Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits

Farhad Hormozdiari1,2*, Steven Gazal1,2, Bryce van de Geijn1,2, Hilary K. Finucane1,2, Chelsea J.-T. Ju3, Po-Ru Loh2,4, Armin Schoech1,2, Yakir Reshef1,2, Xuanyao Liu1,2, Luke O’Connor1,5, Alexander Gusev4,6, Eleazar Eskin3,7 and Alkes L. Price1,2*

There is increasing evidence that many risk loci found using genome-wide association studies are molecular quantitative trait loci (QTLs). Here we introduce a new set of functional annotations based on causal posterior probabilities of fine-mapped molecular cis-QTLs, using data from the Genotype-Tissue Expression (GTEx) and BLUEPRINT consortia. We show that these annotations are more strongly enriched for heritability ($5.84 \times$ for eQTLs; $P = 1.19 \times 10^{-31}$) across 41 diseases and complex traits than annotations containing all significant molecular QTLs ($1.80 \times$ for expression (e)QTLs). eQTL annotations obtained by meta-analyzing all GTEx tissues generally performed best, whereas tissue-specific eQTL annotations produced stronger enrichments for blood- and brain-related diseases and traits. eQTL annotations restricted to loss-of-function intolerant genes were even more enriched for heritability ($17.06 \times; P = 1.20 \times 10^{-35}$). All molecular QTLs except splicing QTLs remained significantly enriched in joint analysis, indicating that each of these annotations is uniquely informative for disease and complex trait architectures.

Although genome-wide association studies (GWAS) have been extremely successful in detecting thousands of risk loci for diseases and traits1–3, our understanding of disease architecture is far from complete because most risk loci lie in non-coding regions of the genome4–9. Leveraging molecular phenotypes, such as gene expression10–14 or chromatin marks15–18, can help us to understand the disease architecture: in particular, previous studies have shown that cis-expression quantitative trait loci (eQTLs) are enriched in GWAS loci as well as genome-wide heritability of several diseases5,6,20, motivating further work on colocalization21–23 and transcriptome-wide association studies (TWAS)24–26. Partitioning heritability using raw genotypes and phenotypes27–31 or summary association statistics32–34 can aid our understanding of disease architectures, but it is currently unclear how to best leverage molecular QTLs from rich resources, such as the GTEx12,14 and BLUEPRINT18 consortia, using these methods.

Here we introduce a set of annotations constructed from eQTL, histone (h)QTL, splicing (s)QTL and methylation (me)QTL data that are very strongly enriched for disease heritability across 41 independent diseases and complex traits. We construct these annotations by applying a fine-mapping method15 (allowing for multiple causal variants at a locus) to compute causal posterior probabilities for each variant to be a causal cis-QTL. We show that our annotations are far more enriched for disease heritability than standard annotations. We further show that our eQTL annotations produce tissue-specific enrichments (despite high cis-genetic correlations of eQTL effect sizes across tissues15,36), and produce much larger enrichments when restricted to loss-of-function intolerant genes from Exome Aggregation Consortium (ExAC)37 data. Finally, we quantify the extent to which annotations constructed from eQTL, hQTL, sQTL and meQTL data provide complementary information about disease.

Results

Overview of methods. Our goal is to construct molecular QTL-based annotations that are maximally enriched for disease heritability. For a given molecular QTL dataset, we construct a probabilistic (continuous-valued) annotation as follows. First, for each molecular phenotype (for example, each gene) with at least one significant (false discovery rate (FDR) <5%) cis-QTLs (1 Mb from transcription start site (TSS)), we compute the causal posterior probability (CPP) of each cis-SNP in the fine-mapped 95% credible set, using our CAusal Variants Identification in Associated Regions (CAVIAR) fine-mapping method15 (see URLs). Then, for each SNP in the genome, we assign an annotation value based on the maximum value of CPP across all molecular phenotypes; SNPs that do not belong to any 95%CredibleSet are assigned an annotation value of 0. We refer to this annotation as MaxCPP. For comparison purposes, we also construct three other molecular QTL-based annotations. First, we construct a binary annotation containing all SNPs that are a significant (FDR < 5%) cis-QTLs for at least one molecular phenotype16; we refer to this annotation as Allcis-QTLs. Second, we construct a binary annotation containing all SNPs that belong to the 95% credible set (see above) for at least one molecular phenotype; we refer to these data as the 95% credible set. Third, we construct a binary annotation containing the most significant SNP for each molecular

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.
phenotype with at least one significant (FDR < 5%) QTL. We refer to this annotation as Topcis-QTL (see Methods).

We applied a previously developed method, stratified linkage disequilibrium (LD) score regression (S-LDSC)\(^3\), to these data, to partition disease heritability using functional annotations. We used two metrics to quantify the contribution of an annotation to disease heritability: enrichment and standardized effect size (\(r\)). Enrichment is defined as the proportion of heritability explained by SNPs in an annotation divided by the proportion of SNPs in the annotation\(^2\); here, we generalize this definition to probabilistic annotations such as MaxCPP. Standardized effect size (\(r\)) is defined as the proportionate change in per-SNP heritability associated with an increase in the value of the annotation by one standard deviation, conditional on other annotations included in the model\(^3\). Unlike enrichment, \(r\) quantifies effects that are unique to the focal annotation (see Methods).

We constructed MaxCPP and other annotations using eQTL data from the GTEx Consortium\(^2\) and eQTL, hQTL, sQTL and meQTL data from the BLUEPRINT Consortium\(^3\) (Table 1; see URLs). We included a broad set of 75 functional annotations from the baselineLD model (Supplementary Table 1) in most analyses. We have made our annotations and partitioned LD scores freely available (see URLs).

**Simulations.** We performed a comprehensive set of simulations to assess whether S-LDSC produces unbiased estimates of the contribution of an annotation to disease heritability for the Allcis-QTL, 95%CredibleSet, Topcis-QTL and MaxCPP annotations. We performed simulations using real genotypes from the UK Biobank, restricting to 749,024 SNPs on chromosome 1 (see Methods). In our main simulation, we simulated gene expression phenotypes for 500 individuals assuming that 10% of cis variants for a gene are causal cis-eQTLs (heritability = 16%), simulated complex trait phenotypes for an independent set of 40,000 individuals assuming that the set of causal variants is exactly the set of causal eQTLs, with independent effect sizes (heritability = 20%), and subsequently assumed that 10% of causal eQTLs are missing from the data analyzed. We also performed secondary simulations under other genetic architectures and assumptions about missing data (see below). We estimated the contribution of each annotation to complex trait heritability using S-LDSC. We performed 400 independent simulations, and averaged results across simulations.

Enrichment estimates and true enrichments for each annotation are shown in Fig. 1 and Supplementary Table 2. We determined that S-LDSC produces unbiased estimates of the contribution of each annotation to complex trait heritability using functional annotations. On the other hand, S-LDSC estimates were slightly conservative for the Allcis-QTL, 95% credible set and MaxCPP annotations. Thus, we restrict our analyses to these three annotations in our analyses of real phenotypes below. Of these three annotations, MaxCPP had the highest enrichment (Fig. 1 and Supplementary Table 2). For comparative purposes, we also computed estimates using genome-wide complex trait analysis (GCTA)\(^27,28\) (see URLs), a method that has previously been applied to assess eQTL enrichment for complex traits\(^29\); we computed GCTA estimates for all annotations except MaxCPP, as GCTA is only applicable to binary annotations. We determined that GCTA estimates generally exhibited greater bias than S-LDSC estimates (Fig. 1 and Supplementary Tables 2, 3). We obtained similar results for both S-LDSC and GCTA using an alternative simulation framework, drawn from our previous work\(^3\). This framework directly uses simulated gene expression to generate complex trait phenotypes (see Methods and Supplementary Table 4).

All functional enrichment methods (GCTA\(^27,28\), BOLT-REML\(^30\), S-LDSC\(^2,13\), and Linkage-Disequilibrium Adjusted Kinships (LDAK)\(^31\)) that we are currently aware of assume that causal disease effect sizes are independent and identically distributed conditional on minor allele frequency (MAF), LD and function annotation values. However, this assumption may be violated for molecular QTL-based annotations when causal variants are sparse (see Methods for an example); in particular, this is a limitation of our S-LDSC method. Indeed, our simulations confirm these biases (Fig. 1 and Supplementary Tables 2–4). We note that other functional enrichment methods are also subject to this limitation. Specifically, GCTA\(^27,28\) is shown by our simulations to exhibit greater biases than S-LDSC (Fig. 1 and Supplementary Tables 2–4). BOLT-REML\(^30\) is a computationally efficient method that produces the same results as GCTA\(^27,28\). For analysis of LDAK\(^31\), please see previous work\(^3\). S-LDSC produces slightly conservative estimates across a comprehensive set of simulations for the Allcis-QTL, 95%CredibleSet and MaxCPP annotations that we consider in our analyses of real phenotypes below.

**Fine-mapped eQTLs are enriched for disease heritability.** We used the GTEx eQTL dataset (Table 1) to construct the Allcis-QTL, 95% credible set and MaxCPP annotations. We constructed annotations using each of the 44 tissues (Supplementary Table 5). We applied S-LDSC to assess the contribution of each annotation to disease heritability for each of the 41 independent diseases and complex trait datasets (average \(N = 320,000\); for six traits we analyzed

### Table 1 | List of molecular QTL datasets analyzed

| Dataset | QTL type | Number of tissues | \(N\) (per tissue) | \(N\) (total) |
|---------|----------|------------------|-------------------|--------------|
| GTEx    | eQTL     | 44               | 70–361            | 7,014        |
| BLUEPRINT | eQTL    | 3                | 169–194           | 555          |
| BLUEPRINT | hQTL (H3K27ac) | 3     | 143–174           | 479          |
| BLUEPRINT | hQTL (H3K4me1) | 3     | 104–173           | 449          |
| BLUEPRINT | sQTL     | 3                | 169–194           | 555          |
| BLUEPRINT | meQTL    | 3                | 132–197           | 525          |

GTEx includes eQTLs for a wide range of tissues. BLUEPRINT includes eQTLs, two hQTLs, sQTLs and meQTLs for three blood cell types. Sample sizes for each tissue are provided in Supplementary Table 5 (GTEx) and Supplementary Table 26 (BLUEPRINT).

---

**Fig. 1 | S-LDSC and GCTA estimates for Topcis-QTL are upward biased in simulations.** a–d. The true enrichment (truth) and S-LDSC and GCTA enrichment estimates for Allcis-QTL (a), 95% credible set (b), Topcis-QTL (c) and MaxCPP (d) annotations. d. GCTA is not applicable to continuous annotations (MaxCPP). These data are shown as the mean of enrichment ± 95% confidence intervals computed for 400 simulations. Numerical results are reported in Supplementary Table 2. See Supplementary Tables 3, 4 for additional simulation scenarios.

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.
two different datasets, leading to a total of 47 datasets analyzed (Supplementary Table 6). We meta-analyzed results across the 47 datasets, which were chosen to be independent (see Methods). We computed enrichment and $\tau^*$ for each annotation, in analyses that included 75 functional annotations (Supplementary Table 1) from the baselineLD model\cite{9}. Results for whole blood, a widely studied tissue, are reported in Fig. 2 and Supplementary Table 7. We determined that the MaxCPP annotation had far higher values of enrichment and $\tau^*$ than Allcis-QTLs or 95%CredibleSet; the enrichment estimates remain much higher for MaxCPP even after accounting for the fact that S-LDSC generally produces more conservative estimates for the fact that S-LDSC generally produces more conservative estimates of enrichment) for blood than for FE-meta-tissue or any other individual tissue (Supplementary Table 12). We then analyzed MaxCPP annotations for blood and FE-meta-tissue jointly (conditional on the baselineLD model). We obtained a significantly positive $\tau^*$ estimate for blood ($\tau^* = 1.17$, s.e. = 0.24; $P = 1.77 \times 10^{-3}$; Fig. 4 and Supplementary Table 13), indicating that fine-mapped blood eQTLs provide additional information about these six diseases conditional on fine-mapped FE-Meta-Tissue eQTLs. We repeated these analyses for the five blood cell traits. When analyzing MaxCPP annotations for blood and FE-Meta-Tissue separately, we obtained higher estimates of $\tau^*$ (and higher or comparable estimates of enrichment) for blood than for FE-meta-tissue or any other individual tissue (Supplementary Table 14). When analyzing MaxCPP annotations for blood and FE-Meta-Tissue jointly, we obtained a significantly positive $\tau^*$ estimate for blood ($\tau^* = 1.17$, s.e. = 0.24; $P = 1.77 \times 10^{-3}$; Fig. 4 and Supplementary Table 13), indicating that fine-mapped blood eQTLs provide additional information about these five traits conditional on fine-mapped FE-meta-tissue eQTLs.

We investigated whether the $\tau^*$ for MaxCPP in each respective tissue varied with sample size. We observed a correlation ($R^2 = 0.69$, $P = 1.36 \times 10^{-12}$) between sample size and $\tau^*$ (Fig. 3 and Supplementary Table 9). We observed a similar pattern in simulations (Supplementary Fig. 2). This suggests that the correlation between sample size and $\tau^*$ in GTEx data is related to statistical power and not because tissue-specific eQTL from tissues with larger sample size are more relevant for the 41 traits that we analyzed. Thus, annotations constructed from tissues with larger sample sizes are more informative for disease and trait architectures.

To maximize sample size, we performed a fixed-effect meta-analysis of eQTL effect sizes across the 10 brain tissues and one nerve tissue (brain and nerve). When analyzing MaxCPP annotations for brain and nerve tissues and FE-Meta-Tissue separately, we obtained higher or comparable estimates of enrichment and $\tau^*$ for brain and nerve tissues than for FE-meta-tissue or any individual tissue (Supplementary Table 15). When analyzing MaxCPP annotations for brain and nerve tissues and FE-Meta-Tissue jointly, we obtained a significantly positive $\tau^*$ estimate for brain and nerve tissues ($\tau^* = 0.28$, s.e. = 0.07; $P = 9.81 \times 10^{-5}$) (Fig. 4 and Supplementary Table 13), indicating that fine-mapped eQTLs for brain and nerve tissues provide additional information about these eight traits conditional on fine-mapped FE-Meta-Tissue eQTLs. The correlations between the MaxCPP annotation values of brain and nerve tissues and the baselineLD model annotations and their LD scores are provided in Supplementary Figs. 6, 7. We repeated these analyses for each trait and each tissue separately, but determined that only three blood cell traits (white blood count, red blood cell distribution width and eosinophil count traits), in conjunction with MaxCPP for blood, attained a significantly positive (FDR < 5%) tissue-specific $\tau^*$ (Supplementary Tables 16, 17). Overall, these results demonstrate that tissue-specific eQTL effects on steady-state expression can be significant for diseases and complex traits, despite the well-documented high cis-genetic correlations of eQTL effect sizes across tissues\cite{12,36}.

MaxCPP signal is concentrated in disease-relevant gene sets. Recent studies have identified gene sets that are depleted for...
An Analysis Nature Genetics

the fact that these genes are depleted of eQTLs. We then analyzed the proportionately strong effect on disease heritability, consistent with the difference. This indicates that eQTLs for these 3,230 genes have a significant enrichment (17.96, s.e. = 1.28; $P = 1.20 \times 10^{-35}$) than MaxCPP(all genes) (5.84, s.e. = 4.90, $P = 4.90 \times 10^{-17}$ for difference). This indicates that eQTLs for these 3,230 genes have a disproportionately strong effect on disease heritability, consistent with the fact that these genes are depleted of eQTLs. We then analyzed MaxCPP(ExAC) and maxCPP(all genes) annotations jointly (conditional on the baselineLD model). We obtained a significantly positive $\tau^*$ for MaxCPP(ExAC) ($\tau^* = 0.41$, s.e. = 0.04; $P = 1.40 \times 10^{-22}$; Fig. 5b and Supplementary Table 19), indicating that MaxCPP(ExAC) provides additional information about disease heritability conditional on maxCPP(all genes). We observed that the effect size ($\tau^*$) for MaxCPP(ExAC) conditional on MaxCPP(all genes) and baselineLD is five times larger, and more statistically significant, than the $\tau^*$ of allSNP(ExAC) conditional on the baselineLD model (Supplementary Table 20). Thus, MaxCPP can increase the power to identify enriched gene sets.

We analyzed four additional gene sets $S$: a set of 1,003 genes that are strongly depleted of missense mutations (Samocha); a set of 2,984 genes with strong selection against protein-truncating variants (Sassa); a set of 1,878 genes predicted to be essential based on CRISPR experiments in a human cancer cell line (Wang); and a set
of 11,983 genes with evidence of allelic heterogeneity in analyses of GTEx gene expression data using our previously developed methods (AH)\(^1\). For each of these gene sets, MaxCPP(S) was strongly enriched in analyses conditional on the baselineLD model, meta-analyzed across 41 independent traits (Fig. 5a and Supplementary Table 18). In addition, for each gene set except the Wang gene set, we obtained a significantly positive \(\tau\) for MaxCPP(S) (after correcting for five gene sets tested) when analyzing MaxCPP(S) and MaxCPP(all genes) jointly (conditional on the baselineLD model) (Fig. 5b, Supplementary Tables 19, 21–25). As with the ExAC gene set, the \(\tau\) for MaxCPP(S) conditional on MaxCPP(all genes) and the baselineLD model were substantially larger than the \(\tau\) of allSNPs(ExAC) conditional on the baselineLD model and were often more statistically significant (Supplementary Table 20), indicating that MaxCPP can increase the power to identify enriched gene sets in which regulatory variants have an important role.

Fine-mapped molecular QTLs are enriched for heritability. We analyzed five molecular QTLs from the BLUEPRINT dataset (Table 1), including eQTLs, hQTLs (H3K27ac and H3K4me1), sQTLs and meQTLs. In each case, we constructed the Allcis-QTL, 95% credible set and MaxCPP annotations using each of the three immune cell types (CD14+ monocytes, CD16+ neutrophils, and naïve CD4+ T cells; Supplementary Table 26) as well as a fixed-effect meta-analysis of molecular QTL effect sizes across the three cell types (FE-Meta-Tissue). We determined that for each QTL dataset the MaxCPP annotation outperformed the Allcis-QTL and 95% credible set annotations (Supplementary Table 27). A histogram of MaxCPP annotation values for each QTL dataset is provided in Supplementary Fig. 8. MaxCPP for each molecular QTL was significantly enriched in an analysis conditional on the baselineLD model, meta-analyzed across the 41 traits: eQTLs (5.44×, s.e. = 0.35; \(P = 3.26 \times 10^{-19}\)), H3K27ac (4.42×, s.e. = 0.37; \(P = 2.59 \times 10^{-19}\)), H3K4me1 (4.27×, s.e. = 0.36; \(P = 1.29 \times 10^{-19}\)), sQTLs (3.61×, s.e. = 0.40; \(P = 1.39 \times 10^{-19}\)) and meQTLs (2.81×, s.e. = 0.19; \(P = 8.36 \times 10^{-22}\); Fig. 6a and Supplementary Table 28); the enrichment for BLUEPRINT eQTLs was almost as large as the enrichment for GTEx eQTLs (5.84×), despite the much smaller total sample size of FE-Meta-Tissue in BLUEPRINT. This indicates that BLUEPRINT sample sizes, though small, are adequately powered for eQTL detection. Consistent with this finding, we observed a high replication rate between cis-QTLs in GTEx and BLUEPRINT (see Supplementary Table 29), confirming that GTEx FE-Meta-Tissue analyses provide increased power relative to GTEx blood (Supplementary Table 28). BLUEPRINT FE-meta-tissue generally attained higher enrichment and \(\tau\) values than MaxCPP computed using each of the three immune cell types individually (Supplementary Table 30), similar to our GTEx results (Supplementary Tables 7, 10). MaxCPP computed using FE-Meta-Tissue also generally outperformed each of the three cell types in a meta-analysis across the six autoimmune diseases (Supplementary Table 31) and a meta-analysis across the five blood cell traits (Supplementary Table 32), in contrast to the stronger enrichments for tissue-specific GTEx blood eQTL annotations for blood cell traits (Supplementary Tables 12, 14). FE-Meta-Tissue generally attained higher enrichments and \(\tau\) than MaxCPP computed using each of the three immune cell types individually (Supplementary Table 30), similar to our GTEx results (Supplementary Tables 7, 10). MaxCPP computed using FE-Meta-Tissue also generally outperformed each of the three cell types in a meta-analysis across the six autoimmune diseases (Supplementary Table 31) and a meta-analysis across the five blood cell traits (Supplementary Table 32), in contrast to tissue-specific results in GTEx (Supplementary Tables 12, 14).

Finally, we jointly analyzed MaxCPP annotations for GTEx eQTLs and each of the five BLUEPRINT molecular QTLs (conditional on the baselineLD model). The purpose of this analysis was to determine whether each of these molecular QTLs provided independent information about disease and complex trait architectures. We determined that \(\tau\) remained statistically significant for all molecular QTLs except sQTLs (Fig. 6b and Supplementary Table 33); a joint analysis of just the five BLUEPRINT molecular QTLs (conditional on the baselineLD model) produced similar findings (Supplementary Table 34). LD scores of the sQTL annotation had the highest correlation with LD scores of the GTEx eQTL and BLUEPRINT eQTL annotations (\(R = 0.56–0.57\); see Supplementary Fig. 5), indicating that much of the informativeness of sQTLs in this analysis is captured by the eQTLs. However, eQTLs, hQTLs (H3K27ac and H3K4me1) and meQTLs are each uniquely informative for disease and complex trait architectures.

Discussion

We have shown that annotations constructed using fine-mapped posterior probabilities for several different molecular QTLs are strongly enriched for disease heritability. These results improve upon two previous studies that made key contributions in showing that annotations constructed using all significant cis-eQTLs were significantly enriched for trait heritability\(^{19,20}\). Our findings provide additional motivation for colocalization studies\(^{21–23}\) and TWAS\(^{24–26}\). Our fine-mapped eQTL annotations were able to detect tissue-specific enrichments for blood- and brain-related traits, despite high cis-genetic correlations\(^{12,16}\) of eQTL effect sizes across tissues and despite the fact that TWAS have generally concluded that their results ‘did not suggest tissue-specific enrichment’\(^{26}\).

We note that a previous study showed that cis-eQTLs often lie close to the TSS or transcription end site (TES)\(^{41}\), motivating us to investigate the orthogonal question of whether cis-eQTLs that lie near the TSS or TES produce more disease signal than cis-eQTLs that do not lie near the TSS or TES; we did not observe such an effect in the GTEx or BLUEPRINT datasets (see Supplementary Tables 35, 36). Notably, our eQTL annotations produced particularly large enrichments when restricted to disease-relevant gene sets, such as loss-of-function intolerant genes from ExAC, highlighting the potential to increase signal in analyses of gene sets that contain regulatory signals by prioritizing fine-mapped cis-eQTLs. Our eQTL annotations may also prove useful in future analyses of gene pathways.

Fig. 5 | Heritability enrichment of fine-mapped eQTLs is concentrated in disease-relevant gene sets. a. b. Meta-analysis results of enrichment (a) and \(\tau\) (b) of MaxCPP(S) for various gene sets. S. We report results conditional on the baselineLD model (dark blue) and results conditional on both the baselineLD model and MaxCPP(all genes) (light blue), meta-analyzed across 41 traits. As expected, \(\tau\) estimates are reduced by conditioning on MaxCPP(all genes), but enrichment estimates are not affected. The y axis is the meta-analyzed value and error bars represent 95% confidence intervals computed over 41 traits. The percentage under each bar indicates the proportion of SNPs in each annotation, defined as the average value of the annotation. Numerical results are reported in Supplementary Table 18.
We also detected strong enrichments using annotations based on other molecular QTLs, with eQTLs, hQTLs and meQTLs all providing complementary information about disease, conditional on each other and on functional annotations from previous studies. These results motivate applying colocalization and TWAS methods to other molecular QTLs; it may also be possible to prioritize other molecular QTLs in gene set analyses by connecting regulatory regions to genes\(^44,45\). Although annotations constructed from sQTLs were not conditionally significant in our analysis, previous work has shown that sQTLs can contain information that is independent from eQTLs\(^7\), motivating further investigation in larger sQTL datasets.

We note several limitations of our work. First, we restrict our analyses to common variants, as S-LDSC is not currently applicable to rare variants\(^45\). Recent work has shown that rare variants can have substantial effects on gene expression\(^50\), motivating ongoing work to extend S-LDSC to rare variants. Second, the CAVIAR fine-mapping method allows up to six causal variants per locus; this may limit power at loci that have more than six causal variants, although this would not lead to spurious signals. We determined that our approach cannot distinguish causal mediation from horizontal pleiotropy (that is, independent effects on molecular QTLs and disease), thus our molecular QTL enrichment results should not be viewed as a quantification of mediated effects. Despite these limitations, our results indicate that fine-mapped QTL annotations are strongly enriched for disease heritability and can help to elucidate the genetic architecture of diseases and complex traits.

**Methods**

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-018-0148-2.

Received: 14 September 2017; Accepted: 27 April 2018; Published online: 25 June 2018

**References**

1. Schizophrenia Working Group of the Psychiatric Genomics Consortium. . Biological insights from 108 schizophrenia-associated genetic loci. Nature \(511\), 421–427 (2014).
2. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature \(533\), 539–542 (2016).
3. Visscher, P. M. et al. 10 years of GWAS discovery: biology, function, and translation. Am. J. Hum. Genet. \(101\), 5–22 (2017).
4. Huidorff, L. A. et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl Acad. Sci. USA \(106\), 9362–9367 (2009).
5. Nicolae, D. L. et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. \(6\), e1000888 (2010).
6. Nica, A. C. et al. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. PLoS Genet. \(6\), e1000895 (2010).
7. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. Science \(337\), 1190–1195 (2012).
8. Trynka, G. et al. Dissecting the effects of colocalizing genomic annotations to functionally prioritize non-coding variants within complex-trait loci. Am. J. Hum. Genet. \(97\), 139–152 (2015).

**URLs**

CAVIAR, http://genetics.cs.uclid.edu/caviar/; GTEx (release v6, dbGaP accession number phs000424.v6.p1), http://www.gtexportal.org; GCTA, http://cnsgenomics.com/software/gcta/#Overview; BLUEPRINT, ftp://ftp.ebi.ac.uk/pub/databases/blueprint/blueprint_Epivar/qtl_as/; baselineLD annotations, https://data.broadinstitute.org/alkesgroupe/LDSCORE/; MaxCPP QTL-based annotations and partitioned LD scores, https://data.broadinstitute.org/alkesgroupe/LDSCORE/LDSC_QTL/; 95% credible set QTL-based annotations and partitioned LD scores, https://data.broadinstitute.org/alkesgroupe/LDSCORE/LDSC_QTL/; 1000 Genomes Project Phase 3 data, ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502; PLINK software, http://www.cog-genomics.org/plink2; BOLT-LMM software, https://data.broadinstitute.org/alkesgroupe/BOLT-LMM; BOLT-LMM summary statistics for UK Biobank traits, https://data.broadinstitute.org/alkesgroupe/UKBB; UK Biobank, http://www.ukbiobank.ac.uk/; UK Biobank Genotyping and QC Documentation, http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UKBiobank_genotyping_QC_documentation-web.pdf; rmeta R package, https://cran.r-project.org/web/packages/rmeta/index.html.
9. Roadmap Epigenomics Consortium. et al. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330 (2015).

10. Wright, F. A. et al. Heritability and genomics of gene expression in peripheral blood. Nat. Genet. 46, 430–437 (2014).

11. Zhang, X. et al. Identification of common genetic variants controlling transcript isoform variation in human whole blood. Nat. Genet. 47, 345–352 (2015).

12. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660 (2015).

13. Zhernakova, D. V. et al. Identification of context-dependent expression quantitative trait loci in whole blood. Nat. Genet. 49, 139–145 (2017).

14. The GTEx Consortium. Genetic effects on gene expression across human tissues. Nature 550, 204–213 (2017).

15. McVicker, G. et al. Identification of genetic variants that affect histone modifications in human cells. Science 342, 747–749 (2013).

16. Waszak, S. M. et al. Population variation and genetic control of modular chromatin architecture in humans. Cell 162, 1039–1050 (2015).

17. Grubert, F. et al. Genetic control of chromatin states in humans involves local and distal chromosomal interactions. Cell 162, 1051–1065 (2015).

18. Chen, L. et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. Cell 167, 1398–1414 (2016).

19. Davis, L. K. et al. Partitioning the heritability of Tourette syndrome and obsessive compulsive disorder reveals differences in genetic architecture. PLoS Genet. 9, e1003864 (2015).

20. Torres, J. M. et al. Cross-tissue and tissue-specific eQTLs: partitioning the heritability of a complex trait. Am. J. Hum. Genet. 95, 521–534 (2014).

21. Hu, X. et al. Regulation of gene expression in autoimmune disease loci and the genetic basis of proliferation in CD4+ effector memory T cells. PLoS Genet. 10, e1004404 (2014).

22. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 10, e1004383 (2014).

23. Hormozdiari, F. et al. Colocalization of GWAS and eQTL signals detects target genes. Am. J. Hum. Genet. 99, 1245–1260 (2016).

24. Gamazon, E. R. et al. A gene-based association method for mapping traits and complex diseases using fast variance-components analysis. PLoS Genet. 9, e1003491 (2013).

25. Flutre, T., Wen, X., Pritchard, J. & Stephens, M. A statistical framework for joint eQTL analysis in multiple tissues. PLoS Genet. 10, e1003486 (2013).

26. Mukambau, M. R. et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nat. Genet. 49, 1602–1612 (2017).

27. Li, Y. L. et al. RNA splicing is a primary link between genetic variation and disease. Science 352, 600–604 (2016).

28. Li, X. et al. The impact of rare variation on gene expression across tissues. Nature 550, 239–243 (2017).

29. Sul, J. H., Han, B., Ye, C., Choi, T. & Eskin, E. Effectively identifying eQTLs from multiple tissues by combining mixed model and meta-analytic approaches. PLoS Genet. 9, e1003491 (2013).

30. Flutre, T., Wen, X., Pritchard, J. & Stephens, M. A statistical framework for joint eQTL analysis in multiple tissues. PLoS Genet. 10, e1003486 (2013).

31. Sul, J. H., Han, B., Ye, C., Choi, T. & Eskin, E. Effectively identifying eQTLs from multiple tissues by combining mixed model and meta-analytic approaches. PLoS Genet. 9, e1003491 (2013).

32. Li, Y. L. et al. RNA splicing is a primary link between genetic variation and disease. Science 352, 600–604 (2016).

33. Li, X. et al. The impact of rare variation on gene expression across tissues. Nature 550, 239–243 (2017).

34. Sul, J. H., Han, B., Ye, C., Choi, T. & Eskin, E. Effectively identifying eQTLs from multiple tissues by combining mixed model and meta-analytic approaches. PLoS Genet. 9, e1003491 (2013).

35. Flutre, T., Wen, X., Pritchard, J. & Stephens, M. A statistical framework for joint eQTL analysis in multiple tissues. PLoS Genet. 10, e1003486 (2013).

36. Liu, X. et al. Functional architectures of local and distal regulation of gene expression in multiple human tissues. Am. J. Hum. Genet. 100, 605–616 (2017).

37. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291 (2016).

38. Gazal, S., Finucane, H. K. & Price, A. I. Reconciling S-LDSC and LDAK functional enrichment estimates. bioRxiv https://doi.org/10.1101/236412 (2018).

39. Cassa, C. A. et al. Estimating the selective effects of heterozygous protein-truncating variants from human exome data. Nat. Genet. 49, 806–810 (2017).

40. Samocha, K. E. et al. A framework for the interpretation of de novo mutation in human disease. Nat. Genet. 46, 944–950 (2014).

41. Wang, T. et al. Identification and characterization of essential genes in the human genome. Science 350, 1096–1101 (2015).

42. Hormozdiari, F. et al. Widespread allelic heterogeneity in complex traits. Am. J. Hum. Genet. 100, 789–802 (2017).

43. Veyrieras, J.-B. et al. High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet. 4, e1000214 (2008).

44. 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).

45. Mumbach, M. R. et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nat. Genet. 49, 1602–1612 (2017).

46. Li, Y. L. et al. RNA splicing is a primary link between genetic variation and disease. Science 352, 600–604 (2016).

47. Li, X. et al. The impact of rare variation on gene expression across tissues. Nature 550, 239–243 (2017).

48. Sul, J. H., Han, B., Ye, C., Choi, T. & Eskin, E. Effectively identifying eQTLs from multiple tissues by combining mixed model and meta-analytic approaches. PLoS Genet. 9, e1003491 (2013).

49. Flutre, T., Wen, X., Pritchard, J. & Stephens, M. A statistical framework for joint eQTL analysis in multiple tissues. PLoS Genet. 10, e1003486 (2013).

50. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat. Genet. 50, 621–629 (2018).

Acknowledgements
We thank S. Raychaudhuri, N. Zaitlen, B. Pasaniuc, M. Nivard, J.-H. Sul and F. Hormozdiari for helpful discussions. This research was funded by NIH grants U01 HG009379, R01 MH101244, R01 MH109978, T32 DK110919 and R01 MH107649. This research was conducted using the UK Biobank Resource under application 16549.

Author contributions
F.H. and A.L.P. designed experiments. F.H. performed experiments. F.H., S.G., B.v.d.G., T.H.F., C.-T.J., P.-R.L., A.S., Y.R., X.L., L.O., A.G. and E.E. analyzed data. F.H. and A.L.P. wrote the manuscript with assistance from all authors.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0148-2.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to F.H. or A.L.P.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Methods

Molecular QTL-based annotations. We construct four annotations for any given QTL dataset using the observed marginal association statistics. The four annotations are MaxCPP, Allcis-QTL, 95% credible set and Topcis-QTL. Each annotation is a vector that assigns a value to each SNP. Let us indicate our annotation for one QTL dataset where \( a_i \) indicates the value assigned to SNP \( j \). For binary annotations (Allics-QTL, 95% credible set and Topcis-QTL) \( a_i \in \{0, 1\} \), and \( a_i = 0 \) indicates that SNP \( j \) is not included in the annotation, whereas \( a_i = 1 \) indicates that SNP \( j \) is included in the annotation. For continuous probabilistic annotations (MaxCPP), \( 0 \leq a_i \leq 1 \).

Let \( S = (s_1, s_2, \ldots) \) indicate an \( m \times g \) matrix of the observed marginal association statistics obtained for each QTL dataset, where \( m \) is the number of SNPs and \( g \) is the number of eGenes (for example, genes that have at least one significant cis-eQTLs). Let \( \theta \) be the vector of marginal association statistics of gene \( j \) for all cis variants. Utilizing \( \theta \) and the LD structure, we can compute the CPP for each variant. CPP is the probability that a variant is causal. Let \( \alpha \) be the posterior probability that SNP \( j \) is causal for gene \( i \). We obtained the CPP values from CAVIAR. In addition to the CPP values, CAVIAR provides a 95% credible set that contains all of the causal variants with a probability of at least 95%. Let \( \tau \) indicate whether SNP \( j \) is in the 95% credible set for gene \( i \) (that is, \( \tau_i = 1 \) indicates that SNP \( j \) is) and \( \theta_i = 0 \) otherwise). We constructed the MaxCPP annotation for SNP \( j \) by computing the maximum value of CPP over all genes for which SNP \( j \) is and the 95% credible set of gene \( i \). More formally, we have: \( a_i = \max_{\tau} \theta_i \), where the maximum is over genes with \( \tau_i = 1 \).

The Allics-QTL annotation is a binary annotation, for which any variant with a marginal association statistic for at least one gene that passes the significance threshold (FDR < 0.05) has an annotation value of 1, and all other variants have an annotation value of 0. Let \( t_i \) indicate whether SNP \( j \) is statistically significant for gene \( i \) (that is, \( t_i = 1 \) when FDR(<0.05) and \( t_i = 0 \) otherwise). More formally, we have: \( a_i = \max_{t} \theta_i \).

The 95% credible set is a binary annotation, any variant that is in a 95% credible set of at least one gene has an annotation value of 1 and all other variants have an annotation value of 0. More formally, we have: \( a_i = \max_{\tau} \theta_i \).

Topcis-QTL is a binary annotation where any variant that is the most significant variant for at least one gene has an annotation value of 1 and all other variants have an annotation value of 0. Let \( \gamma_i \) indicate whether SNP \( j \) is the most significant SNP for gene \( i \) (that is, \( \gamma_i = 1 \) if SNP \( j \) is the most significant SNP among all cis variants for gene \( i \) and \( \gamma_i = 0 \) otherwise). More formally, we have: \( a_i = \max_{\gamma} \theta_i \).

Enrichment and effect size (\( \tau \)) metrics. We used two metrics to measure the importance of an annotation in the context of diseases and complex traits: the enrichment and standardized effect size (\( \tau \)) of the annotation. We used S-LDSC(25) to compute enrichment and standardized effect size (\( \tau \)). Let \( a_i \) indicate the annotation value of SNP \( j \) for the annotation \( c \). S-LDSC estimates \( \tau_c \) using the following equation:

\[
\tau_c = \frac{\sum_{i=1}^{n} a_i \tau_i}{\sum_{i=1}^{n} a_i}
\]

where \( \tau_c \) is the contribution of annotation \( c \) to per-SNP heritability. S-LDSC estimates \( \tau_c \) using the following equation:

\[
E(\tau_c) = N \sum_i^c (l(i) \tau_i + 1)
\]

where \( N \) is the GWAS sample size and \( E(\tau_c) \) is the LD score of SNP \( j \) for the annotation \( c \). S-LDSC computes the LD scores as follow:

\[
l(i, c) = \sum a_i \tau_i
\]

where \( \tau_i \) is the genetic correlation between SNPs \( j \) and \( k \).

Because \( \tau \) depends on trait heritability and the size of the annotation, \( \tau_c \) was previously defined(1) for an annotation as the standardized annotation effect size:

\[
\tau_c = \frac{\tau_c \times \text{s.d.}(c)}{\text{h}^2_c / \text{M}}
\]

where \( \text{s.d.}(c) \) is the standard deviation of annotation \( c \), \( \text{h}^2_c \) is the SNP heritability, and \( \text{M} \) is the number of variants used to compute \( \text{h}^2_c \). In our experiments, \( \text{M} \) is equal to 5,961,159 (see below).

The enrichment of an annotation is defined as the fraction of heritability captured by the annotation divided by the fraction of SNPs in that annotation. We extend the definition of enrichment to continuous probabilistic annotations with values between 0 and 1:

\[
\text{Enrichment} = \frac{\sum_{i=1}^{n} a_i \tau_i}{\sum_{i=1}^{n} a_i}
\]

where \( \sum_{i=1}^{n} a_i \tau_i \) is the heritability captured by the \( \tau \) annotation. We can compute this quantity as follows:

\[
\text{h}^2_c = \sum a_i \tau_i = \sum a_i \iff \sum a_i \tau_i = \frac{\sum_{i=1}^{n} a_i \tau_i}{\sum_{i=1}^{n} a_i}
\]

Although both enrichment and \( \tau \) are computed using a model that includes all annotations, \( \tau \) quantifies effects that are unique to the focal annotation (after conditioning on all other annotations in the model), whereas enrichment quantifies effects that are unique and/or non-unique to the focal annotation. For example, consider a model that includes two annotations, in which the first annotation is a highly disease-informative functional annotation and the second annotation is the first annotation plus a random set of SNPs. Only the first annotation will have a significant \( \tau \), but both annotations will be significantly enriched. We confirmed via simulation that, under a generative model in which only the baseline LD and GTEx-Multi-Tissue MaxCPP annotations directly influence trait heritability, \( \tau \) estimates for the GTEx whole-blood MaxCPP annotations are equal to 0 on average, with a correctly calibrated null distribution of \( P \) values for nonzero \( \tau \) (Supplementary Fig. 11).

We computed the statistical significance level \( P \) value of enrichment for each annotation via block-jackknife, as described in our previous studies(25). We computed the statistical significance \( P \) value(10) of standardized effect size \( \tau \) for each annotation by assuming that \( \tau \) follows a normal distribution with mean 0 and variance of \( \frac{\sigma_\tau^2}{n_c} \).

Simulation framework. Main simulation framework. We simulated both gene expression and trait phenotypes. We utilized UK Biobank genotypes from chromosome 1, which consists of 749,024 variants, for our simulation. We used 40,000 individuals to generate the trait phenotypes and a non-overlapping set of 500 individuals to generate gene expression phenotypes. Let \( \sigma_{ge}^2 \) and \( \sigma_{gt}^2 \) indicate the total heritability of gene expression and trait phenotypes, respectively. We simulated causal trait effect sizes using a polygenic model, \( \beta = N \left( \mu, \sigma_{gt}^2 / \text{M} \right) \), where \( \beta \) is the causal (true) effect size of the \( i \)th causal variant and \( \sigma_{gt}^2 \) is the number of causal variants for the trait. Similarly, we simulated causal gene expression effect sizes using a polygenic model, \( \mu_{ge} = N \left( \mu, \sigma_{ge}^2 / \text{M} \right) \), where \( \mu_{ge} \) is the true (effect size of the \( i \)th causal variant on gene expression of gene \( j \) and \( \mu_{ge} \) is the number of causal variants. We use the following model to simulate gene expression and traits:

\[
y = X_{ge} + \epsilon \quad \epsilon ~ N(0, \sigma_{ge}^2)
\]

where \( \epsilon \) is the environmental and measurement noise, \( y \) is the simulated trait phenotypes, \( \mu_{ge} \) is the gene expression of gene \( j \), and \( X \) is the environmental and measurement noise for gene \( j \). In the case of gene expression, we simulated 1,860 coefficients to represent our simulated gene expression data as 1,860 genes lies on chromosome 1. We simulated three different datasets for which \( n_c \) is set to 1 causal variant, 10 causal variants, or 10% of cis variants that are causal for each gene.

The default value of \( n_c \) is set to 10% of cis variants. In the case of trait phenotypes, we set \( n_c \) to be the union of causal variants for all genes. To simulate causal traits, we assumed that a subset of the causal variants are missing (not measured). Let \( n_c \) indicate the fraction of causal eQTLs that are missing. In our simulated datasets, we set \( n_c \) to 0% (no causal eQTLs are missing), 5% (5% of causal eQTLs are missing) and 10% (10% of causal eQTLs are missing), or 50% (half of the causal eQTLs are missing). The default value of 10% is used for \( n_c \).

Alternative simulation framework. We also considered an alternative simulation framework, drawn from our previous work(1) that directly uses simulated gene expression to generate complex trait phenotypes. We utilized UK Biobank genotypes from chromosome 1, which consists of 749,024 variants, for our simulation. We used 40,000 individuals to generate the trait phenotypes and gene-expression phenotypes. We assume that the trait phenotype is a mixture of direct genotype effects (effects not mediated by gene expression) and gene-expression effects. Let \( \sigma_{ge}^2 \) indicate the phenotype variance explained directly by genotypes, \( \sigma_{gt}^2 \) indicate the gene expression variance explained directly by genotypes and \( \sigma_{ge}^2 \) indicate the phenotypic variance explained by gene expression. We simulated causal gene expression effect sizes using a polygenic model, \( \beta_{ge} = N \left( \mu_{ge}, \sigma_{ge}^2 / \text{M} \right) \), where \( \beta_{ge} \) is the causal (true) effect size of the \( i \)th causal variant on the gene expression.
of gene $j$ and $n_j$ is the number of causal variants. We simulated causal trait effect sizes using a polygenic model, $\beta_j \sim N(0, \frac{\sigma_\beta^2}{n_j})$, where $\beta_j$ is the causal (true) effect size of ith causal variant and $n_j$ is the number of causal variants for simulated trait phenotypes. As above, we simulated expression values for the 1,860 genes on chromosome 1. We simulated the effect size of gene expression on traits using a polygenic model\cite{10}, $\beta_j \sim N(0, \frac{\sigma_\beta^2}{n_j})$, where $\sigma_\beta^2$ is set to $0.05$. We use the following model to simulate the gene expression and traits:

$$y = X\beta + \epsilon$$

After simulating the gene expression and trait phenotypes, we obtained marginal association statistics for each variant using linear regression implemented in the PLINK software\cite{11} (see URLs). In the case of simulated trait phenotypes, we computed the association for each variant with the simulated trait phenotypes ($y \sim x$). In the case of simulated gene expression phenotypes, we computed the marginal statistics for all variants within 1 Mb of the TSS ($y \sim x$), where $x$ is the subset of the X genotype matrix restricted to 500 individuals. We assumed that 10% of the causal variants are missing (not measured), that is, we set $n_j$ to 100. We generated the four annotations as described above. We used CAVIAR\cite{12} to generate the 95% credible set and MaxCPP annotations. We utilized European samples from the 1000 Genomes Project (1000G)\cite{13} (see URLs) to estimate the LD structure required as input into CAVIAR. We applied CAVIAR using setting 1, which allowed up to six causal variants for each gene. We observed that the results for cases of two to six causal variants are similar to six or three causal variants for each gene are not statistically different (Supplementary Figs. 9, 10 and Supplementary Table 37). We note that the 95% credible set is not guaranteed to be unique; in this study, we used a single 95% credible set for each gene for three reasons. First, use of a single 95% credible set is consistent with the output of all existing fine-mapping methods of which we are currently aware. Second, computing all 95% credible sets for each gene is computationally costly. Third, taking the union of all 95% credible sets might reduce the enrichment and $r^2$ across the 47 datasets, as defined using random-effect cross-tissue LDSC\cite{14}. The meta-analyzed values of enrichment and $r^2$ across the 47 datasets were computed using a random-effect meta-analysis, as implemented in the rmeta R package.

Fixed-effect meta-analysis of eQTL effect sizes (FE-Meta-Tissue). Given a set of effect sizes for SNP $(\beta_1, \beta_2, \ldots, \beta_t)$ for $t$ tissues, where $\beta_j$ is the eQTL effect size for tissue $j$, we used fixed-effect meta-analysis (FE-Meta-Tissue) to compute inverse-variance weighted meta-analysis $z$-scores $z_{\beta j}$ as follows:

$$z_{\beta j} = \sqrt{\frac{n_j}{\sum_{j=1}^{t} n_j}} \sum_{j=1}^{t} \frac{\beta_j}{n_j}$$

where $n_j$ is the number of eQTLs and $s.e. (\beta_j)$ is the standard error of $\beta_j$. We note that equation (8) is equivalent to computing a weighted average of $z$-scores.

Our use of FE-Meta-Tissue has two limitations. First, expression levels in two tissues in the same individual are not independent (sample overlap). Second, true effect sizes in two tissues may be different (heterogeneity). We discuss these limitations separately.

Regarding the sample overlap limitation, we determined that noise is mostly uncorrelated across tissues. For example, the correlation of normalized gene expression (read count) between whole blood and brain hippocampus is 0.11. Furthermore, the genetic correlation between these two tissues is around 0.67\cite{15} and the heritability explained by cis-eQTLs is about 0.12. This indicates that the bulk of gene expression correlation is because of genetic correlations. Thus, the noise (environmental and measurement) in expression levels in two tissues in the same individual is close to independent.

Regarding the heterogeneity limitation, we determined that recently developed random-effect cross-tissue eQTL meta-analysis methods\cite{16,17,18,19} are not applicable to our problem. The meta-tissue method\cite{20} is computationally intractable for datasets as large as GTEx (number of tissues and sample size). Other existing methods\cite{21,22}, which are Bayesian methods, do not produce summary statistics (for example, $z$-scores) that are required to compute the MaxCPP annotation. Thus, these methods are not applicable to our work. Previous studies\cite{23,24} have shown that eQTL effects are highly correlated across tissues, suggesting that our fixed-effect meta-analysis approach is likely to be fairly close to optimal.

We note that the above limitations pertain only to power and not to false positives in our setting, which involves building eQTL annotations to apply to independent disease data. Our results (Fig. 2 and Supplementary Tables 7, 27) show that we have improved our results by utilizing FE-Meta-Tissue. Furthermore, utilizing FE-Meta-Tissue increases replication rates for both eQTLs and hQTLs (Supplementary Table 29).

Blood- and brain-related diseases and complex traits. We analyzed six autoimmune diseases: Crohn’s disease\cite{25}, rheumatoid arthritis (ref. 29 and UK Biobank), ulcerative colitis\cite{30}, lupus\cite{31}, celiac\cite{32} and all autoimmune and inflammatory diseases in the UK Biobank). We analyzed five blood cell traits: white blood cell count, red blood cell count, platelet count, eosinophil count and red blood cell distribution width. All of these datasets were obtained from the UK Biobank.

We analyzed seven brain-related diseases and complex traits: age at menarche\cite{33}, body mass index (ref. 34 and UK Biobank), bipolar disorder\cite{35}, depression symptoms, neuroticism (UK Biobank), schizophrenia\cite{36} and smoking status (ref. 37 and UK Biobank), and year of education (ref. 38 and UK Biobank). These traits are a subset of traits from Supplementary Table 6 that have been reported to be brain-enriched\cite{39}.

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

Data availability. The S-LDSC software, baselineLD and MaxCPP QTL-based annotations, and a tutorial on how to use S-LDSC with QTL-based annotations are available online (see URLs).

References

51. 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).
52. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
53. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
54. Loh, P.-R. et al. Mixed model association for Biobank-scale data sets. *Nat. Genet.* **50**, 621–629 (2018).
55. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
56. Urbat, S. M., Wang, G. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. bioRxiv https://doi.org/10.1101/096552 (2017).
57. Park, Y., Sarkar, A. K., Bhutani, K. & Kellis, M. Multi-tissue polygenic models for transcriptome-wide association studies. bioRxiv (2017).
58. Jostins, L. et al. Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
59. Okada, Y. et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).
60. 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).
61. Dubois, P. C. A. et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat. Genet.* **42**, 295–302 (2010).
62. Day, F. R. et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat. Genet.* **49**, 834–841 (2017).
63. Speliotes, E. K. et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
64. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat. Genet.* **43**, 977–983 (2011).
65. The Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42**, 441–447 (2010).
66. Rietveld, C. A. et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467–1471 (2013).
Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size
   - Describe how sample size was determined.
   - Our work is based on summary statistics. We analyzed existing data sets, thus, no statistical method is used to compute the sample size.

2. Data exclusions
   - Describe any data exclusions.
   - Our study was restricted to data sets of European ancestry. No data was excluded from European individuals.

3. Replication
   - Describe whether the experimental findings were reliably reproduced.
   - We analyzed existing data sets. No replication was performed for experimental findings. We use GTEx and BLUEPRINT to replicate our findings.

4. Randomization
   - Describe how samples/organisms/participants were allocated into experimental groups.
   - We analyzed existing data sets. Thus, no randomization was performed.

5. Blinding
   - Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   - We analyzed existing data sets. Thus, blinding was relevant to our work.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters
   - For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).
   - n/a
     - Confirmed

     - The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
     - A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
     - A statement indicating how many times each experiment was replicated
     - The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
     - A description of any assumptions or corrections, such as an adjustment for multiple comparisons
     - The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
     - A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
     - Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.
Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

We use our existing methods which are available online. CAVIAR (v2.0): http://genetics.cs.ucla.edu/caviar/, GTEx (Release v6, dbGaP Accession phs000424.v6.p1): http://www.gtexportal.org/, GCTA: cnsgenomics.com/software/gcta/, BLUEPRINT: ftp://ftp.ebi.ac.uk/pub/databases/blueprint/blueprint_Epivar/qtl_as/, baselineLD annotations: https://data.broadinstitute.org/alkesgroup/LDSCORE/, PLINK software (2.0): https://www.cog-genomics.org/plink2, BOLT-LMM software: https://data.broadinstitute.org/alkesgroup/BOLT-LMM, BOLT-LMM summary statistics for UK Biobank traits: https://data.broadinstitute.org/alkesgroup/UKBB, UK Biobank: http://www.ukbiobank.ac.uk/, UK Biobank Genotyping and QC Documentation: http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UKBiobank_genotyping_QC_documentation-web.pdf

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

There are no restrictions to access the results of this work. However, to access UK Biobank data, one required to have approved application.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies are used in this work.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell lines are used in this work.

b. Describe the method of cell line authentication used.

No eukaryotic cell lines are used in this work.

c. Report whether the cell lines were tested for mycoplasma contamination.

No eukaryotic cell lines are used in this work.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No eukaryotic cell lines are used in this work.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

No animals are used in this work.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Our study involved publicly available data sets (e.g. UK Biobank and existing summary statistics).