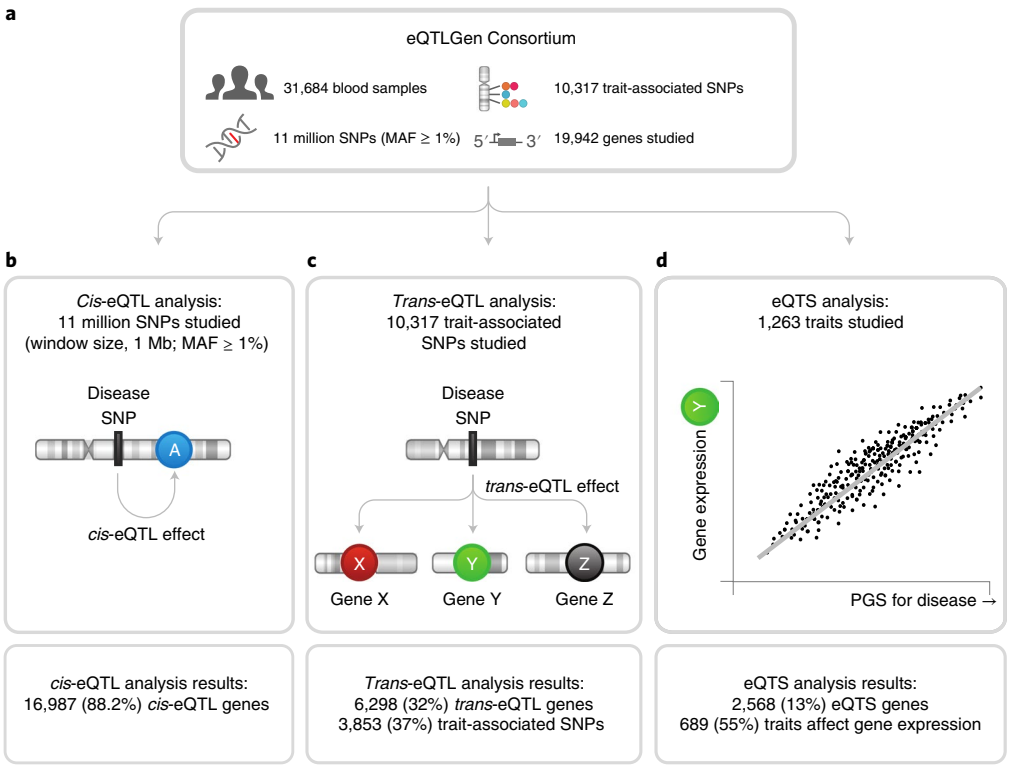


Fig. 1 | Overview of the study. a–d, Overview of discovery analyses and their results.

were generally small (median $r = 0.033$; Supplementary Fig. 8a,e,i and Supplementary Note). Considering the limited sample sizes of available bulk and scRNA-seq *trans*-eQTL datasets ($N = 388$–$1{,}480$; $N = 2{,}905$ for methylation QTL data, all <10% of discovery N), the statistical power to replicate these effects was low (Supplementary Fig. 8g,h), limiting our ability to reliably distinguish cell type-composition effects from intracellular effects. Therefore, we used all *trans*-eQTL in interpretive analyses (Supplementary Note).

To evaluate whether *trans*-eQTL could be explained by direct or indirect transcription factor (TF) action[25], protein–protein interactions (PPIs)[26] or co-regulation between *cis*- and *trans*-eQTL genes, we conducted enrichment analyses (Supplementary Note, Supplementary Fig. 9 and Fig. 3b). *Cis*- and *trans*-eQTL genes emerging from the same SNP were 1.28-fold enriched by TF–target pairs, compared to all other gene pairs ($P = 4.0 \times 10^{-21}$; two-sided Fisher's exact test; Supplementary Fig. 8). When we included additional local genes into *trans*-eQTL loci with Pascal[27], (Supplementary Note and Supplementary Fig. 10) we observed a 1.40-fold enrichment for TFs ($P = 5.6 \times 10^{-36}$; two-sided Fisher's exact test; Fig. 3b). There was also enrichment for genes co-regulated with known TFs (1.38-fold, $P = 5.8 \times 10^{-72}$; two-sided Fisher's exact test; Fig. 3b), genes co-regulated with known target genes (3.57-fold, $P < 1.0 \times 10^{-308}$; two-sided Fisher's exact test; Fig. 3b) and genes co-regulated with both (4.37-fold, $P < 1.0 \times 10^{-308}$; two-sided Fisher's exact test; Fig. 3b), suggesting indirect consequences of transcriptional regulation. We observed a strong 22.3-fold enrichment ($P < 1.0 \times 10^{-308}$; two-sided Fisher's exact test) of co-regulated gene pairs and a 1.45-fold enrichment of PPI[26] pairs ($P = 3.5 \times 10^{-17}$; two-sided Fisher's exact test), including co-regulated subunits of the same protein complex (for example, products of *CPSF1* and *CPSF7*) and receptor–ligand pairs (for example, products of *CSF3* and *CSF3R*). We note that cell type-composition effects likely contribute to the enrichment of co-regulated gene pairs, as there was a depletion of co-regulation among 729 effects nominally replicating in scRNA-seq data (odds ratio (OR) = 0.5, $P = 0.015$; two-sided Fisher's exact test). Additionally, Hi-C chromatin contacts[28] were also enriched for local–distal gene pairs (OR = 1.47, $P = 2.4 \times 10^{-153}$; two-sided Fisher's exact test), suggesting that some *trans*-eQTL could be driven by physical contact (Supplementary Fig. 9). Altogether, 29,207 (49%) of the reported *trans*-eQTL could be assigned a putative biological mechanism (Fig. 3c and Supplementary Data 4). However, the *trans*-eQTL analysis was limited to trait-associated variants; therefore the observed enrichments might not generalize to all *trans*-eQTL.

We identified 1,050 (10.2%) hub SNPs that regulated the expression of more than ten genes (Supplementary Table 6). Of these, 196 (18.6%) had a global upregulating or downregulating effect on downstream genes (two-sided binomial test, Bonferroni-corrected $P < 0.05$, Supplementary Table 6). We identified 507 (48%) hub SNPs showing enrichment for TF- or micro (mi)RNA-binding sites (one-sided Fisher's exact test, Benjamini–Hochberg FDR < 0.05; Supplementary Table 7). For nine of these (five independent loci), we observed that the respective TF was encoded by a gene positioned <1 Mb from the hub SNP, suggesting a TF-mediated mechanism. For example, rs17087335 (ref. [29]) (SNP database (dbSNP) version 137) affects the expression of 88 neuronal genes, possibly through the neuronal repressor encoded by the nearby gene *REST* (Fig. 4 and Supplementary Note).

We also identified 47 GWAS traits for which at least four independent variants affected the same gene in *trans* (Supplementary Tables 8–10) 3.4 times higher than expected by chance ($P = 0.001$; two-tailed two-sample test of equal proportions). For systemic lupus erythematosus (SLE)[30], the expression of *IFIT1*, *IFI44L*, *HERC5*, *IFI6*, *IFI44*, *RSAD2*, *MX1*, *ISG15*, *ANKRD55*, *OAS3*, *OAS2*, *OASL* and *EPSTI1* was affected by at least three SLE-associated genetic variants (FDR < 0.05). These genes are nearly all known to be in the SLE interferon signature[31–33] (Supplementary Table 11), reflecting the involvement of interferon signaling in SLE pathophysiology

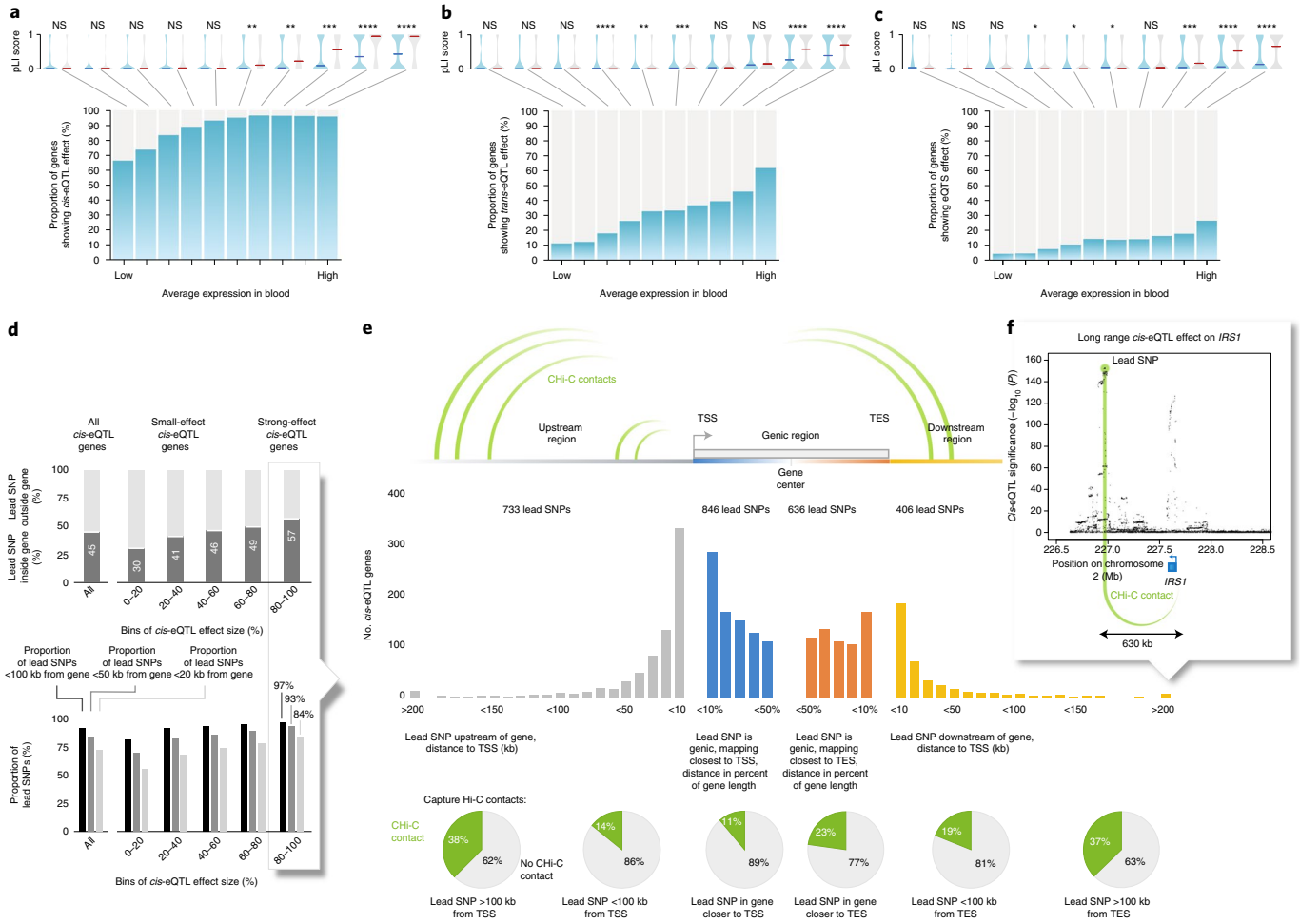


Fig. 2 | Results of *cis*- and *trans*-eQTL analyses. All genes tested in the *cis*-eQTL analysis (**a**), the *trans*-eQTL analysis (**b**) and the eQTS analysis (**c**) were divided into ten bins based on their average expression levels in blood (BIOS cohort). Highly expressed genes without any eQTL effect (gray bars) were less tolerant to loss-of-function variants (two-sided Wilcoxon rank-sum test on probability of being loss-of-function-intolerant (pLI) scores). Median pLI scores per bin are indicated. NS, not significant, $P > 0.05$; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 1 \times 10^{-4}$. **d**, Genes with strong effect sizes are more likely to have a lead SNP located within (top) or close to the gene (bottom). **e**, Lead *cis*-eQTL SNPs overlap CHi-C contacts with TSSs. **f**, Example of the *IRS1* locus.

(Fig. 5). While the *trans*-eQTL analysis identified only one new interferon signature gene, it helped to pinpoint SLE GWAS loci that collectively affect interferon signature genes.

Despite these results, most blood *trans*-eQTL remain unexplained. To aid interpretation, we provide the results of per-phenotype gene ontology-term-enrichment analysis (Supplementary Note and Supplementary Table 12). In the Supplementary Note, we also highlight additional examples of *trans*-eQTL variants associated with age of menarche[34] (*ZNF131* locus), lipid levels[35] (*FADS1–FADS2* locus), inflammatory bowel disease[36] and SLE[37] (*IFIH1* locus), asthma[38] (*GSDMB* locus) and height[39] (*CLOCK* locus) and explore their potential biological mechanisms to show how *trans*-eQTL results can be used to generate hypotheses for further research (Supplementary Fig. 11a–e).

eQTSs identify potential driver genes for polygenic traits. To ascertain the coordinated effects of trait-associated variants on gene expression, we used GWAS summary statistics to calculate PGSs for 1,263 traits in 28,158 samples (Methods and Supplementary Table 13). We reasoned that, when the PGS for a trait correlates with the expression of a gene, *trans*-eQTL effects of individual risk variants (Fig. 6a) converge on that gene, and it can be prioritized as a putative driver of the disease (Fig. 6b).

We identified 18,210 eQTSs (FDR < 0.05) representing 689 unique traits (55% of tested traits) and 2,568 genes (13% of tested genes; Supplementary Data 5 and Fig. 1d). Of these genes, 719 (28%) were not identified in the *trans*-eQTL analysis, emphasizing the value of analyzing eQTSs in addition to *trans*-eQTL (Fig. 6a,b). Median eQTS effect sizes were smaller than those for *cis*-eQTL and similar to those for *trans*-eQTL (median $r = 0.037$; Supplementary Fig. 8a,e,i).

Ten eQTSs replicated in LCL cells (Benjamini–Hochberg FDR < 0.05), and nine also had the same effect direction as that in the discovery dataset (Supplementary Fig. 12a and Supplementary Data 5), while 78 replicated in induced pluripotent stem cells (iPSCs; Benjamini–Hochberg FDR < 0.05) with 71 (91%) showing the same direction (Supplementary Fig. 12b and Supplementary Data 5). We also identified 19 replicating eQTSs (Benjamini–Hochberg FDR < 0.05, same effect direction) in the European subset of GTEx samples (Supplementary Note and Supplementary Data 6) and observed an inflation of replication signal in some tissues, primarily in blood (Supplementary Fig. 6b).

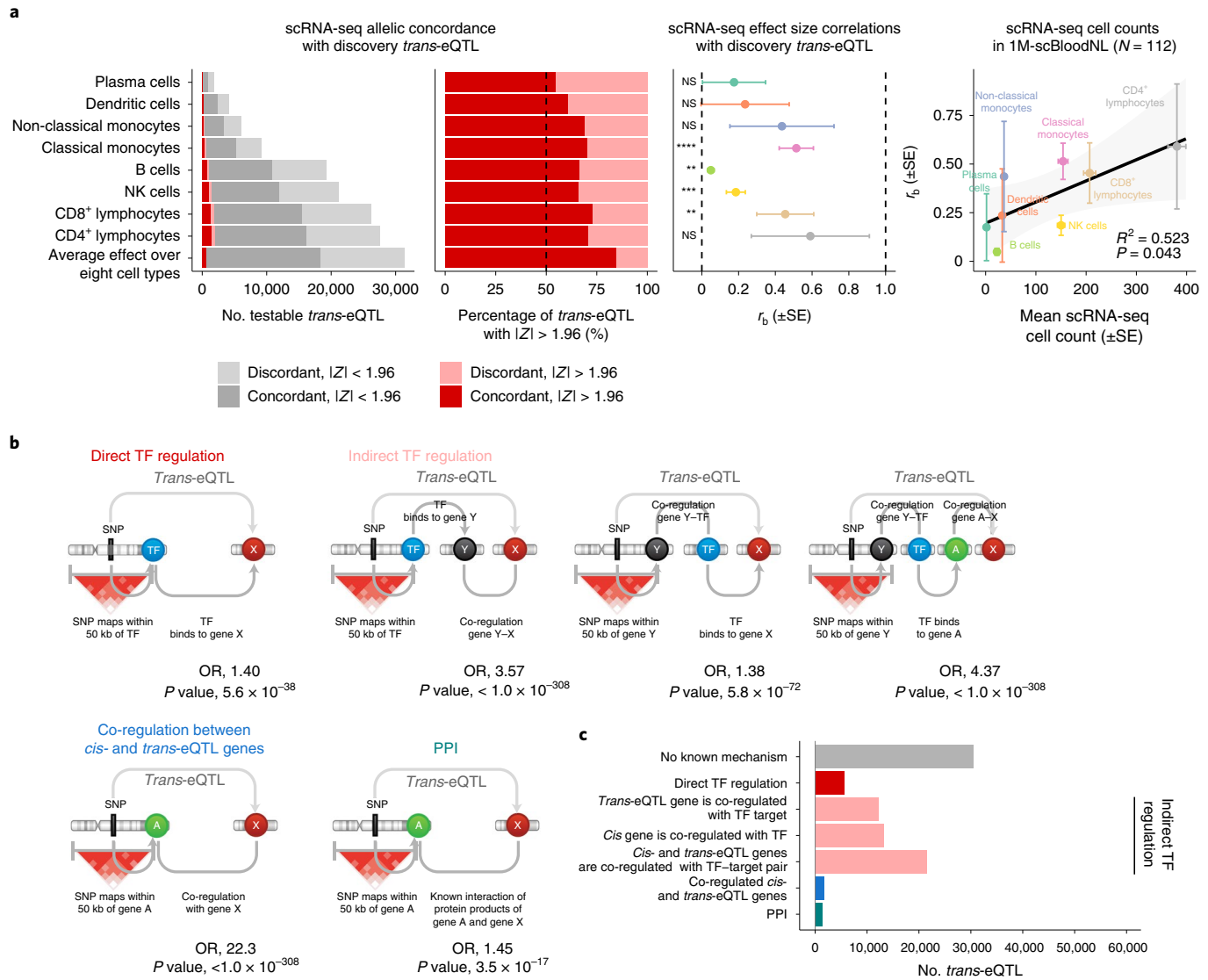


Fig. 3 | *Trans*-eQTL replication in scRNA-seq data and mechanisms leading to *trans*-eQTL. a, Replication analyses in scRNA-seq data from 8 cell types in up to 1,139 individuals. Far left and left, allelic concordances relative to *trans*-eQTL effect direction in the discovery *trans*-eQTL analysis. Middle, correlation estimates (r_b) of *trans*-eQTL effects between the discovery analysis in blood and scRNA-seq blood cell types and corresponding two-sided *P* values (Methods). NS, $P > 0.05$; $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 1 \times 10^{-4}$. Dots indicate r_b and error bars indicate SE for r_b. Right, correlation between cell type counts (mean over the subset of samples from the 1M-scBloodNL cohort; $N = 112$) and r_b estimates. The squared Pearson correlation coefficient and the two-sided *P* value from the Pearson correlation test are shown. Error bars indicate SE for r_b and s.e.m. for cell counts. NK, natural killer. **b**, Enrichment analyses for TF binding, gene co-regulation and PPIs. *Cis*-acting genes were determined by *cis*-eQTL or assigned by the Pascal method (Methods and Supplementary Note). ORs and two-sided *P* values from Fisher's exact test are shown. **c**, All 59,786 *trans*-eQTL stratified by putative mechanism of action. Hi-C enrichment results are not shown as we only observed enrichment when using a lenient threshold for Hi-C contacts (>0 value for contact). Full results are shown in Supplementary Fig. 9.

Most eQTS associations (72.8%) represented blood cell traits (Extended Data Fig. 3 and Supplementary Data 5). For instance, the PGS for mean corpuscular volume[40] correlated positively with the expression of genes specifically expressed in erythrocytes, for example, genes encoding hemoglobin subunits (*HBG1* and *HBG2*, both FDR < 0.05). eQTS genes were most enriched for gene ontology terms involved in cellular secretion, blood cell traits and intercellular signaling (Supplementary Table 14).

Because there was no strong replication signal in non-blood tissues and the majority of eQTSs were observed for blood-related traits, we speculated that these effects were highly tissue specific or cell type specific and that eQTS analysis would yield the most informative results if conducted in the trait-relevant tissue. However, power analyses suggest that the limited replication in other tissues could also be the result of a lack in statistical power explained by small effect sizes of eQTSs (Supplementary Fig. 8i) and moderately sized replication datasets.

Still, in our blood data, we also identified eQTSs for non-blood PGSs, including metabolite and lipid levels, anthropometric traits and diseases such as asthma, celiac disease and coronary artery disease (Supplementary Note, Supplementary Fig. 13a–c and Supplementary Data 5). For example, 11 of the 26 eQTS genes that were associated with the PGS for high-density lipoprotein (HDL[41,42]; all FDR < 0.05; Fig. 6c) levels have previously been linked to lipid or cholesterol metabolism (Supplementary Table 15). *ABCA1* and *ABCG1*, which were positively correlated with the PGS for high

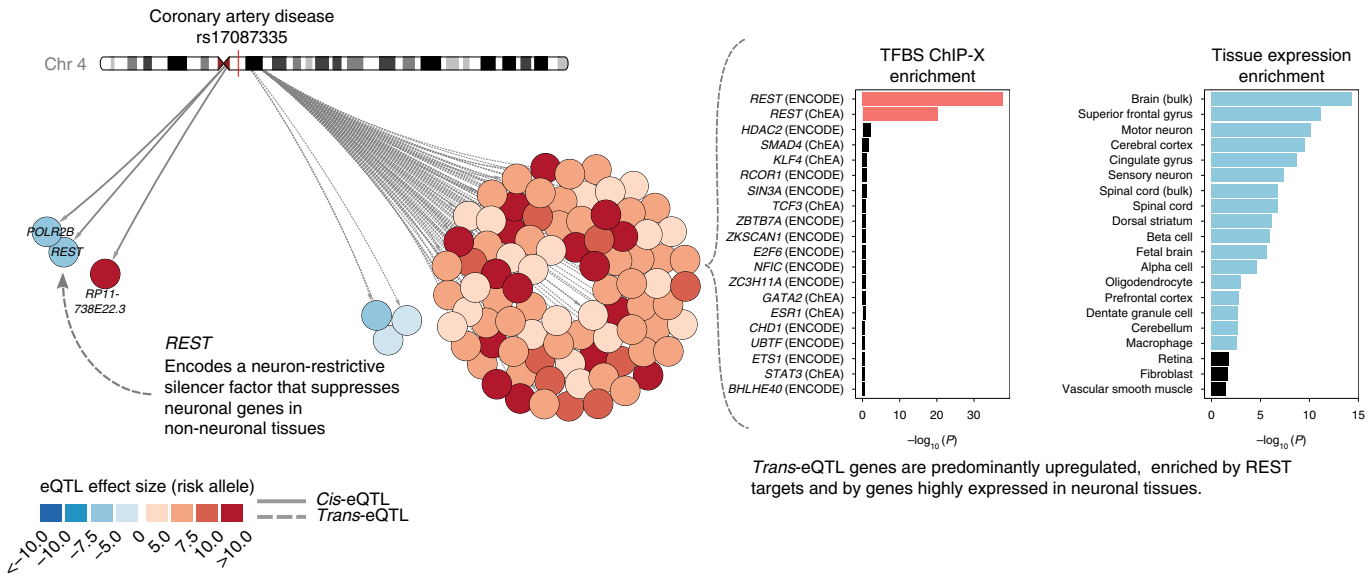


Fig. 4 | The *REST* locus regulates the expression of 88 *trans*-eQTL genes. Left, overview of *cis*- and *trans*-eQTL effects for the coronary artery disease-associated SNP rs17087335. Color of nodes indicates the *trans*-eQTL effect direction and size, relative to the risk allele. Right, *trans*-eQTL genes for the *REST* locus are highly enriched for RE1-silencing TF (REST) targets (TF-binding data from the Encyclopedia of DNA Elements (ENCODE)[57,58] and ChIP Enrichment Analysis (ChEA)[59]) and for the expression of brain-related genes. For each TF and tissue, the length of the bar indicates the $-\log_{10}$ (*P* value) from one-sided Fisher's exact test (Methods). The 20 most significant effects are visualized. Chr, chromosome; TFBS, transcription factor binding site.

HDL levels, mediate the efflux of cholesterol from macrophage foam cells and participate in HDL formation. In macrophages, downregulation of *ABCA1* and *ABCG1* reduces reverse cholesterol transport into the liver by HDL[43] (Fig. 6d). The PGS for high HDL levels was also negatively correlated with expression of *LDLR* (encoding the low-density lipoprotein receptor) (strongest eQTS, $P = 3.35 \times 10^{-20}$), mutations in which are known to cause familial hypercholesterolemia[44]. Similarly, *SREBF2*, the gene encoding the TF sterol-regulatory element-binding transcription factor (SREBP)2, which increases the expression of *LDLR*, was downregulated (strongest eQTS, $P = 3.08 \times 10^{-7}$). A negative correlation between *SREBF2* expression and measured HDL levels has been described before[15], indicating that the eQTS reflects an association with an actual phenotype. Zhernakova et al.[15] proposed a model in which downregulation of *SREBF2* results in lower expression of its target gene *FADS2*. However, we did not observe an HDL eQTS effect on *FADS2* (all eQTS $P > 0.07$), possibly because the indirect effect was too small to detect. We hypothesize that higher blood HDL levels can result in stronger reverse cholesterol transport via HDL (Fig. 6d), which may result in downregulation of *LDLR*[45].

eQTSs can also identify pathways known to be associated with monogenic diseases. For example, PGSs for serine, glycine, the glycine derivative *N*-acetylglycine and creatine[46,47] were negatively associated with the expression of *PHGDH*, *PSAT1* and *AARS* ($P < 5.3 \times 10^{-7}$). PGSs for these traits are driven by SNPs near *CPS1* (2q34), *PHGDH* (1p12) and *PSPH* (7p11.2) (Supplementary Tables 16–18) that influence expression of *PHGDH* and *PSAT1* in *trans*. We nominally replicated these *trans*-eQTL in scRNA-seq data (absolute average $Z > 1.96$ across tested cell types, part of the 729 *trans*-eQTL replicating in the scRNA-seq data; Supplementary Table 4 and Fig. 6e), suggesting that this eQTS is driven by multiple genetic loci but independent of cell type composition. *PHGDH* and *PSAT1* encode enzymes that regulate the synthesis of serine and, in turn, glycine[48], while *N*-acetylglycine and creatine form downstream of glycine[49] (Fig. 6f). Mutations in *PSAT1* and *PHGDH* can result in monogenic conditions with defective serine biosynthesis, which are characterized by low concentrations of serine and glycine in blood and severe neuronal manifestations[50–52]. Unexpectedly, the PGS for higher levels of these amino acids was associated with lower expression of *PHGDH*, *PSAT1* and *AARS*, implying the presence of a negative feedback loop that controls serine synthesis.

Discussion

We performed *cis*-eQTL, *trans*-eQTL and eQTS analyses in 31,684 blood samples, a sixfold increase in sample size over earlier studies[5,9]. Of genes expressed in blood, 88.2% showed a cis-eQTL effect, 32% showed a trans-eQTL effect, and 13% showed an eQTS effect.

Most studies prioritizing genes for complex traits have considered only *cis*-eQTL effects, and our blood *cis*-eQTL can be used for that purpose. However, *cis*-eQTL effects have been estimated to contribute to a limited fraction of the heritability of gene expression, while the combination of many weak *trans*-eQTL effects is estimated to explain the majority[53], emphasizing the importance of distal effects. At the same time, interpretation of *trans*-eQTL in blood remains challenging: limited replication and the influence of blood cell composition suggest that the effects are highly cell type specific. Nevertheless, the replication analyses we carried out on peripheral blood mononuclear cell (PBMC) scRNA-seq data prioritized 729 *trans*-eQTL, and half of the identified *trans*-eQTL were assigned to a putative biological mechanism of action, with transcriptional regulation through TF activity being the most prevalent.

To identify genes that are coordinately affected by multiple independent trait-associated SNPs, we performed eQTS analysis. We identified eQTS associations for 2,568 genes and have outlined several examples in which the associated genes point to interpretable biology. One possible interpretation of these eQTS associations is in the context of the recently proposed omnigenic model[13,14]. As explained by Liu et al.[14], many weak distal effects could converge on trait-relevant 'core' genes, and eQTS analysis might help to prioritize such genes (Supplementary Note). However, an important limitation is that eQTS analysis can also identify genes that are merely co-regulated with trait-relevant ones. Therefore, it remains challenging to systematically evaluate which fraction of the detected

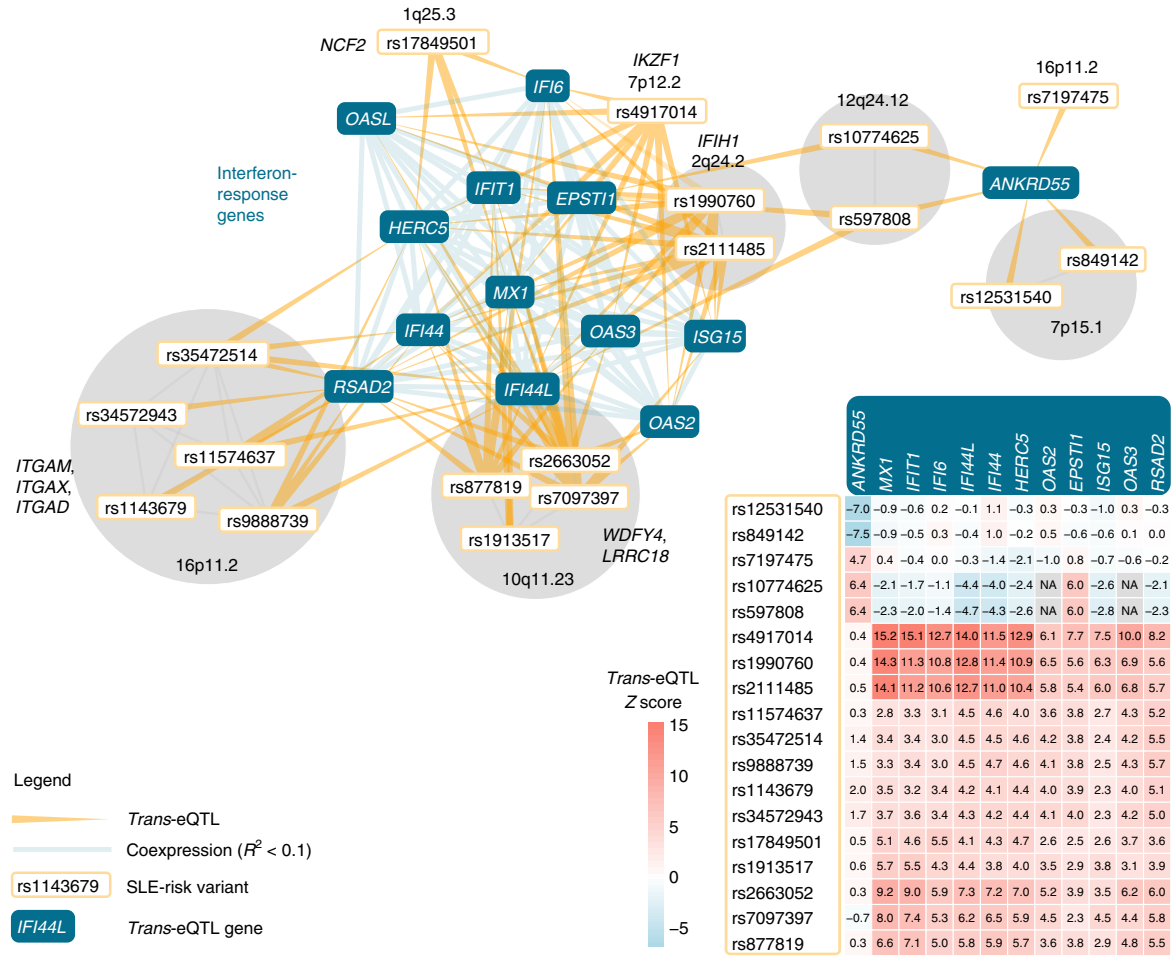


Fig. 5 | SNPs associated with SLE converge on a shared cluster of interferon-response genes. Genes shown are those affected by at least three independent GWAS SNPs. SNPs in the human leukocyte antigen region are not visualized, and SNPs in partial linkage disequilibrium are grouped together. The heatmap indicates the direction and strength of individual *trans*-eQTL effects (*Z* scores) relative to the SLE-risk allele. NA, *trans*-eQTL *Z* score not available due to quality-control filters.

eQTS genes is causal. While our analysis does not formally prove or disprove the validity of the model of Liu et al.[14], and true implications of this model remain to be investigated, our results can serve as a starting point to follow up on eQTS genes and to ascertain their role in complex traits. Our eQTS analysis provides a comprehensive resource for blood that can be used to interpret the effects of PGS on a molecular level.

There are some important limitations that require consideration when using our resource for hypothesis generation. First, we limited our *cis*-eQTL analysis to variants within 1 Mb of the gene center and limited our *trans*-eQTL analysis to variants >5 Mb from genes on the same chromosome. We acknowledge the possibility that these thresholds may have excluded distal *cis*-eQTL (for example, those caused by distal enhancers or chromatin loops) and *trans*-eQTL on nearby genes. We chose these thresholds to ensure that the *trans*-eQTL we observed were not driven by long-range *cis*-eQTL. While our approach might have excluded long-range *cis*-eQTL, we observed that, for 95.6% of genes, the lead *cis*-eQTL SNP mapped within 100 kb of the gene, suggesting that long-range *cis*-eQTL reflect only a small proportion of all *cis*-eQTL. Second, we confined our *trans*-eQTL analyses to a subset of 10,317 variants previously associated with complex phenotypes. As such, a significant *trans*-eQTL for a trait-associated variant does not necessarily mean that the same underlying variant affects both the phenotype and gene expression. Third, PGS estimates were shown to have variable prediction accuracy even when evaluated within the same ancestry. This variability may be caused by differences in sample characteristics (for example, age, sex, socio-economic status) in the original GWAS as well as the dataset in which the PGS is calculated[54]. Such variability may therefore have caused either inflation or deflation of our eQTS effect sizes. Although we present several examples that are interpretable in the context of the respective traits, caution is needed when drawing conclusions on higher-level phenotypes. Instead, our resource should serve as a starting point for further in-depth studies that can reliably connect the reported eQTL and eQTS associations to phenotypes.

Although putative biological mechanisms of action could be assigned to half of the identified *trans*-eQTL, significant replication in scRNA-seq, purified cell type and cell line datasets was very limited. Such low replication rates suggest two likely causes. First, a number of the distal effects are likely driven by inter-individual cell type-composition differences, which occur in any bulk tissue. While such effects could be informative in the context of some complex traits (that is, for autoimmune diseases), the most interesting information lies in intracellular effects. Furthermore, while we corrected for unknown confounders in our analyses, some residual cell-composition effects remain in the data. Therefore, it was not possible to reliably distinguish cell type-dependent effects from intracellular ones. Instead, our results should serve as a prioritized list for in-depth functional studies.

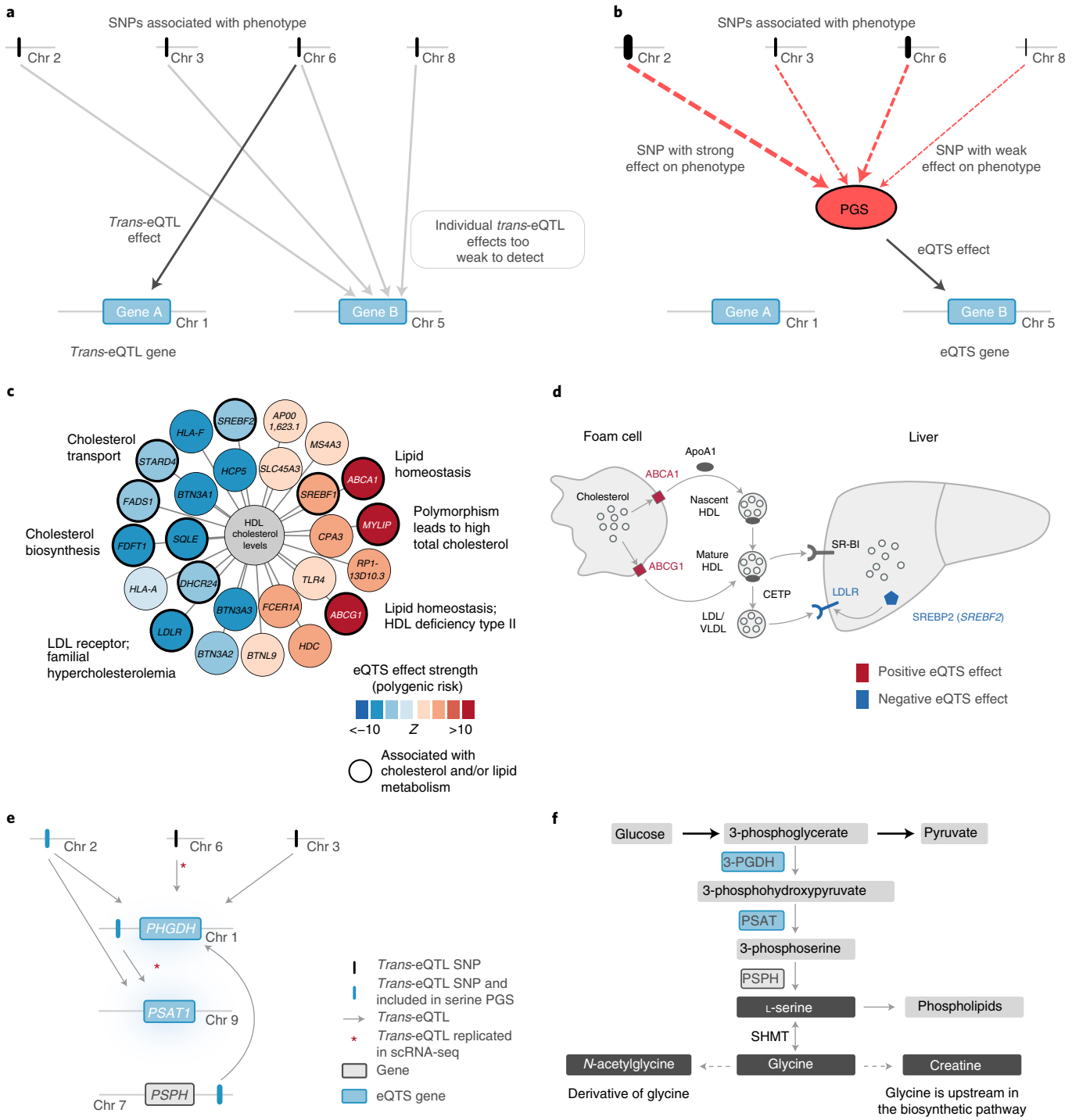


Fig. 6 | eQTS analyses. a, In *trans*-eQTL analysis, individual SNPs are associated with gene expression. **b**, In eQTS analysis, effect sizes and directions of individual trait-associated SNPs are combined into a PGS that is associated with gene expression. Here, we outline the case in which eQTS analysis identifies a gene that is not detectable in the *trans*-eQTL analysis. Other scenarios we observed include gene A also being identified by eQTS analysis, gene B being identified by both methods or the combined effect of PGS yielding no significant eQTS. **c**, The PGS for HDL associates with lipid metabolism genes. **d**, The role of ATP-binding cassette (ABC)A1, ABCG1, LDLR and SREBF2 in cholesterol transport. VLDL, very low-density lipoprotein; CETP, cholesteryl ester transfer protein; SR-BI, scavenger receptor class B member 1. **e**, Both *trans*-eQTL and the serine PGS associate with known serine biosynthesis genes *PHGDH* and *PSAT1*. **f**, Serine biosynthesis pathway. PSPH, phosphoserine phosphatase; 3-PGDH, 3-phosphoglycerate dehydrogenase; SHMT, serine hydroxymethyltransferase.

Second, our discovery analyses were conducted in a sample more than ten times larger than the largest replication datasets available. Because *trans*-eQTL effects are generally weak, this lack of statistical power likely causes low replication rates. Additionally, *trans*-eQTL effects are widely considered to be more cell type specific and tissue specific than local *cis*-eQTL effects[18]. Although this belief might be partly caused by variable *trans*-eQTL strengths in different tissue contexts and the limited power of current *trans*-eQTL studies,

it would also lead to lower replication rates of blood *trans*-eQTL in specific cell types.

Compared to gene expression from bulk tissues, scRNA-seq datasets are less affected by cell type composition and serve as the best current source for replicating, prioritizing and annotating *trans*-eQTL. While we have compiled, to our knowledge, the largest available blood scRNA replication dataset, it was still only 3.6% of the sample size of the discovery study. It is therefore unsurprising that only 35 *trans*-eQTL reached the significance threshold (FDR < 0.05). Nonetheless, 84% of the 729 *trans*-eQTL attaining nominal significance ($P < 0.05$) also showed allelic concordance with data from the discovery analysis, suggesting that there are intracellular effects among our *trans*-eQTL, even if case-by-case distinction of cell type composition and intracellular effects is not yet possible. Upcoming large-scale single-cell eQTL studies[55] (for example, https://www.eqtlgen.org/single-cell.html) as well as highly powered eQTL analyses in non-blood tissues[56] and cell lines will be instrumental in distinguishing intracellular effects from cell type composition.

Full summary statistics for our *cis*-eQTL, *trans*-eQTL and eQTS analyses (http://www.eqtlgen.org) can be used to interpret GWASs, to prioritize putative trait-related genes for in-depth functional studies and to develop methods to perform these tasks. We envision that upcoming statistical tools and frameworks that enable federated analyses in consortia will facilitate conducting highly powered global *trans*-eQTL studies. This will expand the work presented here and enable a better connection between distal effects on gene expression and complex phenotypes.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00913-z.

Received: 20 May 2020; Accepted: 12 July 2021;
Published online: 2 September 2021

References

1. Yao, D. W., O'Connor, L. J., Price, A. L. & Gusev, A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat. Genet.* **52**, 626–633 (2020).
2. O'Connor, L. J. et al. Extreme polygenicity of complex traits is explained by negative selection. *Am. J. Hum. Genet.* **105**, 456–476 (2019).
3. Zeng, J. et al. Signatures of negative selection in the genetic architecture of human complex traits. *Nat. Genet.* **50**, 746–753 (2018).
4. Westra, H. J. et al. Systematic identification of *trans* eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
5. Kirsten, H. et al. Dissecting the genetics of the human transcriptome identifies novel trait-related *trans*-eQTLs and corroborates the regulatory relevance of non-protein coding loci. *Hum. Mol. Genet.* **24**, 4746–4763 (2015).
6. Lloyd-Jones, L. R. et al. The genetic architecture of gene expression in peripheral blood. *Am. J. Hum. Genet.* **100**, 228–237 (2017).
7. Jansen, R. et al. Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Hum. Mol. Genet.* **26**, 1444–1451 (2017).
8. Joehanes, R. et al. Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol.* **18**, 16 (2017).
9. Yao, C. et al. Dynamic role of *trans* regulation of gene expression in relation to complex traits. *Am. J. Hum. Genet.* **100**, 571–580 (2017).
10. Brynedal, B. et al. Large-scale *trans*-eQTLs affect hundreds of transcripts and mediate patterns of transcriptional co-regulation. *Am. J. Hum. Genet.* **100**, 581–591 (2017).
11. Lewis, C. M. & Vassos, E. Prospects for using risk scores in polygenic medicine. *Genome Med.* **9**, 96 (2017).
12. Natarajan, P. et al. Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. *Circulation* **135**, 2091–2101 (2017).
13. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
14. Liu, X., Li, Y. I. & Pritchard, J. K. *Trans* effects on gene expression can drive omnigenic inheritance. *Cell* **177**, 1022–1034 (2019).
15. Zhernakova, D. V. et al. Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139–145 (2017).
16. Bonder, M. J. et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat. Genet.* **49**, 131–138 (2017).
17. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **57**, 289–300 (1995).
18. Aguet, F. et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
19. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
20. Glassberg, E. C., Gao, Z., Harpak, A., Lan, X. & Pritchard, J. K. Evidence for weak selective constraint on human gene expression. *Genetics* **211**, 757–772 (2019).
21. Wu, Y., Zheng, Z., Visscher, P. M. & Yang, J. Quantifying the mapping precision of genome-wide association studies using whole-genome sequencing data. *Genome Biol.* **18**, 86 (2017).
22. Astle, W. J. et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429 (2016).
23. Melé, M. et al. The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015).
24. Qi, T. et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* **9**, 2282 (2018).
25. Marbach, D. et al. Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nat. Methods* **13**, 366–370 (2016).
26. Li, T. et al. A scored human protein–protein interaction network to catalyze genomic interpretation. *Nat. Methods* **14**, 61–64 (2016).
27. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and rigorous computation of gene and pathway scores from SNP-based summary statistics. *PLoS Comput. Biol.* **12**, e1004714 (2016).
28. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
29. Nikpay, M. et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
30. Bentham, J. et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat. Genet.* **47**, 1457–1464 (2015).
31. Davenport, E. E. et al. Discovering in vivo cytokine–eQTL interactions from a lupus clinical trial. *Genome Biol.* **19**, 168 (2018).
32. McBride, J. M. et al. Safety and pharmacodynamics of rontalizumab in patients with systemic lupus erythematosus: results of a phase I, placebo-controlled, double-blind, dose-escalation study. *Arthritis Rheum.* **64**, 3666–3676 (2012).
33. Yao, Y. et al. Development of potential pharmacodynamic and diagnostic markers for anti-IFN-α monoclonal antibody trials in systemic lupus erythematosus. *Hum. Genomics Proteomics* **2009**, 374312 (2009).
34. Perry, J. R. B. et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014).
35. Lemaitre, R. N. et al. Genetic loci associated with plasma phospholipid n-3 fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium. *PLoS Genet.* **7**, e1002193 (2011).
36. Liu, J. Z. et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).
37. Gateva, V. et al. A large-scale replication study identifies *TNIP1*, *PRDM1*, *JAZF1*, *UHRF1BP1* and *IL10* as risk loci for systemic lupus erythematosus. *Nat. Genet.* **41**, 1228–1233 (2009).
38. Moffatt, M. F. et al. A large-scale, consortium-based genomewide association study of asthma. *N. Engl. J. Med.* **363**, 1211–1221 (2010).
39. Wood, A. R. et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
40. Van Der Harst, P. et al. Seventy-five genetic loci influencing the human red blood cell. *Nature* **492**, 369–375 (2012).
41. Teslovich, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
42. Willer, C. J. et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1285 (2013).
43. Wang, X. et al. Macrophage ABCA1 and ABCG1, but not SR-BI, promote macrophage reverse cholesterol transport in vivo. *J. Clin. Invest.* **117**, 2216–2224 (2007).
44. Goldstein, J. L. & Brown, M. S. Binding and degradation of low density lipoproteins by cultured human fibroblasts. Comparison of cells from a normal subject and from a patient with homozygous familial hypercholesterolemia. *J. Biol. Chem.* **249**, 5153–5162 (1974).

45. Singh, A. B., Kan, C. F. K., Shende, V., Dong, B. & Liu, J. A novel posttranscriptional mechanism for dietary cholesterol-mediated suppression of liver LDL receptor expression. *J. Lipid Res.* **55**, 1397–1407 (2014).
46. Kettunen, J. et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat. Commun.* **7**, 11122 (2016).
47. Shin, S. Y. et al. An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46**, 543–550 (2014).
48. El-Hattab, A. W. Serine biosynthesis and transport defects. *Mol. Genet. Metab.* **118**, 153–159 (2016).
49. Leuzzi, V., Alessandrì, M. G., Casarano, M., Battini, R. & Cioni, G. Arginine and glycine stimulate creatine synthesis in creatine transporter 1-deficient lymphoblasts. *Anal. Biochem.* **375**, 153–155 (2008).
50. Hart, C. E. et al. Phosphoserine aminotransferase deficiency: a novel disorder of the serine biosynthesis pathway. *Am. J. Hum. Genet.* **80**, 931–937 (2007).
51. Klomp, L. W. J. et al. Molecular characterization of 3-phosphoglycerate dehydrogenase deficiency—a neurometabolic disorder associated with reduced L-serine biosynthesis. *Am. J. Hum. Genet.* **67**, 1389–1399 (2000).
52. Shaheen, R. et al. Neu-Laxova syndrome, an inborn error of serine metabolism, is caused by mutations in *PHGDH*. *Am. J. Hum. Genet.* **94**, 898–904 (2014).
53. Price, A. L. et al. Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. *PLoS Genet.* **7**, e1001317 (2011).
54. Mostafavi, H. et al. Variable prediction accuracy of polygenic scores within an ancestry group. *eLife* **9**, e48376 (2020).
55. van der Wijst, M. et al. The single-cell eQTLGen consortium. *eLife* **9**, e52155 (2020).
56. Wang, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018).
57. Feingold, E. A. et al. The ENCODE (ENCyclopedia of DNA Elements) Project. *Science* **306**, 636–640 (2004).
58. Myers, R. M. et al. A user's guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biol.* **9**, e1001046 (2011).
59. Lachmann, A. et al. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**, 2438–2444 (2010).

Urmo Võsa[1,2,88] ✉, Annique Claringbould[1,3,4,88] ✉, Harm-Jan Westra[1,3], Marc Jan Bonder[1,5], Patrick Deelen[1,3,6,7], Biao Zeng[8], Holger Kirsten[9,10], Ashis Saha[11], Roman Kreuzhuber[12,13,14], Seyhan Yazar[15], Harm Brugge[1,3], Roy Oelen[1,3], Dylan H. de Vries[1,3], Monique G. P. van der Wijst[1,3], Silva Kasela[2], Natalia Pervjakova[2], Isabel Alves[16,17], Marie-Julie Favé[16], Mawussé Agbessi[16], Mark W. Christiansen[18], Rick Jansen[19], Ilkka Seppälä[20], Lin Tong[21], Alexander Teumer[22,23], Katharina Schramm[24,25], Gibran Hemani[26], Joost Verlouw[27], Hanieh Yaghootkar[28,29,30], Reyhan Sönmez Flitman[31,32], Andrew Brown[33,34], Viktorija Kukushkina[2], Anette Kalnapenkis[2], Sina Rüeger[35], Eleonora Porcu[35], Jaanika Kronberg[2], Johannes Kettunen[36,37,38,39], Bernett Lee[40], Futao Zhang[41], Ting Qi[41], Jose Alquicira Hernandez[15], Wibowo Arindrarto[42], Frank Beutner[43], BIOS Consortium*, i2QTL Consortium, Julia Dmitrieva[49], Mahmoud Elansary[49], Benjamin P. Fairfax[50], Michel Georges[49], Bastiaan T. Heijmans[42], Alex W. Hewitt[51,52], Mika Kähönen[53], Yungil Kim[11,54], Julian C. Knight[50], Peter Kovacs[55], Knut Krohn[56], Shuang Li[1,6], Markus Loeffler[9,10], Urko M. Marigorta[8,47,48], Hailang Mei[57], Yukihide Momozawa[49,58], Martina Müller-Nurasyid[24,25,59], Matthias Nauck[23,60], Michel G. Nivard[61], Brenda W. J. H. Penninx[19], Jonathan K. Pritchard[62,63], Olli T. Raitakari[45,46,64], Olaf Rotzschke[40], Eline P. Slagboom[42], Coen D. A. Stehouwer[65], Michael Stumvoll[66], Patrick Sullivan[67], Peter A. C. 't Hoen[68], Joachim Thiery[10,44], Anke Tönjes[66], Jenny van Dongen[69], Maarten van Iterson[42], Jan H. Veldink[70], Uwe Völker[71], Robert Warmerdam[1,3], Cisca Wijmenga[1], Morris Swertz[6], Anand Andiappan[40], Grant W. Montgomery[41], Samuli Ripatti[72,73,74], Markus Perola[75], Zoltan Kutalik[76], Emmanouil Dermitzakis[32,33,77], Sven Bergmann[31,32], Timothy Frayling[28], Joyce van Meurs[27], Holger Prokisch[78,79], Habibul Ahsan[21], Brandon L. Pierce[21], Terho Lehtimäki[20], Dorret I. Boomsma[69], Bruce M. Psaty[80], Sina A. Gharib[18,81], Philip Awadalla[16], Lili Milani[2], Willem H. Ouwehand[12,13,82], Kate Downes[12,13], Oliver Stegle[5,14,83], Alexis Battle[11,84], Peter M. Visscher[41], Jian Yang[41,85,86], Markus Scholz[9,10], Joseph Powell[15,87,89], Greg Gibson[8,89], Tõnu Esko[2,89] and Lude Franke[1,3,89] ✉

[1]Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [2]Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia. [3]Oncode Institute, Amsterdam, the Netherlands. [4]Structural & Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany. [5]Genome Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany. [6]Genomics Coordination Center, University Medical Centre Groningen, Groningen, the Netherlands. [7]Department of Genetics, University Medical Centre Utrecht, Utrecht, the Netherlands. [8]School of Biological Sciences, Georgia Tech, Atlanta, GA, USA. [9]Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany. [10]LIFE Research Center for Civilization Diseases, University of Leipzig, Leipzig, Germany. [11]Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA. [12]Department of Haematology, University of Cambridge, Cambridge, United Kingdom. [13]NHS Blood and Transplant, Cambridge Biomedical Campus, Cambridge, United Kingdom. [14]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, United Kingdom. [15]Garvan Institute of Medical Research, Garvan-Weizmann Centre for Cellular Genomics, Sydney,

New South Wales, Australia. [16]Computational Biology, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [17]L'institut du thorax, Université de Nantes, CHU Nantes, INSERM, CNRS, Nantes, France. [18]Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA. [19]Department of Psychiatry, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Public Health Research Institute and Amsterdam Neuroscience, Amsterdam, the Netherlands. [20]Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center—Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [21]Department of Public Health Sciences, University of Chicago, Chicago, IL, USA. [22]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [23]DZHK (German Center for Cardiovascular Research), Partner Site Greifswald, Greifswald, Germany. [24]Institute of Genetic Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany. [25]Department of Medicine I, University Hospital Munich, Ludwig Maximilian's University, Munich, Germany. [26]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom. [27]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands. [28]Genetics of Complex Traits, University of Exeter Medical School, Royal Devon & Exeter Hospital, Exeter, United Kingdom. [29]School of Life Sciences, College of Liberal Arts and Science, University of Westminster, London, United Kingdom. [30]Division of Medical Sciences, Department of Health Sciences, Luleå University of Technology, Luleå, Sweden. [31]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [32]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [33]Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland. [34]Population Health and Genomics, University of Dundee, Dundee, United Kingdom. [35]Lausanne University Hospital, Lausanne, Switzerland. [36]Computational Medicine, Faculty of Medicine, University of Oulu, Oulu, Finland. [37]Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland. [38]Biocenter Oulu, University of Oulu, Oulu, Finland. [39]Finnish Institute for Health and Welfare, Helsinki, Finland. [40]Singapore Immunology Network, Agency for Science, Technology and Research, Singapore, Singapore. [41]Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia. [42]Leiden University Medical Center, Leiden, the Netherlands. [43]Heart Center Leipzig, Universität Leipzig, Leipzig, Germany. [44]Institute for Laboratory Medicine, LIFE—Leipzig Research Center for Civilization Diseases, Universität Leipzig, Leipzig, Germany. [45]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. [46]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. [47]Integrative Genomics Lab, CIC bioGUNE, Basque Research and Technology Alliance (BRTA), Bizkaia Science and Technology Park, Derio, Spain. [48]IKERBASQUE, Basque Foundation for Science, Bilbao, Spain. [49]Unit of Animal Genomics, WELBIO, GIGA-R & Faculty of Veterinary Medicine, University of Liège, Liège, Belgium. [50]Wellcome Centre for Human Genetics, University of Oxford, Oxford, United Kingdom. [51]Menzies Institute for Medical Research, School of Medicine, University of Tasmania, Hobart, Tasmania, Australia. [52]Centre for Eye Research Australia, Department of Surgery, University of Melbourne, Melbourne, Victoria, Australia. [53]Department of Clinical Physiology, Tampere University Hospital and Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [54]Genetics and Genomic Science Department, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [55]IFB Adiposity Diseases, Universität Leipzig, Leipzig, Germany. [56]Interdisciplinary Center for Clinical Research, Faculty of Medicine, Universität Leipzig, Leipzig, Germany. [57]Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, the Netherlands. [58]Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan. [59]IBE, Faculty of Medicine, LMU Munich, Munich, Germany. [60]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. [61]Department of Biological Psychology, Faculty of Behaviour and Movement Sciences, Vrije Universiteit, Amsterdam, the Netherlands. [62]Department of Biology, Stanford University, Stanford, CA, USA. [63]Department of Genetics, Stanford University, Stanford, CA, USA. [64]Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland. [65]Department of Internal Medicine and School for Cardiovascular Diseases (CARIM), Maastricht University Medical Center, Maastricht, the Netherlands. [66]Department of Medicine, Universität Leipzig, Leipzig, Germany. [67]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [68]Center for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center Nijmegen, Nijmegen, the Netherlands. [69]Netherlands Twin Register, Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam Public Health Research Institute and Amsterdam Neuroscience, Amsterdam, the Netherlands. [70]UMC Utrecht Brain Center, University Medical Center Utrecht, Department of Neurology, Utrecht University, Utrecht, the Netherlands. [71]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. [72]Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland. [73]Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland. [74]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [75]National Institute for Health and Welfare, University of Helsinki, Helsinki, Finland. [76]Center for Primary Care and Public Health, University of Lausanne, Lausanne, Switzerland. [77]Institute of Genetics and Genomics in Geneva (iGE3), University of Geneva, Geneva, Switzerland. [78]Institute of Neurogenomics, Helmholtz Zentrum München, Neuherberg, Germany. [79]Institute of Human Genetics, Technical University Munich, Munich, Germany. [80]Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA, USA. [81]Department of Medicine, University of Washington, Seattle, WA, USA. [82]Human Genetics, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom. [83]Division of Computational Genomics and Systems Genetics, German Cancer Research Center, Heidelberg, Germany. [84]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA. [85]School of Life Sciences, Westlake University, Hangzhou, China. [86]Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, China. [87]UNSW Cellular Genomics Futures Institute, University of New South Wales, Sydney, New South Wales, Australia. [88]These authors contributed equally: Urmo Võsa, Annique Claringbould. [89]These authors jointly supervised this work: Joseph Powell, Greg Gibson, Tõnu Esko, Lude Franke. *Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: urmo.vosa@gmail.com; anniqueclaringbould@gmail.com; lude@ludesign.nl

BIOS Consortium

Bastiaan T. Heijmans[42], Peter A. C. 't Hoen[68], Joyce van Meurs[27], Rick Jansen[19], Lude Franke[1,3,89], Dorret I. Boomsma[69], Jenny van Dongen[69], Coen D. A. Stehouwer[65], Cisca Wijmenga[1], Eline P. Slagboom[42], Jan H. Veldink[70], Hailang Mei[57], Maarten van Iterson[42], Patrick Deelen[1,3,6,7], Marc Jan Bonder[1,5], Morris A. Swertz[6] and Wibowo Arindrarto[42]

i2QTL Consortium

Marc Jan Bonder[1,5] and Oliver Stegle[5,14,83]

Full list for consortium members appears in Supplementary Note.

Methods

Cohorts. eQTLGen Consortium data consist of 31,684 blood and PBMC samples from 37 datasets, preprocessed in a standardized manner and analyzed by each cohort analyst. In total, 25,482 (80.4%) of the samples were whole-blood samples, and 6,202 (19.6%) were PBMC samples, and the majority of samples were of European ancestry (Supplementary Table 1). Gene expression levels of samples were profiled by Illumina ($N=17,421$, 55%), Affymetrix U219 ($N=2,767$, 8.7%) and Affymetrix Hu-Ex version 1.0 ST ($N=5,075$, 16%) expression arrays and by RNA-seq ($N=6,422$, 20.3%). A summary of each dataset is outlined in Supplementary Table 1. Detailed cohort descriptions can be found in the Supplementary Note. All cohorts participating in this study enrolled participants with informed consent, collected and analyzed data in accordance with ethical and institutional regulations and provided summary statistics for meta-analyses. Information about individual institutional review board approvals is available in the original publications for each cohort (Supplementary Note) or in the cohort-specific section in the Supplementary Note.

Each of the cohorts carried out genotype and expression data preprocessing, PGS calculation and *cis*-eQTL, *trans*-eQTL and eQTS mapping following the steps outlined in online analysis plans specific for each platform (https://github.com/molgenis/systemsgenetics/wiki/eQTL-mapping-analysis-cookbook-(eQTLGen), https://github.com/molgenis/systemsgenetics/wiki/eQTL-mapping-analysis-cookbook-for-RNA-seq-data, https://github.com/molgenis/systemsgenetics/wiki/QTL-mapping-analysis-cookbook-for-Affymetrix-expression-arrays) or with slight alterations as described in Supplementary Table 1 and the Supplementary Note. All but one cohort (Framingham Heart Study) included unrelated individuals into the analysis.

Information about replication datasets is detailed in the Supplementary Note.

Genotype data preprocessing. Primary preprocessing and quality control of genotype data were conducted by each cohort, as specified in the original publications and in the Supplementary Note. The majority of cohorts used genotypes imputed to the 1000 Genomes phase 1 version 3 (1000G p1v3) or a newer reference panel. Genotype Harmonizer[60] version 1.4.9 (https://github.com/molgenis/systemsgenetics/wiki/Genotype-Harmonizer) was used to harmonize all genotype datasets to match the Genetic Investigation of ANthropometric Traits (GIANT) consortium 1000G p1v3 all ancestries (ALL) reference panel (ftp://share.sph.umich.edu/1000genomes/fullProject/2012.03.14/GIANT.phase1_release_v3.20101123.snps_indels_svs.genotypes.refpanel.ALL.vcf.gz.tgz) and to fix potential strand issues for A–T and C–G SNPs. Each cohort tested SNPs with minor allele frequency (MAF) > 0.01, Hardy–Weinberg P value >0.0001, call rate >0.95 and MACH $r^2 > 0.5$. Reported SNP identifiers are in dbSNP version 137.

Expression data preprocessing. *Illumina arrays.* Illumina array expression datasets were profiled by HT-12v3, HT-12v4 and HT-12v4 WGDASL arrays. Before analysis, all probe sequences from the manifest files of those platforms were remapped to the GRCh37.p10 human genome build and transcriptome using the SHRiMP version 2.2.3 aligner[61] (http://compbio.cs.toronto.edu/shrimp/), allowing two mismatches. Probes mapping to multiple locations in the genome were removed from further analyses.

For Illumina arrays, the raw unprocessed expression matrix was exported from GenomeStudio. First, two PCs were calculated on quantile-normalized and $\log_2$-transformed expression data and plotted to identify and exclude outlier samples. Data were normalized in several steps: quantile normalization, $\log_2$ transformation, probe centering and scaling by subtracting from each expression value the mean of the respective probe and then dividing it with the standard deviation of the respective probe. Genes showing no variance were removed. Next, the first four multidimensional scaling components, calculated based on unimputed and pruned genotypes using PLINK version 1.07 (ref. [62]), were regressed out of the expression matrix to account for population stratification. We further removed up to 20 of the first expression-based PCs that were not associated with any SNPs, as these capture non-genetic variation in expression. After regressing out these covariates, the residual gene expression matrix was used for eQTL mapping. Each cohort also ran MixupMapper[63] software to identify incorrectly labeled genotype–expression combinations and resolved any sample mix-ups (https://github.com/molgenis/systemsgenetics/wiki/Resolving-mixups).

Affymetrix arrays. Affymetrix-array-based datasets used expression data previously preprocessed and controlled for quality as indicated in the Supplementary Note.

RNA-seq. Alignment, initial quality control and quantification differed slightly across datasets, as described in the Supplementary Note. Each cohort removed outliers as described above and then used trimmed mean of M-values normalization[64] and a counts-per-million filter to include genes with >0.5 counts per million in at least 1% of samples. Subsequent steps were identical to those in the Illumina processing, with some exceptions for BIOS Consortium datasets (Supplementary Note).

Empirical probe matching. To integrate different expression platforms for the purpose of meta-analysis, we developed an empirical probe-matching approach. We used pruned SNPs to conduct per-platform meta-analyses for all Illumina arrays, for all RNA-seq datasets and for each Affymetrix dataset separately, using summary statistics from analyses without correction for PCs. For each platform, this yielded an empirical *trans*-eQTL Z-score matrix, as well as ten permuted Z-score matrices in which links between genotypes and expression files were shuffled. These permuted Z-score matrices reflect the gene–gene or probe–probe correlation structure.

We then used RNA-seq permuted Z-score matrices as a gold standard reference and calculated, for each gene, Pearson correlation coefficients with all the other genes, yielding a correlation profile for each gene. We then repeated the same analysis for the Illumina meta-analysis and the two different Affymetrix platforms. Finally, we correlated correlation profiles from each array platform with correlation profiles from RNA-seq data. If there were multiple probes detecting the expression of one gene, we selected the probe showing the highest Pearson correlation with the corresponding gene in the RNA-seq data and treated these as matching expression features in the combined meta-analyses. This yielded 19,942 genes that were detected in RNA-seq datasets and tested in the combined meta-analyses. Genes and probes were matched to Ensembl version 71 (ref. [65]) (ftp://ftp.ensembl.org/pub/release-71/gtf/homo_sapiens/Homo_sapiens.GRCh37.71.gtf.gz) stable gene IDs and HGNC symbols in all analyses.

Meta-analysis procedure. Results presented in this study were meta-analyzed using a weighted Z-score method[66] in which Z scores are weighted by the square root of the sample size of the cohort. For *cis*-eQTL and *trans*-eQTL meta-analyses, this resulted in a final sample size of up to 31,684. The combined eQTS meta-analysis included only unrelated individuals from the Framingham Heart Study, resulting in a combined sample size of up to 28,158. Considering that our analysis contained many different gene expression and genotyping platforms, we limited our meta-analysis to associations present in at least two cohorts to reduce platform-specific effects. Specifics for each meta-analysis (*cis*-eQTL, *trans*-eQTL, eQTS) are detailed below.

Cross-platform replications. To test the performance of the empirical probe-matching approach, we conducted discovery *cis*, *trans* and eQTS meta-analyses for each expression platform (RNA-seq, Illumina, Affymetrix U219 and Affymetrix Hu-Ex version 1.0 ST arrays; array probes matched to 19,942 genes by empirical probe matching). For each discovery analysis, we conducted replication analyses in the three remaining platforms.

***Cis*-eQTL mapping.** *Cis*-eQTL mapping was performed in each cohort using a pipeline described previously[4]. In brief, the pipeline takes a window 1 Mb upstream and 1 Mb downstream around each SNP to select genes or expression probes to test, based on the center position of the gene or probe. Associations between these SNP–gene combinations are calculated using Spearman correlation. Next, every cohort performed ten permutations. In each permutation, links between genotype and expression identifiers were shuffled before recalculating all associations. Both non-permuted results and each round of permuted results were meta-analyzed across cohorts.

Multiple-testing correction for *cis*-eQTL mapping. For our multiple-testing procedure, we used meta-analyzed permutations to calculate the overall FDR as previously described[4]. In short, we reasoned that the large numbers of correlated SNPs and genes present in the *cis*-eQTL results might cause inflated estimates (that is, highly correlated SNPs associated with a specific gene would result in equal permuted P values for that particular gene). To circumvent this issue, we first selected the lowest-association P value per gene in each of the permuted and non-permuted meta-analyses. The resulting lists of P values were sorted and, per P value in the non-permuted data, we determined the proportion of P values equal to or below this value in both permuted and non-permuted data. We then determined our FDR estimate as the proportion of permuted P values over the proportion of non-permuted P values. If a specific eQTL from the full set was not among the set of per-gene lowest-association P values, this eQTL was assigned the higher FDR value corresponding to the next eQTL available among the set of lead variants per gene. We refer to this procedure as 'gene-level' FDR but note that FDR estimates should be evaluated 'analysis wide', because the ultimate distribution of permuted P values used to calculate our FDR estimates was derived for all tested genes, rather than per gene. *Cis*-eQTL with a gene-level FDR < 0.05 (corresponding to $P < 2.02 \times 10^{-5}$) that were tested in more than one cohort were deemed significant.

***Trans*-eQTL mapping.** *Trans*-eQTL mapping was performed using a previously described pipeline[4] while testing a subset of 10,317 SNPs associated with complex traits. We required the distance between the SNP and the center of the gene to be >5 Mb. To maximize the power to identify *trans*-eQTL effects, expression matrices were corrected for the results of summary-statistic-based or iterative conditional *cis*-eQTL-mapping analyses (Supplementary Note) before *trans*-eQTL mapping. For this, lead SNPs for significant (FDR < 0.05) conditional *cis*-eQTL were regressed out from the expression matrix. Finally, we removed potential false positive *trans*-eQTL caused by reads cross-mapping with *cis* regions (Supplementary Note).

Selection of SNPs for trans-eQTL mapping. Genetic risk factors were downloaded from three public repositories: the EBI GWAS Catalog[67] (https://www.ebi.ac.uk/gwas/, downloaded 21 November 2016), the National Institutes of Health (NIH) GWAS Catalog and Immunobase (https://genetics.opentargets.org/immunobase, accessed 26 April 2016), applying a significance threshold of $P \leq 5 \times 10^{-8}$. Additionally, we added 2,706 genome-wide significant GWAS SNPs from a blood trait GWAS[23]. SNP coordinates were lifted to hg19 using the 'liftOver' command from R package rtracklayer version 1.34.1 and subsequently standardized to match the GIANT 1000G p1v3 ALL reference panel. This yielded 10,562 SNPs (Supplementary Table 4). We tested associations between all risk factors and genes that were at least 5 Mb away to ensure that that they did not tag a *cis*-eQTL effect. In total, 10,317 trait-associated SNPs were tested in *trans*-eQTL analyses.

eQTS mapping. *Polygenic-score trait inclusion.* Full association summary statistics were downloaded from several publicly available resources (Supplementary Table 13). PGSs are most predictive when individuals in the original GWAS are of similar ancestry as individuals for whom the PGS is calculated. Because most of the individuals in our meta-analyses were of European ancestry, we investigated the information presented on websites or abstracts of corresponding publications and omitted GWASs performed exclusively in non-European cohorts. Filters applied to separate data sources are indicated in the Supplementary Note. All the dbSNP rs numbers and directions of effects were standardized to match GIANT 1000G p1v3 identifiers and minor alleles. SNPs with opposite alleles compared to GIANT alleles were flipped. SNPs with A–T and C–G alleles, tri-allelic SNPs, indels and SNPs with unknown or different alleles from those in GIANT 1000G p1v3 were removed from the analysis. Genomic control was applied to all *P* values for datasets not genotyped by Immunochip or Metabochip. Additionally, genomic control was skipped for one dataset that did not have full associations available and for all the datasets from the GIANT consortium because genomic control had already been applied for these. In total, 1,263 summary-statistic files were added to the analysis. Information about summary-statistic files can be found in the Supplementary Note and Supplementary Table 13.

Polygenic-score calculation. A custom Java program, GeneticRiskScoreCalculator version 0.1.0c (https://github.com/molgenis/systemsgenetics/tree/master/GeneticRiskScoreCalculator), was used for calculating several PGSs in parallel. Independent effect SNPs for each summary-statistic file were identified by double-clumping, first using a 250-kb distance and subsequently a 5-Mb distance with a linkage-disequilibrium threshold $R^2 = 0.1$. Weighted PGSs were calculated by summing risk alleles for each independent SNP weighted by its GWAS effect size (β or log(OR) from the GWAS study). Five GWAS *P*-value thresholds ($P < 5 \times 10^{-8}$, 1×10^{-5}, 1×10^{-4}, 1×10^{-3} and 1×10^{-2}) were used for constructing PGSs for each summary-statistic file. The human leukocyte antigen region (chr6:25,000,000–35,000,000) was omitted from calculations, and PGSs were scaled to fall between 0 and 2 for compatibility with the QTL-mapping pipeline.

Pruning SNPs and PGSs. To identify a set of independent SNPs, we conducted linkage-disequilibrium-based pruning as implemented in PLINK 1.9 with the setting '–indep-pairwise 50 5 0.1'. This yielded 4,586 uncorrelated SNPs ($R^2 < 0.1$, GIANT 1000G p1v3 ALL).

To identify the set of uncorrelated PGSs, ten permuted *trans*-eQTL *Z*-score matrices from the combined *trans*-eQTL analysis were first confined to the pruned set of SNPs. These matrices were then used to identify 3,042 uncorrelated genes based on *Z*-score correlations (absolute Pearson $R < 0.05$). Next, permuted eQTS *Z*-score matrices were confined to uncorrelated genes and used to calculate pairwise correlations between all genetic risk scores to define a set of 1,873 uncorrelated PGSs (Pearson $R^2 < 0.1$).

Multiple-testing correction in *trans*-eQTL and eQTS mapping. To calculate FDR estimates for *trans*-eQTL and eQTSs we compared each *P* value from the non-permuted meta-analysis with all *P* values from ten meta-analyzed permutation rounds. We note that this differs from the permutation strategy used in the *cis*-eQTL analysis, because here we used *P* values from all SNP–gene combinations, not just the smallest *P* value for each gene. Nevertheless, the 10,317 SNPs tested for *trans*-eQTL contained many linked variants. To establish a conservative FDR estimate, we therefore used the pruned set of 4,586 SNPs to perform a meta-analysis for both non-permuted and permuted datasets. We derived FDR estimates from these limited meta-analyses by sorting lists of *P* values and determining the proportion of *P* values in non-permuted and permuted datasets for each given *P* value in the non-permuted dataset. We then applied these FDR estimates to *trans*-eQTL results from all 10,317 genetic trait-associated SNPs. If a specific eQTL from the full set was not tested in the meta-analysis conducted on the pruned set, this eQTL was assigned the higher FDR value corresponding to the next eQTL tested in the pruned set. We used an FDR threshold of 0.05 (corresponding to $P < 8.3 \times 10^{-6}$) to declare a *trans*-eQTL effect significant. Similarly, in the eQTS analysis, we used a set of 1,873 uncorrelated (Pearson $R^2 < 0.1$) PGSs and performed an analogous FDR calculation. The FDR threshold for eQTSs corresponded to $P < 3.02 \times 10^{-6}$. We analyzed only SNP–gene or PGS–gene pairs that were tested in at least two cohorts.

Replication of *trans*-eQTL and eQTSs in bulk datasets. Information about replication cohorts and their respective settings for replication analyses is outlined in the Supplementary Note. If applicable, summary statistics from different replication datasets for the same cell type or tissue were meta-analyzed using a weighted *Z*-score method[66]. Benjamini–Hochberg FDR[17] was used to adjust replication-analysis *P* values for multiple testing. We required FDR < 0.05 and the same effect direction with discovery to declare effect replicating. The R package 'pwr' (https://cran.r-project.org/web/packages/pwr/index.html) was used to conduct power analyses for replication datasets.

Single-cell RNA-seq analyses. *Single-cell RNA-seq cohorts and data.* For the replication of *trans*-eQTL in scRNA-seq data, we used unpublished data of PBMCs from 1,139 unrelated individuals in two cohorts generated using the 10x Chromium platform: OneK1K ($N = 982$) and 1M-scBloodNL ($N = 157$). Data were processed using the Cell Ranger Single Cell Software Suite version 3.0.2 (https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger) and aligned using STAR[68] implementation within the Cell Ranger Single Cell Software Suite. Cells were demultiplexed, and doublets were removed before performing cell type classification. We combined the data in a meta-analysis within each of the eight available cell types: B cells, $CD4^+$ T cells, $CD8^+$ T cells, classical monocytes, non-classical monocytes, dendritic cells, natural killer cells and plasma cells. See the Supplementary Note for methodological details.

Replication of trans-eQTL effects. We tested the replication of 59,786 discovery *trans*-eQTL only if the *trans*-eQTL gene was sufficiently expressed (that is, it had a missing sample fraction that was at most 20% in the large OneK1K dataset), leaving between 1,917 and 27,582 eQTL to be studied, depending on cell type. We estimated the inflation of signal by calculating the λ inflation relative to the inverse χ^2 cumulative distribution function of 0.5. *Trans*-eQTL with FDR < 0.05 in any cell type were deemed significantly replicating. To obtain a better idea of replication across cell types, we calculated the average *Z* score across cell types. We selected effects with an absolute average *Z* score > 1.96 (equivalent to $P < 0.05$) to calculate the allelic concordance with the discovery *trans*-eQTL.

Correlation of trans-eQTL effects. To test the correlation between *trans*-eQTL effects in discovery and replication datasets, we used the r_b approach[24], which accounts for errors in estimated eQTL effects such that the estimate of correlation is less dependent on sample sizes. First, we derived the estimate of the *trans*-eQTL effect (β) and the SE of β (SE_β) from the *Z* score and the MAF of significant *trans*-eQTL, using the following formulae from Zhu et al.[69]:

$$\beta = z \left(\sqrt{\left(2p\,(1-p)\left(n + z^2 \right) \right)} \right)^{-1}$$

$$SE_\beta = \left(\sqrt{\left(2p\,(1-p)\left(n + z^2 \right) \right)} \right)^{-1},$$

where *p* is the MAF, *n* is the sample size, and *z* is the meta-analysis *Z* score. MAF was computed from 26,609 eQTLGen samples (excluding the FHS) for discovery analysis and from 1,139 replication samples for scRNA-seq replication analyses. For analyses in purified cell types and cell lines (LCL, iPSC) for which allele frequencies were not available, we used the MAF as observed in the eQTLGen study instead.

To include independent effects in the analysis, for each *trans*-eQTL gene, we included only the strongest significant discovery effect in each 2-Mb window. Statistics of r_b and $SE(r_b)$ were calculated as detailed in Qi et al.[24] assuming no sample overlap between discovery and replication datasets. Because we were only seeking to correlate the effects of identified *trans*-eQTL, we did not use any reference discovery dataset for selecting *trans*-eQTL to estimate r_b and hence did not consider potential ascertainment bias, although such bias is likely to be small. To calculate a *P* value, the *Z* score was first calculated by dividing r_b by $SE(r_b)$ and then squared to calculate the χ^2 statistic. The *P* value was then derived from the χ^2 distribution with one degree of freedom.

Comparison of pLI metrics over gene expression bins. All genes were divided into ten bins based on the average expression value of the TMM-normalized and $\log_2$-transformed expression matrix from the BIOS cohort. If the number of tested genes was not divisible by ten, the lowest bin was set to include more genes than the rest of the bins. ExAC pLI metrics[19] were acquired from ftp://ftp.broadinstitute.org/pub/ExAC_release/release0.3.1/functional_gene_constraint/fordist_cleaned_exac_r03_march16_z_pli_rec_null_data.txt. For every expression bin, genes for which the pLI metric was available were tested for differences by two-sided Wilcoxon rank-sum test.

Transcription factor and tissue enrichment analyses for the *REST* locus. We downloaded curated sets of known TF targets and tissue-expressed genes from the Enrichr[70,71] website (http://amp.pharm.mssm.edu/Enrichr/). TF–target gene sets were assayed by ChIP-X experiments from the ChEA[59] and ENCODE[57,58] projects. Tissue-expressed genes were based on the ARCHS4 database[72]. Gene sets were

processed, mapped to entrez IDs with the R package clusterProfiler version 3.10.1 (ref. [73]) (http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html) and tested for over-representation by one-sided Fisher's exact test as implemented in the R package GeneOverlap version 1.18.0 (https://www.bioconductor.org/packages/release/bioc/html/GeneOverlap.html) by using 19,942 genes tested in the *trans*-eQTL analysis as background. Multiple-testing correction was conducted using the Benjamini–Hochberg method[17].

Biological mechanisms explaining *trans*-eQTL. To better understand biological mechanisms underlying *trans*-eQTL, we performed a number of enrichment analyses. We converted *trans*-eQTL to a gene-by-gene matrix via three methods: using Pascal[27], using *cis*-eQTL information and combining both (Supplementary Note). For the enrichments, we calculated whether there was significant overlap with known TF–target pairs[25] (http://www.RegulatoryCircuits.org), gene co-regulation patterns (Supplementary Note), PPIs[26] (https://www.intomics.com/inbio/map/api/get_data?file=InBio_Map_core_2016_09_12.tar.gz) and Hi-C contacts in LCL cells[28] using a two-sided Fisher's exact test.

Capture Hi-C overlap for *cis*-eQTL. To assess whether *cis*-eQTL lead SNPs overlapped with chromosomal contact as measured by Hi-C data, we used promoter CHi-C data[74] downloaded from CHiCP[75] (https://www.chicp.org/). We took the lead eQTL SNPs, overlapped these with CHi-C data and studied the 10,428 *cis*-eQTL genes for which these data were available. We tested whether the CHi-C target mapped within 5 kb of the lead SNP. Of the 803 *cis*-eQTL genes for which the lead SNP mapped more than 100 kb away from the TSS or TES, 223 overlapped with CHi-C data (27.8%). Of 9,625 *cis*-eQTL genes for which the lead SNP mapped within 100 kb from the TSS or TES, 1,641 overlapped with CHi-C data (17.0%). To test whether these observed overlaps happened by chance, we performed the same analysis while flipping the location of the CHi-C target relative to the location of the bait and tested the difference with a two-tailed two-sample test of equal proportions.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Primary genotype and gene expression data were analyzed by individual cohorts participating in the study, and our study analyzed summary statistics. Full summary statistics of eQTLGen *cis*-eQTL, *trans*-eQTL and eQTS meta-analyses are available on the eQTLGen website, http://www.eqtlgen.org, which was built using the MOLGENIS framework[76]. We also provide *cis*-eQTL files formatted for use in SMR, MAFs and replication statistics for *cis*-eQTL, *trans*-eQTL and eQTSs. Per-cohort summary statistics for discovery cohorts can be made available after approval of an analysis proposal in eQTLGen and with agreement of the cohort PIs; contact corresponding authors for further information. Trait-associated variants were collected from the EBI GWAS Catalog (https://www.ebi.ac.uk/gwas/, accessed on 21 November 2016), the NIH GWAS Catalog (now hosted by the EBI GWAS Catalog, https://www.ebi.ac.uk/gwas/) and Immunobase (http://www.immunobase.org, accessed 26 April 2016; now hosted by Open Targets at https://genetics.opentargets.org/immunobase). Sources of numerous GWAS summary statistics used for eQTS analyses are outlined in the Supplementary Note and Supplementary Table 13. ExAC pLI scores used for Fig. 2 originate from ftp://ftp.broadinstitute.org/pub/ExAC_release/release0.3.1/functional_gene_constraint/fordist_cleaned_exac_r03_march16_z_pli_rec_null_data.txt. Genotype reference files used for harmonizing discovery datasets for meta-analysis originate from ftp://share.sph.umich.edu/1000genomes/fullProject/2012.03.14/GIANT.phase1_release_v3.20101123.snps_indels_svs.genotypes.refpanel.ALL.vcf.gz.tgz. The gene model used for gene annotations originates from Ensembl version 71 (ftp://ftp.ensembl.org/pub/release-71/gtf/homo_sapiens/Homo_sapiens.GRCh37.71.gtf.gz). FANTOM TF annotations used for eQTS enrichment analyses originate from http://fantom.gsc.riken.jp/5/sstar/Browse_Transcription_Factors_hg19. ChIP-seq data used for *cis*-eQTL overlap originate from https://www.chicp.org/. PPI data used for *trans*-eQTL mechanism enrichment analyses originate from https://www.intomics.com/inbio/map/api/get_data?file=InBio_Map_core_2016_09_12.tar.gz. Hi-C data used for *trans*-eQTL mechanism enrichment are deposited in the GEO (GM12878, GEO accession GSE63525). Curated gene sets used for enrichment analyses (gene ontology sets, ENCODE ChIP-X and CheA ChIP-X TF targets, TRANSFAC and JASPAR PWMs, ARCHS4 tissue expression, TargetScan miRNA target predictions, TarBase miRNA validated targets) were downloaded from the Enrichr website (https://maayanlab.cloud/Enrichr/#stats). Gene expression summaries and metadata from GTEx version 7 originate from https://gtexportal.org/home/. Gene expression summaries from BIOS are available in the BIOS Omics Atlas (http://bbmri.researchlumc.nl/atlas/#data). Per-cohort individual-level genotype and gene expression data are governed by respective biobanks and access can be requested according to procedures established by each biobank, with relevant restrictions applying as imposed by the IRB or local legislation. Data-access procedures established for the BIOS Consortium are available at https://www.bbmri.nl/acquisition-use-analyze/bios. Source data are provided with this paper.

Code availability

Individual cohorts participating in the study followed analysis plans as specified in our analysis cookbooks (https://github.com/molgenis/systemsgenetics/wiki/eQTL-mapping-analysis-cookbook-(eQTLGen), https://github.com/molgenis/systemsgenetics/wiki/eQTL-mapping-analysis-cookbook-for-RNA-seq-data, https://github.com/molgenis/systemsgenetics/wiki/QTL-mapping-analysis-cookbook-for-Affymetrix-expression-arrays) or with slight alterations as described in the Methods and the Supplementary Note. Tools and source codes used for genotype harmonization, identification of sample mix-ups, eQTL mapping, meta-analyses and calculation of PGSs are available at https://github.com/molgenis/systemsgenetics/. Tools used for primary analyses were written in Java (versions 6–8, https://www.java.com/). PLINK version 1.0.7 (https://zzz.bwh.harvard.edu/plink/) and version 1.90 (https://www.cog-genomics.org/plink/1.9/) was used for clumping and pruning. Downstream analyses and plots were performed and constructed with R (versions 3.4.4, 3.6.1 and 4.0.0, https://cran.r-project.org/) using packages data.table version 1.12 (https://cran.r-project.org/web/packages/data.table/), tidyverse version 1.2.1 (https://cran.r-project.org/web/packages/tidyverse/), broom version 0.5.1 (https://cran.r-project.org/web/packages/broom/), pheatmap version 1.0.12 (https://cran.r-project.org/web/packages/pheatmap/) and GeneOverlap version 1.18.0 (https://bioconductor.org/packages/release/bioc/html/GeneOverlap.html). Power analyses were conducted with the R package pwr version 1.3-0 (https://cran.r-project.org/web/packages/pwr/). scRNA-seq analyses were performed using the Cell Ranger Single Cell Software Suite version 3.0.2 (https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger) and its implementation of STAR aligner. The ToppGene web tool (https://toppgene.cchmc.org/) was used for some interpretative enrichment analyses, as well as the GeneNetwork web tool (https://genenetwork.nl/). The Decon2 framework (https://github.com/molgenis/systemsgenetics/tree/master/Decon2) was used for predicting cell counts in BIOS data. We formatted our *cis*-eQTL into the BESD format using SMR (https://cnsgenomics.com/software/smr/#Overview).

References

60. Deelen, P. et al. Genotype Harmonizer: automatic strand alignment and format conversion for genotype data integration. *BMC Res. Notes* **7**, 901 (2014).
61. Rumble, S. M. et al. SHRiMP: accurate mapping of short color-space reads. *PLoS Comput. Biol.* **5**, e1000386 (2009).
62. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
63. Westra, H. J. et al. MixupMapper: correcting sample mix-ups in genome-wide datasets increases power to detect small genetic effects. *Bioinformatics* **27**, 2104–2111 (2011).
64. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
65. Zerbino, D. R. et al. Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).
66. Zaykin, D. V. Optimally weighted *Z*-test is a powerful method for combining probabilities in meta-analysis. *J. Evol. Biol.* **24**, 1836–1841 (2011).
67. MacArthur, J. et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
68. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
69. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
70. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
71. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
72. Lachmann, A. et al. Massive mining of publicly available RNA-seq data from human and mouse. *Nat. Commun.* **9**, 1366 (2018).
73. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
74. Javierre, B. M. et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384 (2016).
75. Schofield, E. C. et al. CHiCP: a web-based tool for the integrative and interactive visualization of promoter capture Hi-C datasets. *Bioinformatics* **32**, 2511–2513 (2016).
76. Swertz, M. A. et al. The MOLGENIS toolkit: rapid prototyping of biosoftware at the push of a button. *BMC Bioinformatics* **11**, S12 (2010).

Acknowledgements

The cohorts participating in this study list their acknowledgements in the cohort-specific sections of the Supplementary Note. This work is supported by a grant from the European Research Council (ERC, ERC Starting Grant agreement number 637640 ImmRisk), a VIDI grant (917.14.374) and a VICI grant from the Netherlands Organisation for Scientific Research (NWO) to L.F. This work has been supported by the European Regional Development Fund and the program Mobilitas Pluss (MOBTP108) to U.Võsa. The project was supported by the 'De Drie Lichten'

foundation in the Netherlands with a grant to A.C. M.G.N. is supported by ZonMw grants 849200011 and 531003014 from the Netherlands Organisation for Health Research and Development, a VENI grant from the NWO (VI.Veni.191G.030) and a Jacobs Foundation research fellowship. H.Y. is funded by a Diabetes UK RD Lawrence fellowship (17/0005594). This project received funding from the ERC under the European Union's Horizon 2020 research and innovation program (grant agreement no. 772376 (EScORIAL)) to J.H.V. T.E. and A.K. were supported by the Estonian Research Council grant PRG (PRG1291). A.Battle was supported by NIH grant R01MH109905, NIH grant R01HG008150 (NHGRI; Non-Coding Variants Program) and NIH grant R01MH101814 (NIH Common Fund; GTEx Program). M.G.P.v.d.W. was funded by the Nederlandse Organisatie voor Wetenschappelijk onderzoek, NWO-Veni 192.029. This work was supported by NIH grants R21ES024834 (B.Pierce), R01ES020506 (B.Pierce), R01ES023834 (B.Pierce), R35ES028379 (B.Pierce) and R01CA107431 (H.A.). This work was supported by the Sigrid Juselius Foundation (J.Kettunen) and funds from the Academy of Finland (grant numbers 297338 and 307247) (J.Kettunen) and the Novo Nordisk Foundation (grant number NNF17OC0026062) (J.Kettunen). S.Ripatti was supported by the Academy of Finland Centre of Excellence in Complex Disease Genetics (grant no. 312062). M.G. was supported by EU Horizon 2020 (grant 733100 for SYSCID) and a grant from the Excellence of Science (FNRS and FWO) (grant no. 30770923). We acknowledge support from the BBMRI-NL (Biobanking and Biomolecular Resources Research Infrastructure 184.021.007 and 184.033.111), Spinozapremie (NWO 56-464-14192), the ERC (ERC Advanced 230374) and the KNAW Academy Professor Award (PAH/6635) to D.I.B. G.H. works in a unit that receives funding from the UK MRC (MC_UU_12013/1&2&5) and the University of Bristol. S.B. was supported by the Swiss National Science Foundation (310030-152724). B.M.P. was supported by CHARGE infrastructure grant number HJ105756 for the HVH cohort. This work was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept (grant 01ZX1906B) and by LIFE (Leipzig Research Center for Civilization Diseases), Universität Leipzig (which is funded by the European Union, by the European Regional Development Fund and by the Free State of Saxony within the framework of the excellence initiative to H.K. and M.Scholz). We thank the UMCG Genomics Coordination Center, the MOLGENIS team, the UG Center for Information Technology and the UMCG research IT program and their sponsors, in particular the BBMRI-NL for data storage, high-performance computing and web hosting infrastructure. The BBMRI-NL is a research infrastructure financed by the NWO (grant number 184.033.111). We thank K. McIntyre for editing the manuscript text.

Author contributions

U.Võsa. and A.C. coordinated consortium analyses, ran meta-analyses, interpreted data, performed downstream analyses and drafted and revised the manuscript. H.-J.W., M.J.B. and P.D. developed software used in the analyses, performed downstream analyses and participated in manuscript writing and revisions. L.F. and T.E. conceived the study. L.F. supervised the project, ran downstream analyses and participated in manuscript writing and revisions. B.Z., H.K., A.S., S.K., N.P., I.A., M.-J.F., M.A., M.W.C., R.J., I.S., L.T., A.Teumer., K.S., J.V., H.Y., V.K., A.K., J. Kettunen, J.P. and B.L. ran consortium analyses in their respective cohorts. A.S., R.K., S.K., G.H., R.S. and A.Brown ran replication analyses in their respective cohorts. A.A., G.W.M., S.Ripatti, M.P., E.D., S.B., T.F., J.v.M., H.P., H.A., B.Pierce., T.L., D.I.B., B.M.P., S.A.G., P.A., L.M., W.H.O., K.D., O.S., A.Battle, M.Scholz, G.G., T.E., W.A., F.B., J.D., M.E., B.P.F., M.G., B.T.H., M.K., Y.K., J.C.K., P.K., K.K., M.L., U.M.M., H.M., Y.M., M.M.-N., M.Nauck, M.G.N., B.W.J.H.P., O.T.R., O.Rotzschke, E.P.S., C.D.A.S., M.Stumvoll, P.S., P.A.C.'t.H., J.T., A.Tönjes, J.v.D., M.v.I., J.H.V., U.Völker and C.W. provided data used in the study. B.Z., H.K., Z.K., J.Kronberg, S.Rüeger, E.P., S.L., J.Y., F.Z., P.M.V., J.P., T.Q., R.W., H.K., M.Scholz and G.G. participated in downstream analyses. S.Y., H.B., R.O., D.H.d.V. and M.G.P.v.d.W. ran replication analyses in scRNA-seq cohorts. A.W.H., J.A.H. and J.P. generated scRNA-seq replication data. H.K., A.Teumer., M.G., M.G.N., J.P., Z.K., J.Y., P.M.V., M.Scholz, G.G., J.P., S.A.G. and P.A.C.'t.H. contributed to writing and revising the manuscript. J.K.P. provided Supplementary Equations for interpretation of results. H.B. and M.Swertz created the website to host results. H-J.W., M.J.B. and P.D. contributed equally to this work. The BIOS Consortium contributed the subset of whole-blood data used in discovery analyses. The i2QTL Consortium contributed *trans*-eQTL and eQTS replication analyses of iPSCs.

Competing interests

B.M.P. serves on the Steering Committee for the Yale Open Data Access Project funded by Johnson & Johnson. This activity is unrelated to this work. The rest of the authors declare no competing interests.

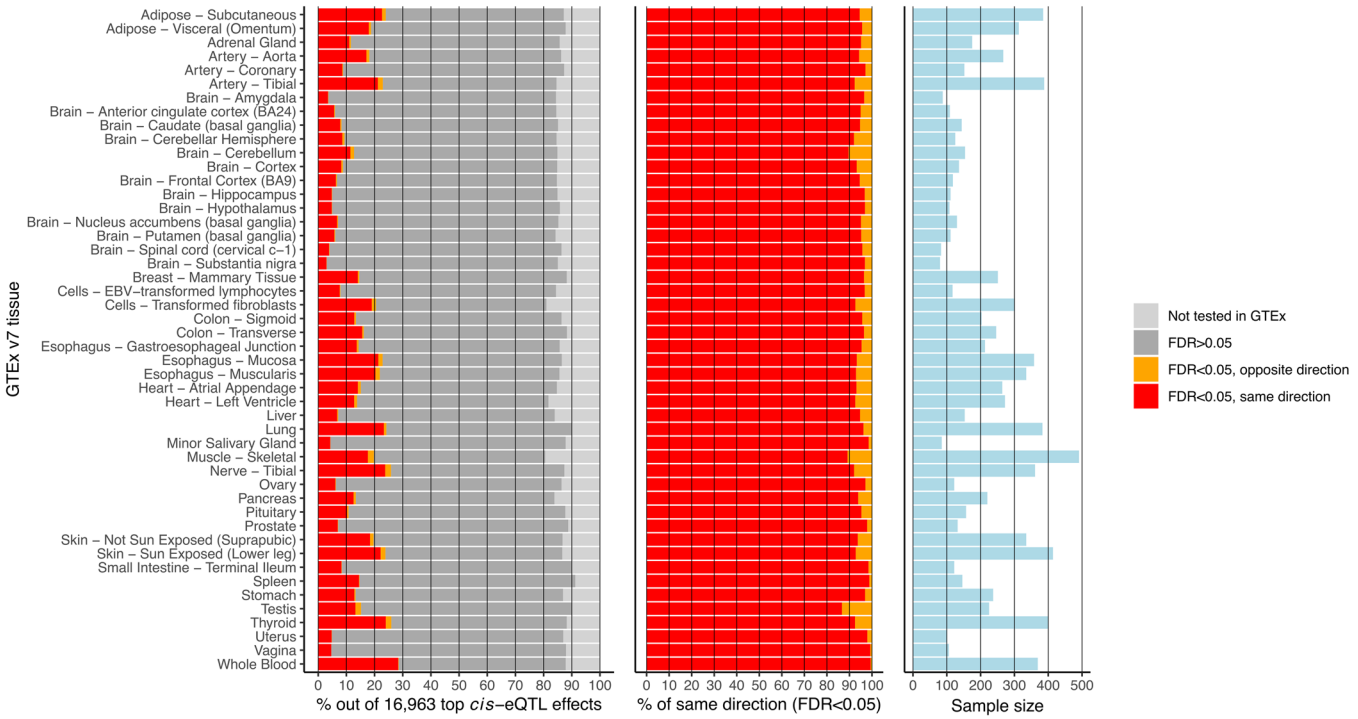
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-021-00913-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00913-z.

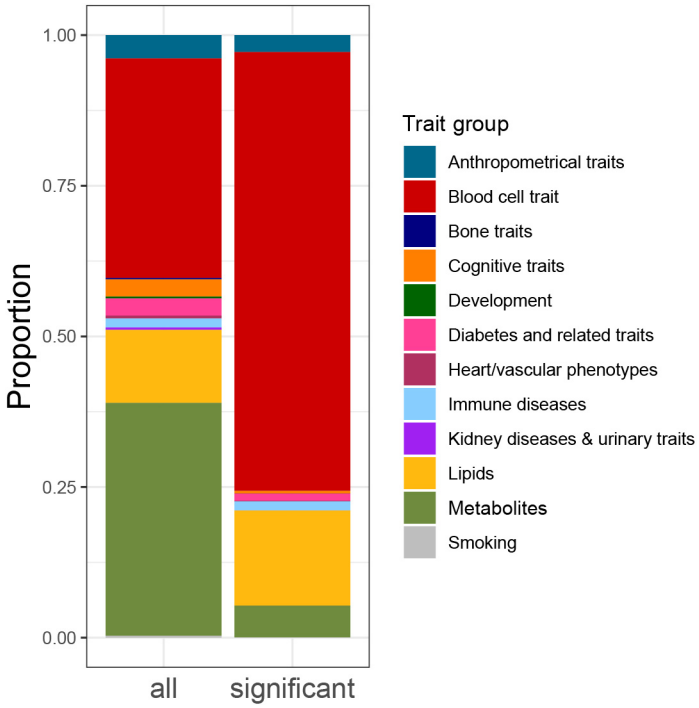
Correspondence and requests for materials should be addressed to Urmo Võsa, Annique Claringbould or Lude Franke.

Peer review information *Nature Genetics* thanks Eric Gamazon, Douglas Yao, and Vijay Sankaran for their contribution to the peer review of this work.



Extended Data Fig. 1 | *Cis*-eQTL replication in GTEx v7 tissues. *Cis*-eQTL replication in GTEx v7 tissues. For this analysis, the most significant *cis*-eQTL SNP for each gene was tested in the available post-mortem tissues in GTEx v7. Since GTEx was part of our discovery meta-analysis, the *cis*-eQTL discovery analysis was repeated while excluding GTEx whole blood, identifying 16,963 lead *cis*-eQTL effects that were subsequently replicated in each GTEx tissue. Left: while the majority of the 16,963 *cis*-eQTL were tested in the GTEx replication study, a relatively small fraction had an FDR < 0.05. Middle: of those *cis*-eQTL showing a replication FDR < 0.05, allelic directions were highly consistent with the discovery meta-analysis. Right: sample sizes of GTEx tissues. Limited replication rates at FDR < 0.05 were probably due to the relatively small sample size per GTEx tissue.

Extended Data Fig. 2 | Dot-plot showing the locations of the *trans*-eQTL effects identified in discovery meta-analysis and their association P-values (-$\log_{10}$ scale). Dot-plot showing the locations of the *trans*-eQTL effects identified in discovery meta-analysis (weighted Z-score meta-analysis on Spearman correlation) and their respective two-sided association P-values in -$\log_{10}$ scale. SNP positions are shown on the x-axis and gene locations on the y-axis, each dot shows one significant *trans*-eQTL effect (FDR < 0.05). Vertical bands appear where a single genomic locus affects many genes in *trans*, while horizontal bands illustrate genes affected by many SNPs.



Extended Data Fig. 3 | Overview of GWAS trait classes in eQTS analysis. Overview of tested and significant (FDR < 0.05) GWAS trait classes in eQTS analysis.