Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation

Eric R. Gamazon1,2,21*, Ayellet V. Segrè3,4,21*, Martijn van de Bunt3,6,21, Xiaoquan Wen7, Hualin S. Xi8, Farhad Hormozdiari9,10, Halit Ongen11,12,13, Anuar Konkashbaev1, Eske M. Derks14, François Aguet3, Jie Quan8, GTEx Consortium15, Dan L. Nicolae16,17,18, Eleazar Eskin9, Manolis Kellis3,19, Gad Getz3,20, Mark I. McCarthy3,6, Emmanuel T. Dermitzakis11,12,13, Nancy J. Cox1 and Anuar Konkashbaev1, Eske M. Derks3,20, Mark I. McCarthy3,4,21*, Martijn van de Bunt.*e-mail: egamazon@uchicago.edu; asegre@broadinstitute.org

We apply integrative approaches to expression quantitative loci (eQTLs) from 44 tissues from the Genotype-Tissue Expression project and genome-wide association study data. About 60% of known trait-associated loci are in linkage disequilibrium with a cis-eQTL, over half of which were not found in previous large-scale whole blood studies. Applying polygenic analyses to metabolic, cardiovascular, anthropometric, autoimmune, and neurodegenerative traits, we find that eQTLs are significantly enriched for trait associations in relevant pathogenic tissues and explain a substantial proportion of the heritability (40–80%). For most traits, tissue-shared eQTLs underlie a greater proportion of trait associations, although tissue-specific eQTLs have a greater contribution to some traits, such as blood pressure. By integrating information from biological pathways with eQTL target genes and applying a gene-based approach, we validate previously implicated causal genes and pathways, and propose new variant and gene associations for several complex traits, which we replicate in the UK Biobank and BioVU.

A primary goal of the Genotype-Tissue Expression (GTEX) project is to elucidate the biological basis of genome-wide association study (GWAS) findings for a range of complex traits, by measuring eQTLs in a broad collection of normal human tissues. Several recent papers have described the GTEx v6p data, where cis-eQTLs were mapped for 44 tissues from a total of 449 individuals (70–361 samples per tissue) using a single-tissue method that detects eQTLs in each tissue separately, and a multi-tissue method that increases the power to detect weak-effect eQTLs. Here, we leverage the extensive resource of regulatory variation from multiple tissues to elucidate the causal genes for various GWAS loci and to assess their tissue specificity (Fig. 1a). We highlight the challenges of using eQTL data for the functional interpretation of GWAS findings and identification of tissue of action. Using several polygenic approaches (Table 1), we provide comprehensive analyses of the contribution of eQTLs to trait variation. Finally, by integrating eQTL with pathway analysis, and replication in DNA biobanks tied to electronic health records (UK Biobank1 and BioVU; see URLs), we propose new trait associations and causal genes for follow-up analyses for a range of complex traits.

Results
Relevance of eQTLs from 44 tissues to trait associations. We tested the extent to which cis-eQTLs (using the ‘best eQTL per eGene’ at a genome-wide false discovery rate (FDR) ≤0.05 per tissue) from each of the 44 tissues were enriched for trait associations (GWAS P ≤0.05) using eQTLEnrich (Methods, Supplementary Fig. 1). Testing 18 complex traits (metabolic, cardiovascular, anthropometric, autoimmune, and neurodegenerative, listed in Supplementary Table 1) with available GWAS summary statistics, we found significant enrichment for trait associations amongst eQTLs (Bonferroni-adjusted P <6.3×10−5) for 11% of 792 tissue-trait pairs tested, with a median fold-enrichment per trait ranging from 1.19 to 5.75 (Fig. 1b, Supplementary Table 2), and different tissues significant per trait (Supplementary Fig. 2). The enrichment results also suggest hundreds of modest-effect associations amongst
eQTLs in various tissues for all traits tested (Supplementary Fig. 3, Supplementary Table 2). While the adjusted fold-enrichment (Methods) is unaffected by differences in number of eQTLs per tissue (Supplementary Fig. 4), increased enrichment was observed for GWAS with larger sample sizes, such as for height (\(N > 250,000\)), where there is greater detection power (Fig. 1c). Enrichment amongst eQTLs was also found for less-powered GWAS, such as HOMA-IR (\(N \sim 37,000\)), where no variants passed genome-wide significance (Supplementary Fig. 5). The tissues in which eQTLs were most strongly enriched for trait associations included relevant tissues, such as aortic artery for systolic blood pressure (SBP), coronary artery for coronary artery disease (CAD), skeletal muscle for type 2 diabetes (T2D), colon for Crohn’s disease, and hippocampus for Alzheimer’s disease (Fig. 1d, Supplementary Table 2). However, the most enriched tissues per trait also included less biologically obvious tissues, suggesting either shared regulation with the actual tissues of action or new pathogenic tissues. Notably, eQTLs in (commonly studied) whole blood were enriched for associations with about half of the traits tested (\(P < 6.3 \times 10^{-5}\); for example, ulcerative colitis, low-density lipoprotein cholesterol (LDL), and

---

**Fig. 1** Incorporating eQTLs from 44 tissues into GWAS of complex traits. **a**, Schematic diagram demonstrating how eQTL annotation from various tissues can be used to propose one or more potential causal genes whose regulation is either tissue-specific (orange) or tissue-shared (blue) for a trait-associated (GWAS) variant. A gene close to the trait-associated variant (gray) may have an eQTL that is not in linkage disequilibrium (LD) with the trait-associated variant. **b**, Fold-enrichment of eQTLs (FDR \(\leq 0.05\)) with GWAS \(P \leq 0.05\) compared to a null distribution of GWAS \(P\) values (Methods), shown for 44 tissues by 18 complex traits (abbreviations in Supplementary Table 1). Red circles, tissue-trait pairs that pass Bonferroni correction (\(P < 6.3 \times 10^{-5}\); 89 out of 792 tissue-trait pairs tested); dashed line, median fold-enrichment of all significant tissue-trait pairs. The ‘best eQTL per eGene’ set per tissue was used. **c**, Quantile-quantile (Q-Q) plot of variant association \(P\) values from a large GWAS meta-analysis of height (\(n = 253,288\)) for all variants tested (black), and for eQTLs in tissues most highly enriched for height associations: pituitary (green), stomach (peach), and esophagus muscularis (brown). All significant variant-gene eQTL pairs were plotted. **d**, Top-ranked tissues based on their adjusted fold-enrichment of trait associations amongst eQTLs (compared to the best eQTL for all non-significant eGenes) that pass Bonferroni correction (\(P < 6.3 \times 10^{-5}\)) for T2D (\(n = 69,033\)), Alzheimer’s disease (AD, \(n = 54,162\)), CAD (\(n = 184,405\)), and SBP (\(n = 69,395\)) (Methods, Supplementary Table 2). Estimated lower and upper bound 95% confidence intervals for the adjusted fold-enrichment are shown (Methods). EBV, Epstein–Barr virus. Chr, chromosome.
high-density lipoprotein cholesterol; Supplementary Table 2), demonstrating the utility of blood for broadly studying the underlying genetic mechanisms of some associations, but also emphasizing the importance of studying gene regulation in a biologically diverse set of disease-relevant tissues.

Applying a Bayesian-based enrichment method that accounts for eQTL effect size and considers all significant variant-gene set of disease-relevant tissues.

## Table 1 | Summary of polygenic methods used to test contribution of eQTLs to trait variation

| Method | Goal | Description and assumptions | Limitations | eQTL set used | GWAS data types |
|--------|------|-----------------------------|-------------|---------------|----------------|
| eQTLEnrich, rank- and permutation-based GWAS-eQTL enrichment method | Tests whether eQTLs from a given tissue are significantly enriched for trait associations more than would be expected by chance and estimates adjusted fold-enrichment. | Estimates the probability of observing a given fold-enrichment of top-ranked trait associations (for example, GWAS $P \leq 0.05$) amongst eQTLs in a given tissue, relative to the fold-enrichment of non-significant eVariants (adjusted fold-enrichment), using a null distribution derived from multiple randomly sampled variants matched on MAF, distance to TSS, and local linkage disequilibrium. Per GWAS tested, tissues are ranked based on their adjusted fold-enrichment. | Adjusted fold-enrichment is correlated with GWAS sample size. | Best eQTL per eGene | Variant association $P$ values |
| TORUS, Bayesian and Maximum Likelihood Estimation (MLE) approach for quantifying GWAS-eQTL enrichment | Estimates an enrichment parameter that represents the relationship between the log odds ratio of the trait associations being causal and their eQTL effect size. | Estimates the relationship between the (absolute value of) single variant eQTL $z$-scores and the corresponding log odds of a variant being causally associated with the complex trait of interest. A confident positive estimate of the log odds ratio indicates the increased odds of a variant being causally associated with the trait with stronger effect of eQTL association. Uses $z$-scores from all gene-variant pairs for a given tissue, and assumes a single causal trait association per linkage disequilibrium block (following the assumption of fgwas). | Enrichment parameter estimation (especially standard error) is correlated with tissue sample size of eQTLs. | All variant-gene pairs tested | Variant association test statistics |
| $\pi$, method | Estimates the fraction of eQTLs in a given tissue that are likely to be associated with a given complex trait. | Estimates the fraction of true trait associations amongst eQTLs in a given tissue, using the $\pi_1$ statistic, which assumes a standard uniform distribution for the null distribution and independence between variants. | Results not robust to small variant sets. | Best eQTL per eGene | Variant association $P$ values |
| Summary statistics-based heritability estimation | Estimates the relative contribution of eQTLs in aggregate to the heritability of complex traits, using linkage disequilibrium score regression applied to publicly available GWAS summary statistics. | Estimates the per-variant effect of the trait association by an annotated eQTL versus an unannotated variant. A larger difference indicates a higher degree of enrichment of contribution of eQTLs to trait associations. | Works optimally when the per-variant variance is not correlated with the linkage disequilibrium score. | All significant variant-gene pairs | Variant association test statistics |
| Mixed-effects model heritability estimation | Estimates proportion of complex trait variance explained by eQTL variants in aggregate using GWAS genotype data. | Estimates the heritability attributable to eQTL variants using the Restricted Maximum Likelihood approach. The approach assumes a normal distribution of trait effect sizes for the eQTL variants and uses a genetic similarity matrix generated from the eQTL variants. | Requires genotype data. | All significant variant-gene pairs | Individual genotype data |

$^a$See URLs for links to methods software.

### Results not robust to GWAS sample size.

Fewer tissue-trait pairs when restricting to tissue-specific eQTLs (Supplementary Table 4, Supplementary Fig. 7b) than with all eQTLs (Supplementary Table 2). Among the top results were adipose-specific eQTLs for diastolic blood pressure (DBP) and aorta-specific eQTLs for SBP, proposing different tissue-specific processes that may underlie DBP and SBP.

### Cis-eQTL characterization of known trait associations.

Since regulatory effects are enriched for top-ranked trait associations, we asked how many of the genome-wide significant associations ($P < 5 \times 10^{-8}$) from the NHGRI-EBI GWAS catalog might be acting via eQTLs, and in what tissues. We annotated 5,895 genome-wide significant associations ($P < 5 \times 10^{-8}$; hereafter ‘trait-associated variants’), identified primarily in samples of European descent (Supplementary Table 5), with GTex eQTLs from single-tissue...
Further, a common assumption is that the nearest gene to the trait-associated variant is the probable causal gene. However, for only ~50% of trait-associated variants in linkage disequilibrium with at least 1 eQTL was the target gene the nearest gene, illustrating the limitations of proximity-based assignment in identifying potentially causal genes. In addition, the distance of eQTLs in linkage disequilibrium with trait-associated variants to the transcription start site (TSS) of their target gene was significantly greater than that of all other eQTLs (Wilcoxon rank sum \( P = 3.0 \times 10^{-18} \)), and more likely to be downstream of the TSS (Fig. 2d, Supplementary Fig. 8).

Since eQTLs are ubiquitous in the genome, linkage disequilibrium between an eQTL and trait-associated variant can occur by chance. Hence, we applied two co-localization methods, Regulatory Trait Concordance13,14 and eCAVIAR15 (Supplementary Note), to three traits: SBP, DBP, and CAD. Out of 21 (SBP), 19 (DBP), and 37 (CAD) associated variants (\( P < 5 \times 10^{-8} \)), which are in linkage disequilibrium with an eQTL, there is co-localization support for 67%, 58%, and 32% of the loci, respectively, by at least 1 of the methods (Supplementary Table 9, Supplementary Fig. 9). Some high-confidence genes suggested by high-linkage disequilibrium and supported by both co-localization methods include rs1412444-LIPA and rs6544713-ABCG8 for CAD, rs173771-NPR3 and rs17477177 with CCDC71L and CTB-30L5.1 (a lincRNA) for SBP, and rs2521501-MAN2A2 for both SBP and DBP (results and significant tissues in Supplementary Table 9). For CAD, the lead variant (rs6544713)16, located in the intron of ABCG8, is in almost complete linkage disequilibrium (\( r^2 = 0.99 \)) with the best eQTL for ABCG8 (rs4245791; Fig. 3a), which is specific to transverse colon (Fig. 3b) and has a 2.45-fold effect on expression17 (ALT versus REF allele). ABCG8 plays a critical role in cholesterol metabolism by limiting intestinal dietary sterol uptake and by secreting sterol into bile. Recessive mutations in ABCG8 cause sitosterolemia, a disorder characterized by premature atherosclerosis and abnormal sterol accumulation18. The minor T-allele at rs6544713 is associated with lower expression of ABCG8 in transverse colon (Fig. 3c), and increased CAD risk and higher LDL levels18. The three top eQTLs for ABCG8, which are in strong linkage disequilibrium with the CAD-associated variant rs6544713 (\( r^2 > 0.95 \)), overlap gastrointestinal and liver enhancers based on Roadmap Epigenomics Project19 data.

Breadth versus depth of tissues in eQTL analysis of GWAS loci.

Most eQTL analyses have been limited to a few readily accessible tissues (primarily blood), although with large sample sizes (900–5,000). A specific goal of the GTEx study, in contrast, was to survey a wide range of (often inaccessible) tissues from the body, although with necessarily smaller sample sizes. To assess the relative value of breadth in sample type versus depth of sample size in the functional characterization of trait associations, we compared cis-eQTLs found in at least 1 of the 44 tissues to those discovered in 2 large cis-eQTL studies of whole blood (Depression Genes and Networks (DGN)20,21 \( n = 922 \); Westra et al.\(^2\) \( n = 5,531 \)). We found that 80% of all ‘best eQTL per eGene’ variants and 63% of all eGenes found in ≥1 tissue in GTEx were not found in DGN, an RNA sequencing-based study (FDR <0.05; Methods, Fig. 3d). Of just the subset of eQTLs in linkage disequilibrium (\( r^2 > 0.8 \)) with 467 independent trait-associated variants from the GWAS catalog, 62% were not found in DGN, and, of these, 82% were not significant in GTEx whole blood (Fisher’s exact \( P = 3.3 \times 10^{-22} \); Fig. 3d). Due to differences in analytical methods, we also inspected the overlap at the eGene level. Importantly, 47% of all eGenes identified in GTEx across the 44 tissues were not found in DGN, of which 81% were identified only in non-blood tissues in GTEx (Fisher’s exact test \( P = 1.1 \times 10^{-15} \); Fig. 3d). In contrast, only 3% of DGN eGenes were not detected in GTEx in any of the 44 tissues, even though DGN detected 1.3-fold more eGenes than GTEx in whole blood. Notably, the GTEx eQTLs not found in DGN, in particular...
non-blood eQTLs, tended to be more tissue-specific than GTEx eQTLs that were also found in the larger DGN blood study (Wilcoxon rank sum \(P=1.0 \times 10^{-16}\); Fig. 3e). Similar patterns were observed with the much larger, microarray-based study by Westra and colleagues (Methods, Supplementary Figs. 10 and 11). Hence, while larger studies provide better discovery power for a specific tissue of interest, there is great value to the diversity of tissues in proposing new biological hypotheses, especially tissue-specific ones, for a considerable number of trait associations (examples listed in Supplementary Tables 10 and 11).

**Trait heritability attributable to cis-eQTLs.** To quantify the proportion of genetic contribution to trait variation (heritability) that may be attributed to regulatory variation from across the 44 tissues, we applied (summary statistics-based) linkage disequilibrium score regression (LDSR)\(^{24}\) to 15 of the 18 traits tested for enrichment above, with available GWAS meta-analysis effect sizes (Supplementary Table 1, Methods). Using all significant (single-tissue) eQTL variant-gene pairs from the 44 tissues, we found that while the eQTLs comprise on average 33% of the variants tested in all GWAS analyses, they explained 52.1% of the variant-based heritability, showing a 1.6-fold concentration of heritability (Methods; Fig. 4a, Supplementary Table 12). The combined set of eQTLs explains from 38.0 ± 2.7% (for body mass index (BMI)) to 78.2 ± 15.2% (for Alzheimer’s disease) of the traits’ heritability (Supplementary Table 12), of which 10–16% are tissue-specific eQTLs (Methods,
Supplementary Table 13). By restricting our analysis to the top 10 eQTLs per eGene, which are likely to be enriched for causal variants\(^1\), proportionately, we found an even greater contribution of eQTLs to the variant-based heritability (3.2-fold concentration of heritability); Fig. 4a, Supplementary Table 14). Considering the proportion of heritability was for conserved genomic regions, and the lowest for non-conserved genomic features\(^2\), we found the highest concentration of heritability attributed to those eQTLs that target ‘tissue-specific genes’ (that is, genes showing higher expression in a given tissue than in all other tissues; Methods) using LDSR, and found it to be a limited fraction of the heritability attributed to all eQTLs (Fig. 4d, Supplementary Table 17). Biologically plausible patterns of tissue-specific heritability concentration were observed across the different traits analyzed (Supplementary Fig. 12, Supplementary Note).

To conduct tissue-specific assessment of the eQTL contribution to heritability, we evaluated the proportion of heritability attributed to eQTLs across tissues for several Wellcome Trust Case Control Consortium traits\(^3\), where GWAS sample size is identical for all traits and genotype data are available, and also found biologically plausible (tissue- and trait-dependent) patterns of co-localization.

Fig. 3 | Proposing causal genes in inaccessible tissues. a, LocusZoom\(^4\) plot showing that the lead variant at ABCG5/8 locus for CAD (\(n = 184,405\)) and LDL cholesterol (\(n = 95,454\)) (rs6544713; purple diamond) is in linkage disequilibrium (\(r^2 = 0.99\), and colocalizes, with an eQTL signal for ABCG8 in transverse colon, using eCAVIAR and Regulatory Trait Concordance. No other gene in the locus was implicated based on linkage disequilibrium or co-localization. b, Forest PM-plot\(^5\) of single-tissue eQTL \(-\log_{10}(P\text{-value})\) against the METASOFT posterior probability, \(m\text{-value}\) (indicating multi-tissue support), demonstrating that rs6544713-ABCG8 eQTL is specific to transverse colon. c, Box plot showing correlation between rs6544713 and normalized expression in transverse colon, corrected for covariates used in cis-eQTL analysis. Box edges depict interquartile range, whiskers 1.5 times the interquartile range, and center lines the median. Minor T-allele, associated with lower expression, is associated with increased CAD risk and higher LDL\(^6\). d, Fraction of best eQTL per eGene (‘eQTLs’) or ‘eGenes’ significant in at least one GTEx tissue identified (yellow and purple) or not identified (blue and red) in DGN blood study at FDR \(\leq 0.05\), further stratified by being significant (FDR \(\leq 0.05\)) (blue and yellow) or non-significant (red and purple) in GTEx blood. e, Distribution of number of significant tissues per ‘best eQTL per eGene’ (FDR \(\leq 0.05\)) sets in DGN blood. We compared all (21,643) eQTLs in GTEx (‘All’) to the subset of eQTLs in linkage disequilibrium (\(r^2 \geq 0.8\)) with a GWAS variant (‘GWAS’; 471 independent trait-associated variants from GWAS catalog). 

To conduct tissue-specific assessment of the eQTL contribution to heritability, we evaluated the proportion of heritability attributed to those eQTLs that target ‘tissue-specific genes’ (that is, genes showing higher expression in a given tissue than in all other tissues; Methods) using LDSR, and found it to be a limited fraction of the heritability attributed to all eQTLs (Fig. 4d, Supplementary Table 17). Biologically plausible patterns of tissue-specific heritability concentration were observed across the different traits analyzed (Supplementary Fig. 12, Supplementary Note).

Since the estimated proportion of heritability is modestly correlated with GWAS sample size (which explains \(R^2 = 2.3–13.7\%\) of variance in LDSR-derived heritability; Supplementary Fig. 13c,f), we investigated the pattern of heritability attributed to eQTLs across tissues for several Wellcome Trust Case Control Consortium traits\(^3\), where GWAS sample size is identical for all traits and genotype data are available, and also found biologically plausible (tissue- and trait-dependent) patterns of co-localization.
we found that Crohn’s disease and ulcerative colitis clustered traits showed high $\pi$ in only a subset of tissues (Supplementary Table 19), suggesting that hundreds of trait associations, and up to 1,551 trait associations across all tissue-trait pairs tested). Consistent with the eQTLEnrich results, the anthropometric (height and BMI) and autoimmune (Crohn’s disease and ulcerative colitis) traits showed high $\pi$ in most tissues, while other traits showed high $\pi$ in only a subset of tissues (Supplementary Fig. 15). Clustering traits on the basis of $\pi$ across tissues (Methods), we found that Crohn’s disease and ulcerative colitis clustered together (Pearson’s $r = 0.39$, $P = 0.008$), suggesting that eQTLs may contribute substantially to the known genetic correlation between these traits; waist-to-hip ratio clustered with T2D, more strongly than with BMI (Pearson’s $r = 0.37$, $P = 0.01$ versus Pearson’s $r = 0.12$, $P = 0.44$), consistent with reports that waist-to-hip ratio is a better predictor of T2D than BMI; and CAD clustered with SBP, a known CAD risk factor (Supplementary Fig. 15).

Similar to the eQTLEnrich analysis, the tissues with highest estimated $\pi$ contained relevant pathogenic tissues, such as hippocampus for Alzheimer’s disease and skeletal muscle for T2D, but also less obvious tissues, such as the reproductive tissues. We therefore examined the relative contribution of tissue-specific eQTLs (significant in at most 10% of tissues) versus tissue-shared eQTLs (significant in over 90% of tissues) to trait associations (Methods). Most traits showed, on average, higher absolute numbers and higher rates of trait associations ($\pi$) among tissue-shared eQTLs (median $\pi = 9.3\%$, range: 0–88%) relative to tissue-specific eQTLs (median $\pi = 5.6\%$, range: 0–87%) (Fig. 5c, Supplementary Fig. 17a, Supplementary Table 19). Thus, at least some of the less obvious tissues with high $\pi$ are capturing some component of shared regulation with the actual pathogenic tissues. On the other hand, two-hour glucose tolerance levels (2hGlu), SBP, and DBP showed on average a larger number of tissue-specific versus tissue-shared heritability (Supplementary Fig. 14, Supplementary Table 18 and Supplementary Note).

Using eQTLs to discover new trait associations and genes. Since many more associations are likely to underlie trait variation than those currently passing genome-wide significance (for example, Fig. 1b, Supplementary Fig. 3), we tested whether we could use eQTLs to identify novel associations, and to propose causal genes and potential tissues of action for these associations. We estimated the true positive rate ($\pi$, statistic) of trait associations amongst eQTLs (using the ‘best eQTL per eGene’ sets) in the 44 tissues for the 18 traits tested above (Methods). The average $\pi$ across the 44 tissues per trait ranged from 2.9% to 45.5% for the 18 traits (Fig. 5a, Supplementary Table 19), suggesting that hundreds of trait associations, and up to 1,551 trait associations across all tissue-trait pairs tested). Consistent with the eQTLEnrich results, the anthropometric (height and BMI) and autoimmune (Crohn’s disease and ulcerative colitis) traits showed high $\pi$ in most tissues, while other traits showed high $\pi$ in only a subset of tissues (Supplementary Fig. 15). Clustering traits on the basis of $\pi$ across tissues (Methods), we found that Crohn’s disease and ulcerative colitis clustered together (Pearson’s $r = 0.39$, $P = 0.008$), suggesting that eQTLs may contribute substantially to the known genetic correlation between these traits; waist-to-hip ratio clustered with T2D, more strongly than with BMI (Pearson’s $r = 0.37$, $P = 0.01$ versus Pearson’s $r = 0.12$, $P = 0.44$), consistent with reports that waist-to-hip ratio is a better predictor of T2D than BMI; and CAD clustered with SBP, a known CAD risk factor (Supplementary Fig. 15).

Similar to the eQTLEnrich analysis, the tissues with highest estimated $\pi$ contained relevant pathogenic tissues, such as hippocampus for Alzheimer’s disease and skeletal muscle for T2D, but also less obvious tissues, such as the reproductive tissues. We therefore examined the relative contribution of tissue-specific eQTLs (significant in at most 10% of tissues) versus tissue-shared eQTLs (significant in over 90% of tissues) to trait associations (Methods). Most traits showed, on average, higher absolute numbers and higher rates of trait associations ($\pi$) among tissue-shared eQTLs (median $\pi = 9.3\%$, range: 0–88%) relative to tissue-specific eQTLs (median $\pi = 5.6\%$, range: 0–87%) (Fig. 5c, Supplementary Fig. 17a, Supplementary Table 19). Thus, at least some of the less obvious tissues with high $\pi$ are capturing some component of shared regulation with the actual pathogenic tissues. On the other hand, two-hour glucose tolerance levels (2hGlu), SBP, and DBP showed on average a larger number of tissue-specific versus tissue-shared
Tissue-specific eQTLs among 44 tissues.

Fig. 5 | Estimated true positive rate of trait associations amongst eQTLs in 44 tissues. a, Distribution of estimated true positive rate ($\pi$, statistic) of trait associations (considering the full spectrum of GWAS $P$ values) amongst eQTLs across 44 tissues shown for 18 complex traits (abbreviations in Supplementary Table 1). $\pi_1$ ($\pi$, statistic), estimated true positive rate. b, Estimated number of true trait associations that are eQTLs in each of the 44 tissues, computed for 18 complex traits by multiplying $\pi_1$ by the number of eQTLs identified per GWAS. The median number of trait associations ranges from 0 to 554, with a median of 80 trait associations per tissue-trait pair (dashed line) and a maximum of 1,551 for Crohn's disease. These are lower bound estimates due to incomplete overlap of variants between the GTEx and GWAS studies (Methods). c, Distribution of estimated true positive rate ($\pi$, statistic) of trait associations amongst tissue-specific eQTLs (yellow; significant in about $\leq$10% of tissues including the given tissue, based on METASOFT) versus tissue-shared eQTLs (pink; significant in $\geq$90% of tissues and the given tissue, based on METASOFT) computed for 44 tissues by 18 traits. The 'best eQTL per eGene' set per tissue was used for all analyses (Supplementary Table 19). a, c, b, The boxes depict the interquartile range, center lines show the median, and ‘+’ represents the outliers.

eQTLs amongst their trait associations (Methods; Supplementary Fig. 17b, Supplementary Table 19). This result persisted after normalizing for differences in number of tissue-specific and tissue-shared eQTLs in each tissue (Supplementary Fig. 17c) and was not dependent on GWAS sample size (Supplementary Fig. 18).

To identify the true positive trait associations that contribute to the observed enrichment, we searched for target genes of eQTLs with top-ranked GWAS $P$ values ($P \leq 0.05$) that are enriched in biological pathways or functionally related gene sets, such as genes that share mouse knock-out phenotypes. We applied eGeneEnrich (Methods) to several tissue-trait pairs (Supplementary Table 20) for a number of traits (Alzheimer’s disease, CAD, LDL, SBP, and T2D) that showed significant enrichment based on eQTLEnrich or $\pi$, estimates, both of which are not affected by tissue sample size (Supplementary Figs. 4 and 16; Supplementary Tables 2 and 19). Multiple gene sets were nominally enriched (eGeneEnrich adjusted $P < 0.05$) for each tissue-trait pair tested (Supplementary Table 20). The proposed causal genes and corresponding best eQTLs were then tested for replication in large-scale biobanks (see below).

To identify tissue-specific processes, we also applied eGeneEnrich to target genes of tissue-specific eQTLs. We analyzed the target genes of aorta-specific eQTLs with SBP $P < 0.05$ (that showed one of the strongest tissue-specific eQTL-GWAS enrichments; Supplementary Table 4), using a GWAS meta-analysis of 69,000 individuals, and found significant enrichment in gene sets related to body weight and the cardiovascular system. These gene sets suggested, for example, an aorta-specific eQTL acting on two protein-coding genes, GUCY1A3 and GUCY1B3, and a non-coding gene, RP11-588K22.2, as a novel association with SBP (Fig. 6a,b). Notably, the best aorta eQTL for GUCY1B3 (rs4691707) was recently reported as genomewide significant in a 5-fold larger GWAS meta-analysis of ~342,000 individuals, but aorta would have not been prioritized as a tissue of action, based solely on the expression of GUCY1B3 or GUCY1A3 across tissues (Fig. 6c, Supplementary Fig. 19).

We tested for independent support for the proposed causal target genes from the discovery gene set analysis (eGeneEnrich adjusted $P < 0.05$) in two large-scale repositories—UK Biobank, a prospective study with extensive phenotypic data, and BioVU, an electronic health records-linked DNA biobank (Methods). First, using the gene-level association method, PrediXcan, we evaluated the contribution of the genetic component of gene expression to trait variance in the UK Biobank for two traits with sufficient sample size: SBP and myocardial infarction, a proxy for CAD (Methods). The eGeneEnrich–proposed causal genes for SBP in aorta artery or
**Fig. 6** Discovery and replication of novel associations and genes. **a**, PM-plot of best eQTL for GUCY1B3 in artery aorta (n = 197) (rs4691707) showing -log_{10} (P value) from single-tissue eQTL analysis versus the multi-tissue m-value. **b**, rs4691707 is also an eQTL for GUCY1A3, although less specific to artery aorta, being significant (m-value ≥0.9) also in nerve tibial (n = 256) and thyroid (n = 278). **c**, Violin plots of GUCY1B3 expression across 44 tissues. Overlaid boxes indicate interquartile ranges and center-lines the median. Artery aorta is not the top-ranked tissue for GUCY1B3 based on expression -value. **d-e**, Box plots of PrediXcan P values (−log_{10}) with UK Biobank GWAS for SBP and aorta artery genes (**d**) and myocardial infarction (MI) and coronary artery genes (**e**), comparing eGeneEnhch-proposed causal genes to remaining genes expressed in the corresponding tissues. For both traits, proposed genes show significantly lower P values, as assessed by Wilcoxon rank sum one-tailed test (P = 1.5 × 10^{-7} for **d**, P = 5.8 × 10^{-5} for **e**). The boxes indicate interquartile ranges, whiskers 1.5 x interquartile range, center-lines median values, and ‘+’ represents the outliers. **f**, Q-Q plot of replication association P values from UK Biobank GWAS of SBP for artery aorta eQTLs (purple), enriched for SBP associations in a discovery GWAS, compared to 100 null variant sets (gray; empirical P < 0.01). **g**, Q-Q plot of replication association P values from a UK Biobank GWAS of myocardial infarction for coronary artery eQTLs (orange), enriched for CAD associations in a discovery GWAS, compared to 100 null variant sets (gray; empirical P < 0.01).
myocardial infarction in coronary artery (Table 2, Supplementary Table 20) each had significantly lower replication $P$ values than the remaining genes analyzed by PrediXcan in the specific tissue (Wilcoxon rank sum one-tailed test $P = 1.5 \times 10^{-7}$ for SBP and $P = 5.8 \times 10^{-3}$ for myocardial infarction; Fig. 6d; e; Supplementary Table 21). At FDR $\leq 0.05$, 33 (58%) of the proposed causal genes replicated for SBP, some of which have been previously implicated, such as $FURIN$ ($P = 6.94 \times 10^{-15}$), a gene important for the renin-angiotensin system and sodium-electrolyte balance$^{37,38}$, $ARHGAP42$ ($P = 1.66 \times 10^{-29}$), shown to contribute to variation in blood pressure by modulating vascular resistance$^{39}$, and $GUCY1B3$ ($P = 2.65 \times 10^{-19}$), implicated in the development of hypertension in mice, and 15 (28%) proposed genes replicated for CAD (Supplementary Table 21). The significant association of the expression of $HLA-C$ ($P = 2.96 \times 10^{-5}$) with myocardial infarction lends further support to an important role for a chronic inflammatory process in the development of atherosclerosis$^{40,41}$.

Second, we tested for replication of association of the best eQTL variants for the proposed causal genes (eGenes) (Supplementary Table 20) in the UK Biobank. The proposed aorta eQTLs were more likely to be replicated for SBP than matched null variants with GWAS $P < 0.05$ (Fig. 6f; fold-enrichment $= 11.9$, empirical $P < 0.01$; Methods), and similarly for coronary artery eQTLs and myocardial infarction (Fig. 6g; fold-enrichment $= 4.9$, empirical $P < 0.01$ for myocardial infarction; Methods), implicating robust novel variant-level associations for SBP and CAD (list of eQTLs with replication $P < 0.05$ and those that pass Bonferroni correction in Supplementary Tables 22 and 23).

Finally, we found substantial replication (17%) of the eGeneEnrich-proposed genes in the specific tissue for the remaining GWAS traits (Alzheimer's disease, LDL, and T2D, as well as SBP and CAD) by applying PrediXcan to related clinical phenotypes in BioVU (Supplementary Table 20, Supplementary Note), most of which are new associations (Supplementary Table 6). Taken together, these results demonstrate a new and robust framework for identifying true positive associations, at both the gene and variant levels, for complex traits.

Discussion

Characterizing the biological mechanisms underlying genetic variants associated with disease predisposition and other complex traits has proven to be an enormous, but critical, challenge. Here, we conducted integrative analyses of eQTL and GWAS data for a broad spectrum of complex traits. Using a diverse set of tissues, we assessed the contribution of regulatory variants to trait variation through several approaches, including enrichment analysis, heritability analysis, and true positive rate estimation, and investigated the relative contribution of tissue-specific eQTLs. Our analyses demonstrate a substantial polygenic contribution from eQTLs, including tissue-shared and tissue-specific ones, to a range of complex traits. A broader sampling of cell types with larger sample sizes promises greater resolution of the impact of regulatory variants on disease risk and trait variation.

We observed a 5-fold increase in the number of known trait-associated variants in linkage disequilibrium with at least 1 best eQTL per eGene in the 44 tissues compared to the GTEx pilot phase with 9 tissues. Notably, for over half of these trait-associated variants, more than one target gene, in one or more tissues, was suggested by the linked eQTLs, raising the possibility that more than one causal gene, and possibly tissue, might underlie many of the associations. This pattern was also observed from co-localization analysis (also shown for v6p in ref. 2). Measuring eQTLs in individual cell types might increase resolution and narrow down the list of candidate genes and cell types. Furthermore, gene- and causal inference-based methods (such as PrediXcan$^{35}$ or a Mendelian Randomization approach$^{39}$) and additional functional validation (such as with CRISPR-mediated genome editing$^{42,44}$) will be important in determining the causal genes at trait-associated loci. The proposed causal gene for trait-associated variants on the basis of the

### Table 2 | Complex trait causal genes proposed by gene set enrichment and PrediXcan analyses of top-ranked eQTL target genes

| Trait | eQTL tissue | eGene | No. of significant gene sets$^a$ | PrediXcan UK Biobank q-value |
|-------|-------------|-------|-------------------------------|-----------------------------|
| SBP   | Aorta artery$^a$ | FURIN | 22                            | $1.16 \times 10^{-22}$          |
| SBP   | Aorta artery$^a$ | ARHGAP42 | 1                             | $1.39 \times 10^{-27}$          |
| SBP   | Aorta artery$^a$ | GUCY1A3 | 23                            | $2.05 \times 10^{-18}$          |
| SBP   | Aorta artery$^a$ | GUCY1B3 | 31                            | $1.11 \times 10^{-18}$          |
| SBP   | Aorta artery$^a$ | PKRARB2 | 33                            | $5.71 \times 10^{-17}$          |
| SBP   | Aorta artery$^a$ | CSK | 25                             | $7.27 \times 10^{-15}$          |
| SBP   | Aorta artery | ACADVL | 6                             | $7.35 \times 10^{-12}$          |
| SBP   | Aorta artery | PRDM6 | 2                             | $6.23 \times 10^{-11}$          |
| SBP   | Aorta artery | SLC4A7 | 12                            | $3.46 \times 10^{-6}$           |
| SBP   | Aorta artery | MED8 | 1                             | $1.54 \times 10^{-6}$           |
| SBP   | Aorta artery | ARVCF | 1                             | $1.68 \times 10^{-6}$           |
| SBP   | Aorta artery | MED19 | 1                             | $3.81 \times 10^{-5}$           |
| SBP   | Aorta artery | ATF1 | 1                             | $1.30 \times 10^{-4}$           |
| SBP   | Aorta artery | HFE | 2                             | $1.40 \times 10^{-3}$           |
| SBP   | Aorta artery | PCDHA4 | 1                             | $1.40 \times 10^{-3}$           |
| SBP   | Aorta artery | FBNL7 | 1                             | $1.86 \times 10^{-3}$           |
| SBP   | Aorta artery | GTF2IRD1 | 35                           | $2.40 \times 10^{-4}$           |
| SBP   | Aorta artery | MRAS | 5                             | $5.74 \times 10^{-4}$           |
| SBP   | Aorta artery | RTN4 | 1                             | $4.72 \times 10^{-4}$           |
| SBP   | Aorta artery | GRID1 | 9                             | $5.85 \times 10^{-3}$           |
| SBP   | Aorta artery | FSCN2 | 12                            | $7.20 \times 10^{-3}$           |
| SBP   | Aorta artery | TCF4 | 1                             | $1.40 \times 10^{-3}$           |
| SBP   | Aorta artery | JPH2 | 1                             | $1.64 \times 10^{-2}$           |
| SBP   | Aorta artery | TMEM8B | 1                             | $2.57 \times 10^{-3}$           |
| SBP   | Aorta artery | DCHS1 | 9                             | $2.98 \times 10^{-2}$           |
| SBP   | Aorta artery | ULK2 | 1                             | $3.71 \times 10^{-2}$           |
| CAD   | Coronary artery | PHACTR1 | 1                             | $2.00 \times 10^{-3}$           |
| CAD   | Coronary artery | HLA-C | 4                             | $2.24 \times 10^{-3}$           |
| CAD   | Coronary artery | ANAPC3 | 1                             | $3.31 \times 10^{-2}$           |
| CAD   | Coronary artery | CDC25A | 4                             | $3.31 \times 10^{-2}$           |
| CAD   | Coronary artery | CEP63 | 2                             | $3.31 \times 10^{-2}$           |
| CAD   | Coronary artery | CT5K | 6                             | $3.31 \times 10^{-2}$           |
| CAD   | Coronary artery | HLA-DOB | 4                             | $3.31 \times 10^{-2}$           |
| CAD   | Coronary artery | GSTT2 | 2                             | $3.95 \times 10^{-2}$           |
| CAD   | Coronary artery | NME1 | 4                             | $3.95 \times 10^{-2}$           |
| CAD   | Coronary artery | SRD5A3 | 1                             | $3.95 \times 10^{-2}$           |
| CAD   | Coronary artery | NPHP3 | 3                             | $4.04 \times 10^{-4.02}$         |
| CAD   | Coronary artery | BAG6 | 4                             | $4.81 \times 10^{-2}$           |
| CAD   | Coronary artery | DDT | 1                             | $4.81 \times 10^{-2}$           |
| CAD   | Coronary artery | DTL | 1                             | $4.81 \times 10^{-2}$           |
| CAD   | Coronary artery | RPS28 | 2                             | $4.81 \times 10^{-2}$           |

$^a$The list of gene sets, from four different databases, in which the eQTL target genes were enriched, based on eGeneEnrich (adjusted $P < 0.05$; Methods), along with additional results, can be found in Supplementary Table 21. See Methods (“Replication framework using large-scale biobanks”) for description of the statistical approach (PrediXcan) used for the replication analysis. $^b$Denotes aorta-specific eQTLs (significant in at most four tissues other than aorta).
strongest eQTL-derived target gene was, notably, often discordant (~50%) with proximity-based assignment, reinforcing the importance of eQTL analysis for prioritizing causal genes.

Our study implicates non-coding target genes, in particular lincRNAs and antisense genes that are polyadenylated, for about 15% of trait associations. This is of particular interest as many non-coding RNAs have regulatory functions (for example, associated with chromatin-modifying complexes\(^\text{18}\)), and participate in regulatory networks\(^\text{18}\). This suggests that among the trait-associated variants acting via non-coding RNA targets, some may be trans-eQTLs. For the complex traits tested, eQTLs explain a substantial proportion of the genetic contribution to trait variation (10–50% per tissue), only a small fraction of which is due to eQTLs acting on tissue-specific genes. The proportion of heritability explained by all eQTLs (40–80%) is likely to increase with greater tissue sample size, which will lead to improved detection of eQTLs with weaker regulatory effects and additional independent eQTL signals per gene. The observation that tissue-shared eQTLs comprise a larger fraction of the trait associations than tissue-specific eQTLs for many of the tissue-trait pairs tested poses challenges in distinguishing pathogenic tissues from shared regulation among tissues. Alternatively, it also suggests that the underpinnings of many non-coding trait associations may be decipherable even if the actual pathogenic tissue is not available. Integrating additional layers of information, such as the tissue-specificity of eQTLs\(^\text{18,19}\), expression of transcriptional regulators, or broader cellular network effects on the locus in different cell types, may assist in detecting relevant tissue(s) of action.

While tissue-shared regulation appears to underlie an appreciable proportion of the genetic component of complex traits, we find multiple examples for which the trait associations are tissue-specific eQTLs that were not found in previous, much larger whole blood eQTL studies. Our polygenic analyses also demonstrate the importance of a broad sampling of tissues; for some traits, enrichment for trait associations amongst eQTLs is most prominent only in a subset of difficult-to-acquire tissues.

By integrating prior biological knowledge (of pathways and mouse phenotype ontologies) with top-ranked trait-associated eQTLs in relevant tissues, followed by additional analysis for independent support in large-scale DNA biobanks, we were able to propose and replicate potentially causal genes and novel trait associations. Our work suggests that gene-based approaches that test the contribution of the genetically determined expression to trait variation\(^\text{35}\), coupled with better understanding of biological networks in a diverse set of tissues, promise to greatly enhance the functional interpretation of GWAS findings and identification of disease-relevant genes.

**Methods**

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-018-0154-4.

**References**

1. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45, 580–585 (2013).
2. GTEx Consortium. Genetic effects on gene expression across human tissues. Nature 550, 204–213 (2017).
3. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics 33, 1479–1485 (2016).
4. Han, B. & Eskin, E. Random-effects model aimed at discovering associations using genome-wide association study data. Am. J. Hum. Genet. 88, 586–598 (2011).
5. Sudlow, C. et al. UK Biobank: An open access resource for identifying the underlying causes of complex disease. PLoS Med. 12, e1001779 (2015).
6. Denny, J. C. et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat. Biotechnol. 31, 1102–1110 (2013).
7. 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).
8. Dupuis, J. et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat. Genet. 42, 105–116 (2010).
9. Wen, X. Molecular QTL discovery incorporating genomic annotations using Bayesian false discovery rate estimation. Ann. Appl. Stat. 10, 1619–1638 (2016).
10. Wen, X., Lee, Y., Luca, F. & Pique-Regi, R. Efficient integrative multi-SNP association analysis via deterministic approximation of posteriors. Am. J. Hum. Genet. 98, 1114–1129 (2016).
11. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science 348, 648–660 (2015).
12. Pandey, G. K. et al. The risk-associated long noncoding RNA NBRAT-1 controls neuroblastoma progression by regulating cell proliferation and neuronal differentiation. Cancer Cell 26, 722–737 (2014).
13. Nica, A. C. et al. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. PLoS Genet. 6, e1000895 (2010).
14. Ongen, H. et al. Estimating the causal tissues for complex traits and diseases. Nat. Genet. 49, 1676–1683 (2017).
15. Hormozdiari, F. et al. Colocalization of GWAS and eQTL signals detects target genes. Am. J. Hum. Genet. 99, 1245–1260 (2016).
16. CARDiGRAMplusC4D Consortium. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat. Genet. 45, 25–33 (2013).
17. Mohammad, P., Castel, S. E., Brown, A. A. & Lappalainen, T. Quantifying the regulatory effect size of cis-acting genetic variation using allelic fold change. Genome Res. 27, 1872–1884 (2017).
18. Berge, K. E. et al. Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. Science 290, 1771–1775 (2000).
19. Kathiresan, S. et al. Common variants at 30 loci contribute to polygenic risk of dyslipidemia. Nat. Genet. 41, 56–65 (2009).
20. Ward, L. D. & Kellis, M. Haplotypev4: Systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and diseases. Nucleic Acids Res. 44, D877–D881 (2016).
21. Battle, A. et al. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 932 individuals. Genome Res. 24, 14–24 (2014).
22. Kukurba, K. R. et al. Impact of the X chromosome and sex on regulatory variation. Genome Res. 26, 768–777 (2016).
23. Westra, H. J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. 45, 1238–1243 (2013).
24. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47, 291–295 (2015).
25. Brown, A. A. et al. Predicting causal variants affecting expression by using whole-genome sequencing and RNA-seq from multiple human tissues. Nat. Genet. 49, 1747–1751 (2017).
26. Ficinica, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1235 (2015).
27. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661–678 (2007).
28. Eichler, E. E. et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat. Rev. Genet. 11, 446–450 (2010).
29. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. Proc. Natl Acad. Sci. USA 100, 9440–9445 (2003).
30. Qiao, O. & Nyamador, R. Is the association of type II diabetes with waist circumference or waist-to-hip ratio stronger than that with body mass index? Eur. J. Clin. Nutr. 64, 30–34 (2010).
31. Cheng, C. H. et al. Waist-to-hip ratio is a better anthropometric index than body mass index for predicting the risk of type 2 diabetes in Taiwanese population. Nutr. Res. 30, 585–593 (2010).
32. Emerging Risk Factors Collaboration. Major lipids, apolipoproteins, and risk of vascular disease. JAMA 302, 1993–2000 (2009).
33. International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 478, 103–109 (2011).
34. Ehret, G. B. et al. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. Nat. Genet. 48, 1171–1184 (2016).
35. Gamazon, E. R. et al. A gene-based association method for mapping traits using reference transcriptome data. Nat. Genet. 47, 1091–1098 (2015).
36. Barbeira, A. N. et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat. Commun. 9, 1825 (2018).
37. Ganesh, S. K. et al. Loci influencing blood pressure identified using a cardiovascular gene-centric array. Hum. Mol. Genet. 22, 1663–1678 (2013).
38. Li, N. et al. Associations between genetic variations in the FURIN gene and lipid traits: a genome-wide association scan results. Bioinformatics 26, 2336–2337 (2010).
39. Rippe, C. et al. Hypertension reduces soluble guanylyl cyclase expression in the mouse aorta via the Notch signaling pathway. Sci. Rep. 7, 1334 (2017).
40. Davies, R. W. et al. A genome-wide association study for coronary artery disease identifies a novel susceptibility locus in the major histocompatibility complex. Circ. Cardiovasc. Genet. 5, 217–225 (2012).
41. Laboute, C., Herbin, O., Mallat, Z. & Tedgui, A. Adaptive immunity in atherosclerosis: Mechanisms and future therapeutic targets. Nat. Rev. Cardiol. 8, 348–358 (2011).
42. Zou, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. 48, 481–487 (2016).
43. Jinde, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821 (2012).
44. Barrangou, R. & Doudna, J. A. Applications of CRISPR technologies in research and beyond. Nat. Biotechnol. 34, 933–941 (2016).
45. Khalil, A. M. et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl Acad. Sci. USA 106, 11667–11672 (2009).
46. Bai, Y., Dai, X., Harrison, A. P. & Chen, M. RNA regulatory networks in animals and plants: A long noncoding RNA perspective. Brief. Funct. Genomics 14, 91–101 (2015).
47. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat. Genet. 50, 621–629 (2018).
48. Pruijm, R. J. et al. LocusZoom: Regional visualization of genome-wide association scan results. Bioinformatics 26, 2336–2337 (2010).
49. Kang, E. Y. et al. ForestPMPlot: A flexible tool for visualizing heterogeneity between studies in meta-analysis. G3 (Bethesda) 6, 1793–1798 (2016).
50. Nkay, M. et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat. Genet. 47, 1121–1130 (2015).

Acknowledgements
We thank the DIAGRAM, MAGIC, GIANT, GLGC, CARDIoGRAM, ICBP, IGAP and IBIDS consortia for making their GWAS meta-analysis summary statistics publicly available. This work was conducted using the UK Biobank Resource (application number 25331). E.R.G. acknowledges support from R01-MH101820, R01-MH090937, R01-MH113362, and R01-CA157823 and benefited immensely from a Fellowship at Clare Hall, University of Cambridge. A.V.S., F.A., M.K.G., and K.G.A. acknowledge support from the NIH contract HHSN268201000029C to The Broad Institute, Inc. M.v.d.B. acknowledges support by a Novo Nordisk postdoctoral fellowship run in partnership with the University of Oxford. F.H. and E.E. are supported by NIH grants R01-MH101782 and R01-ES022282. X.W. acknowledges support from NIH grants R01-HG007022 and R01-AR042742. M.I.McC. is a Wellcome Senior Investigator supported by Wellcome (098381, 099352, 116130, 201341) and the NIH (U01-DK105535, R01-MH101814). E.T.D. acknowledges support from the Swiss National Science Foundation, the European Research Council, the NIH-NIMH, and the Louis Jeantet Foundation. N.J.C. is supported by R01-MH113362, R01-MH101820, and R01-MH090937. The datasets used for part of the replication analysis were obtained from Vanderbilt University Medical Center's BioVU, which is supported by numerous sources: institutional funding, private agencies, and federal grants. These include the NIH funded Shared Instrumentation Grant S10 RR022443, and CTSA grants UL1-TR002243, UL1-TR000445, and UL1-RR024975. Genomic data are also supported by investigator-led projects that include U01-HG004798, R01-N032830, R22-GM092618, R01-GM115305, U01-HG006378, U19-HL65962, and R01-HD074711, and additional funding sources listed at https://victr.vanderbilt.edu/pub/biovu/.

Author contributions
E.R.G. and A.V.S. jointly designed the study and led the analysis. E.R.G., A.V.S., M.v.d.B., and K.G.A. wrote the manuscript. E.R.G., A.V.S., M.v.d.B., X.W., H.S.X., F.H., H.O., A.K., E.M.D., F.A., and J.Q. performed the statistical analysis. E.R.G., A.V.S., M.v.d.B., X.W., H.S.X., F.H., E.M.D., D.L.N., E.E., M.K.G., M.I.McC, E.T.D., N.J.C., and K.G.A. interpreted the results of the analysis. All authors contributed to the critical review of the manuscript.

Competing interests
M.I.McC. serves on advisory panels for Pfizer and NovoNordisk. He has received honoraria from Pfizer, NovoNordisk, Sanofi-Aventis, and Eli-Lilly, and research funding from Pfizer, Eli-Lilly, Merck, Takeda, Sanofi Aventis, AstraZeneca, NovoNordisk, Servier, Jansen, Boehringer Ingelheim, and Roche. M.v.d.B is an employee of Novo Nordisk. H.S.X. and J.Q. are employees of Pfizer.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0154-4.
Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to E.R.G. or A.V.S.
Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Methods
All statistical tests based on theoretical distributions were two-sided, unless noted otherwise.

GTEx project. All eQTLs used in the paper were computed from 44 tissues in GTEx release v6p. Complete descriptions of the donor enrollment and consent process, and the biospecimen procurement methods, sample fixation, and histopathological review procedures were previously described1. Description of single-tissue and multi-tissue eQTL analyses can be found in the Supplementary Note.

eQTL analyses of trait-associated variants. eQTL annotations of genome-wide significant associations with complex traits. To assess the utility of GTEx eQTLs (release v6p) for providing functional insights into trait-associated variants, we used all genome-wide significant associations (P ≤ 5 x 10^{-8}) from the NHGRI-EBI GWAS catalog version 1.0.1, release 2016-07-10 (see URLs), which contains significant associations from published GWAS studies for 659 distinct diseases or traits (referred to as 'trait-associated variants'). Since entries in the complete GWAS catalog could have multiple GWAS signal; (2) similar fold-enrichment values are computed for 100–100,000 randomly sampled sets (with replacement) of null-eVariants of equal size to the eQTL set, using bootstrapping of randomly sampled sets of null-eVariants with replacement, matching on the 3 potential confounding factors above. We note that our definition of null-eVariants (FDR > 0.05) for this method should yield a conservative estimate of the adjusted fold-enrichment.

Polygenic analyses of top-ranked trait associations using eQTLs. GWAS meta-analysis data. Polygenic analysis is an approach aimed at relating phenotypic variation to multiple genetic variants simultaneously. It differs from conventional single-variant tests of association by allowing large numbers of loci (potentially in the thousands) to be tested for their contribution to the genetic architecture of phenotype. We analyzed 18 complex traits with available GWAS summary statistics, as well as several extensively studied Wellcome Trust Case Control Consortium phenotypes, for which genotype and phenotype data are available. These phenotypes span a wide range of complex traits, including metabolically cardiovascular, anthropometric, autoimmune, and neurodegenerative phenotypes (Supplementary Table 1), allowing us to conduct comprehensive polygenic analyses (Table 1) of their genetic basis, using the eQTLs from the single-tissue and multi-tissue analyses.

Tissue-specific and tissue-shared eQTLs. For the GWAS-eQTL fold-enrichment and eQTL analyses, tissue-specific eQTLs were defined as eQTLs with m-value > 0.9 in METASOFT and/or FDR ≤ 0.05 in the single-tissue analysis in 1–5 tissues (up to 10% of tissues; the most highly similar tissues, except brain, are in sets of 2–3), including the tissue of interest, and tissue-shared eQTLs were defined as eQTLs with m-value ≥ 0.9 in METASOFT and/or FDR ≤ 0.05 in the single-tissue analysis in 4–44 tissues (over 90% of tissues), including the tissue of interest (Supplementary Fig. 7A).

Rank- and permutation-based GWAS-eQTL fold-enrichment analysis. To test whether a set of eQTLs in a given tissue is enriched for subthreshold (for example, 5 x 10^{-6} < P ≤ 0.05) common variant associations with a given complex disease or trait, more than would be expected by chance, we developed the following rank- and permutation-based method, called eQTLEnrich. Specifically, for a given GWAS and for each of the 44 tissues with eQTLs, the most significant (best) cis-eQTL per eGene was retrieved (to control for linkage disequilibrium between the multiple variants tested per gene), and the GWAS variant association P values for each set of eQTLs were extracted (eQTLs affecting more than 1 gene are considered only once). The distribution of GWAS P values for each set of eQTLs is then tested for enrichment of highly ranked trait associations compared to an empirical null distribution sampled from non-significant variant-gene expression associations (FDR > 0.05), also called null-eVariants, as follows: (1) a fold-enrichment is computed for each GWAS-tissue pair as the fraction of GWAS with GWAS variant P ≤ 0.05 compared to expectation (5% of GWAS); (2) similar fold-enrichment values are computed for 100–100,000 randomly sampled sets (with replacement) of null-eVariants of equal size to the eQTL set, matching on potential confounding factors (using 10 quantile bins): distance of eQTL to TSS of the target gene, MAP; and number of proxy variants (at r^2 > 0.1), representing local linkage disequilibrium (see Supplementary Fig. 1); (3) an enrichment P value is then computed as the fraction of permutations with similar or higher fold-enrichment than the observed value; (4) an adjusted fold-enrichment (column H in Supplementary Table 2) is computed by dividing the fold-enrichment for a specific GWAS-tissue pair by the fold-enrichment of all null-eVariants with GWAS P < 0.05 for the tissue-trait pair. The adjusted fold-enrichment is used as the enrichment test-statistic for ranking tissues per trait, because it is not dependent on tissue sample size (variance in adjusted fold-enrichment explained by tissue sample size is R^2 = 0.04%), while the enrichment P value is weakly correlated with tissue sample size (variance in the P value explained by tissue sample size is R^2 = 0.64%; Supplementary Fig. 4). Lower and upper bound 95% confidence intervals were estimated using bootstrapping of randomly sampled sets of null-eVariants with replacement, matching on the 3 potential confounding factors above. We note that our definition of null-eVariants (FDR > 0.05) for this method should yield a conservative estimate of the adjusted fold-enrichment.

eQTLEnrich was applied to 18 GWAS meta-analyses (Supplementary Table 1) using eQTLs from the single-tissue analysis at FDR ≤ 0.05 (Supplementary Table 2) and tissue-specific eQTLs (defined above; Supplementary Table 4). Significant GWAS-tissue pairs were assessed using Bonferroni correction, correcting for total number of GWAS-tissue pairs tested (P < 6.3 x 10^{-5}). The adjusted fold-enrichment of the tissue-specific eQTLs is only weakly dependent on tissue samples size or number of eQTLs analyzed, and not dependent on GWAS sample size (Supplementary Fig. 20).

Gene set enrichment analysis (GSEA) of top-ranked eQTL target genes using eGeneEnrich. When enrichment for trait associations (subthreshold to genome-wide significant) is found amongst a set of eQTLs, GSEA can help detect the true relationship between the top-ranked eQTLs by testing the assumption that causal genes affecting a given trait will tend to cluster in a limited number of biological processes. To this end, we developed a GSEA approach, called eGeneEnrich, that tests whether the top-ranked target genes of eQTLs with GWAS P values below a given cutoff (P ≤ 0.05 used here) for a given trait-tissue pair are enriched for genes in predefined gene sets, compared to a null distribution that only randomly genes are enriched in the given gene set (adjusted below (Supplementary Table 2) and method described in refs. 45)). For each gene set gs and a set of eQTLs, I (FDR ≤ 0.05), we computed the probability (hypergeometric) of observing at least k target
genes of eQTLs l with GWAS $P < 0.05$ out of a total of m eGenes with GWAS $P < 0.05$ that belong to gene set $g$, given that $n$ out of $N$ target genes of all (eQTLs and null-eVariants) ‘best-eQTL per gene’ eQTLs belong to the gene set $g$:

$$P_{g|n}(X = k) = \binom{m}{n} \binom{N-n}{k} \binom{N}{n}$$

To account for potential bias that may arise from the subset of genes expressed in a given tissue, we computed an $eGeneEnrich$ adjusted $P$ value, that is, an empirical GSEA $P$ value, which is the fraction of 1,000–10,000 randomly sampled target genes from a null set of variants, $r$ (null eVariants and eQTLs with GWAS $P > 0.05$) of equal sizes, the eQTL set $l$, that have the same or more significant probability, $P_{l}$, than the observed probability, $P_{g|n}(X = k)$.

We tested a range of sets of functionally related genes with $\geq 10$ genes expressed in the given tissue, including metabolic and signaling pathways, gene ontology, and mouse phenotype ontology, starting with: 674 gene sets from REACTOME (downloaded from MSigDB v5.1), 186 gene sets from KEGG (downloaded from KEGG in 2010), 1,942 gene ontologies (see URLs), and 3,792 mouse phenotype ontologies (downloaded from Mouse Genome Informatics, MGI in 2013; see URLs). Bonferroni correction was applied per resource, correcting for number of gene sets tested that contained $\geq 1$ target gene of a best eQTL per eGene with GWAS $P < 0.05$. The method was applied to GWAS meta-analyses for SBP, T2D, LDL, CAD, and Alzheimer’s disease, and a number of tissues chosen based on significant eQTL enrichment for trait associations or high $\pi$, statistic and their relevance to the trait.

Replication framework using large-scale biobanks. To evaluate the role a gene may play in the etiology of a trait, we used PrediXcan$^{57}$. Evaluating the genetically determined component of gene expression in an independent dataset for contribution to trait variance may facilitate replication of proposed causal genes. Specifically, from the weights $\tilde{f}_j$ derived from the gene expression model$^{15}$ and the number of effect alleles $X_j$ at the variant $j$, the genetically determined component of gene expression was estimated as follows:

$$\tilde{g}_i = \sum_j N_j \tilde{f}_j$$

An observed association between the estimated genetic component of gene expression and a trait provides a causal direction of effect, as with eQTLs.

To test for independent support for the proposed causal genes for given trait–tissue pairs from the $eGeneEnrich$ analysis, we utilized GWAS data from two large-scale biobanks. For replication analysis of proposed genes using the 500K UK Biobank$^5$, we performed (variant-level) GWAS of SBP (phenotype code = 48080, SBP, automated reading; $n = 473,460$) and myocardial infarction (phenotype code = 20002, 1075, non-cancer illness code, self-reported: heart attack/myocardial infarction; number of cases = 10,866, number of controls = 428,004), using the mixed model association method, BOLT-LMM$^{58}$, and applied PrediXcan using summary statistics$^{59}$.

The two phenotypes were chosen for their available large sample size. Replication of a gene was tested in the same discovery tissue (aorta for SBP and coronary artery for CAD), and significance was assessed using the q-value approach (FDR $P < 0.05$) of equal size to the eQTL set $l$, that have the same or more significant $P$ values (from 0 to 1).

To test for higher replication rate for the proposed genes compared to that of the remaining genes with gene expression (in the given tissue context), we compared the distribution of replication $P$ values for the proposed genes to that of the remaining genes with gene expression

$$T_{S_{SN},s} = \pi_{tissue-specific} \times \frac{\text{N}_{eQTLs} (s)}{\text{FDR} \leq 0.05} \times \frac{\text{N}_{eQTLs} (s)}{\text{N}_{eQTLs} (a)}$$

MixPres$^{50}$ below 0.01 were set to 0.01. The statistic provides a measure of eQTLs tissue-specificity per tissue and controls for the effect of GWAS sample size and number of eQTLs tested per tissue (Supplementary Figs. 17b and 18).

We also sought variant-level replication of the associations of the best eQTLs for the $eGeneEnrich$-proposed genes using the BOLT-LMM results for SBP and myocardial infarction in the UK Biobank. To determine whether our framework for finding true positive associations yields significantly improved replication rates, we generated an empirical distribution from 100 sets of null variants of equal size to the input set, matching on distance of the eQTL to the TSS of the proposed gene, MAF, and number of linkage disequilibrium-proxy variants (at $r^2 > 0.5$). In addition, the null variants were chosen from the best eQTLs for non-significant eGenes (FDR $P > 0.05$) and were required to show a nominal GWAS association $P < 0.05$.

We sought to replicate the proposed gene-tissue pairs for all remaining tissues (Alzheimer’s disease, $n = 473,460$), as well as SBP and CAD, from the $eGeneEnrich$ analysis using BioVU$^{60}$. For each gene–tissue pair, we estimated the genetic component of gene expression in the implicated tissue in 18,620 BioVU samples using PrediXcan$^{57}$, enabling testing of gene association with the trait despite the lack of directly measured gene expression on the samples.

Estimation of true positive trait associations amongst eQTLs using $\pi$, statistic.

We calculated the proportion ($\pi$) of true positive trait associations amongst the set of ‘best eQTL per eGene’ (FDR $P < 0.05$) for each of the 44 tissues (computed with the single-tissue analysis) for 18 complex traits (Supplementary Table 1), by applying Storey’s method$^{61}$ (qvalue R package 2.4.2, default options) to the GWAS association $P$ values for each tissue-trait pair (Supplementary Table 19). The $\pi$, statistic considers the full distribution of GWAS P values (from 0 to 1). We used the 'best eQTL per eGene' to control for potential confounding effects due to linkage disequilibrium between the multiple variants tested per eGene. The $\pi$, statistic was not correlated with number of 'best eQTL per eGene' analysed per tissue-trait pair ($r = 0.05$, Supplementary Fig. 16b). Furthermore, the tissue sample size explained only a small percentage of the variability ($R^2 = 0.1\%$) in the $\pi$, statistic (Supplementary Fig. 16a). The $\pi$, statistic was not correlated with GWAS sample size after excluding the height GWAS meta-analysis, which is an outlier with respect to its much larger sample size compared to the other meta-analyses (Pearson’s $r = 0.00$, Supplementary Fig. 16c,d).

We performed the following steps in the web tutorial to estimate the relative contribution of eQTLs to the heritability of complex traits. To estimate the overall contribution to heritability from the eQTLs to 15 complex traits with available GWAS meta-analysis variant effect sizes, LDSR was applied to 3 different sets of eQTLs aggregated across all 44 tissues: (1) all significant variant-gene pair eQTLs (FDR $\leq 0.05$) from the single-tissue analysis, (2) all tissue-specific eQTLs based on multi-tissue analysis (defined above), and (3) a more stringent set of only the top 10 eQTLs per eGene in each of the traits (Supplementary Tables 12–14). We also used $\pi$, statistic of small eQTL sets (with $\leq 10$ eQTLs) and the tissue–shared eQTLs, anchored to the tissue $t$ (as defined above) (Supplementary Figs. 7a and 17a). The $\pi$, statistic of small eQTL sets (with $\leq 10$ eQTLs) was set to ‘NA’ (Not Applicable).

We calculated a tissue-specificity measure per tissue-trait pair $T_{S_{SN},s}$, defined as the estimated number of tissue-specific eQTLs that are true positive trait associations based on $\pi_{tissue-shared}$, divided by the estimated number of tissue-shared eQTLs that are true positive trait associations based on $\pi_{tissue-shared}$ for tissue $t$:

$$\pi_{tissue-shared}$$
Code availability. Code for methods applied in the paper can be downloaded from the URLs above.

Data availability. The protected data for the GTEx project (for example, genotype and RNA-sequence data) are available via access request to dbGaP accession number phs000424.v6.p1. Processed GTEx data (for example, gene expression and eQTLs) are available on the GTEx portal: https://gtexportal.org. The NHGRI-EBI GWAS Catalog version 1.0.1, release 2016-07-10 was downloaded from www.ebi.ac.uk/gwas. The URLs of the summary statistics datasets of all the GWAS meta-analyses analyzed in the paper can be found in Supplementary Table 1.

References
51. Carithers, L. J. et al. A novel approach to high-quality postmortem tissue procurement: The GTEx Project. Biopreserv. Biobank 13, 311–319 (2015).
52. Chang, C. C. et al. Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015).
53. Sudmant, P. H. et al. An integrated map of structural variation in 2,504 human genomes. Nature 526, 75–81 (2015).
54. Morris, A. P. et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat. Genet. 44, 981–990 (2012).
55. Segre, A. V. et al. Pathways targeted by antidiabetes drugs are enriched for multiple genes associated with type 2 diabetes risk. Diabetes 64, 1470–1483 (2015).
56. Loh, P. R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat. Genet. 47, 284–290 (2015).
57. Yang, R. Y. et al. A systematic survey of human tissue-specific gene expression and splicing reveals new opportunities for therapeutic target identification and evaluation. bioRxiv https://doi.org/10.1101/311563 (2018).