Barbeira *et al. Genome Biology* (2021) 22:49
https://doi.org/10.1186/s13059-020-02252-4

Genome Biology

RESEARCH Open Access

Exploiting the GTEx resources to decipher the mechanisms at GWAS loci

Alvaro N. Barbeira[1†], Rodrigo Bonazzola[1†], Eric R. Gamazon[2,3,4,5†], Yanyu Liang[1†], YoSon Park[6,7†], Sarah Kim-Hellmuth[8,9,10], Gao Wang[11], Zhuoxun Jiang[1], Dan Zhou[2], Farhad Hormozdiari[12,13], Boxiang Liu[14], Abhiram Rao[14], Andrew R. Hamel[12,15], Milton D. Pividori[1], François Aguet[12], GTEx GWAS Working Group, Lisa Bastarache[16,17], Daniel M. Jordan[18,19,20], Marie Verbanck[18,19,20,21], Ron Do[18,19,20], GTEx Consortium, Matthew Stephens[11], Kristin Ardlie[12], Mark McCarthy[22], Stephen B. Montgomery[23,24], Ayellet V. Segrè[12,15], Christopher D. Brown[6], Tuuli Lappalainen[9,10], Xiaoquan Wen[25] and Hae Kyung Im[1*]

*Correspondence: haky@uchicago.edu
†Alvaro N. Barbeira, Rodrigo Bonazzola, Eric R. Gamazon, Yanyu Liang, and YoSon Park contributed equally to this work.
[1]Section of Genetic Medicine, Department of Medicine, The University of Chicago, Chicago, IL, USA
Full list of author information is available at the end of the article

Abstract

The resources generated by the GTEx consortium offer unprecedented opportunities to advance our understanding of the biology of human diseases. Here, we present an in-depth examination of the phenotypic consequences of transcriptome regulation and a blueprint for the functional interpretation of genome-wide association study-discovered loci. Across a broad set of complex traits and diseases, we demonstrate widespread dose-dependent effects of RNA expression and splicing. We develop a data-driven framework to benchmark methods that prioritize causal genes and find no single approach outperforms the combination of multiple approaches. Using colocalization and association approaches that take into account the observed allelic heterogeneity of gene expression, we propose potential target genes for 47% (2519 out of 5385) of the GWAS loci examined.

Introduction

In the last decade, the number of reproducible genetic associations with complex human traits that have emerged from genome-wide association studies (GWAS) has substantially grown. Many of the identified associations lie in non-coding regions of the genome, suggesting that they influence disease pathophysiology and complex traits via gene regulatory changes. Integrative studies of molecular quantitative trait loci (QTL) [1] have established gene expression as a key intermediate molecular phenotype, and improved functional interpretation of GWAS findings, spanning immunological diseases [2], various cancers [3, 4], lipid traits [5, 6], and a broad array of other complex traits.

Large-scale international efforts such as the Genotype-Tissue Expression (GTEx) Consortium have provided an atlas of the regulatory landscape of gene expression and splicing

variation in a broad collection of primary human tissues [7–9]. Nearly all protein-coding genes in the genome now have at least one local variant associated with expression changes and the majority also have common variants affecting alternative splicing (FDR < 5%) [9]. In parallel, there has been an explosive growth in the number of genetic discoveries across a large number of traits, prompting the development of integrative approaches to characterize the function of GWAS findings [10–14]. Nevertheless, our understanding of underlying biological mechanisms for most complex traits substantially lags behind the improved efficiency of the discovery of genetic associations, made possible by large-scale biobanks and GWAS meta-analyses.

One of the primary tools for the functional interpretation of GWAS associations has been the integrative analysis of molecular QTLs. Colocalization approaches that seek to establish shared causal variants (e.g., eCaviar [15], *enloc* [16], and *coloc* [17]), enrichment analysis (S-LDSC [18] and QTLEnrich [11]), or mediation and association methods (SMR [12], TWAS [13], and PrediXcan [19]) have provided important insights, but they are often used in isolation, and there have been limited prior assessments of power and error rates associated with each [20]. Their applications often fail to provide a comprehensive, biologically interpretable view across multiple methods, traits, and tissues or offer guidelines that are generalizable to other contexts. Thus, a comprehensive assessment of expression and splicing QTLs for their contributions to disease susceptibility and other complex traits requires the development of novel methodologies with improved resolution and interpretability.

Here, we present methods and resources that help elucidate how genetic variants associated with gene expression (cis-eQTLs) or splicing (cis-sQTLs) contribute to, or mediate, the functional mechanisms underlying a wide array of complex diseases and quantitative traits. Since splicing QTLs have largely been understudied, we perform a comprehensive integrative study of this class of QTLs, in a broad collection of tissues, and disease associations. We provide predictions of functional mechanisms for 74 distinct complex traits from 87 GWA study results and demonstrate independent validation and evaluation of findings using likely causal gene-disease relationships in the Online Mendelian Inheritance of Man (OMIM) database. Notably, we find widespread dose-dependent effects of cis-QTLs on traits through multiple lines of evidence. We examine the importance of considering, or correcting for, false functional links attributed to GWAS loci due to neighboring but distinct causal variants. We call this confounding LD contamination for the remainder of the paper. To identify predicted causal effects among the complex trait-associated QTLs, we conduct systematic evaluation across different methods. Furthermore, we provide guidelines for employing complementary methods to map the regulatory mechanisms underlying genetic associations with complex traits.

Mapping the regulatory landscape of complex traits

The final GTEx data release (v8) included 54 primary human tissues, 49 of which included at least 70 samples with both whole genome sequencing (WGS) and tissue-specific RNA-seq data. A total of 15,253 samples from 838 individuals were used for cis-QTL mapping (Fig. 1) [9]. In addition to the expression quantitative trait loci (eQTL) mapping, we also evaluated genetic variation associated with alternative splicing (sQTL) and their impact on complex traits.

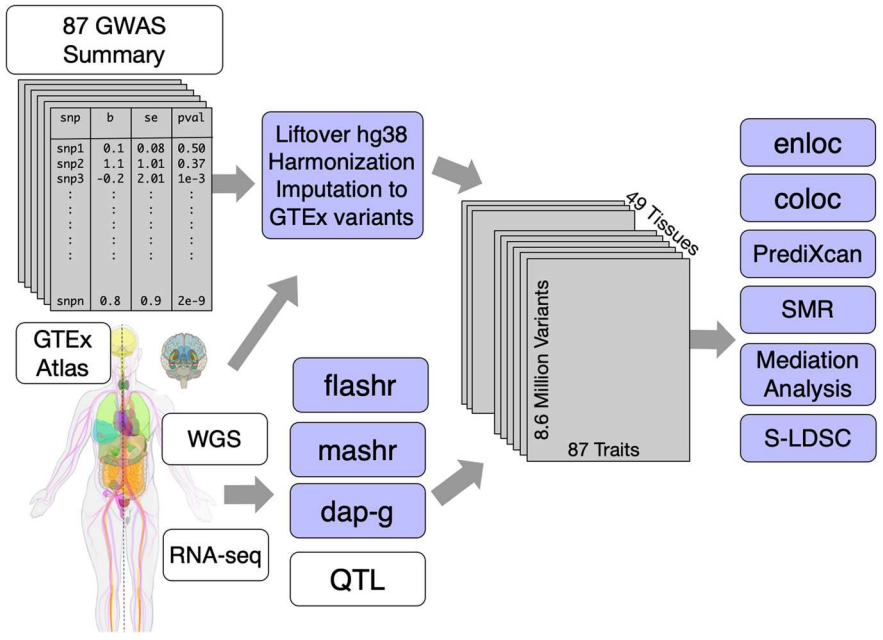

Fig. 1 Overview of workflow for mapping complex trait-associated QTLs. Full variant association summary statistics results from 114 GWAS were downloaded, standardized, and imputed to the GTEx v8 WGS variant calls (maf > 0.01) for analyses. A total of 8.87 million imputed and genotyped variants were investigated to identify trait-associated QTLs. A total of 49 tissues, 87 studies (74 distinct traits), and 23,268 protein-coding genes and lncRNAs remained after stringent quality assurance protocols and selection criteria. A wide array of complex trait classes, including cardiometabolic, anthropometric, and psychiatric traits, were included

We downloaded and processed 114 publicly available GWAS datasets with genome-wide variant association summary statistics (here onwards, summary statistics). After data harmonization, format standardization, missing data imputation, and other quality assurance steps (Additional file 1: Fig. S1, Fig. S2, and Fig. S3), we retained 87 datasets representing 74 distinct complex traits including cardiometabolic, hematologic, neuropsychiatric, and anthropometric traits (Additional file 1: Fig. S4). We provide the full list of datasets used in our study and all processing scripts as a resource to the community (Additional file 2: Table S1 and Additional file 1: Table S2).

Using these resources, we sought to identify likely causal associations among these gene- and alternatively spliced transcript-associated variants (eVariants and sVariants, respectively). For this purpose, we applied colocalization, enrichment, and association analyses, and provide a resource to enable investigations into gene prioritization approaches for disease associations.

Gene expression and alternative splicing dysregulations have been proposed as the underlying mechanism of the association signals in many diseases [5, 11, 21–24]. Similar to previous reports [8], we observed robust and widespread enrichment of eQTLs and sQTLs among disease-associated variants (Fig. 2). This observation suggests a causal role for expression and splicing regulation in complex traits. Figure 2 also illustrates the dangers of using a naive approach to assigning causal genes to GWAS variants that are associated with expression or splicing, especially when using loose *p* value thresholds. For example, with a *p* value threshold of 0.05, over 97% of common variants will be assigned some gene in some tissue associated at that level.

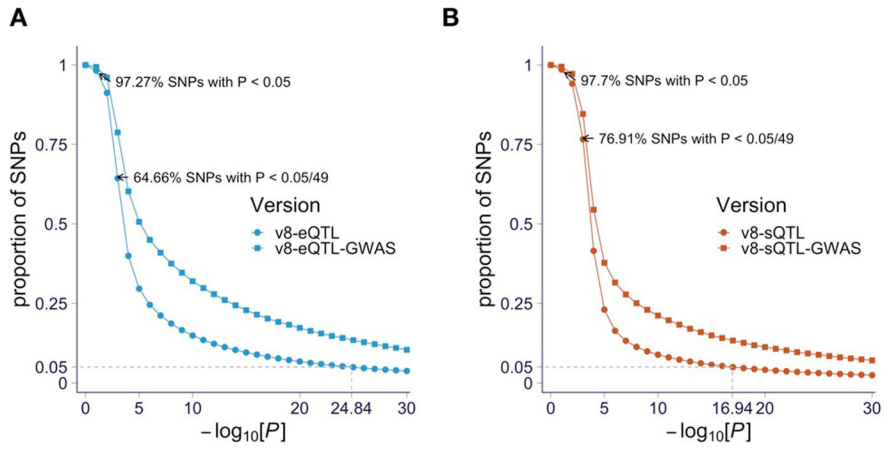

Fig. 2 Expression and splicing QTL enrichment among GWAS variants. The proportion of genetic variants associated with gene expression (**a**) and splicing (**b**) of at least one gene in at least one tissue for each *p* value cutoff (on *x*-axis in $-\log_{10}(p)$ scale) is shown. The proportions for all tested variants are shown as circles, and the proportions for the GWAS catalog variants are shown as squares

Dose-dependent regulatory effects of expression and alternative splicing on complex traits

Nevertheless, enrichment studies can be confounded by many unknown factors. Therefore, we sought to gather stronger evidence for a causal link by testing whether there is a dose-dependent effect of expression and splicing QTLs on complex traits. Figure 3a illustrates schematically our approach. We examined whether expression or splicing associated variants (referred to as e/sVariants for the remainder of the paper) with higher impact on gene expression or splicing lead to higher impact on a complex trait, i.e., a larger GWAS effect (Fig. 3a). The impact of the regulation of a gene on a trait is quantified by the slope β_{gene}. That is, a null hypothesis of no dose-dependent effect is equivalent to $\beta_{\text{gene}} = 0$.

To reduce unnecessary noise in the analysis, we included only the most likely causal e/sVariant within each credible set as determined by the e/sQTL fine-mapping (denoted "fine-mapped variants" throughout the remainder of the paper; see Methods on QTL fine-mapping).

First, we quantified dose-dependent effect of expression and splicing regulation on the trait as the average mediating effect size, $\bar{\beta}$. We calculated this average effect using the Pearson correlation between the absolute values of the molecular and complex trait effect sizes ($\text{cor}(|\gamma|, |\delta|)$) across all fine-mapped variants (for any gene) for each trait-tissue pair. As hypothesized, we found, consistently across all tissue-trait pairs, a positive correlation between the GWAS and QTL effects, which was significantly larger than the permuted null with matched local LD. The average correlations were 0.18 (s.e. = 0.004, $p < 1 \times 10^{-30}$) and 0.25 (s.e. = 0.006, $p < 1 \times 10^{-30}$) for expression and splicing, respectively with the distribution of the median correlation across tissues for each trait shown in Fig. 3b. Averages and standard errors were calculated taking into account correlation between tissues, and *p* values were calculated against permuted null with matched local LD (Supplementary Text). The non-negative permuted correlation values indicate that local LD contributed to inflate the estimated mediation effect. These results provide the first line of evidence of the dose-response effect.

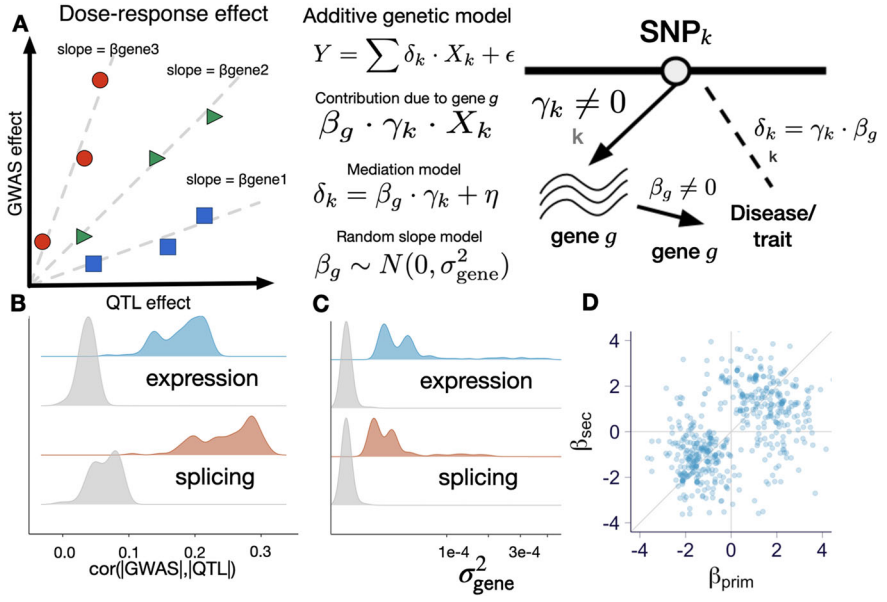

Fig. 3 Dose-dependent effects of QTLs on complex traits. Here, all analyses were performed with fine-mapped variants (QTL with highest posterior inclusion probability). **a** Schematic representation of dose-response model. **b** Correlation between QTL and GWAS effects, Cor($|\hat{\delta}|, |\hat{\gamma}|$). Gray distribution represents permuted null with matched local LD. Each data point corresponds to the median correlation for the trait across 49 tissues. **c** Average mediated effects from mediation model (σ^2_{gene}, median across tissues). Gray distribution represents permuted null with matched local LD. **e** Mediated effects of secondary vs. primary eQTLs of genes with colocalization probability (rcp) >0.10. in whole blood, genes for all 87 traits are shown

To test and account for mediation effect heterogeneity (different slope/dosage sensitivity for different genes), we modeled the gene-specific mediation effect, β_g, as a random variable following a normal distribution $\beta_g \sim \mathcal{N}(0, \sigma^2_{\text{gene}})$. Under this random-effects model, the null hypothesis can be stated as $\sigma^2_{\text{gene}} = 0$ (Supplementary Text; Fig. 3c). As shown in Fig. 3c, these effects were significantly larger than expected from the permuted null (expression $p = 1.8 \times 10^{-9}$; splicing $p = 2.5 \times 10^{-7}$). These results indicate that strong genetic effects on expression or splicing are more likely to have a strong association to complex traits, adding strong support to a dose-dependent relationship between gene regulation and downstream traits.

Importantly, by averaging across all genes, the estimates, from both the average and the random-effects approach, of the mediating effect are robust to confounding due to LD, as discussed in the Supplementary Text.

Another way to account for mediation effect heterogeneity is to make use of the allelic series of independent eQTLs identified for over half of the eGenes [9]. We examined whether the mediating effect ($\beta = \delta/\gamma$) inferred from the primary eQTL (β_{prim}) was consistent with the one inferred from the secondary eQTL (β_{sec}). Among the independent eQTLs for a given gene, we called primary the one with the larger effect size. We considered only fine-mapped eQTLs given the low power to detect multiple independent sQTLs. We confirmed this concordance, as reported by the GTEx consortium [9], demonstrating that the correlation between the primary and secondary mediating effects is larger than expected given the LD between them. To better visualize this concordance, we plotted the estimated mediating effects of primary against the secondary eQTLs (whole

blood shown here but other tissues look similar) in Fig. 3d and showed that they cluster in the first and third quadrants. All gene-trait pairs with relatively high regional colocalization probability (rcp > 0.10, see colocalization details below) are shown here to facilitate visualization, but the clustering around the diagonal line was observed even without the filtering. This provides a third confirmatory evidence for the widespread dose-dependent effects of eQTLs on complex traits.

Note that genes with discordant effects within the allelic series would be harder to detect and suggest more complex causal relationship or context specificity.

Causal gene prediction and prioritization

In addition to genome-wide analyses that shed light on the molecular architecture of complex traits, QTL analysis of GWAS data can identify potential causal genes and molecular changes in individual GWAS loci. Towards this end, we performed association analysis with genetically predicted regulation and colocalization (Fig. 4a). After evaluating the performance of *coloc* and *enloc* [16, 17], we chose *enloc* as our primary approach, due to its use of hierarchical models to estimate colocalization priors [16] and its ability to account for multiple causal variants. The *coloc* assumption of a single causal variant drastically reduces performance especially in large QTL datasets such as GTEx with widespread allelic heterogeneity. For a more extensive discussion on the benefits of Bayesian colocalization methods and comparison of *enloc* to other colocalization approaches including SMR-HEIDI, see [25]. We estimated the posterior regional colocalization probability (rcp), using *enloc*, for 12,072,964 tissue-gene-GWAS locus-trait tuples and 67,943,800 tissue-splicing event-GWAS locus-trait tuples. For the tally of colocalized genes, we used rcp > 0.5 as a stringent cutoff as demonstrated below with the low colocalization probabilities of height loci using two different datasets.

In total, we identified 3477 (15% of 23,963) unique genes colocalizing with GWAS hits (rcp > 0.5) across all traits and tissues analyzed. Similarly, 3157 splicing events (1% out of 310,042) colocalized with GWAS hits, corresponding to 1226 genes with at least one colocalized splicing event (5% of 23,963).

Colocalization of e/sQTLs with GWAS variants provides important causal support for molecular traits. However, we found their estimates to be overly conservative. To illustrate this point, we tested the colocalization of height with itself, using two large-scale studies of individuals of European-ancestry individuals: GIANT [26] and UK Biobank. We started by performing fine-mapping of both GWAS results using *susier* [27]. Notably, only 416 (39%) of GIANT's fine-mapped credible sets overlapped with the corresponding UK Biobank credible sets. We estimated the colocalization probability as the sum of the product of posterior inclusion probabilities of variants for each of the 1069 independent credible sets in GIANT, which is similar to the approach used by eCAVIAR [15]. Two thirds of the GIANT credible sets (66.2%) had a colocalization probability below 0.01, and about half (48.9%) had a colocalization probability below 0.001. In other words, two thirds of the loci found by GIANT would be considered not to be colocalized with UK Biobank's loci when using a seemingly very loose colocalization probability cutoff of 0.01. Given the larger sample size of the UK Biobank GWAS (n = 337,119 UKB GWAS vs. n = 253,288 for GIANT), the low colocalization cannot be attributed to lack of power. This result is likely due in part to the sensitivity to small LD differences between different EUR populations that make up large GWAS meta-analysis cohorts such as GIANT. Our

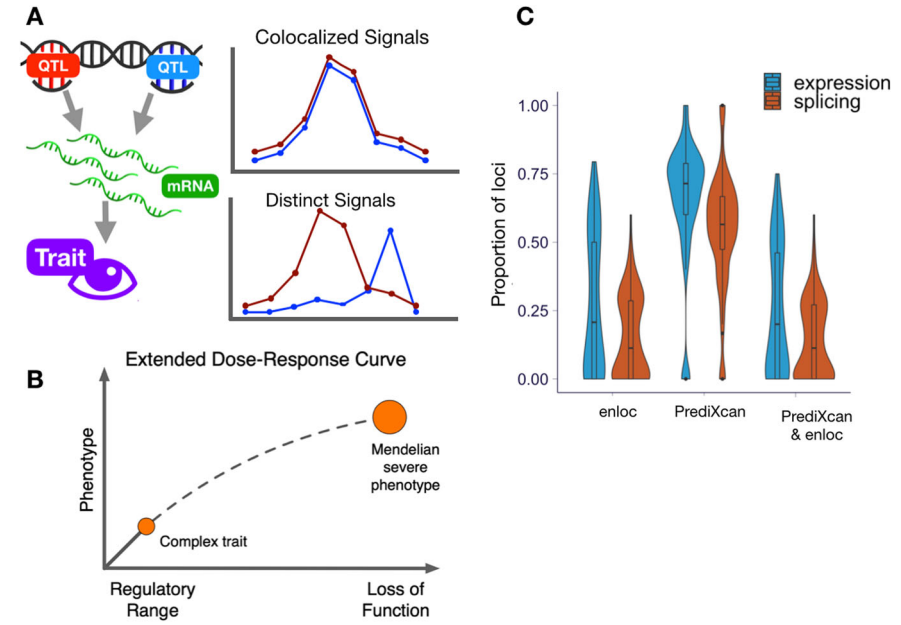

Fig. 4 Identifying and validating predicted causal genes. **a** Schematic representation of association and colocalization approaches. **b** Schematic representation of extrapolating the dose-response curve to the Mendelian end of phenotypic variation spectrum [37]. **c** Proportion of GWAS-associated loci per trait that contain colocalized and PrediXcan-associated signals for expression and splicing

analysis illustrates the fact that colocalization probability estimates are highly conservative and may miss many causal genes, and low colocalization probability should not be interpreted as evidence of lack of a causal link between the molecular phenotype and the GWAS trait. Notice that this limitation is not inherent to the colocalization method itself but the limitation of currently available large-scale GWAS meta-analysis results.

A complementary approach to colocalization is to estimate the GWAS trait association with genetically predicted gene expression or splicing [19]. The GTEx v8 data provides an important expansion of these analyses, allowing generation of prediction models in 49 tissues with whole genome sequencing data to impute gene expression and splicing variation. We trained prediction models using a variety of approaches and selected the top performing one based on precision, recall, and other metrics [28]. Briefly, the optimal model uses fine-mapping probabilities for feature selection and exploits global patterns of tissue sharing of regulation (Supplementary Text) to improve prediction. In-depth comparison of these fine-mapped models with Elastic Net-based and CTIMP [29] models is described in [28]. The analysis presented here uses these improved models (fine-mapped-*mashr*) instead of Elastic Net as reported in the main GTEx publication [9]. Multi-SNP prediction models were generated for a total of 686,241 gene-tissue and 1,816,703 splicing event-tissue pairs. The larger sample size and improved models led to an increase in the number of expression models to a median across tissues of 14,062, from a median of 4776 GTEx v7 Elastic Net models (median increase at 191%, Additional file 1: Fig. S5). Splicing models are available only for the v8 release.

Next, we computed the association between an imputed molecular phenotype (expression or splicing) and a trait to estimate the genic effect on the trait, using the summary

statistics-based PrediXcan [24]. Given the widespread tissue sharing of regulatory variation [8], we also computed MultiXcan scores to integrate patterns of associations from multiple tissues and increase statistical power [10]. Out of the 22,518 genes tested with PrediXcan, 6407 (28%) showed a significant association with at least one of the 87 traits at Bonferroni-corrected p value threshold ($p < 0.05/686,241$, where the denominator is the number of gene-tissue pairs tested; Additional file 1: Fig. S6). For splicing, about 15% (20,364 of 138,890) of tested splicing events showed a significant association ($p < 0.05/1,816,703$, where the denominator is the number of intron-tissue pairs tested). Nearly all traits (94%; 82 out of 87) showed at least one significant gene-level PrediXcan association in at least one tissue (Additional file 1: Figs. S7 and S8); the median number of associated genes across traits was 974. This resource of PrediXcan associations can be used to prioritize a list of putatively causal genes for follow-up studies.

To replicate the PrediXcan expression associations in an independent dataset, BioVU, which is a large-scale biobank tied to Electronic Health Records [30, 31], we selected seven traits with predicted high statistical power. Out of 947 gene-tissue-trait discoveries tested, 458 unique gene-tissue-trait triplets (48%) showed replication in this independent biobank (PrediXcan association $p < 0.05$; see Supplementary Text). Further confirming this statistical replication in BioVU, we used the PheWAS [32] catalog as the silver standard and found an AUC curve of 0.62. [33].

Altogether, these results provide abundant links between gene regulation and GWAS loci. To further quantify this, we split the genome into approximately LD-independent blocks [34] and identified blocks with a significant GWAS variant for each trait (at Bonferroni threshold adjusted for number of variants $0.05/8.8 \times 10^6 \sim 5.7 \times 10^{-9}$); we refer to any such region-trait pair by "GWAS locus." We calculated the proportion of GWAS loci that contain a significantly associated gene via PrediXcan or a colocalized gene via *enloc* (rcp > 0.5). Briefly, the LD blocks are defined by analyzing empirical patterns of LD observed in 1000 Genomes [35] and variants in different regions are unlikely to be correlated, thus providing us with a data-driven criterion to distinguish independent genomic signals.

Across the traits, 72% (3899/5385) of GWAS loci had a PrediXcan expression association in the same LD block, of which 55% (2125/3899) had evidence of colocalization with an eQTL; for splicing, 62% (3345/5385) had a PrediXcan association of which 34% (1135/3345) colocalized with an sQTL (Additional file 1: Table S3). From the combined list of eGenes and sGenes, 47% of loci have a gene with both *enloc* and PrediXcan support. The distribution of the proportion of associated and colocalized GWAS loci across 87 traits is summarized in Fig. 4c; for a typical complex trait, about 20% of GWAS loci contained a colocalized, significantly associated gene while 11% contained a colocalized, significantly associated splicing event. These results propose function for a large number of GWAS loci, but most loci remain without candidate genes, highlighting the need to expand the resolution of transcriptome studies.

A recent report estimates that the proportion of trait variance explained by the assayed transcriptome is on average 11% [36]. Even though this number is not directly comparable with the proportion of loci with support from PrediXcan and *enloc*, some discussion is warranted. Differences may arise with our analysis from the fact that (1) GTEx v8 doubles the number of samples with both genotype and RNA-seq relative to v7, (2) we include links based on splicing in addition to expression, (3) a variant may act through

both regulation of expression levels and other undetected mechanisms (pleiotropy), and (4) attenuation bias may reduce the estimates given the error in eQTL effect sizes.

Of note, two members of the sterolin family, *ABCG5* and *ABCG8*, showed highly significant predicted causal associations using both PrediXcan and *enloc* for LDL-C levels and self-reported high cholesterol levels. *ABCG8* showed more significant associations in both datasets (chr2: 43838964–43878466; UKB self-reported high cholesterol: $-\log10(p_{\text{PrediXcan}})$ = 38.43, rcp = 0.985; GLGC LDL-C: $-\log10(p_{\text{PrediXcan}})$ = 71.40, rcp = 0.789), compared to *ABCG5* (chr2: 43812472–43838865; $-\log10(p_{\text{PrediXcan}})$ = 36.85, rcp = 0.941; $-\log10(p_{\text{PrediXcan}})$ = 80.80, rcp = 0.705). Mutations in either of the two ATP-binding cassette (ABC) half-transporters, *ABCG5* and *ABCG8*, lead to reduced secretion of sterols into bile and, ultimately, obstruct cholesterol and other sterols exiting the body [38]. In mice with disrupted *Abcg5* and *Abcg8* (G5G8-/-), a 2- to 3-fold increase in the fractional absorption of dietary plan sterols and extremely low biliary cholesterol levels was observed, indicating that disrupting these genes contributes greatly to plasma cholesterol levels [39]. The overexpression of human *ABCG5* and *ABCG8* in transgenic Ldlr-/- mice resulted in 30% reduction in hepatic cholesterol levels and 70% reduced atherosclerotic legion in the aortic root and arch [40] after 6 months on a Western diet.

Several other lipid-associated loci were also consistently predicted as causal across OMIM, the rare variant derived set, PrediXcan and *enloc*. Rare protein-truncating variants in *APOB* have been previously associated with reduced LDL-C and triglyceride levels and reduced coronary heart disease risk [41]. Interestingly, *APOB* has been predicted as a causal gene in four related traits, coronary artery disease, LDL-C levels, triglyceride levels, and self-reported high cholesterol levels. Among the four traits, PrediXcan showed the highest association to LDL-C levels ($-\log10(p_{\text{PrediXcan}})$ = 130.89; rcp = 0.485) while self-reported high cholesterol showed the strongest evidence using *enloc* at nearly maximum posterior probability ($-\log10(p_{\text{PrediXcan}})$ = 93.66; rcp = 0.969). Although *APOB* has been suggested as a better molecular indicator of predicted cardiac events in place of LDL-C levels [42, 43], its translation has been surprisingly slow in clinical practice [44]. Here, we provide an additional support for the crucial role *APOB* may play in predicting lipid traits.

Performance for identifying "ground truth" genes

To compare the ability of different approaches to identify the causal gene that mediates the association between GWAS loci and the traits, we sought to curate sets of "ground truth" genes using information that is independent of GWAS results (Additional file 1: Fig. S9). We call these sets "silver standards" as a reminder of their imperfect nature. The first silver standard was based on the OMIM (Online Mendelian Inheritance in Man) database [45], and the second one was based on publicly available rare variant tests from exome-wide association studies [46–48], resulting in 1592 OMIM gene-trait pairs and 101 rare variant-based gene-trait pairs (Additional file 3: Table S4, Additional file 4: Table S5).

The rationale behind the choice of the OMIM database is the comorbidity among Mendelian and complex diseases suggesting that genes whose loss of function cause Mendelian diseases also manifest in milder phenotypic variation when modified to a lesser degree by regulatory variation [49, 50]. In other words, that the dose-response curve at the regulatory range may be extrapolated to the rare, loss-of-function end (Fig. 4b). The rationale behind the use of the rare variant association study results is the excess of

deleterious rare variants associated with complex traits in genes that are in the vicinity of common variants associated with the same trait [46, 51, 52]. Note that rare variant associations are nearly independent of common variants due to the allele frequency difference between them.

For the analysis, we partitioned the genome into approximately independent LD blocks [34] and considered all the blocks where a silver standard gene was available for the trait. Since only genes in the vicinity of an index gene can be discovered with cis-regulatory information, we only considered the LD blocks with a GWAS significant variant (Additional file 1: Fig. S10). This selection resulted in 228 OMIM gene-trait pairs (28 distinct traits) and 80 rare variant-associated gene-trait pairs (5 distinct traits) that are located within the same LD block as the GWAS locus for a matched trait.

Both PrediXcan and *enloc* based on expression and splicing showed good sensitivity and specificity for identifying the silver standard genes as demonstrated by the ROC curves in Fig. 5a, b. These are well above the gray random guess lines indicating the predictive ability of these methods to find causal genes (see comparison with permuted null in Additional file 1: Fig. S11).

For applications such as target selection for drug development or follow-up experiments, another relevant metric is the precision or, equivalently, positive predictive value (PPV)—the probability that the gene-trait link is causal given that it is called significant

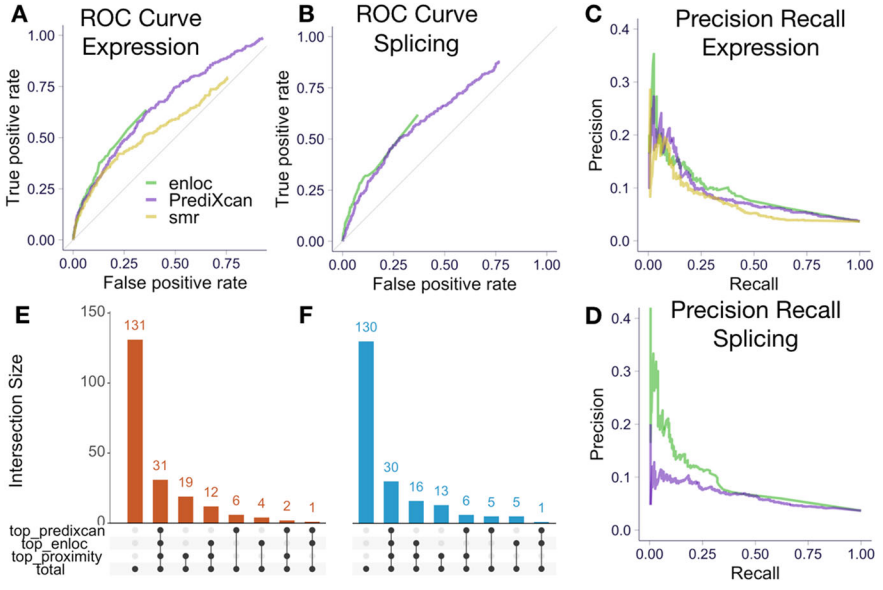

Fig. 5 Causal gene identification performance. ROC curves of *enloc* and PrediXcan statistics to identify the "causal" genes (OMIM silver standard) using expression (**a**) and splicing (**b**) are shown. Precision recall curves of *enloc* and PrediXcan to identify silver standard genes using expression (**c**) and splicing (**d**) (we show the precision in the range 0 to 0.4 to improve visualization). The number of GWAS loci (LD block-trait pairs) where the OMIM gene was ranked at the top by proximity, *enloc*, and PrediXcan using expression (**e**) and splicing (**f**). In 131 loci out of 206, the OMIM gene was not ranked at the top by either proximity, significance, or colocalization. In thirty one of the loci, the OMIM gene was ranked first by all three criteria. In nineteen loci, the OMIM gene was closest gene (to the top GWAS variant) but not the top gene by PrediXcan significance nor *enloc*'s colocalization probability

or colocalized. Precision recall curves for expression- and splicing-based predictions are shown in Fig. 5c, d. With more stringent threshold (towards the left in the recall axis), higher precision is obtained.

For example, 8.7% of genes with PrediXcan significant genes ($p < 0.05/49 \times$ number of gene/trait pairs) were OMIM genes and 14.8% of genes with high colocalization probability (rcp > 0.5) were also OMIM genes for matched traits.

Multiple factors contribute to the rather low precision. One of them is the widespread molecular pleiotropy [9], i.e., multiple genes affected by the same trait-associated variants. Another factor reducing the overall causal gene detection performance is the inherent bias of the OMIM gene list. Our current understanding of gene function is biased towards protein-coding variants with very large effects, as reflected in the list of OMIM genes. Genes associated to rare severe disease tend to be depleted of regulatory variation [53, 54], which will decrease the performance of a QTL-based method [54].

Among the 206 loci with at least one OMIM gene (a few loci contained multiple OMIM genes), an OMIM gene was the closest to the top GWAS SNP in 31.6% of the loci, it was the most colocalized in 24.8% of the loci, and it was the most significant in 20.4% of the loci (Fig. 5e, f).

To further investigate whether this predictive power could be improved by combining multiple criteria, we performed a joint logistic regression of OMIM gene status on (1) the proximity of the top GWAS variant to the nearest gene (distance to the gene body), (2) posterior probability of colocalization, and (3) PrediXcan association significance between QTL and GWAS variants. To make the scale of the three features more comparable, we used their respective ranking. When genes did not have an enloc or PrediXcan score, they were assigned to the last position in the ranking. All three features were significant predictors of OMIM gene status, with better ranked genes more likely to be OMIM genes (proximity $p = 2.0 \times 10^{-2}$, *enloc* $p = 6.1 \times 10^{-3}$, PrediXcan $p = 2.5 \times 10^{-4}$), indicating that each method provides an additional source of causal evidence even after conditioning on the others. Similar results were obtained using splicing colocalization and association scores and the rare variant-based silver standard, as shown in Additional file 1: Table S6. These results provide further empirical evidence that a combination of colocalization and association methods will perform better than individual ones. The significance of the proximity score even after accounting for significance and colocalization indicates missing regulatory events, i.e., mechanisms that may be uncovered by assaying other tissue or cell type contexts, larger samples, and other molecular traits, underscoring the need to expand the size and breadth of QTL studies. Proximity criterion also helps resolve cases when QTL data indicates multiple genes with similar significance.

Predicted OMIM genes included well-known findings such as *PCSK9* for LDLR, with *PCSK9* significant and colocalized for relevant GWAS traits (LDL-C levels, coronary artery disease, and self-reported high cholesterol), and *Interleukins* and *HLA* subunits for asthma, both significant and colocalized for related immunological traits. Significantly associated and colocalized genes that predicted OMIM genes also included *FLG* (eczema), *TPO* (hypothyroidism), and *NOD2* (inflammatory bowel disease) (see Additional file 1: Table S4 for complete list). Analysis with rare variant-based silver standard yielded similar conclusions (Supplementary Text; Additional file 1: Fig. S12).

Tissue enrichment of GWAS signals

The broad sharing of regulatory variation across tissues and the reduced significance of tissue-specific eQTLs make causal tissue identification challenging. To address this problem, we devised a novel approach to identify tissues of relevance for the etiology of complex traits. We investigated the patterns of tissue specificity and tissue sharing of PrediXcan association results across 49 tissues. For each trait-gene pair, the PrediXcan z-score can be represented as a 49×1 vector with each entry being the gene-level z-score in the corresponding tissue (if the prediction model of the gene is not available in that tissue, we filled in zero). To explore the tissue specificity of the PrediXcan z-score vector, we proceeded by assigning the z-score vector to a tissue-pattern category and tested whether certain tissue-pattern categories were over-represented among colocalized PrediXcan genes as compared to non-colocalized genes. We used the FLASH factors identified from matrix factorization applied to the cis-eQTL effect size matrix, as PrediXcan and cis-eQTL shared similar tissue-sharing pattern (Supplementary Text). To obtain a set of detailed and biologically interpretable tissue-pattern categories from the 31 FLASH factors, we manually merged them into 18 categories as shown in Additional file 1: Fig. S13. For each trait, we projected the z-score vector of each gene to one of the 31 FLASH factors (as described in Section 9 of Additional file 1) so that the gene was assigned to the corresponding tissue-pattern category. We defined a "positive" set of genes as the ones with PrediXcan p value that meets Bonferroni significance at $\alpha = 0.05$ in at least one tissue and *enloc* rcp > 0.01 in at least one tissue, which could be thought as a set of candidate genes affecting the trait through expression level. We chose a rather low threshold used for the rcp due to the stringent conservative nature of colocalization probabilities. We also constructed a "negative" set of genes with *enloc* rcp = 0, which could be thought as a set of genes whose expressions were unlikely to affect the trait. We proceeded to test whether certain tissue-pattern categories were enriched in "positive" set as compared to "negative" set. Since the main focus of this analysis was tissue-specific patterns, we excluded *Factor1* (the cross-tissue factor) and *Factor25* (likely to be a tissue-shared factor capturing tissues with large sample size). Additionally, we excluded *Factor7* (testis), as it was unlikely to be the mediating tissue but might introduce false positives. We tested the enrichment of each tissue-pattern category by Fisher's exact test ("positive"/"negative" sets and in/not in tissue-pattern category). Among 87 traits, 82 traits had *enloc* signal and the enrichment of these was calculated accordingly.

Using the pattern of tissue classes of non-colocalized genes (rcp = 0) as the expected null, we assessed whether significantly associated and colocalized genes (PrediXcan significant and rcp > 0.01) were over-represented in certain tissue classes (Fig. 6). Consistent with previous reports [11, 55], we identified several instances in which the most significant tissue is supported by current biological knowledge. For example, blood cell count traits were enriched in whole blood, neuroticism and fluid intelligence in brain/pituitary, hypothyrodism in thyroid, coronary artery disease in artery, and cholesterol-related traits in liver. Taken together, these results show the potential of leveraging regulatory variation to help identify tissues of relevance for complex traits.

Discussion

We performed in-depth examination of the phenotypic consequences of the genetic regulation of the transcriptome and provide data-driven analytical approaches to benchmark

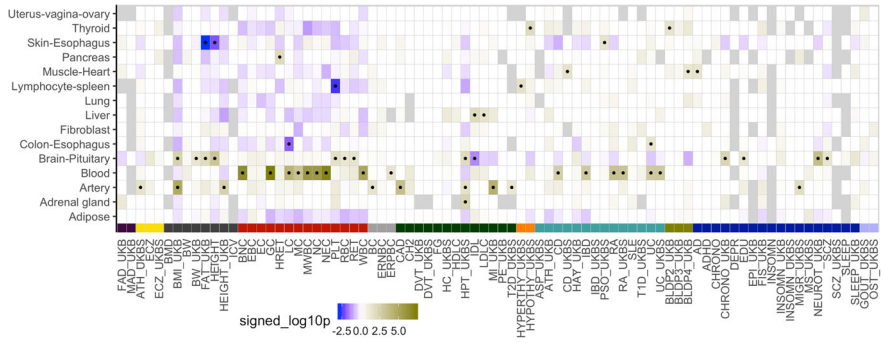

Fig. 6 Identifying trait-relevant tissues using tissue-specific enrichment. Enrichment of tissue-specific association and colocalization compared to the pattern of tissue specificity of non-colocalized genes. Over-representation of the tissue class for PrediXcan-significant and colocalized genes is indicated by dark yellow while depletion is indicated by blue. Black dots label the tissue class-trait pairs passing the nominal *p* value significance threshold of 0.05. Abbreviation: Table S2. Trait category colors: Fig. S4

methods that assign function to GWAS loci and best-practice guidelines for using the GTEx resources to interpret GWAS results. We provide a systematic empirical demonstration of the widespread dose-dependent effect of expression and splicing on complex traits, i.e., variants with larger impact at the molecular level have larger impact at the trait level. Furthermore, we found that target genes in GWAS loci identified by *enloc* and PrediXcan were predictive of OMIM genes for matched traits, implying that for a proportion of the genes, the dose-response curve can be extrapolated to the rare and more severe end of the genotype-trait spectrum. The observation that common regulatory variants target genes also implicated by rare coding variants underscores the extent to which these different types of genetic variants converge to mediate a spectrum of similar pathophysiological effects and may provide a powerful approach to drug target discovery.

We implemented association and colocalization methods that leverage the observed allelic heterogeneity of expression traits. After extensive comparison using two independent sets of silver standard gene-trait pairs, we conclude that combining *enloc*, PrediXcan, and proximity ranking outperforms the individual approaches. The significance of the proximity ranking is a sign of the "missing regulability" emphasizing the need to expand the resolution, sample size, and range of contexts of transcriptome studies as well as to examine other molecular mechanisms.

We caution that the increased power offered by this release of the GTEx resources also brings higher risk of false links due to LD contamination and that naive use of eQTL or sQTL association *p* values to assign function to a GWAS locus can be misleading. Colocalization approaches can be used to weed out LD contamination, but given the lack of LD references from source studies, they can also be overtly conservative. General purpose reference LD from publicly available sources are not sufficient for fine-mapping and colocalization approaches, which can be highly sensitive to LD misspecification when only summary results are used [56]. The GWAS community has made great progress in recognizing the need to share summary results, but to take full advantage of these data, improved sharing of LD information from the source study as well as from large sequencing reference datasets is also required.

Finally, we generated several resources that can open the door for addressing key questions in complex trait genomics. We present a catalog of gene-level associations, including

potential target genes for nearly half of the GWAS loci investigated here that provides a rich basis for studies on the functional mechanisms of complex diseases and traits. We provide a database of optimal gene expression imputation models that were built on the fine-mapping probabilities for feature selection and that leverage the global patterns of tissue sharing of regulation to improve the weights. These imputation models of expression and splicing, which to date has been challenging to study, provide a foundation for transcriptome-wide association studies of the human phenome—the collection of all human diseases and traits—to further accelerate discovery of trait-associated genes. Collectively, these data thus represent a valuable resource, enabling novel biological insights and facilitating follow-up studies of causal mechanisms.

Authors

* alphabetic order

Lead Analysts* *Equal contribution* Alvaro N Barbeira, Rodrigo Bonazzola, Eric R Gamazon, Yanyu Liang, YoSon Park

Analysts* François Aguet, Lisa Bastarache, Ron Do, Gao Wang, Andrew R Hamel, Farhad Hormozdiari, Zhuoxun Jiang, Daniel Jordan, Sarah Kim-Hellmuth, Boxiang Liu, Milton D Pividori, Abhiram Rao, Marie Verbanck, Dan Zhou

GTEx GWAS Working Group* François Aguet, Kristin Ardlie, Alvaro N Barbeira, Rodrigo Bonazzola, Christopher D Brown, Lin Chen, Eric R Gamazon, Kevin Gleason, Andrew R Hamel, Farhad Hormozdiari, Hae Kyung Im, Sarah Kim-Hellmuth, Tuuli Lappalainen, Yanyu Liang, Boxiang Liu, Dan L Nicolae, Yoson Park, Milton D Pividori, Abhiram Rao, John M. Rouhana, Ayellet V Segrè, Xiaoquan Wen

Senior Leadership* Kristin Ardlie, Christopher D. Brown, Hae Kyung Im, Tuuli Lappalainen, Mark McCarthy, Stephen Montgomery, Ayellet V Segrè, Matthew Stephens, Xiaoquan Wen

Manuscript Writing Group* Eric R Gamazon, Hae Kyung Im, Tuuli Lappalainen, Yanyu Liang, YoSon Park

Corresponding Author* Hae Kyung Im

GTEx Consortium

Laboratory and Data Analysis Coordinating Center (LDACC): François Aguet[1], Shankara Anand[1], Kristin G Ardlie[1], Stacey Gabriel[1], Gad Getz[1,2], Aaron Graubert[1], Kane Hadley[1], Robert E Handsaker[3,4,5], Katherine H Huang[1], Seva Kashin[3,4,5], Xiao Li[1], Daniel G MacArthur[4,6], Samuel R Meier[1], Jared L Nedzel[1], Duyen Y Nguyen[1], Ayellet V Segrè[1,7], Ellen Todres[1]

Analysis Working Group (funded by GTEx project grants): François Aguet[1], Shankara Anand[1], Kristin G Ardlie[1], Brunilda Balliu[8], Alvaro N Barbeira[9], Alexis Battle[10,11], Rodrigo Bonazzola[9], Andrew Brown[12,13], Christopher D Brown[14], Stephane

E Castel[15,16], Don Conrad[17,18], Daniel J Cotter[19], Nancy Cox[20], Sayantan Das[21], Olivia M de Goede[19], Emmanouil T Dermitzakis[12,22,23], Barbara E Engelhardt[24,25], Eleazar Eskin[26], Tiffany Y Eulalio[27], Nicole M Ferraro[27], Elise Flynn[15,16], Laure Fresard[28], Eric R Gamazon[29,30,31,20], Diego Garrido-Martín[32], Nicole R Gay[19], Gad Getz[1,2], Aaron Graubert[1], Roderic Guigó[32,33], Kane Hadley[1], Andrew R Hamel[7,1], Robert E Handsaker[3,4,5], Yuan He[10], Paul J Hoffman[15], Farhad Hormozdiari[34,1], Lei Hou[35,1], Katherine H Huang[1], Hae Kyung Im[9], Brian Jo[24,25], Silva Kasela[15,16], Seva Kashin[3,4,5], Manolis Kellis[35,1], Sarah Kim-Hellmuth[15,16,36], Alan Kwong[21], Tuuli Lappalainen[15,16], Xiao Li[1], Xin Li[28], Yanyu Liang[9], Daniel G MacArthur[4,6], Serghei Mangul[26,37], Samuel R Meier[1], Pejman Mohammadi[15,16,38,39], Stephen B Montgomery[28,19], Manuel Muñoz-Aguirre[32,40], Daniel C Nachun[28], Jared L Nedzel[1], Duyen Y Nguyen[1], Andrew B Nobel[41], Meritxell Oliva[9,42], YoSon Park[14,43], Yongjin Park[35,1], Princy Parsana[11], Ferran Reverter[44], John M Rouhana[7,1], Chiara Sabatti[45], Ashis Saha[11], Ayellet V Segrè[1,7], Andrew D Skol[9,46], Matthew Stephens[47], Barbara E Stranger[9,48], Benjamin J Strober[10], Nicole A Teran[28], Ellen Todres[1], Ana Viñuela[49,12,22,23], Gao Wang[47], Xiaoquan Wen[21], Fred Wright[50], Valentin Wucher[32], Yuxin Zou[51]

Analysis Working Group (not funded by GTEx project grants): Pedro G Ferreira[52,53,54], Gen Li[55], Marta Melé[56], Esti Yeger-Lotem[57,58]

Leidos Biomedical - Project Management: Mary E Barcus[59], Debra Bradbury[60], Tanya Krubit[60], Jeffrey A McLean[60], Liqun Qi[60], Karna Robinson[60], Nancy V Roche[60], Anna M Smith[60], Leslie Sobin[60], David E Tabor[60], Anita Undale[60]

Biospecimen collection source sites: Jason Bridge[61], Lori E Brigham[62], Barbara A Foster[63], Bryan M Gillard[63], Richard Hasz[64], Marcus Hunter[65], Christopher Johns[66], Mark Johnson[67], Ellen Karasik[63], Gene Kopen[68], William F Leinweber[68], Alisa McDonald[68], Michael T Moser[63], Kevin Myer[65], Kimberley D Ramsey[63], Brian Roe[65], Saboor Shad[68], Jeffrey A Thomas[68,67], Gary Walters[67], Michael Washington[67], Joseph Wheeler[66]

Biospecimen core resource: Scott D Jewell[69], Daniel C Rohrer[69], Dana R Valley[69]

Brain bank repository: David A Davis[70], Deborah C Mash[70]

Pathology Mary E Barcus[59], Philip A Branton[71], Leslie Sobin[60]

ELSI study: Laura K Barker[72], Heather M Gardiner[72], Maghboeba Mosavel[73], Laura A Siminoff[72]

Genome Browser Data Integration & Visualization: Paul Flicek[74], Maximilian Haeussler[75], Thomas Juettemann[74], W James Kentv[75], Christopher M Lee[75], Conner C Powell[75], Kate R Rosenbloom[75], Magali Ruffier[74], Dan Sheppard[74], Kieron Taylor[74], Stephen J Trevanion[74], Daniel R Zerbino[74]

eGTEx groups: Nathan S Abell[19], Joshua Akey[76], Lin Chen[42], Kathryn Demanelis[42], Jennifer A Doherty[77], Andrew P Feinberg[78], Kasper D Hansen[79], Peter F Hickey[80], Lei Hou[35,1], Farzana Jasmine[42], Lihua Jiang[19], Rajinder Kaul[81,82], Manolis Kellis[35,1], Muhammad G Kibriya[42], Jin Billy Li[19], Qin Li[19], Shin Lin[83], Sandra E Linder[19], Stephen B Montgomery[28,19], Meritxell Oliva[9,42], Yongjin Park[35,1], Brandon L Pierce[42], Lindsay F Rizzardi[84], Andrew D Skol[9,46], Kevin S Smith[28], Michael Snyder[19], John Stamatoyannopoulos[81,85], Barbara E Stranger[9,48], Hua Tang[19], Meng Wang[19]

NIH program management: Philip A Branton[71], Latarsha J Carithers[71,86], Ping Guan[71], Susan E Koester[87], A. Roger Little[88], Helen M Moore[71], Concepcion R Nierras[89], Abhi K Rao[71], Jimmie B Vaught[71], Simona Volpi[90]

Affiliations 1. The Broad Institute of MIT and Harvard, Cambridge, MA, USA
2. Cancer Center and Department of Pathology, Massachusetts General Hospital, Boston, MA, USA
3. Department of Genetics, Harvard Medical School, Boston, MA, USA
4. Program in Medical and Population Genetics, The Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA
5. Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA
6. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA
7. Ocular Genomics Institute, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, USA
8. Department of Biomathematics, University of California, Los Angeles, Los Angeles, CA, USA
9. Section of Genetic Medicine, Department of Medicine, The University of Chicago, Chicago, IL, USA v 10. Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA
11. Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA
12. Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland
13. Population Health and Genomics, University of Dundee, Dundee, Scotland, UK
14. Department of Genetics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA
15. New York Genome Center, New York, NY, USA
16. Department of Systems Biology, Columbia University, New York, NY, USA
17. Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA
18. Department of Pathology & Immunology, Washington University School of Medicine, St. Louis, MO, USA
19. Department of Genetics, Stanford University, Stanford, CA, USA
20. Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
21. Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA
22. Institute for Genetics and Genomics in Geneva (iGE3), University of Geneva, Geneva, Switzerland

23. Swiss Institute of Bioinformatics, Geneva, Switzerland
24. Department of Computer Science, Princeton University, Princeton, NJ, USA
25. Center for Statistics and Machine Learning, Princeton University, Princeton, NJ, USA
26. Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, USA
27. Program in Biomedical Informatics, Stanford University School of Medicine, Stanford, CA, USA
28. Department of Pathology, Stanford University, Stanford, CA, USA
29. Data Science Institute, Vanderbilt University, Nashville, TN, USA
30. Clare Hall, University of Cambridge, Cambridge, UK
31. MRC Epidemiology Unit, University of Cambridge, Cambridge, UK
32. Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Barcelona, Catalonia, Spain
33. Universitat Pompeu Fabra (UPF), Barcelona, Catalonia, Spain
34. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
35. Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA
36. Statistical Genetics, Max Planck Institute of Psychiatry, Munich, Germany
37. Department of Clinical Pharmacy, School of Pharmacy, University of Southern California, Los Angeles, CA, USA
38. Scripps Research Translational Institute, La Jolla, CA, USA
39. Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA
40. Department of Statistics and Operations Research, Universitat Politècnica de Catalunya (UPC), Barcelona, Catalonia, Spain
41. Department of Statistics and Operations Research and Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA
42. Department of Public Health Sciences, The University of Chicago, Chicago, IL, USA
43. Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA
44. Department of Genetics, Microbiology and Statistics, University of Barcelona, Barcelona, Spain
45. Departments of Biomedical Data Science and Statistics, Stanford University, Stanford, CA, USA
46. Department of Pathology and Laboratory Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA
47. Department of Human Genetics, University of Chicago, Chicago, IL, USA
48. Center for Genetic Medicine, Department of Pharmacology, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA
49. Department of Twin Research and Genetic Epidemiology, King's College London, London, UK
50. Bioinformatics Research Center and Departments of Statistics and Biological Sciences, North Carolina State University, Raleigh, NC, USA
51. Department of Statistics, University of Chicago, Chicago, IL, USA
52. Department of Computer Sciences, Faculty of Sciences, University of Porto, Porto,

Portugal
53. Instituto de Investigação e Inovação em Sauúde, Universidade do Porto, Porto, Portugal
54. Institute of Molecular Pathology and Immunology, University of Porto, Porto, Portugal
55. Columbia University Mailman School of Public Health, New York, NY, USA
56. Life Sciences Department, Barcelona Supercomputing Center, Barcelona, Spain
57. Department of Clinical Biochemistry and Pharmacology, Ben-Gurion University of the Negev, Beer-Sheva, Israel
58 National Institute for Biotechnology in the Negev, Beer-Sheva, Israel
59. Leidos Biomedical, Frederick, MD, USA
60. Leidos Biomedical, Rockville, MD, USA
61. UNYTS, Buffalo, NY, USA
62. Washington Regional Transplant Community, Annandale, VA, USA
63. Therapeutics, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA
64. Gift of Life Donor Program, Philadelphia, PA, USA
65. LifeGift, Houston, TX, USA
66. Center for Organ Recovery and Education, Pittsburgh, PA, USA
67. LifeNet Health, Virginia Beach, VA. USA v 68. National Disease Research Interchange, Philadelphia, PA, USA v 69. Van Andel Research Institute, Grand Rapids, MI, USA
70. Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA
71. Biorepositories and Biospecimen Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD, USA
72. Temple University, Philadelphia, PA, USA
73. Virgina Commonwealth University, Richmond, VA, USA
74. European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK v 75. Genomics Institute, UC Santa Cruz, Santa Cruz, CA, USA
76. Carl Icahn Laboratory, Princeton University, Princeton, NJ, USA
77. Department of Population Health Sciences, The University of Utah, Salt Lake City, UT, USA
78. Schools of Medicine, Engineering, and Public Health, Johns Hopkins University, Baltimore, MD, USA
79. Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA
80. Department of Medical Biology, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia
81. Altius Institute for Biomedical Sciences, Seattle, WA, USA v 82. Division of Genetics, University of Washington, Seattle, WA, USA
83. Department of Cardiology, University of Washington, Seattle, WA, USA
84. HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA
85. Genome Sciences, University of Washington, Seattle, WA, USA
86. National Institute of Dental and Craniofacial Research, Bethesda, MD, USA
87. Division of Neuroscience and Basic Behavioral Science, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA
88. National Institute on Drug Abuse, Bethesda, MD, USA

89. Office of Strategic Coordination, Division of Program Coordination, Planning and Strategic Initiatives, Office of the Director, National Institutes of Health, Rockville, MD, USA
90. Division of Genomic Medicine, National Human Genome Research Institute, Bethesda, MD, USA

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-020-02252-4.

Additional file 1: Supplementary Materials including detailed methods, tables, and figures
Additional file 2: The metadata of the full list of 114 GWASs
Additional file 3: Presumed causal genes included in the OMIM database
Additional file 4: Genes suggested as causal by rare variant association studies
Additional file 5: BioVU table
Additional file 6: OMIM genes included in the analysis
Additional file 7: Rare variant silver standard genes included in the analysis
Additional file 8: PrediXcan and enloc results for predicted causal genes selected based on OMIM
Additional file 9: PrediXcan and enloc results for presumed causal genes in the rare variant based silver standard
Additional file 10: Review history

Peer review information
Tim Sands was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Acknowledgements
We thank the donors and their families for their generous gifts of organ donation for transplantation, and tissue donations for the GTEx research project; Mariya Khan and Christopher Stolte for the illustration in Fig. 1; and Laura Vairus for the illustrations in Figs. 2 and 3.
We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for these analyses. The investigators within IGAP contributed to the design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report. IGAP was made possible by the generous participation of the control subjects, the patients, and their families. http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php

Review history
The review history is available as Additional file 10.

Authors' contributions
ANB, RB, ERG, YL, YP designed the computational experiments, led and performed major components of the data wrangling, statistical analysis, visualization, and interpretation of the results; they wrote the manuscript. The GTEx GWAS Working Group discussed and interpreted the analysis results and proposed computational experiments. The GTEx Consortium collected the samples and provided pre-processed RNAseq and WGS data. SKH, GW, ZJ, DZ, FH, BL, AR, ARH, MDP, and FA contributed data analysis and figures. LB, DMJ, MV, RD provided processed data and figures. MS, KA, MM, SBM, AVS, CDB, TL, XW supervised portions of the analyses. HKI supervised the full project, designed the computational experiments, performed analyses, and wrote the manuscript. All authors read and approved the final manuscript.

Funding
The consortium was funded by GTEx program grants: HHSN268201000029C (F.A., K.G.A., A.V.S., X.Li., E.T., S.G., A.G., S.A., K.H.H., D.Y.N., K.H., S.R.M., J.L.N.), 5U41HG009494 (F.A., K.G.A.), 10XS170 (Subcontract to Leidos Biomedical) (W.F.L., J.A.T., G.K., A.M., S.S., R.H., G.Wa., M.J., M.Wa., L.E.B., C.J., J.W., B.R., M.Hu., K.M., L.A.S., H.M.G., M.Mo., L.K.B.), 10XS171 (Subcontract to Leidos Biomedical) (B.A.F., M.T.M., E.K., B.M.G., K.D.R., J.B.), 10ST1035 (Subcontract to Leidos Biomedical) (S.D.J., D.C.R., D.R.V.), R01DA006227-17 (D.C.M., D.A.D.), Supplement to University of Miami grant DA006227 (D.C.M., D.A.D.), HHSN261200800001E (A.M.S., D.E.T., N.V.R., J.A.M., L.S., M.E.B., L.Q., T.K., D.B., K.R., A.U.), R01MH101814 (M.M-A., V.W., S.B.M., R.G., E.T.D., D.G-M., A.V.), U01HG007593 (S.B.M.), R01MH101822 (C.D.B.), U01HG007598 (M.O., B.E.S.), R01MH107666 (H.K.I.), P30DK020595 (H.K.I.). E.R.G. is supported by the National Human Genome Research Institute (NHGRI) under Award Numbers 1R35HG010718 and 1R01HG011138 and by the National Heart, Lung, and Blood Institute (NHLBI) under Award Number 1R01HL133559. E.R.G. has also significantly benefitted from a Fellowship at Clare Hall, University of Cambridge (UK) and is grateful to the President and Fellows of the college for a stimulating intellectual home. S.K.-H. is supported by the Marie-Sklodowska Curie fellowship H2020 Grant 706636. R.D.: R35GM124836, R01HL139865, 15CVGPSD27130014. D.M.J.: T32HL00782. Y.Pa. is supported by the NHGRI award R01HG10067. A.R.H. was supported by the Massachusetts Lions Eye Research Fund Grant. Computation was performed at the high performance cluster of the Center for Research Informatics at the University of Chicago, funded by the Biological Sciences Division and CTSA UL1TR000430. Additional computation was performed with resources provided by the University of Chicago Research Computing Center.

Availability of data and materials

Genotype-Tissue Expression (GTEx) project's raw whole transcriptome and genome sequencing data are available via dbGaP accession number phs000424.v8.p2 [57]. All processed GTEx data are available via GTEx portal (http://gtexportal.org/). All the code used for the reproducible analysis is available, under MIT license, on Zenodo with the access code DOI https://doi.org/10.5281/zenodo.4321149 [58] and GitHub https://github.com/hakyimlab/gtex-gwas-analysis [59]. The softwares for imputed summary results, *enloc*, *coloc*, PrediXcan, MultiXcan, *dap-g*, prediction models are available at links there in. 1000 Genomes Project Reference for LDSC, https://data.broadinstitute.org/alkesgroup/LDSCORE/1000G_Phase3_plinkfiles.tgz; 1000 Genomes Project Reference with regression weights for LDSC, https://data.broadinstitute.org/alkesgroup/LDSCORE/1000G_Phase3_weights_hm3_no_MHC.tgz; BioVU, https://victr.vanderbilt.edu/pub/biovu/?sid=194; eCAVIAR, https://github.com/fhormoz/caviar; QTLEnrich, https://github.com/segrelabgenomics/eQTLEnrich; flashr, https://gaow.github.io/mnm-gtex-v8/analysis/mashr_flashr_workflow.html#flashr-prior-covariances; Gencode, https://www.gencodegenes.org/releases/26.html; GTEx GWAS subgroup repository, https://github.com/broadinstitute/gtex-v8; GTEx portal, http://gtexportal.org; Hail, https://github.com/hail-is/hail; HapMap Reference for LDSC, https://data.broadinstitute.org/alkesgroup/LDSCORE/w_hm3.snplist.bz2; LD score regression (LDSD regression), https://github.com/bulik/ldsc; MetaXcan, https://github.com/hakyimlab/MetaXcan; Mouse Phenotype Ontology, http://www.informatics.jax.org/vocab/mp_ontology; NHGRI-EBI GWAS catalog, https://www.ebi.ac.uk/gwas/; picard, http://picard.sourceforge.net/; PLINK 1.90, https://www.cog-genomics.org/plink2; PrediXcan, https://github.com/hakyim/PrediXcan; pyliftover, https://pypi.org/project/pyliftover/; Storeyś qvalue R package, https://github.com/StoreyLab/qvalue; Summary GWAS imputation, https://github.com/hakyimlab/summary-gwas-imputation; TORUS, https://github.com/xqwen/torus; UK Biobank GWAS, http://www.nealelab.is/uk-biobank/; UK Biobank, http://www.ukbiobank.ac.uk/

Ethics approval and consent to participate

Not applicable.

Competing interests

F.A. is an inventor on a patent application related to TensorQTL; S.E.C. is a co-founder, chief technology officer, and stock owner at Variant Bio; E.R.G. is on the Editorial Board of Circulation Research, and does consulting for the City of Hope/Beckman Research Institute; E.T.D. is chairman and member of the board of Hybridstat LTD.; B.E.E. is on the scientific advisory boards of Celsius Therapeutics and Freenome; G.G. receives research funds from IBM and Pharmacyclics, and is an inventor on patent applications related to MuTect, ABSOLUTE, MutSig, MSMuTect, MSMutSig, POLYSOLVER, and TensorQTL. G.G. is a founder, consultant and holds privately held equity in Scorpion Therapeutics; S.B.M. is on the scientific advisory board of MyOme; D.G.M. is a co-founder with equity in Goldfinch Bio, and has received research support from AbbVie, Astellas, Biogen, BioMarin, Eisai, Merck, Pfizer, and Sanofi-Genzyme; H.K.I. has received speaker honoraria from GSK and AbbVie; T.L. is a scientific advisory board member of Variant Bio with equity and Goldfinch Bio. P.F. is a member of the scientific advisory boards of Fabric Genomics, Inc., and Eagle Genomes, Ltd. P.G.F. is a partner of Bioinf2Bio.

Author details

[1]Section of Genetic Medicine, Department of Medicine, The University of Chicago, Chicago, IL, USA. [2]Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. [3]Data Science Institute, Vanderbilt University, Nashville, TN, USA. [4]Clare Hall, University of Cambridge, Cambridge, UK. [5]MRC Epidemiology Unit, University of Cambridge, Cambridge, UK. [6]Department of Genetics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA. [7]Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA. [8]Statistical Genetics, Max Planck Institute of Psychiatry, Munich, Germany. [9]New York Genome Center, New York, NY, USA. [10]Department of Systems Biology, Columbia University, New York, NY, USA. [11]Department of Human Genetics, University of Chicago, Chicago, IL, USA. [12]The Broad Institute of MIT and Harvard, Cambridge, MA, USA. [13]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [14]Department of Biology, Stanford University, Stanford, CA, 94305, USA. [15]Ocular Genomics Institute, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, USA. [16]Department of Biomedical Informatics, Department of Medicine, Vanderbilt University, Nashville, TN, USA. [17]Center for Human Genetics Research, Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, TN, USA. [18]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [19]Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [20]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [21]Université de Paris - EA 7537 BIOSTM, Paris, France. [22]University of Oxford, Oxford, UK. [23]Department of Genetics, Stanford University, Stanford, CA, USA. [24]Department of Pathology, Stanford University, Stanford, CA, USA. [25]Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA.

Received: 6 June 2020 Accepted: 18 December 2020
Published online: 26 January 2021

References

1. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. 2010;6(4):1000888. https://doi.org/10.1371/journal.pgen.1000888.
2. Guo H, Fortune MD, Burren OS, Schofield E, Todd JA, Wallace C. Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases. Hum Mol Genet. 2015;24(12):3305–13. https://doi.org/10.1093/hmg/ddv077. https://doi.org/http://oup.prod.sis.lan/hmg/article-pdf/24/12/3305/1720369/ddv077.pdf.
3. Wu L, Shi W, Long J, Guo X, Michailidou K, Beesley J, Bolla MK, Shu X-O, Lu Y, Cai Q, Al-Ejeh F, Rozali E, Wang Q, Dennis J, Li B, Zeng C, Feng H, Gusev A, Barfield RT, Andrulis IL, Anton-Culver H, Arndt V, Aronson KJ, Auer PL, Barrdahl M, Baynes C, Beckmann MW, Benitez J, Bermisheva M, Blomqvist C, Bogdanova NV, Bojesen SE, Brauch H,

Brenner H, Brinton L, Broberg P, Brucker SY, Burwinkel B, Caldes T, Canzian F, Carter BD, Castelao JE, Chang-Claude J, Chen X, Cheng T-YD, Christiansen H, Clarke CL, Collee M, Cornelissen S, Couch FJ, Cox D, Cox A, Cross SS, Cunningham JM, Czene K, Daly MB, Devilee P, Doheny KF, Dork T, Dos-Santos-Silva I, Dumont M, Dwek M, Eccles DM, Eilber U, Eliassen AH, Engel C, Eriksson M, Fachal L, Fasching PA, Figueroa J, Flesch-Janys D, Fletcher O, Flyger H, Fritschi L, Gabrielson M, Gago-Dominguez M, Gapstur SM, Garcia-Closas M, Gaudet MM, Ghoussaini M, Giles GG, Goldberg MS, Goldgar DE, Gonzalez-Neira A, Guenel P, Hahnen E, Haiman CA, Hakansson N, Hall P, Hallberg E, Hamann U, Harrington P, Hein A, Hicks B, Hillemanns P, Hollestelle A, Hoover RN, Hopper JL, Huang G, Humphreys K, Hunter DJ, Jakubowska A, Janni W, John EM, Johnson N, Jones K, Jones ME, Jung A, Kaaks R, Kerin MJ, Khusnutdinova E, Kosma V-M, Kristensen VN, Lambrechts D, Le Marchand L, Li J, Lindstrom S, Lissowska J, Lo W-Y, Loibl S, Lubinski J, Luccarini C, Lux MP, MacInnis RJ, Maishman T, Kostovska IM, Mannermaa A, Manson JE, Margolin S, Mavroudis D, Meijers-Heijboer H, Meindl A, Menon U, Meyer J, Mulligan AM, Neuhausen SL, Nevanlinna H, Neven P, Nielsen SF, Nordestgaard BG, Olopade OI, Olson JE, Olsson H, Peterlongo P, Peto J, Plaseska-Karanfilska D, Prentice R, Presneau N, Pylkas K, Rack B, Radice P, Rahman N, Rennert G, Rennert HS, Rhenius V, Romero A, Romm J, Rudolph A, Saloustros E, Sandler DP, Sawyer EJ, Schmidt MK, Schmutzler RK, Schneeweiss A, Scott RJ, Scott CG, Seal S, Shah M, Shrubsole MJ, Smeets A, Southey MC, Spinelli JJ, Stone J, Surowy H, Swerdlow AJ, Tamimi RM, Tapper W, Taylor JA, Terry MB, Tessier DC, Thomas A, Thone K, Tollenaar RAEM, Torres D, Truong T, Untch M, Vachon C, Van Den Berg D, Vincent D, Waisfisz Q, Weinberg CR, Wendt C, Whittemore AS, Wildiers H, Willett WC, Winqvist R, Wolk A, Xia L, Yang XR, Ziogas A, Ziv E, Dunning AM, Pharoah PDP, Simard J, Milne RL, Edwards SL, Kraft P, Easton DF, Chenevix-Trench G, Zheng W. A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. Nat Genet. 2018;50(7):968–78. https://doi.org/10.1038/s41588-018-0132-x.

4. Gong J, Mei S, Liu C, Xiang Y, Ye Y, Zhang Z, Feng J, Liu R, Diao L, Guo A-Y, Miao X, Han L. PancanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. Nucleic Acids Res. 2018;46(D1):971–6. https://doi.org/10.1093/nar/gkx861.
5. Pashos EE, Park Y, Wang X, Raghavan A, Yang W, Abbey D, Peters DT, Arbelaez J, Hernandez M, Kuperwasser N, Li W, Lian Z, Liu Y, Lv W, Lytle-Gabbin SL, Marchadier DH, Rogov P, Shi J, Slovik KJ, Stylianou IM, Wang L, Yan R, Zhang X, Kathiresan S, Duncan SA, Mikkelsen TS, Morrisey EE, Rader DJ, Brown CD, Musunuru K. Large, diverse population cohorts of hiPSCs and derived hepatocyte-like cells reveal functional genetic variation at blood lipid-associated loci. Cell Stem Cell. 2017;20(4):558–70. https://doi.org/10.1016/j.stem.2017.03.017.
6. Caliskan M, Manduchi E, Rao HS, Segert JA, Beltrame MH, Trizzino M, Park Y, Baker SW, Chesi A, Johnson ME, Hodge KM, Leonard ME, Loza B, Xin D, Berrido AM, Hand NJ, Bauer RC, Wells AD, Olthoff KM, Shaked A, Rader DJ, Grant SFA, Brown CD. Genetic and epigenetic fine mapping of complex trait associated loci in the human liver. Am J Hum Genet. 2019;105(1):89–107. https://doi.org/10.1016/j.ajhg.2019.05.010.
7. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, Compton CC, DeLuca DS, Peter-Demchok J, Gelfand ET, Guan P, Korzeniewski GE, Lockhart NC, Rabiner CA, Rao AK, Robinson KL, Roche NV, Sawyer SJ, Segrè AV, Shive CE, Smith AM, Sobin LH, Undale AH, Valentino KM, Vaught J, Young TR, Moore HM, et al. A novel approach to high-quality postmortem tissue procurement: the GTEx project. Biopreservation and Biobanking. 2015;13(5):311–9. https://doi.org/10.1089/bio.2015.0032.
8. GTEx Consortium, Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, Mohammadi P, Park Y, Parsana P, Segrè AV, Strober BJ, Zappala Z, Cummings BB, Gelfand ET, Hadley K, Huang KH, Lek M, Li X, Nedzel JL, Nguyen DY, Noble MS, Sullivan TJ, Tukiainen T, MacArthur DG, Getz G, Addington A, Guan P, Koester S, Little AR, Lockhart NC, Moore HM, Rao A, Struewing JP, Volpi S, Brigham LE, Hasz R, Hunter M, Johns C, Johnson M, Kopen G, Leinweber WF, Lonsdale JT, McDonald A, Mestichelli B, Myer K, Roe B, Salvatore M, Shad S, Thomas JA, Walters G, Washington M, Wheeler J, Bridge J, Foster BA, Gillard BM, Karasik E, Kumar R, Miklos M, Moser MT, Jewell SD, Montroy RG, Rohrer DC, Valley D, Mash DC, Davis DA, Sobin L, Barcus ME, Branton PA, Abell NS, Balliu B, Delaneau O, Frésard L, Gamazon ER, Garrido-Martín D, Gewirtz ADH, Gliner G, Gloudemans MJ, Han B, He AZ, Hormozdiari F, Li X, Liu B, Kang EY, McDowell IC, Ongen H, Palowitch JJ, Peterson CB, Quon G, Ripke S, Saha A, Shabalin AA, Shimko TC, Sul JH, Teran NA, Tsang EK, Zhang H, Zhou Y-H, Bustamante CD, Cox NJ, Guigó R, Kellis M, McCarthy MI, Conrad DF, Eskin E, Li G, Nobel AB, Sabatti C, Stranger BE, Wen X, Wright FA, Ardlie KG, Dermitzakis ET, Lappalainen T, Aguet F, Ardlie KG, Cummings BB, Gelfand ET, Getz G, Hadley K, Handsaker RE, Huang KH, Kashin S, Karczewski KJ, Lek M, Li X, MacArthur DG, Nedzel JL, Nguyen DT, Noble MS, Segrè AV, Trowbridge CA, Tukiainen T, Abell NS, Balliu B, Barshir R, Basha O, Battle A, Bogu GK, Brown A, Brown CD, Castel SE, Chen LS, Chiang C, Conrad DF, Cox NJ, Damani FN, Davis JR, Delaneau O, Dermitzakis ET, Engelhardt BE, Eskin E, Ferreira PG, Frésard L, Gamazon ER, Garrido-Martín D, Gewirtz ADH, Gliner G, Gloudemans MJ, Guigo R, Hall IM, Han B, He Y, Hormozdiari F, Howald C, Kyung Im H, Jo B, Yong Kang E, Kim Y, Kim-Hellmuth S, Lappalainen T, Li G, Li X, Liu B, Mangul S, McCarthy MI, McDowell IC, Mohammadi P, Monlong J, Montgomery SB, Muñoz-Aguirre M, Ndungu AW, Nicolae DL, Nobel AB, Oliva M, Ongen H, Palowitch JJ, Panousis N, Papasaikas P, Park Y, Parsana P, Payne AJ, Peterson CB, Quan J, Reverter F, Sabatti C, Saha A, Sammeth M, Scott AJ, Shabalin AA, Sodaei R, Stephens M, Stranger BE, Strober BJ, Sul JH, Tsang EK, Urbut S, van de Bunt M, Wang G, Wen X, Wright FA, Xi HS, Yeger-Lotem E, Zappala Z, Zaugg JB, Zhou Y-H, Akey JM, Bates D, Chan J, Chen LS, Claussnitzer M, Demanelis K, Diegel M, Doherty JA, Feinberg AP, Fernando MS, Halow J, Hansen KD, Haugen E, Hickey PF, Hou L, Jasmine F, Jian R, Jiang L, Johnson A, Kaul R, Kellis M, Kibriya MG, Lee K, Li B. Genetic effects on gene expression across human tissues. Nature. 2017;550:204.
9. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369(6509):1318–30.
10. Barbeira AN, Pividori M, Zheng J, Wheeler HE, Nicolae DL, Im HK. Integrating predicted transcriptome from multiple tissues improves association detection. PLOS Genet. 2019;15(1):1–20. https://doi.org/10.1371/journal.pgen.1007889.
11. Gamazon ER, Segrè AV, van de Bunt M, Wen X, Xi HS, Hormozdiari F, Ongen H, Konkashbaev A, Derks EM, Aguet F, Quan J, GTEx Consortium, Nicolae DL, Eskin E, Kellis M, Getz G, McCarthy MI, Dermitzakis ET, Cox NJ, Ardlie KG. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. Nat Genet. 2018;50(7):956–67. https://doi.org/10.1038/s41588-018-0154-4.

12. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, Montgomery GW, Goddard ME, Wray NR, Visscher PM, Yang J. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet. 2016;48(5):481–7. https://doi.org/10.1038/ng.3538.
13. Gusev A, Lee SH, Trynka G, Finucane H, Vilhjálmsson BJ, Xu H, Zang C, Ripke S, Bulik-Sullivan B, Stahl E, Kähler AK, Hultman CM, Purcell SM, McCarroll SA, Daly M, Pasaniuc B, Sullivan PF, Neale BM, Wray NR, Raychaudhuri S, Price AL. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. 2016;48:245–52. https://doi.org/10.1038/ng.3506.
14. Wen X. Molecular QTL discovery incorporating genomic annotations using Bayesian false discovery rate control. Ann Appl Stat. 2016;10(3):1619–38. https://doi.org/10.1214/16-AOAS952.
15. Hormozdiari F, van de Bunt M, Segre AV, Li X, Joo JWJ, Bilow M, Sul JH, Sankararaman S, Pasaniuc B, Eskin E. Colocalization of GWAS and eQTL signals detects target genes. Am J Hum Genet. 2016;99(6):1245–60. https://doi.org/10.1016/j.ajhg.2016.10.003.
16. Wen X, Pique-Regi R, Luca F. Integrating molecular QTL data into genome-wide genetic association analysis: probabilistic assessment of enrichment and colocalization. PLoS Genet. 2017;13(3):1006646. https://doi.org/10.1371/journal.pgen.1006646.
17. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10(5):1–15. https://doi.org/10.1371/journal.pgen.1004383.
18. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson N, Daly MJ, Price AL, Neale BM. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47:291.
19. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, Eyler AE, Denny JC, GTEx Consortium, Nicolae DL, Cox NJ, Im HK. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet. 2015;47:1091.
20. Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, Ermel R, Ruusalepp A, Quertermous T, Hao K, Bjorkegren JLM, Im HK, Pasaniuc B, Rivas MA, Kundaje A. Opportunities and challenges for transcriptome-wide association studies. Nat Genet. 2019;51(4):592–9. https://doi.org/10.1038/s41588-019-0385-z.
21. Takata A, Matsumoto N, Kato T. Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. Nat Commun. 2017;8(1):14519. https://doi.org/10.1038/ncomms14519.
22. Saferali A, Yun JH, Parker MM, Sakornsakolpat P, Chase RP, Lamb A, Hobbs BD, Boezen MH, Dai X, de Jong K, Beaty TH, Wei W, Zhou X, Silverman EK, Cho MH, Castaldi PJ, Hersh CP, Investigators C, the International COPD Genetics Consortium Investigators. Analysis of genetically driven alternative splicing identifies FBXO38 as a novel COPD susceptibility gene. PLoS Genet. 2019;15(7):1–19. https://doi.org/10.1371/journal.pgen.1008229.
23. Li YI, van de Geijn B, Raj A, Knowles DA, Petti AA, Golan D, Gilad Y, Pritchard JK. RNA splicing is a primary link between genetic variation and disease. Science. 2016;352(6285):600–4. https://doi.org/10.1126/science.aad9417.
24. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, Torstenson ES, Shah KP, Garcia T, Edwards TL, Stahl EA, Huckins LM, Aguet F, Ardlie KG, Cummings BB, Gelfand ET, Getz G, Hadley K, Handsaker RE, Huang KH, Kashin S, Karczewski KJ, Lek M, Li X, MacArthur DG, Nedzel JL, Nguyen DT, Noble MS, Segrè AV, Trowbridge CA, Tukiainen T, Abell NS, Balliu B, Barshir R, Basha O, Battle A, Bogu GK, Brown A, Brown CD, Castel SE, Chen LS, Chiang C, Conrad DF, Damani FN, Davis JR, Delaneau O, Dermitzakis ET, Engelhardt BE, Eskin E, Ferreira PG, Frésard L, Gamazon ER, Garrido-Martín D, Gewirtz ADH, Gliner G, Gloudemans MJ, Guigo R, Hall IM, Han B, He Y, Hormozdiari F, Howald C, Jo B, Kang EY, Kim Y, Kim-Hellmuth S, Lappalainen T, Li G, Li X, Liu B, Mangul S, McCarthy MI, McDowell IC, Mohammadi P, Monlong J, Montgomery SB, Muñoz-Aguirre M, Ndungu AW, Nobel AB, Oliva M, Ongen H, Palowitch JJ, Panousis N, Papasaikas P, Park YS, Parsana P, Payne AJ, Peterson CB, Quan J, Reverter F, Sabatti C, Saha A, Sammeth M, Scott AJ, Shabalin AA, Sodaei R, Stephens M, Stranger BE, Strober BJ, Sul JH, Tsang EK, Urbut S, Van De Bunt M, Wang G, Wen X, Wright FA, Xi HS, Yeger-Lotem E, Zappala Z, Zaugg JB, Zhou YH, Akey JM, Bates D, Chan J, Claussnitzer M, Demanelis K, Diegel M, Doherty JA, Feinberg AP, Fernando MS, Halow J, Hansen KD, Haugen E, Hickey PF, Hou L, Jasmine F, Jian R, Jiang L, Johnson A, Kaul R, Kellis M, Kibriya MG, Lee K, Li JB, Li Q, Lin J, Lin S, Linder S, Linke C, Liu Y, Maurano MT, Molinie B, Nelson J, Neri FJ, Park Y, Pierce BL, Rinaldi NJ, Rizzardi LF, Sandstrom R, Skol A, Smith KS, Snyder MP, Stamatoyannopoulos J, Tang H, Wang L, Wang M, Van Wittenberghe N, Wu F, Zhang R, Nierras CR, Branton PA, Carithers LJ, Guan P, Moore HM, Rao A, Vaught JB, Gould SE, Lockart NC, Martin C, Struewing JP, Volpi S, Addington AM, Koester SE, Little AR, Brigham LE, Hasz R, Hunter M, Johns C, Johnson M, Kopen G, Leinweber WF, Lonsdale JT, McDonald A, Mestichelli B, Myer K, Roe B, Salvatore M, Shad S, Thomas JA, Walters G, Washington M, Wheeler J, Bridge J, Foster BA, Gillard BM, Karasik E, Kumar R, Miklos M, Moser MT, Jewell SD, Montroy RG, Rohrer DC, Valley DR, Davis DA, Mash DC, Undale AH, Smith AM, Tabor DE, Roche NV, McLean JA, Vatanian N, Robinson KL, Sobin L, Barcus ME, Valentino KM, Qi L, Hunter S, Hariharan P, Singh S, Um KS, Matose T, Tomaszewski MM, Barker LK, Mosavel M, Siminoff LA, Traino HM, Flicek P, Juettemann T, Ruffier M, Sheppard D, Taylor K, Trevanion SJ, Zerbino DR, Craft B, Goldman M, Haeussler M, Kent WJ. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun. 2018. https://doi.org/10.1038/s41467-018-03621-1.
25. Hukku A, Pividori M, Luca F, Pique-Regi R, Im HK, Wen X. Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations. bioRxiv. 2020. https://doi.org/10.1101/2020.07.01.182097.
26. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan J, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, Day FR, Duan Y, Fall T, Fehrmann R, Ferreira T, Jackson AU, Karjalainen J, Lo KS, Locke AE, Mägi R, Mihailov E, Porcu E, Randall JC, Scherag A, Vinkhuyzen AAE, Westra H-J, Winkler TW, Workalemahu T, Zhao JH, Absher D, Albrecht E, Anderson D, Baron J, Beekman M, Demirkan A, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Fraser RM, Goel A, Gong J, Justice AE, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Lui JC, Mangino M, Leach IM, Medina-Gomez C, Nalls MA, Nyholt DR, Palmer CD, Pasko D, Pechlivanis S, Prokopenko I, Ried JS, Ripke S, Shungin D, Stančáková A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Afzal U, Ärnlöv J, Arscott

GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Blüher M, Bolton JL, Böttcher Y, Boyd HA, Bruinenberg M, Buckley BM, Buyske S, Caspersen IH, Chines PS, Clarke R, Claudi-Boehm S, Cooper M, Daw EW, De Jong PA, Deelen J, Delgado G, Denny JC, Dhonukshe-Rutten R, Dimitriou M, Doney ASF, Dörr M, Eklund N, Eury E, Folkersen L, Garcia ME, Geller F, Giedraitis V, Go AS, Grallert H, Grammer TB, Gräßler J, Grönberg H, de Groot LCPGM, Groves CJ, Haessler J, Hall P, Haller T, Hallmans G, Hannemann A, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hemani G, Henders AK, Hillege HL, Hlatky MA, Hoffmann W, Hoffmann P, Holmen O, Houwing-Duistermaat JJ, Illig T, Isaacs A, James AL, Jeff J, Johansen B, Johansson A, Jolley J, Juliusdottir T, Junttila J, Kho AN, Kinnunen L, Klopp N, Kocher T, Kratzer W, Lichtner P, Lind L, Lindstrom J, Lobbens S, Lorentzon M, Lu Y, Lyssenko V, Magnusson PKE, Mahajan A, Maillard M, McArdle WL, McKenzie CA, McLachlan S, McLaren PJ, Menni C, Merger S, Milani L, Moayyeri A, Monda KL, Morken MA, Müller G, Müller-Nurasyid M, Musk AW, Narisu N, Nauck M, Nolte IM, Nöthen MM, Oozageer L, Pilz S, Rayner NW, Renstrom F, Robertson NR, Rose LM, Roussel R, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Schunkert H, Scott RA, Sehmi J, Seufferlein T, Shi J, Silventoinen K, Smit JH, Smith AV, Smolonska J, Stanton AV, Stirrups K, Stott DJ, Stringham HM, Sundström J, Swertz MA, Syvänen A-C, Tayo BO, Thorleifsson G, Tyrer JP, van Dijk S, van Schoor NM, van der Velde N, van Heemst D, van Oort FVA, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Waldenberger M, Wennauer R, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q, Arveiler D, Bakker SJL, Beilby J, Bergman RN, Bergmann S, Biffar R, Blangero J, Boomsma DI, Bornstein SR, Bovet PA. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet. 2014. https://doi.org/10.1038/ng.3097.

27. Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine mapping. J R Stat Soc Ser B (Stat Methodol). 2020;82(5):1273–300.
28. Barbeira AN, Melia OJ, Liang Y, Bonazzola R, Wang G, Wheeler HE, et al. Fine-mapping and QTL tissue-sharing information improves the reliability of causal gene identification. Genet Epidemiol. 2020;44(8):854–67.
29. Hu Y, Li M, Lu Q, Weng H, Wang J, Zekavat SM, Yu Z, Li B, Gu J, Muchnik S, et al. A statistical framework for cross-tissue transcriptome-wide association analysis. Nat Genet. 2019;51(3):568–76.
30. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balser JR, Masys DR. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin Pharmacol Ther. 2008. https://doi.org/10.1038/clpt.2008.89.
31. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, Field JR, Pulley JM, Ramirez AH, Bowton E, Basford MA, Carrell DS, Peissig PL, Kho AN, Pacheco JA, Rasmussen LV, Crosslin DR, Crane PK, Pathak J, Bielinski SJ, Pendergrass SA, Xu H, Hindorff LA, Li R, Manolio TA, Chute CG, Chisholm RL, Larson EB, Jarvik GP, Brilliant MH, Mccarty CA, Kullo IJ, Haines JL, Crawford DC, Masys DR, Roden DM. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol. 2013;31(12):1102. https://doi.org/10.1038/nbt.2749.
32. Bastarache L, Carroll RJ, Ritchie MD, Zink R, Field JR, Mosley JD, Pulley JM, Ramirez AH, Bowton E, Basford MA, Carrell DS, Peissig PL, Kho AN, Pacheco JA, Rasmussen LV, Crosslin DR, Crane PK, Pathak J, Bielinski SJ, Pendergrass SA, Xu H, Li R, Hindorff LA, Manolio TA, Chute CG, Larson EB, Chisholm RL, Brilliant MH, Jarvik GP, McCarty CA, Kullo IJ, Crawford DC, Haines JL, Masys DR, Roden DM, Denny JC. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol. 2013;31(12):1102–10. https://doi.org/10.1038/nbt.2749.
33. Pividori M, Rajagopal PS, Barbeira A, Liang Y, Melia O, Bastarache L, Park Y, GTEx Consortium, Wen X, Im HK. PhenomeXcan: mapping the genome to the phenome through the transcriptome. Sci Adv. 2020;6(37):2083. https://doi.org/10.1126/sciadv.aba2083.
34. Berisa T, Pickrell JK. Approximately independent linkage disequilibrium blocks in human populations. Bioinformatics. 2016;32(2):283–5. https://doi.org/10.1093/bioinformatics/btv546.
35. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. Nature. 2015;526(7571):68–74. https://doi.org/10.1038/nature15393.
36. Yao DW, O'Connor LJ, Price AL, Gusev A. Quantifying genetic effects on disease mediated by assayed gene expression levels. Nat Genet. 20201–8.
37. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. Nat Publ Group. 2013;12(8):581–94. https://doi.org/10.1038/nrd4051.
38. Kidambi S, Patel SB. Cholesterol and non-cholesterol sterol transporters: ABCG5, ABCG8 and NPC1L1: a review. Xenobiotica. 2008;38(7-8):1119–39. https://doi.org/10.1080/00498250802007930.
39. Yu L, Hammer RE, Li-Hawkins J, von Bergmann K, Lutjohann D, Cohen JC, Hobbs HH. Disruption of Abcg5 and Abcg8 in mice reveals their crucial role in biliary cholesterol secretion. Proc Natl Acad Sci. 2002;99(25):16237–42. https://doi.org/10.1073/pnas.252582399.
40. Wilund KR, Yu L, Xu F, Hobbs HH, Cohen JC. High-level expression of ABCG5 and ABCG8 attenuates diet-induced hypercholesterolemia and atherosclerosis in Ldlr-/- mice. J Lipid Res. 2004;45(8):1429–36. https://doi.org/10.1194/jlr.M400167-JLR200.
41. Peloso GMP, Nomura A, Khera AV, Chaffin M, Won H-H, Ardissino D, Danesh J, Schunkert H, Wilson JG, Samani N, Erdmann J, McPherson R, Watkins H, Saleheen D, McCarthy S, Teslovich TM, Leader JB, Kirchner HL, Marrugat J, Nohara A, Kawashiri M, Tada H, Dewey FE, Carey A., Baras DJ, Kathiresan S. Rare protein-truncating variants in APOB, lower low-density lipoprotein cholesterol, and protection against coronary heart disease. Circ Genom Precis Med. 2019. https://doi.org/10.1161/CIRCGEN.118.002376.
42. Walldius G, Jungner I. Apolipoprotein B and apolipoprotein A-I: risk indicators of coronary heart disease and targets for lipid-modifying therapy. J Intern Med. 2004;255(2):188–205. https://doi.org/10.1046/j.1365-2796.2003.01276.x.
43. Contois JH, McConnell JP, Sethi AA, Csako G, Devaraj S, Hoefner DM, Warnick GR. Apolipoprotein B and cardiovascular disease risk: position statement from the AACC Lipoproteins and Vascular Diseases Division Working Group on Best Practices. Clin Chem. 2009;55(3):407–19. https://doi.org/10.1373/clinchem.2008.118356.
44. Leslie M. To help save the heart, is it time to retire cholesterol tests? Science (New York, N.Y.) 2017;358(6368):1237–8. https://doi.org/10.1126/science.358.6368.1237.

45. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 2005. https://doi.org/10.1093/nar/gki033.
46. Marouli E, Graff M, Medina-Gomez C, Lo KS, Wood AR, Kjaer TR, Fine RS, Lu Y, Schurmann C, Highland HM, et al. Rare and low-frequency coding variants alter human adult height. Nature. 2017;542(7640):186.
47. Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, Saleheen D, Emdin C, Alam D, Alves AC, et al. Exome-wide association study of plasma lipids in> 300,000 individuals. Nat Genet. 2017;49(12):1758.
48. Locke AE, Steinberg KM, Chiang CW, Service SK, Havulinna AS, Stell L, et al. Exome sequencing of Finnish isolates enhances rare-variant association power. Nature. 2019;572(7769):323–328.
49. Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA. Clan genomics and the complex architecture of human disease. Cell. 2011;147(1):32–43. https://doi.org/10.1016/j.cell.2011.09.008.
50. Blair DR, Lyttle CS, Mortensen JM, Bearden CF, Jensen AB, Khiabanian H, Melamed R, Rabadan R, Bernstam EV, Brunak S, Jensen LJ, Nicolae D, Shah NH, Grossman RL, Cox NJ, White KP, Rzhetsky A. A nondegenerate code of deleterious variants in Mendelian loci contributes to complex disease risk. Cell. 2013;155(1):70–80. https://doi.org/10.1016/j.cell.2013.08.030.
51. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, Ma C, Fontanillas P, Moutsianas L, McCarthy DJ, Rivas MA, Perry JRB, Sim X, Blackwell TW, Robertson NR, Rayner NW, Cingolani P, Locke AE, Tajes JF, Highland HM, Dupuis J, Chines PS, Lindgren CM, Hartl C, Jackson AU, Chen H, Huyghe JR, van de Bunt M, Pearson RD, Kumar A, Muller-Nurasyid M, Grarup N, Stringham HM, Gamazon ER, Lee J, Chen Y, Scott RA, Below JE, Chen P, Huang J, Go MJ, Stitzel ML, Pasko D, Parker SCJ, Varga TV, Green T, Beer NL, Day-Williams AG, Ferreira T, Fingerlin T, Horikoshi M, Hu C, Huh I, Ikram MK, Kim B-J, Kim Y, Kim YJ, Kwon M-S, Lee J, Lee S, Lin K-H, Maxwell TJ, Nagai Y, Wang X, Welch RP, Yoon J, Zhang W, Barzilai N, Voight BF, Han B-G, Jenkinson CP, Kuulasmaa T, Kuusisto J, Manning A, Ng MCY, Palmer ND, Balkau B, Stancakova A, Abboud HE, Boeing H, Giedraitis V, Prabhakaran D, Gottesman O, Scott J, Carey J, Kwan P, Grant G, Smith JD, Neale BM, Purcell S, Butterworth AS, Howson JMM, Lee HM, Lu Y, Kwak S-H, Zhao W, Danesh J, Lam VKL, Park KS, Saleheen D, So WY, Tam CHT, Afzal U, Aguilar D, Arya R, Aung T, Chan E, Navarro C, Cheng C-Y, Palli D, Correa A, Curran JE, Rybin D, Farook VS, Fowler SP, Freedman BI, Griswold M, Hale DE, Hicks PJ, Khor C-C, Kumar S, Lehne B, Thuillier D, Lim WY, Liu J, van der Schouw YT, Loh M, Musani SK, Puppala S, Scott WR, Yengo L, Tan S-T, Taylor HAJ, Thameem F, Wilson GS, Wong TY, Njolstad PR, Levy JC, Mangino M, Bonnycastle LL, Schwarzmayr T, Fadista J, Surdulescu GL, Herder C, Groves CJ, Wieland T, Bork-Jensen J, Brandslund I, Christensen C, Koistinen HA, Doney ASF, Kinnunen L, Esko T, Farmer AJ, Hakaste L, Hodgkiss D, Kravic J, Lyssenko V, Hollensted M, Jorgensen ME, Jorgensen T, Ladenvall C, Justesen JM, Karajamaki A, Kriebel J, Rathmann W, Lannfelt L, Lauritzen T, Narisu N, Linneberg A, Melander O, Milani L, Neville M, Orho-Melander M, Qi L, Qi Q, Roden M, Rolandsson O, Swift A, Rosengren AH, Stirrups K, Wood AR, Mihailov E, Blancher C, Carneiro MO, Maguire J, Poplin R, Shakir K, Fennell T, DePristo M, de Angelis MH, Deloukas P, Gjesing AP, Jun G, Nilsson P, Murphy J, Onofrio R, Thorand B, Hansen T, Meisinger C, Hu FB, Isomaa B, Karpe F, Liang L, Peters A, Huth C, O'Rahilly SP, Palmer CNA, Pedersen O, Rauramaa R, Tuomilehto J, Salomaa V, Watanabe RM, Syvanen A-C, Bergman RN, Bharadwaj D, Bottinger EP, Cho YS, Chandak GR, Chan JCN, Chia KS, Daly MJ, Ebrahim SB, Langenberg C, Elliott P, Jablonski KA, Lehman DM, Jia W, Ma RCW, Pollin TI, Sandhu M, Tandon N, Froguel P, Barroso I. The genetic architecture of type 2 diabetes. Nature. 2016;536(7614):41–7. https://doi.org/10.1038/nature18642.
52. Keinan A, Clark AG. Recent Explosive Human Population Growth Has Resulted in an Excess of Rare Genetic Variants. Science. 2012;336(6082):740–3.
53. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434–443.
54. Mohammadi P, Castel SE, Cummings BB, Einson J, Sousa C, Hoffman P, Donkervoort S, Jiang Z, Mohassel P, Foley AR, Wheeler HE, Im HK, Bonnemann CG, MacArthur DG, Lappalainen T. Genetic regulatory variation in populations informs transcriptome analysis in rare disease. Science. 2019;366(6463):351–6. https://doi.org/10.1126/science.aay0256.
55. Ongen H, Brown AA, Delaneau O, Panousis NI, Nica AC, Dermitzakis ET, Consortium G, et al. Estimating the causal tissues for complex traits and diseases. Nat Genet. 2017;49(12):1676.
56. Benner C, Havulinna AS, Jarvelin M-R, Salomaa V, Ripatti S, Pirinen M. Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. Am J Hum Genet. 2017;101(4):539–51. https://doi.org/10.1016/j.ajhg.2017.08.012.
57. GTEx Consortium. Genotype-Tissue Expression Project (GTEx). dbGaP. 2020. Available from: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000424.v8.p2.
58. Barbeira AN, Bonazzola R, Gamazon ER, Liang Y, Park Y, Kim-Hellmuth S, et al. hakyimlab/gtex-gwas-analysis: zenodo-release.v1.0. Zenodo. 2020. Available from: https://doi.org/10.5281/zenodo.4321149.
59. Barbeira AN, Bonazzola R, Gamazon ER, Liang Y, Park Y, Kim-Hellmuth S, et al. GTEx v8 GWAS analysis. GitHub. 2020. Available from: https://github.com/hakyimlab/gtex-gwas-analysis.

Publisher's Note