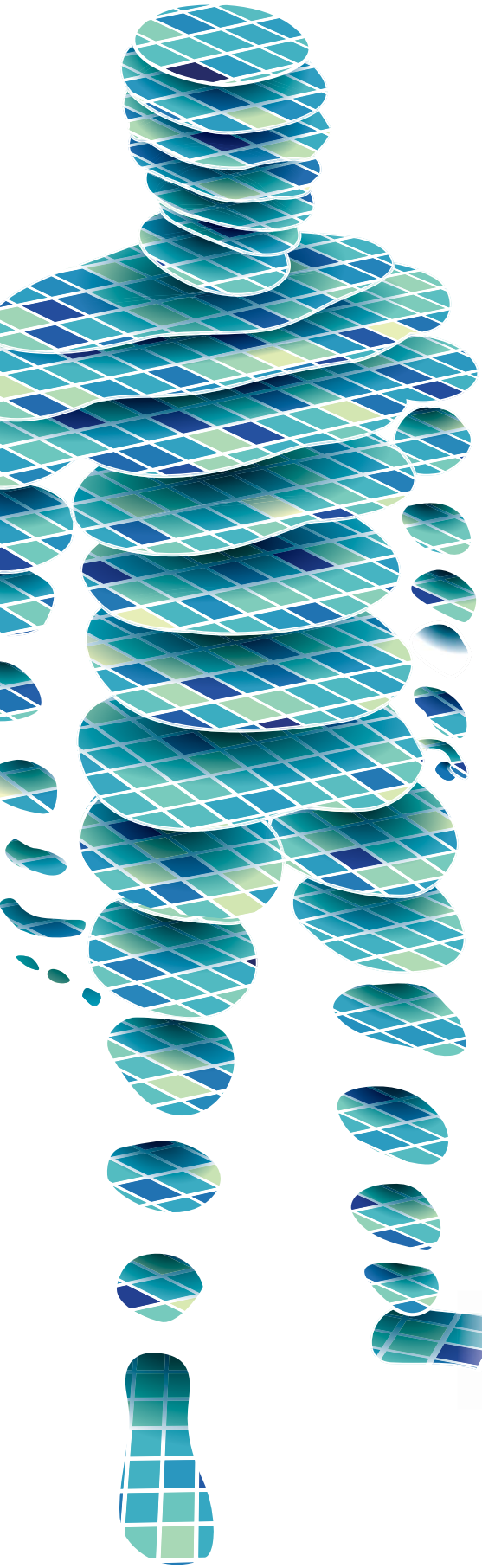

RESEARCH ARTICLE

HUMAN GENOMICS

The GTEx Consortium atlas of genetic regulatory effects across human tissues

The GTEx Consortium*

The Genotype-Tissue Expression (GTEx) project was established to characterize genetic effects on the transcriptome across human tissues and to link these regulatory mechanisms to trait and disease associations. Here, we present analyses of the version 8 data, examining 15,201 RNA-sequencing samples from 49 tissues of 838 postmortem donors. We comprehensively characterize genetic associations for gene expression and splicing in cis and trans, showing that regulatory associations are found for almost all genes, and describe the underlying molecular mechanisms and their contribution to allelic heterogeneity and pleiotropy of complex traits. Leveraging the large diversity of tissues, we provide insights into the tissue specificity of genetic effects and show that cell type composition is a key factor in understanding gene regulatory mechanisms in human tissues.

The characterization and interpretation of the function of the millions of genetic variants across the human genome remain a pressing need in human genetics. Understanding the effects of genetic variation is essential for identifying the molecular mechanisms of genetic risk for complex traits and diseases, which are mainly driven by noncoding loci with largely uncharacterized regulatory functions. To address this challenge, several projects have built comprehensive annotations of genome function across tissues and cell types (*1*, *2*) and mapped the effects of regulatory variation across large numbers of individuals, primarily from whole blood and blood cell types (*3*–*5*). The Genotype-Tissue Expression (GTEx) project provides an essential intersection where variant function can be studied across a wide range of both tissues and individuals.

The GTEx project was launched in 2010 with the aim of building a catalog of genetic effects on gene expression across a large number of human tissues to elucidate the molecular mechanisms of genetic associations with complex diseases and traits and to improve our understanding of regulatory genetic variation (*6*). The project set out to collect biospecimens from ~50 tissues from up to ~1000 postmortem donors and to create standards and protocols for optimizing postmortem tissue collection and donor recruitment (*7*, *8*), biospecimen processing (*7*), and data sharing (www.gtexportal.org).

Following the GTEx pilot (*9*) and midstage results (*10*), we present a final analysis of the version 8 (v8) data release from the GTEx Consortium. We provide a catalog of genetic regulatory variants affecting gene expression and splicing in cis and trans across 49 tissues and describe patterns and mechanisms of tissue and cell type specificity of genetic regulatory effects. Through integration of GTEx data with genome-wide association studies (GWASs), we characterize mechanisms of how genetic effects on the transcriptome mediate complex trait associations.

Quantitative trait locus (QTL) discovery

The GTEx v8 dataset, after quality control (*11*), consists of 838 donors and 17,382 samples from 52 tissues and two cell lines. In the analysis of this study, we used 49 tissues or cell lines that had at least 70 individuals for which both RNA sequencing (RNA-seq) and genotype data from whole-genome sequencing (WGS) were available, for a total of 15,201 samples from 838 donors (Fig. 1A and figs. S1 and S2). Of the 838 donors, 715 (85.3%) were European American, 103 (12.3%) African American, and 12 (1.4%) Asian American, with 16 (1.9%) reporting Hispanic or Latino ethnicity; 557 (66.4%) donors were male and 281 (33.5%) female (fig. S1). WGS was performed for each donor to a median depth of 32×, resulting in the detection of 43,066,422 single-nucleotide variants after quality control and phasing [10,008,325 with minor allele frequency (MAF) ≥ 0.01] and 3,459,870 small indels (762,535 with MAF ≥ 0.01) (fig. S3 and table S1) (*11*). The mRNA of each of the tissue samples was sequenced to a median depth of 82.6 million reads, and alignment, quantification, and quality control were performed as described in (*11*) (figs. S4 to S6).

The resulting data provide a broad survey of individual- and tissue-specific gene expression, enabling a comprehensive view of the

*A full list of the GTEx authors and their affiliations is available at the end of this article.
Corresponding authors: François Aguet (francois@broadinstitute.org); Kristin G. Ardlie (kardlie@broadinstitute.org); Tuuli Lappalainen (tlappalainen@nygenome.org)

V. ALTOUNIAN/SCIENCE, IN COLLABORATION WITH CHRISTIAN STOLTE (DATA) GTEX CONSORTIUM

Downloaded from http://science.sciencemag.org/ on October 31, 2020

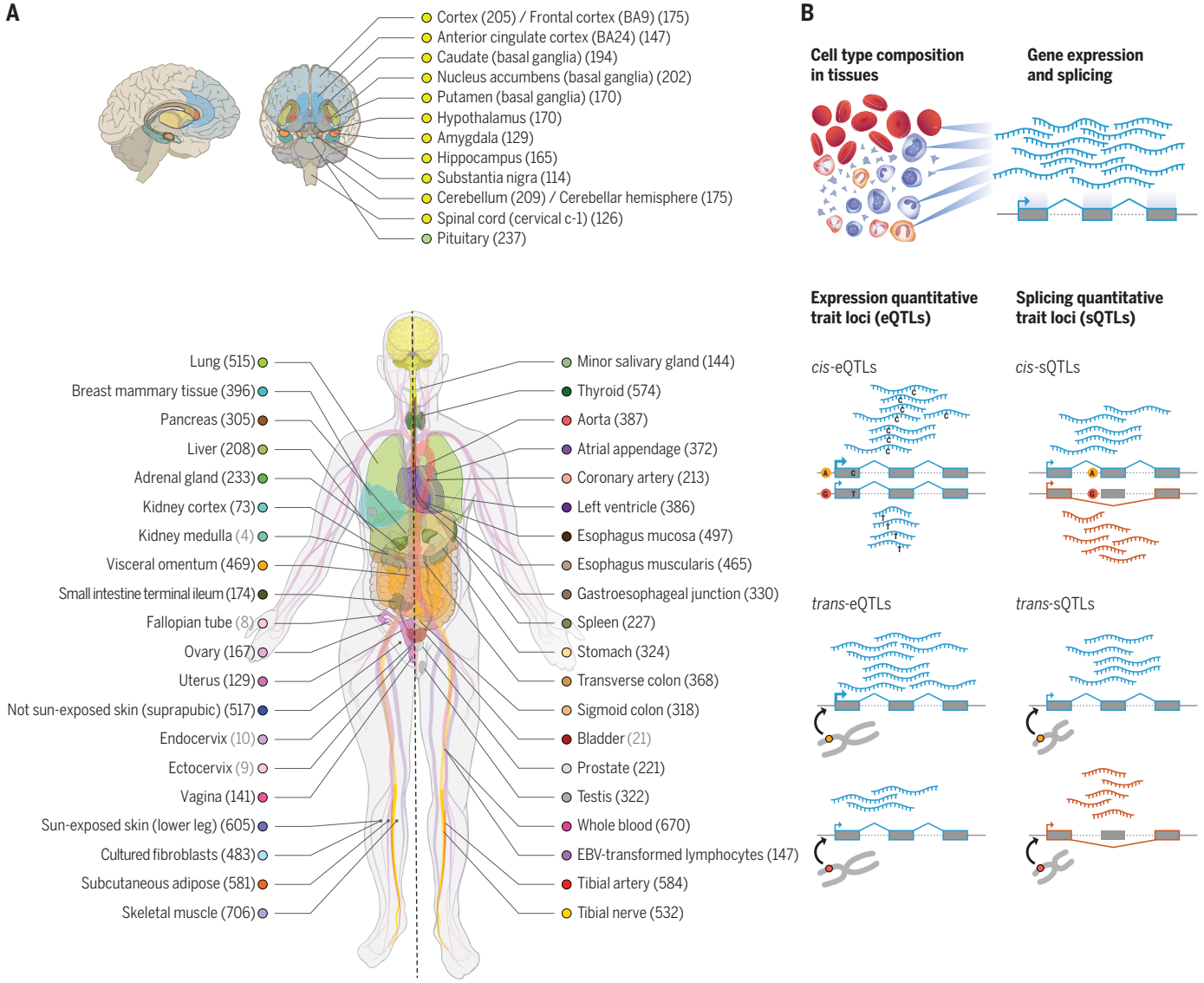

Fig. 1. Sample and data types in the GTEx v8 study. (**A**) Illustration of the 54 tissue types examined (including 11 distinct brain regions and two cell lines), with sample numbers from genotyped donors in parentheses and color coding indicated in the adjacent circles. Tissues with 70 or more samples were included in QTL analyses. (**B**) Illustration of the core data types used throughout the study. Gene expression and splicing were quantified from bulk RNA-seq of heterogeneous tissue samples, and local and distal genetic effects (cis-QTLs and trans-QTLs, respectively) were quantified across individuals for each tissue.

impact of genetic variation on gene regulation (Fig. 1B). We mapped genetic loci that affect the expression (eQTL) or splicing (sQTL) of protein-coding and long intergenic noncoding RNA (lincRNA) genes, both in cis and trans. Genes with an eQTL or sQTL are called eGenes and sGenes, respectively, and the corresponding significant variants are called eVariants and sVariants, respectively.

Across all tissues, we discovered cis-eQTLs [5% false discovery rate (FDR) per tissue (*11*), with 1% FDR results shown in fig. S7] for 18,262 protein-coding and 5006 lincRNA genes [23,268 genes with a cis-eQTL (i.e., cis-eGenes) corresponding to 94.7% of all protein-coding and 67.3% of all lincRNA genes detected in at least one tissue], with a total of 4,278,636 genetic variants (43% of all variants with MAF ≥ 0.01) that were significant in at least one tissue (cis-eVariants) (Fig. 2A, figs. S7 and S8, and table S2). The discovered eQTLs had a high replication rate in external datasets (figs. S12 and S13). Cis-eQTLs for all long noncoding RNAs (lncRNAs), which include lincRNAs and other types, are characterized in (*12*). The genes lacking a cis-eQTL were enriched for those lacking expression in the tissues analyzed by GTEx, including genes involved in early development (fig. S9). While most of the discovered cis-eQTLs had small effect sizes measured as allelic fold change (aFC), across tissues an average of 22% of cis-eQTLs had a greater than twofold effect on gene expression (fig. S14). We mapped sQTLs in cis with intron excision ratios from LeafCutter (*11*, *13*) and discovered 12,828 (66.5%) protein-coding and 1600 (21.5%) lincRNA genes (14,424 total) with a cis-sQTL (5% FDR per tissue) in at least one tissue (cis-sVariants) (Fig. 2A and table S2; with 1% FDR results shown in fig. S7). As expected (*10*), cis-QTL discovery was highly correlated with the sample size for each tissue [Spearman's rank correlation coefficient (ρ) = 0.95 for cis-eQTLs and 0.92 for cis-sQTLs]. The increased cis-eQTL discovery in larger tissues is primarily driven by additional power to discover small effects, with discovery of cis-eGenes with a greater than twofold effect saturating

Downloaded from http://science.sciencemag.org/ on October 31, 2020

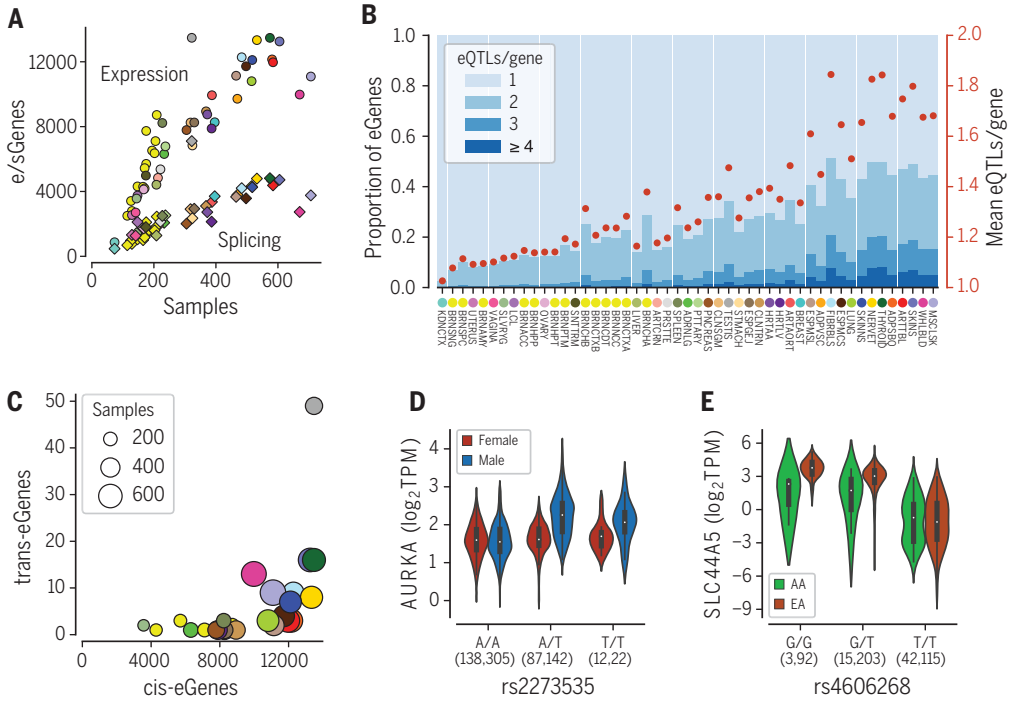

Fig. 2. QTL discovery. (**A**) The number of genes with a cis-eQTL (eGenes) or cis-sQTL (sGenes) per tissue, as a function of sample size. See Fig. 1A for the legend of tissue colors. (**B**) Allelic heterogeneity of cis-eQTLs depicted as proportion of eGenes with one or more independent cis-eQTLs (blue stacked bars; left *y* axis) and as a mean number of cis-eQTLs per gene (red dots; right *y* axis). The tissues are ordered by sample size. (**C**) The number of genes with a trans-eQTL as a function of the number of cis-eGenes. (**D**) Sex-biased cis-eQTL for *AURKA* in skeletal muscle, where rs2273535-T is associated with increased *AURKA* expression in males ($P = 9.02 \times 10^{-27}$) but not in females ($P = 0.75$). (**E**) Population-biased cis-eQTL for *SLC44A5* in esophagus mucosa [aFC = −2.85 and −4.82 and in African Americans (AA) and European Americans (EA), respectively; permutation *P* value = 1.2×10^{-3}]. TPM, transcripts per million.

at ~1500 genes in tissues with >200 samples (fig. S14).

Previous studies have shown widespread allelic heterogeneity of gene expression in cis, that is, multiple independent causal eQTLs per gene (*4*, *14*, *15*). We mapped independent cis-eQTLs and cis-sQTLs using stepwise regression, where the 5% FDR threshold for significance was defined by the single cis-QTL mapping (*10*). We observed widespread allelic heterogeneity, with up to 50% of eGenes having more than one independent cis-eQTL in the tissues with the largest sample sizes (Fig. 2B and fig. S10). Our analysis captured a lower rate of allelic heterogeneity for cis-sQTLs, which could be a result of both underlying biology and lower power in cis-sQTL mapping (fig. S10). These results highlight gains in cis-eQTL mapping with increasing sample sizes, even when the discovery of new eGenes in specific tissues starts to saturate.

Interchromosomal trans-eQTL mapping yielded 143 trans-eGenes (121 protein-coding and 22 lincRNA at 5% FDR assessed at the gene level, separately for each gene type), after controlling for false positives due to read misalignment (*11*, *16*) (table S13). The number of trans-eGenes discovered per tissue is correlated with sample size (Spearman's ρ = 0.68) and to the number of cis-eQTLs (Spearman's ρ = 0.77), with outlier tissues such as testis contributing disproportionately to both cis and trans (Fig. 2C). We identified a total of 49 trans-eGenes in testis, 47 of which were found in no other tissue even at FDR 50%. Greater than twofold effect sizes on trans-eGene expression were observed for 19% of trans-eQTLs (fig. S14). Trans-sQTL mapping yielded 29 trans-sGenes (5% FDR per tissue), including a replication of a previously described trans-sQTL (*3*) and visual support of the association pattern in several loci (*11*) (fig. S11 and table S14). These results suggest that while trans-sQTL mapping is challenging, we can discover robust genetic effects on splicing in trans.

We produced allelic expression (AE) data using two complementary approaches (*11*). In addition to the conventional AE data for each heterozygous genotype, we produced AE data by haplotype, integrating data from multiple heterozygous sites in the same gene, yielding 153 million gene-level measurements (≥8 reads) across all samples (*17*). Allelic expression reflects differential regulation of the two haplotypes in individuals that are heterozygous for a regulatory variant in cis; indeed, cis-eQTL effect size is strongly correlated with allelic expression (median Spearman's ρ = 0.82) (*10*). We hypothesized that cis-sQTLs could also partially contribute to allelic imbalance, even if only for parts of transcripts. However, there is drastically less signal of increased allelic imbalance among individuals heterozygous for cis-sQTLs (median Spearman's ρ = −0.05) (fig. S15), which indicates that AE data primarily capture cis-eQTL effects and that genetic splicing variation in cis is not strongly reflected in gene-level AE data.

Genetic regulatory effects across populations and sexes

Variability in human traits and diseases between sexes and population groups likely partially results from differences in genetic effects (*18*–*20*). To study whether genetic regulatory variants manifest such variability, we analyzed variable cis-eQTL effects between males and females, as well as between individuals of European ancestry and those of African ancestry. Because external replication datasets are sparse, we developed an AE approach for validation with an orthogonal data type from the same samples (*17*): Allelic imbalance in individuals heterozygous for the cis-eQTL allows individual-level quantification of the cis-eQTL effect size (*21*) and can be correlated with the interaction terms used in cis-eQTL analysis

Downloaded from http://science.sciencemag.org/ on October 31, 2020

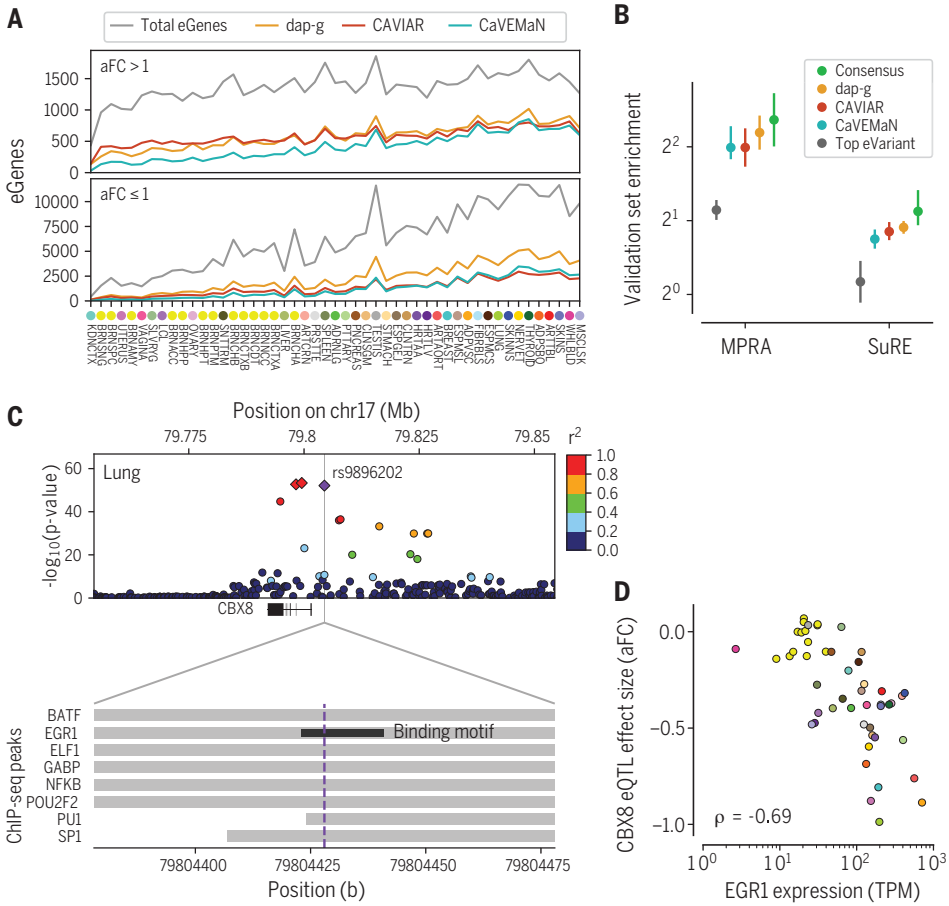

Fig. 3. Fine-mapping of cis-eQTLs. (**A**) Number of eGenes per tissue with variants fine-mapped with >0.5 posterior probability of causality, using three methods. The overall number of eGenes with at least one fine-mapped eVariant increases with sample size for all methods. However, this increase is in part driven by better statistical power to detect small effect size cis-eQTLs (aFC ≤ 1 in $\log_2$ scale; see also fig. S14) with larger sample sizes, and the proportion of well fine-mapped eGenes with small effect sizes increases more modestly with sample size (bottom versus top panels), indicating that such cis-eQTLs are generally more difficult to fine-map. (**B**) Enrichment of variants among experimentally validated regulatory variants, shown for the cis-eVariant with the best *P* value (top eVariant), and those with posterior probability of causality >0.8 according to each of the three methods individually or all of them (consensus). Error bars: 95% confidence interval (CI). (**C**) The cis-eQTL signal for *CBX8* is fine-mapped to a credible set of three variants (red and purple diamonds), of which rs9896202 (purple diamond) overlaps a large number of transcription factor binding sites in ENCODE chromatin immunoprecipitation sequencing (ChIP-seq) data and disrupts the binding motif of *EGR1*. (**D**) The potential role of EGR1 binding driving this cis-eQTL is further supported by correlation between *EGR1* expression and the *CBX8* cis-eQTL effect size across tissues.

to validate modifier effects of the cis-eQTL association (fig. S16).

To characterize sex-differentiated genetic effects on gene expression in GTEx tissues, we mapped sex-biased cis-eQTLs (sb-eQTLs). Analyzing the set of all conditionally independent cis-eQTLs, we identified eQTLs with significantly different effects between sexes by fitting a linear regression model and testing for a significant genotype-by-sex (G×S) interaction (*11*). Across the 44 GTEx tissues shared between sexes, we identified 369 sb-eQTLs (FDR ≤ 25%), characterized further in (*22*). Sex-biased eQTL discovery had a modest correlation with tissue sample size (Spearman's ρ = 0.39, *P* = 0.03), with most sb-eQTLs discovered in breast but others also discovered in muscle, skin, and adipose tissues.

In some cases, the cis-eQTL signal—identified with males and females combined—seems to be driven exclusively by one sex. For example, the cis-eQTL association of rs2273535 with the gene *AURKA* in skeletal muscle (cis-eQTL $P = 6.92 \times 10^{24}$) is correlated with sex ($P_{G\times S} = 9.28 \times 10^{-12}$, Storey $q_{G\times S} = 1.07 \times 10^{-7}$, AE validation $P = 1.15 \times 10^{-11}$) and present only in males (Fig. 2D and fig. S17). *AURKA* is a member of the serine and threonine kinase family involved in mitotic chromosomal segregation that has been widely studied as a risk factor in several cancers (*23*–*26*) and has recently been shown to be involved in muscle differentiation (*27*).

We also characterized population-biased cis-eQTLs (pb-eQTLs), where a variant's molecular effect on gene expression differs between individuals of European and African ancestry, controlling for differences in allele frequency, linkage disequilibrium (LD), and covariates (*11*). Analyzing 31 tissues with sample sizes >20 in both populations, we mapped genes with a different eQTL effect size measured by aFC. After applying stringent filters to remove differences potentially explained by LD or other artifacts (fig. S18A), we identified 178 pb-eQTLs for 141 eGenes (FDR ≤ 25%) that show a moderate degree of validation in allele-specific expression data (fig. S18, C and D, and table S10).

While some of the pb-eQTL effects are tissue specific, there are also effects that are shared across most tissues (fig. S18E). Figure 2E shows an example of a pb-eQTL for the *SLC44A5* gene involved in transport of sugars and amino acids, which is expressed at different levels in the epidermis of lighter skin and darker skin (reconstructed in vitro) (*28*, *29*). In Europeans, the derived allele of rs4606268 decreases expression of the gene in esophagus mucosa (aFC = −4.82), but this effect is significantly lower in African Americans (aFC = −2.85, permutation *P* value = 1.2×10^{-3}, AE validation *P* = 0.002) (fig. S18C).

Altogether, despite the relaxed FDR, we discovered only a few hundred sex- or population-biased cis-eQTLs out of tens of thousands of cis-eQTLs in GTEx, which indicates that there are few regulatory variants with major modifier effects and that these associations continue to be challenging to identify without a much larger sample size. However, the discovered effects can provide insights into sex- or population-specific regulatory effects on gene expression. Importantly, factors correlated with sex or population—for example, cell type composition or environmental exposures—may contribute to sex- or population-biased cis-eQTLs. These effects are described in detail in (*22*).

Fine-mapping

A major challenge of all genetic association studies is to distinguish the causal variants from their LD proxies. We applied three different

Downloaded from http://science.sciencemag.org/ on October 31, 2020

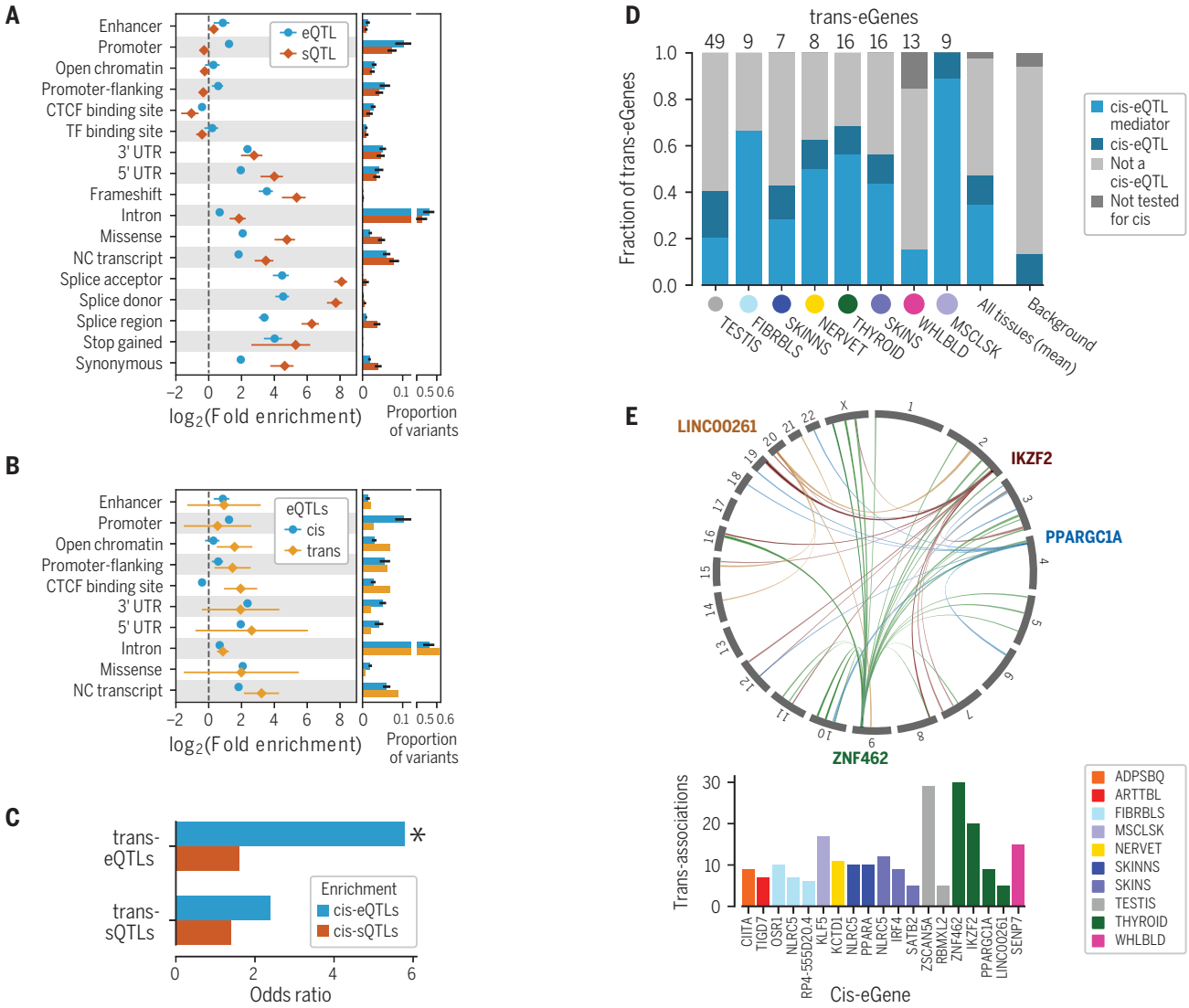

Fig. 4. Functional mechanisms of genetic regulatory effects. QTL enrichment in functional annotations for (**A**) cis-eQTLs and cis-sQTLs and for (**B**) trans-eQTLs. cis-QTL enrichment is shown as mean ± SD across tissues; trans-eQTL enrichment as 95% CI. UTR, untranslated region. (**C**) Enrichment of lead trans-eVariants or trans-sVariants that have been tested for cis-QTL effects also being significant cis-eVariants or cis-sVariants in the same tissue, respectively. Asterisk denotes significant enrichment, $P < 10^{-21}$. (**D**) Proportion of trans-eQTLs that are significant cis-eQTLs or mediated by cis-eQTLs. (**E**) Trans associations of cis-mediating genes identified through colocalization (PP4 > 0.8 and nominal association with discovery trans-eVariant $P < 10^{-5}$). (Top) Associations for four thyroid cis-eQTLs (indicated by gene names); (bottom) cis-mediating genes with five or more colocalizing trans-eQTLs.

statistical fine-mapping methods—CaVEMaN (*30*), CAVIAR (*31*), and dap-g (*32*)—to infer likely causal variants of cis-eQTLs in each tissue (Fig. 3A) (*11*). For many cis-eQTLs, the causal variant can be mapped with a high probability to a handful of candidates. The 90% credible set for each cis-eQTL consists of variants that include the causal variant with 90% probability; using dap-g, we identified a median of six variants in the 90% credible set for each cis-eQTL (fig. S19). Furthermore, 9.3% of the cis-eQTLs have a variant with a posterior probability >0.8 according to dap-g, indicating a single likely causal variant for those cis-eQTLs. We defined a consensus set of 24,740 cis-eQTLs across all tissues (7709 unique variants), for which the posterior probability was >0.8 across all three methods (fig. S20). Fine-mapped variants were significantly more enriched among experimentally validated causal variants from MPRA (*33*) and SuRE (*34*) compared with the lead eVariant across all eGenes (Fig. 3B). The highest enrichment was observed for the consensus set, although with overlapping confidence intervals (Fig. 3B). This demonstrates how careful fine-mapping facilitates the identification of likely causal regulatory variants.

Knowing the likely causal variant enables greater insights into the molecular mechanisms of individual eQTLs, including the mechanisms of their tissue-specific effects. Figure 3C shows an example of an eQTL for the gene *CBX8* that colocalizes with breast cancer risk and birth weight (posterior probability = 0.68 for both in lung). One of the three variants in the confident set overlaps the binding site and disrupts the motif of the transcription factor *EGR1* (*1*) (fig. S21). The role of *EGR1* as an upstream driver of this eQTL is further supported by a cross-tissue correlation of the effect size of the eQTL and the expression level of *EGR1* (Spearman's ρ = −0.69) (Fig. 3D).

Functional mechanisms of QTL associations

Quantitative trait data from multiple molecular phenotypes, integrated with the regulatory annotation of the genome (table S3), offer a

Downloaded from http://science.sciencemag.org/ on October 31, 2020

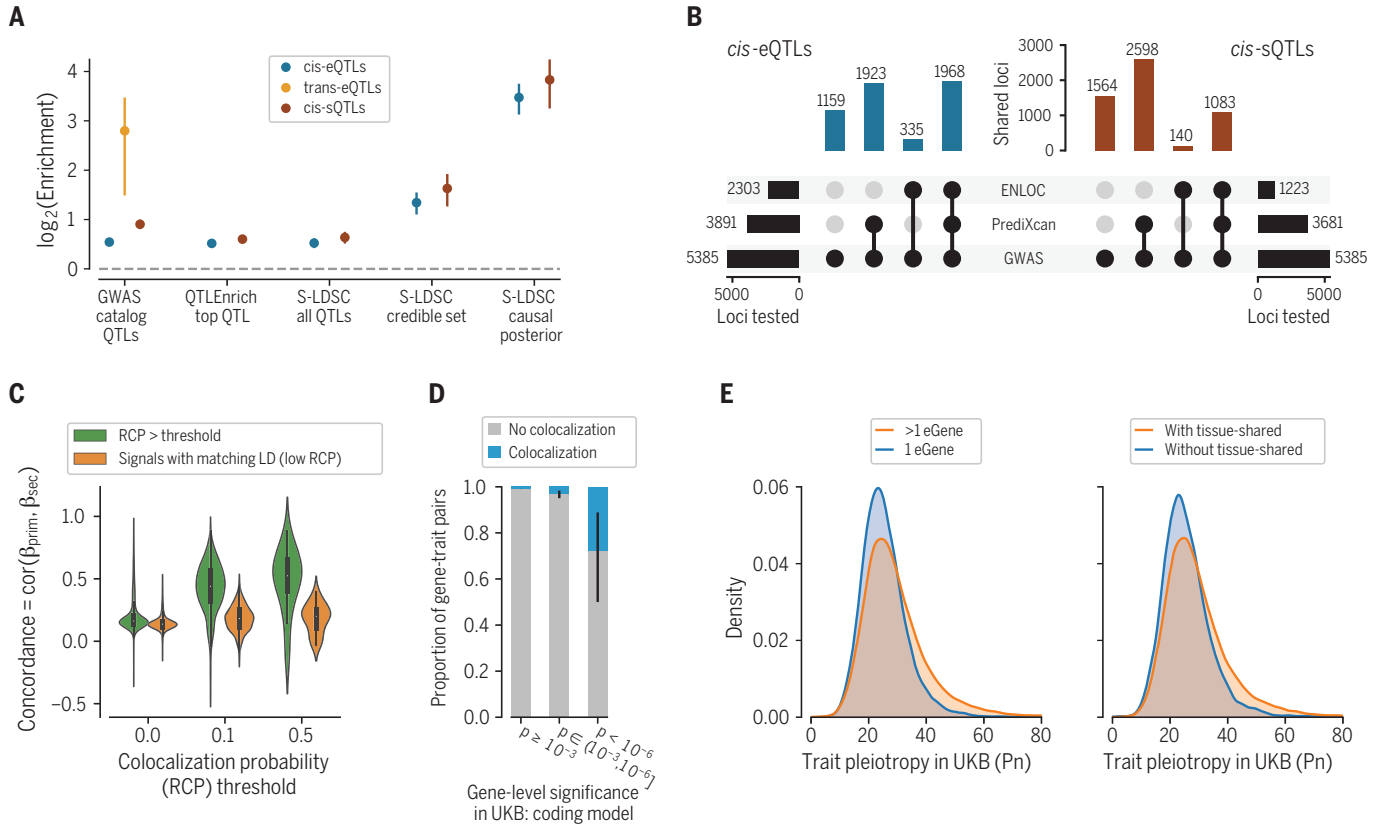

Fig. 5. Regulatory mechanisms of GWAS loci. (**A**) GWAS enrichment of cis-eQTLs, cis-sQTLs, and trans-eQTLs measured with different approaches: enrichment calculated from GWAS summary statistics of the most significant cis-QTL per eGene or sGene with QTLEnrich and LD score regression with all significant cis-QTLs (S-LDSC all QTLs), simple QTL overlap enrichment with all GWAS catalog variants, and LD score regression with fine-mapped cis-QTLs in the 95% credible set (S-LDSC credible set) and using posterior probability of causality as a continuous annotation (S-LDSC causal posterior). Enrichment is shown as mean and 95% CI. (**B**) Number of GWAS loci linked to eGenes or sGenes through colocalization (ENLOC) and association (PrediXcan), aggregated across tissues. (**C**) Concordance of mediated effects among independent cis-eQTLs for the same gene, shown for different levels of regional colocalization probability (RCP) (*32*), which is used as a proxy for the gene's causality. As the null, we show the concordance for LD matched genes without colocalization. (**D**) Proportion of colocalized cis-eQTLs with a matching phenotype for genes with different levels of rare variant trait association in the UK Biobank (UKB). (**E**) Horizontal GWAS trait pleiotropy score distribution for cis-eQTLs that regulate multiple versus a single gene (left) and for cis-eQTLs that are tissue-shared versus specific.

powerful way to understand the molecular mechanisms and phenotypic consequences of genetic regulatory effects. As expected, cis-eQTLs and cis-sQTLs are enriched in functional elements of the genome (Fig. 4A). Although the strongest enrichments are driven by variant classes that lead to splicing changes or nonsense-mediated decay, these account for relatively few variants. Cis-sQTLs are enriched almost entirely in transcribed regions, whereas cis-eQTLs are enriched in both transcribed regions and transcriptional regulatory elements. Previous studies (*4*, *35*) have indicated that cis-eQTL and cis-sQTL effects on the same gene are typically driven by different genetic variants. This observation is corroborated by the GTEx v8 data, where the overlap of cis-eQTL credible sets of likely causal variants, from CAVIAR analysis, have only a 12% overlap with cis-sQTL credible sets (fig. S22). Functional enrichment of overlapping and nonoverlapping cis-eQTLs and cis-sQTLs, using stringent LD filtering, showed that the patterns characteristic for each type—such as enrichment of cis-eQTLs in enhancers and cis-sQTLs in splice sites—are even stronger for distinct loci (fig. S22).

We hypothesized that eVariants and their target eGenes in cis are more likely to be in the same topologically associated domains (TADs) that allow chromatin interactions between more distant regulatory regions and target gene promoters (*36*). To test this supposition, we analyzed TAD data from ENCODE (*1*) and cis-eQTLs from matching GTEx tissues (table S3). Compared to matching random variant-gene pairs and controlling for distance from the transcription start site, cis-eVariant and cis-eGene pairs were significantly enriched for being in the same TAD [median odds ratio (OR) 4.55; all $P < 10^{-12}$] (fig. S23).

Trans-eQTLs are enriched in regulatory annotations that suggest both pre- and post-transcriptional mechanisms (Fig. 4B). Unlike cis-eQTLs, trans-eQTLs are enriched in CTCF binding sites, suggesting that disruption of CTCF binding may underlie distal genetic regulatory effects, potentially via its effect on interchromosomal chromatin interactions (*36*). Trans-eQTLs are also partially driven by cis-eQTLs (*37*, *38*), with a significant enrichment of lead trans-eVariants among cis-eVariants in the same tissue (5.9×; two-sided Fisher's exact test, $P = 5.03 \times 10^{-22}$) (Fig. 4C). A lack of analogous enrichment suggests that cis-sQTLs are less important contributors to trans-eQTLs (P = 0.064), and trans-sVariants had no significant enrichment of either cis-eQTLs (P = 0.051) or cis-sQTLs (P = 0.53). A further demonstration of the important contribution of cis-eQTLs to trans-eQTLs is that, on the basis of mediation analysis, 77% of lead trans-eVariants that are also cis-eVariants (corresponding to 31.6% of all lead trans-eVariants) appear to act through

Downloaded from http://science.sciencemag.org/ on October 31, 2020

the cis-eQTL (Fig. 4D and fig. S24). Colocalization of cis-eQTLs and trans-eQTLs was widespread and often tissue specific, with Fig. 4E showing cis-eQTLs with at least 10 nominally significant colocalized trans-eQTLs each [posterior probability of colocalization (PP4) > 0.8 and trans-eQTL $P < 10^{-5}$], pinpointing how local effects on gene expression can potentially lead to downstream regulatory effects across the genome (fig. S25 and table S16). The many remaining trans-eQTLs that do not coincide with a cis-eQTL may arise owing to mechanisms including undetected cis effects in specific cell types or conditions, protein coding changes, effects on cell type heterogeneity, or more complex causality such as a variant that influences a trait with downstream consequences on gene expression.

Genetic regulatory effects mediate complex trait associations

To analyze the role of regulatory variants in genetic associations for human traits, we first asked whether variants in the GWAS catalog were enriched for significant QTLs compared with all variants tested for QTLs (*11*). We observed a 1.46-fold enrichment for cis-eQTLs (63% versus 43%) and 1.86-fold enrichment for cis-sQTLs (37% versus 20%). The enrichment was even stronger for trans-eQTLs [6.97-fold (0.029% versus 0.0042%)], consistent with other analyses (*39*) (Fig. 5A, fig. S26, tables S5 and S6). Cell type proportion may influence detection of trans-eQTLs in heterogeneous tissues and may also be reflected in GWAS associations for blood cell count phenotypes and other complex traits. To minimize the possible impact of cell type heterogeneity on these enrichment statistics, we excluded blood cellularity traits and repeated these analyses. The resulting enrichments were 5.21-fold for trans-eQTLs, 1.43-fold for cis-eQTLs, and 1.81-fold for cis-sQTLs, largely preserving the patterns observed using the full set of GWAS traits.

This approach does not leverage the full power of GWAS and QTL association statistics, nor does it account for LD contamination, a situation wherein the causal variants for QTL and GWAS signals are distinct but LD between the two causal variants can suggest a false functional link (*40*). Therefore, for subsequent analyses (below) we selected 87 GWASs representing a broad array of binary and continuous complex traits that have summary results available in the public domain (*11*, *41*). To match the ancestry of the GWASs, analyses were performed using cis-QTL statistics calculated from the European subset of GTEx donors (fig. S29). The analyses were performed for all pairwise combinations of 87 phenotypes and 49 tissues and are summarized using an approach that accounts for similarity between tissues and variable standard errors of the QTL effect estimates, driven mainly by tissue sample size (fig. S27 and tables S4 and S11) (*11*).

To analyze the mediating role of cis-regulation of gene expression on complex traits (*35*, *42*), we used two complementary approaches, QTLEnrich (*43*) and stratified LD score regression (S-LDSC) (*11*, *44*). To rule out the possibility that enrichment is driven by specific features of cis-QTLs such as allele frequency, distance to the transcription start site, or local level of LD [number of LD proxy variants; coefficient of determination (R^2) ≥ 0.5], we used QTLEnrich. We found a 1.46-fold (SE = 0.006) and 1.56-fold (SE = 0.007) enrichment of trait associations among best cis-eQTLs and cis-sQTLs, respectively, adjusting for enrichment among matched null variants (Fig. 5A and table S7). The fact that these enrichment estimates differ little from those derived from the GWAS catalog overlap (above), even after accounting for the potential confounders, indicates how relatively robust these estimates are. Next, we used S-LDSC, adjusting for functional annotations (*44*), to confirm the robustness of these results and to analyze how GWAS enrichment is affected by the causal eVariant or sVariant being typically unknown (*11*). We computed the heritability enrichment of all cis-QTLs, fine-mapped cis-QTLs (in 95% credible set and posterior probability > 0.01 from dap-g), and fine-mapped cis-QTLs with maximum posterior inclusion probability as continuous annotation (*45*) (Fig. 5A). The largest increase in GWAS enrichment was for likely causal cis-QTL variants [11.1-fold (SE = 1.2) for cis-eQTLs and 14.2-fold (SE = 2.4) for cis-sQTLs, for the continuous annotation], which is strong evidence of shared causal effects of cis-QTLs and GWAS, and for the importance of fine-mapping.

Joint enrichment analysis of cis-eQTLs and cis-sQTLs shows an independent contribution to complex trait variation from both (fig. S28) (*11*), consistent with their limited overlap (fig. S22). The relative GWAS enrichments of cis-sQTLs and cis-eQTLs were similar (Fig. 5A; not significant for the robust QTLEnrich and LDSC analyses), but the larger number of cis-eQTLs discovered (Fig. 2) suggests a greater aggregated contribution of cis-eQTLs.

While these enrichment methods are powerful for genome-wide estimation of the QTL contribution to GWAS signals, they are not informative of regulatory mechanisms in individual loci. Thus, to provide functional interpretation of the 5385 significant GWAS associations in 1167 loci from approximately independent LD blocks (*46*) across the 87 complex traits, we performed colocalization with ENLOC (*32*) to quantify the probability that the cis-QTL and GWAS signals share the same causal variant. We also assessed the association between the genetically regulated component of expression or splicing and complex traits with PrediXcan (*11*, *41*, *47*). Both methods take multiple independent cis-QTLs into account, which is critical in large cis-eQTL studies with widespread allelic heterogeneity, such as GTEx. Of the 5385 GWAS loci, 43 and 23% were colocalized with a cis-eQTL and cis-sQTL, respectively (Fig. 5B). A large proportion of colocalized genes coincide with significant PrediXcan trait associations with predicted expression or splicing (median of 86 and 88% across phenotypes, respectively) (figs. S30 to S33 and tables S8 and S15), with the full resource available in (*41*). While colocalization does not prove a causal role of a QTL in any given locus nor a genome-wide proportion of GWAS loci driven by eQTLs, these results do suggest target genes and their potential molecular changes for thousands of GWAS loci, sometimes including both cis and trans targets (fig. S34).

Having multiple independent cis-eQTLs for a large number of genes allowed us to test whether mediated effects of primary and secondary cis-eQTLs on phenotypes—the ratio of GWAS and cis-eQTL effect sizes—are concordant. To ensure that concordance is not driven by residual LD between primary and secondary signals, we used LD-matched cis-eGenes with low colocalization probability as controls (*11*, *41*) and observed a significant increase in primary and secondary cis-eQTL concordance for colocalized genes (correlated t test $P < 10^{-30}$) (Fig. 5C). Additionally, colocalization of a cis-eQTL increased the colocalization of an independent cis-sQTL in the same locus (OR = 4.27, Fisher's exact test $P < 10^{-16}$) and, correspondingly, colocalization of a cis-sQTL increased cis-eQTL colocalization (OR = 4.54, Fisher's exact test $P < 10^{-16}$) (figs. S35 and S36). These observations indicate that multiple regulatory effects for the same gene often mediate the same complex trait associations. Furthermore, genes with suggestive rare variant trait associations in the UK Biobank (*48*) have a substantially increased proportion of colocalized eQTLs for the same trait (Fig. 5D and fig. S37), showing concordant trait effects from rare coding and common regulatory variants (*49*). These genes, as well as those with multiple colocalizing cis-QTLs, represent bona fide disease genes with multiple independent lines of evidence.

The growing number of genome and phenome studies has revealed extensive pleiotropy, where the same variant or locus associates with multiple organismal phenotypes (*50*). We sought to analyze how this phenomenon can be driven by gene regulatory effects. First, we calculated the number of cis-eGenes of each fine-mapped and LD-pruned cis-eVariant per tissue at local false sign rate (LFSR) < 5%, with cross-tissue smoothing of effect sizes with *mashr* (*11*, *51*). We observed that a median of 57% of variants were associated with more than one gene per tissue, typically co-occurring across tissues, indicating widespread regulatory pleiotropy. Using a binary classification of cis-eVariants with regulatory pleiotropy defined

Downloaded from http://science.sciencemag.org/ on October 31, 2020

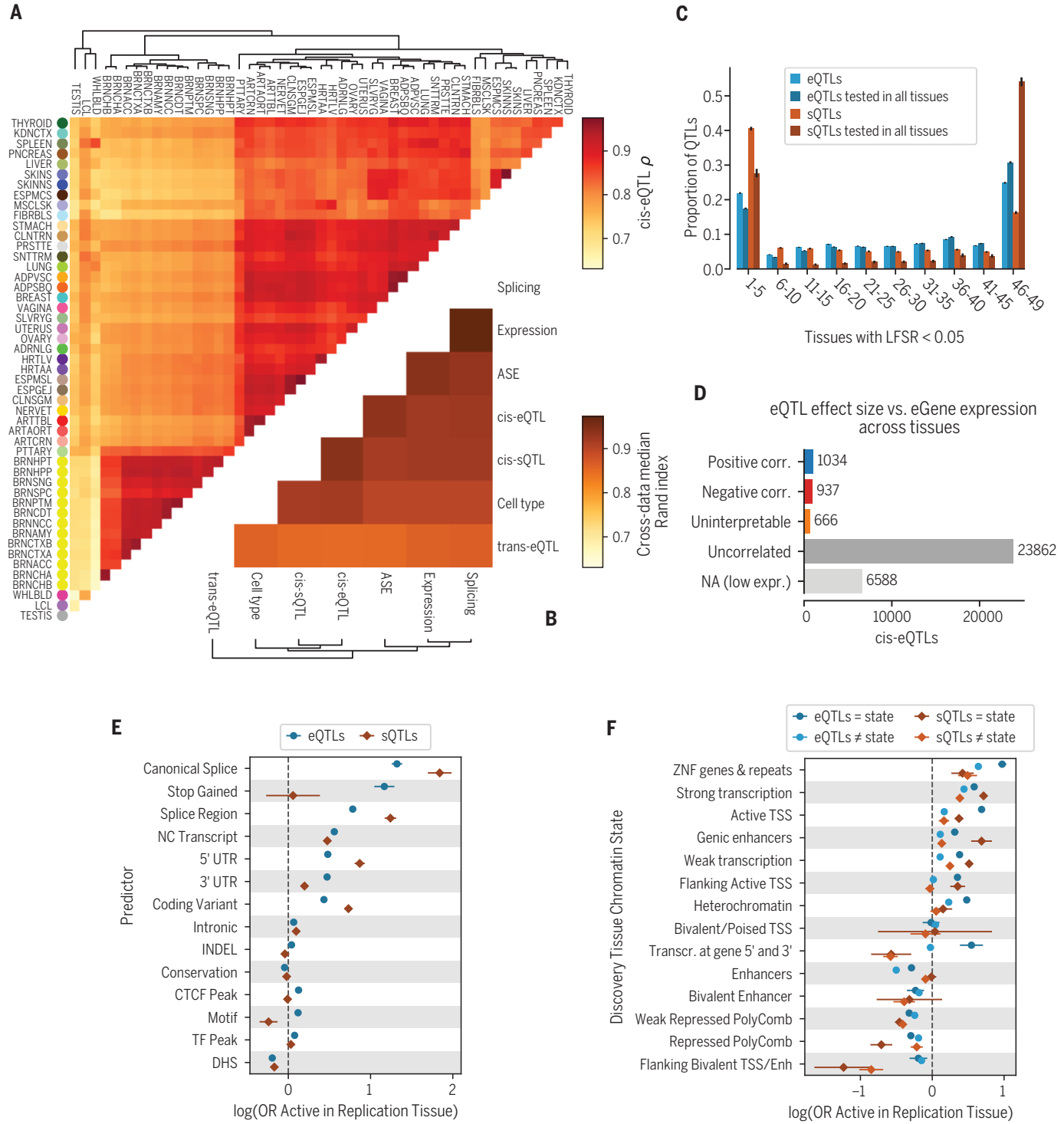

Fig. 6. Tissue specificity of cis-QTLs. (**A**) Tissue clustering with pairwise Spearman correlation of cis-eQTL effect sizes. (**B**) Similarity of tissue clustering across core data types quantified using median pairwise Rand index calculated across tissues. (**C**) Tissue activity of cis expression and splicing QTLs, where an eQTL was considered active in a tissue if it had a *mashr* local false sign rate (LFSR, equivalent to FDR) of <5%. This is shown for all cis-QTLs and only those that could be tested in all 49 tissues (red and blue). (**D**) Spearman correlation (corr.) between cis-eQTL effect size and eGene expression level across tissues. cis-eQTL counts are shown for those not tested owing to low expression (low expr.) level; tested but without significant (FDR < 5%) correlation (uncorrelated); a significant correlation, but effect sizes crossed zero, which made the correlation direction unclear (uninterpretable); positively correlated; and negatively correlated. (**E** and **F**) The effect of genomic function on cis-QTL tissue sharing modeled using logistic regression with functional annotations (E) and chromatin state (F). CTCF peak, motif, TF peak, and DHS (DNase I hypersensitive site) indicate whether the cis-QTL lies in a region annotated as having one of these features in any of the Ensembl Regulatory Build tissues. For chromatin states, model coefficients are shown for the discovery and replication tissues that have the same or different chromatin states. INDEL, insertion or deletion; ZNF, zinc finger; TSS, transcription start site; Transcr., transcription; Enh, enhancers.

Downloaded from http://science.sciencemag.org/ on October 31, 2020

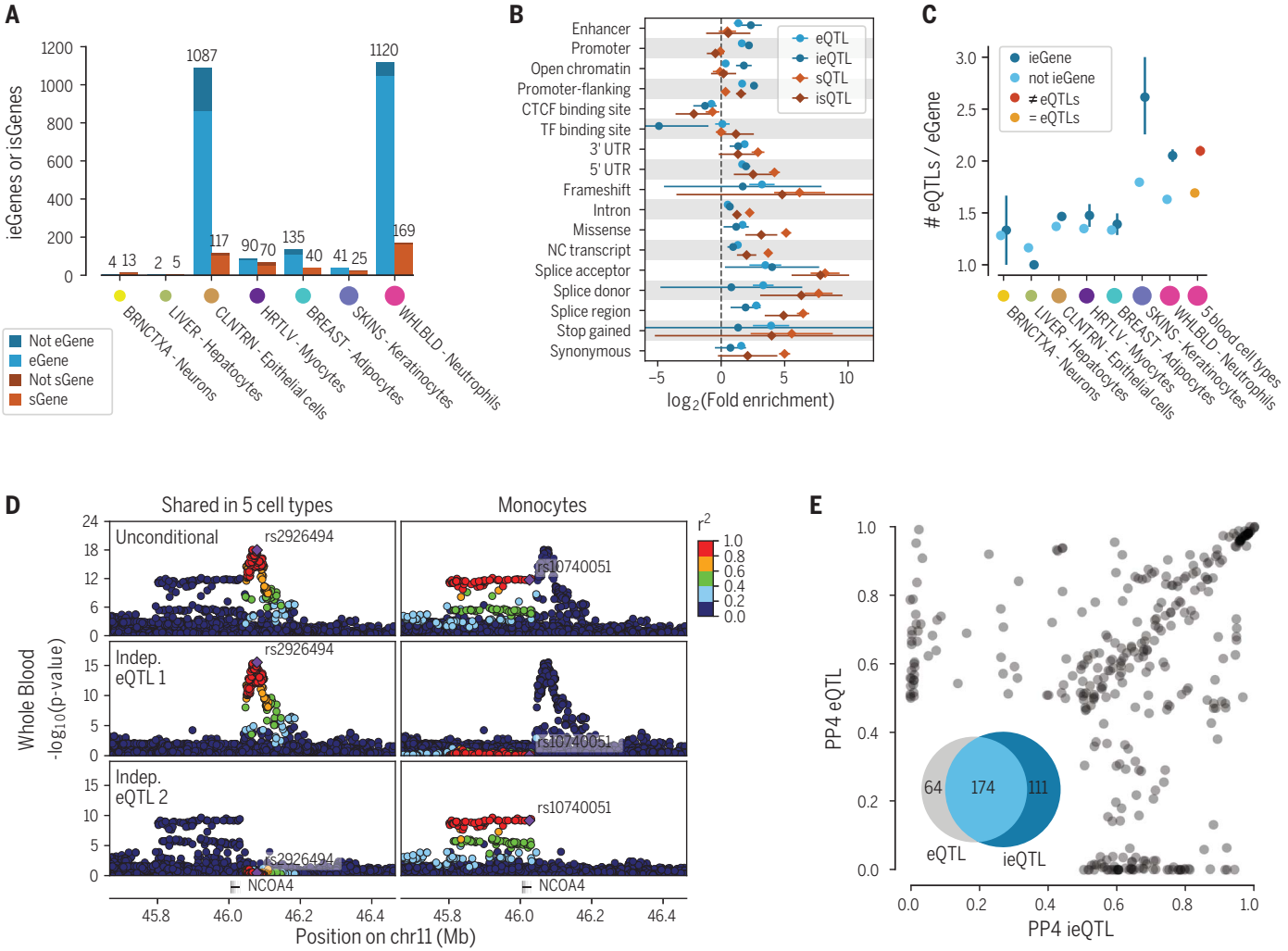

Fig. 7. Cell type interaction cis-eQTLs and cis-sQTLs. (**A**) Number of cell type interaction cis-eQTLs and cis-sQTLs (ieQTLs and isQTLs, respectively) discovered in seven tissue–cell type pairs, with shading indicating whether the ieGene or isGene was discovered by cis-eQTL or cis-sQTL analysis in bulk tissue. Colored dots are proportional to sample size. (**B**) Functional enrichment of neutrophil ieQTLs and isQTLs compared with cis-eQTLs and cis-sQTLs from whole blood. (**C**) Proportion of conditionally independent cis-eQTLs per eGene, for eGenes that do or do not have ieQTLs in GTEx, and for eGenes that have shared (=eQTLs) or nonshared (≠eQTLs) cis-eQTLs across five sorted blood cell types. (**D**) Whole blood cis-eQTL *P* value landscape for *NCOA4*, for the standard analysis (unconditional; top row) and for two independent cis-eQTLs (bottom rows). In a dataset of five sorted cell types (*56*), analyses of all cell types yielded a lead eVariant, rs2926494 (left), which is in high LD with the first independent cis-eQTL but not the second. The lead variant in monocyte cis-eQTL analysis, rs10740051, is in high LD with the second conditional cis-eQTL, indicating that this cis-eQTL is active specifically in monocytes. Thus, the full GTEx whole blood cis-eQTL pattern and allelic heterogeneity is composed of cis-eQTLs that are active in different cell types. (**E**) COLOC posterior probability (PP4) of GWAS colocalization with whole blood ieQTLs and eQTLs of the same eGene. Three hundred forty-nine gene-trait combinations across 132 genes and 36 GWAS traits showed evidence of colocalization (PP4 > 0.5) with an ieQTL and/or eQTL.

as those associated with more than one gene, we observed that they are more significantly associated with complex traits compared with matched cis-eVariants (fig. S38). This could be due to the fact that if a variant regulates multiple genes, there is a higher probability that at least one of them affects a GWAS phenotype.

However, cis-eVariants with regulatory pleiotropy also have higher GWAS complex trait pleiotropy (*50*) than cis-eVariants with effects on a single gene (Fig. 5E). This observation suggests a mechanism for complex trait pleiotropy of genetic effects where the expression of multiple genes in cis, rather than a single eGene effect, translates into diverse downstream physiological effects. Furthermore, GWAS pleiotropy is higher for tissue-shared (*41*) than tissue-specific cis-eQTLs, indicating that regulatory effects affecting multiple tissues are more likely to translate to diverse physiological traits (Fig. 5E).

Tissue specificity of genetic regulatory effects

The GTEx data provide an opportunity to study patterns and mechanisms of tissue specificity of the transcriptome and its genetic regulation. Pairwise similarity of GTEx tissues was quantified from gene expression and splicing, as well as allelic expression, eQTLs in cis and trans, and cis-sQTLs (Fig. 6A and fig. S41) (*11*). These estimates show consistent patterns of tissue relatedness, indicating that the biological processes that drive transcriptome similarity also control tissue sharing of genetic effects (Fig. 6B). As seen in earlier versions of the GTEx data (*9*, *10*), the brain regions form a separate cluster, and testis, lymphoblastoid cell lines, whole blood, and sometimes liver tend to be

Downloaded from http://science.sciencemag.org/ on October 31, 2020

outliers, while most other organs have a notably high degree of similarity to each other. This indicates that blood is not an ideal proxy for most tissues and that some other relatively accessible tissues, such as skin, may better capture molecular effects in other tissues.

The overall tissue specificity of QTLs (*11*) follows a U-shaped curve, recapitulating previous GTEx analyses (*9*, *10*), where genetic regulatory effects tend to be either highly tissue specific or highly shared (Fig. 6C), with trans-eQTLs being more tissue specific than cis-eQTLs (fig. S40). Cis-sQTLs appear to be significantly more tissue specific than cis-eQTLs when considering all mapped cis-QTLs, but this pattern is reversed when considering only those cis-QTLs where the gene or splicing event is quantified in all tissues (Fig. 6C and fig. S39). These observations indicate that splicing measures are more tissue specific than gene expression, but genetic effects on splicing tend to be more highly shared, which is consistent with pairwise tissue-sharing patterns (fig. S41). These opposite patterns are important for understanding effects that disease-causing splicing variants may have across tissues and for validation of splicing effects in cell lines that rarely are an exact match to cells in vivo.

Next, we analyzed the sharing of AE across multiple tissues of an individual, which is a metric of sharing of any heterozygous regulatory variant effects in that individual. Variation in AE has been useful for analysis of rare, potentially disease-causing variants (*52*). Using a clustering approach (*11*), we found that in 97.4% of the cases, AE across all tissues forms a single cluster. This suggests that in AE analysis, different tissues are often relatively good proxies for one another, provided that the gene of interest is expressed in the probed tissue (fig. S42).

We next computed the cross-tissue correlation of eQTL effect size and eGene expression level—often a proxy for gene functionality—and discovered that 1971 cis-eQTLs (7.4%; FDR 5%) had a significant and robust correlation between eGene expression and cis-eQTL effect size across tissues (Fig. 6D and fig. S43). These correlated cis-eQTLs are split nearly evenly between negative (937) and positive (1034) correlations. Thus, the tissues with the highest cis-eQTL effect sizes are equally likely to be among tissues with higher or lower expression levels for the gene. Trans-eQTLs show a different pattern, typically being observed in tissues with high expression of the trans-eGene relative to other tissues (fig. S43).

These observations raise the question of how to prioritize the relevant tissues for eQTLs in a disease context. To address this, we chose a subset of GWAS traits with a strong prior indication for the likely relevant tissue(s) (table S12). Analyzing colocalized cis-eQTLs for 1778 GWAS loci (*11*), we discovered that the relevant tissues were significantly enriched in having high expression and effect sizes (paired Wilcoxon sign test $P < 1.5 \times 10^{-4}$), but the relatively weak signal indicates that pinpointing the likely relevant tissue for GWAS loci is challenging (figs. S44 and S45 and table S9). These results indicate that both effect sizes and gene expression levels are important for interpreting the tissue context where an eQTL may have downstream phenotypic effects.

The diverse patterns of QTL tissue specificity raise the question of what molecular mechanisms underlie the ubiquitous regulatory effects of some genetic variants and the highly tissue-specific effects of others. To gain insight into this question, we modeled cis-eQTL and cis-sQTL tissue specificity using logistic regression as a function of the lead eVariant's genomic and epigenomic context (*11*). Cis-QTLs where the top eVariant was in a transcribed region had overall higher sharing than those in classical transcriptional regulatory elements, indicating that genetic variants with post- or cotranscriptional expression or splicing effects have more ubiquitous effects (Fig. 6E). Canonical splice and stop-gained variant effects had the highest probability of being shared across tissues, which may benefit disease-focused studies relying on likely gene-disrupting variants.

We also considered whether varying regulatory activity between tissues contributed to tissue specificity of genetic effects, and we found that shared chromatin states between the discovery and query tissues were associated with increased probability of cis-eQTL sharing and vice versa (Fig. 6F). cis-eQTLs and cis-sQTLs followed similar patterns. Because cis-sQTLs are more enriched in transcribed regions and likely arise via posttranscriptional mechanisms (Fig. 4A), this is likely to contribute to their higher overall degree of tissue sharing (Fig. 6C). In comparison to cis-eQTLs, cis-sQTLs are more often located in regions where regulatory effects are shared.

These data indicate a possible means by which we can predict whether a cis-eQTL observed in a GTEx tissue is active in another tissue of interest, using the variant's annotation and properties in the discovery tissue (*11*). After incorporating additional features including cis-QTL effect size, distance to transcription start site, and eGene and sGene expression levels, we obtain reasonably good predictions of whether a cis-QTL is active in a query tissue (median area under the curve = 0.779 and 0.807, minimum = 0.703 and 0.721, maximum = 0.807 and 0.875 for cis-eQTLs and cis-sQTLs, respectively) (fig. S46). These results suggest that it is possible to extrapolate the GTEx cis-eQTL catalog to additional tissues and potentially developmental stages, where population-scale data for QTL analysis are particularly difficult to collect.

From tissues to cell types

The GTEx tissue samples consist of heterogeneous mixtures of multiple cell types. Hence, the RNA extracted and QTLs mapped from these samples reflect a composite of genetic effects that may vary across cell types and may mask cell type–specific mechanisms. To characterize the effect of cell type heterogeneity on analyses from bulk tissue, we used the xCell method (*53*) to estimate the enrichment of 64 reference cell types from the bulk expression profile of each sample (*11*). Although these results need to be interpreted with caution given the scarcity of validation data (*54*), the resulting enrichment scores were generally biologically meaningful with, for example, myocytes enriched in heart left ventricle and skeletal muscle; hepatocytes enriched in liver; and various blood cell types enriched in whole blood, spleen, and lung, which harbors a large leukocyte population (fig. S47). Interestingly, the pairwise relatedness of GTEx tissues derived from their cell type composition is highly correlated with tissue sharing of regulatory variants (cis-eQTL versus cell type composition Rand index = 0.92) (Fig. 6B and figs. S48 and S41), suggesting that similarity of regulatory variant activity between tissue pairs may often be due to the presence of similar cell types and not necessarily shared regulatory networks within cells. This observation highlights the key role that characterizing cell type diversity will have for understanding not only tissue biology but genetic regulatory effects as well.

Enrichment of many cell types shows interindividual variation within a given tissue, partially owing to tissue sampling variation between individuals. This variation can be leveraged to identify cis-eQTLs and cis-sQTLs with cell type specificity by including an interaction between genotype and cell type enrichment in the QTL model (*11*, *55*). We applied this approach to seven tissue–cell type pairs with robustly quantified cell types in the tissue where each cell type was most enriched (Fig. 7A) [an additional 36 pairs are described in (*54*)]. The largest numbers of cell type interaction cis-eQTLs and cis-sQTLs (ieQTLs and isQTLs, respectively) were 1120 neutrophil ieQTLs and 169 isQTLs in whole blood and 1087 epithelial cell ieQTLs and 117 isQTLs in transverse colon (Fig. 7A). Of these ieQTLs, 76 and 229, respectively, corresponded to an eGene for which no QTL was detected in bulk tissue.

We validated these effects using published eQTLs from purified blood cell types (*56*), where neutrophil eQTLs had higher neutrophil ieQTL effect sizes than eQTLs from other blood cell types (fig. S49). For other cell types, external replication data was not available. Thus, we verified the robustness of the ieQTLs by the allelic expression validation approach that was used for sex- and population-biased

Downloaded from http://science.sciencemag.org/ on October 31, 2020

cis-eQTL analyses: For ieQTL heterozygotes, we calculated the Spearman correlation between cell type enrichment and ieQTL effect size from AE data and observed a high validation rate (*54*). Note that ieQTLs and isQTLs should not be considered cell type–specific QTLs, because the enrichment of any cell type may be (anti)correlated with other cell types (fig. S50). While full deconvolution of cis-eQTL effects driven by specific cell types remains a challenge for the future, ieQTLs and isQTLs can be interpreted as being enriched for cell type–specific effects.

In most subsequent analyses to characterize the properties of ieQTLs and isQTLs, we focused on neutrophil ieQTLs, which are numerous and supported by external replication data. Functional enrichment analyses of these QTLs show that they largely follow the enrichment patterns observed for bulk tissue cis-QTLs (Fig. 7B). However, ieQTLs are more strongly enriched in promoter-flanking regions and enhancers, which are known to be major drivers of cell type–specific regulatory effects (*2*). Epithelial cell ieQTLs yielded similar patterns (fig. S51).

We hypothesized that the widespread allelic heterogeneity observed in the bulk tissue cis-eQTL data could be partially driven by an aggregate signal from cis-eQTLs that are each active in a different cell type present in the tissue. Indeed, the number of cis-eQTLs per gene is higher for ieGenes than for standard eGenes, especially in skin and blood (Fig. 7C). While differences in power could contribute to this pattern, it is corroborated by eGenes that have independent cis-eQTLs ($R^2 < 0.05$) in five purified blood cell types (*56*) also showing an increased amount of allelic heterogeneity in GTEx whole blood (Fig. 7, C and D). Thus, quantifying cell type specificity can provide mechanistic insights into the genetic architecture of gene expression and may be leveraged to improve the resolution of complex patterns of allelic heterogeneity wherein we can distinguish effects manifesting in different cell types.

Next, we analyzed how cell type interaction cis-QTLs contribute to the interpretation of regulatory variants underlying complex disease risk. GWAS colocalization analysis of neutrophil ieQTLs (*11*) revealed multiple loci (111, ~32%) that colocalize only with ieQTLs and not with whole blood cis-eQTLs (Fig. 7E), although 75% (42 of 56) of the corresponding eGenes have both cis-eQTLs and ieQTLs. Improved resolution into allelic heterogeneity appears to contribute to colocalization exclusively with eQTLs. For example, the absence of colocalization between a platelet count GWAS signal and bulk tissue cis-eQTL for *SPAG7* appears to be due to the whole blood signal being an aggregate of multiple independent signals (fig. S52). The neutrophil ieQTL analysis uncovers a specific signal that mirrors the GWAS association, suggesting that platelet counts are affected by *SPAG7* expression only in one or several specific cell types. Thus, in addition to previously undetected colocalizations pinpointing potential causal genes, ieQTL analysis has the potential to provide insights into cell type–specific mechanisms of complex traits.

Outlook

The GTEx v8 data release represents a deep survey of both intra- and interindividual transcriptome variation across a large number of tissues. With 838 donors and 15,201 samples—approximately twice the size of the v6 release used in the previous set of GTEx Consortium papers—we have created a comprehensive resource of genetic variants that influence gene expression and splicing in cis. This substantially expands and updates the GTEx catalog of sQTLs, doubles the number of eGenes per tissue, and saturates the discovery of eQTLs with greater than twofold effect sizes in ~40 tissues. The fine-mapping data of GTEx cis-eQTLs provide a set of thousands of likely causal functional variants. While trans-QTL discovery and the characterization of sex- and population-specific genetic effects are still limited by sample size, analyses of the v8 data provide important insights into each.

Cell type interaction cis-eQTLs and cis-sQTLs, mapped with computational estimates of cell type enrichment, constitute an important extension of the GTEx resource to effects of cell types within tissues. The highly similar tissue-sharing patterns across these data types suggest shared biology from cell type composition to transcriptome variation and genetic regulatory effects. Our results indicate that shared cell types between tissues may be a key factor behind tissue sharing of genetic regulatory effects, which will constitute a key challenge to tackle in the future. Finally, GWAS colocalization with cis-eQTLs and cis-sQTLs provides rich opportunities for further functional follow-up and characterization of regulatory mechanisms of GWAS associations.

Given the very large number of cis-eQTLs, the extensive allelic heterogeneity—multiple independent regulatory variants affecting the same gene—is unsurprising. With well-powered cis-QTL mapping, it becomes possible and important to describe and disentangle these effects; the assumption of a single causal variant in a cis-eQTL locus no longer holds true for datasets of this scale. Similarly, we highlight cis-eQTL and cis-sQTL effects on the same gene, typically driven by distinct causal variants (*4*, *35*). The joint complex trait contribution of independent cis-eQTLs and cis-sQTLs and that of cis-eQTLs and rare coding variants for the same gene highlights how different genetic variants and functional perturbations can converge at the gene level to similar physiological effect. This orthogonal evidence pinpoints highly likely causal disease genes, and these associations could be leveraged to build allelic series, a powerful tool for estimating dosage-risk relationship for the purposes of drug development (*57*).

Finally, we provide mechanistic insights into the cellular causes of allelic heterogeneity, showing the separate contributions from cis-eQTLs active in different cell types to the combined signal seen in a bulk tissue sample. With evidence that this increased cellular resolution improves colocalization in some loci, cell type–specific analyses appear particularly promising for finer dissection of genetic association data.

Integration of GTEx QTL data and functional annotation of the genome provides powerful insights into the molecular mechanisms of transcriptional and posttranscriptional regulation that affect gene expression levels and splicing. A large proportion of cis-eQTL effects are driven by genetic perturbations in classical regulatory elements of promoters and enhancers. However, the magnitude of these enrichments is perhaps unexpectedly modest, which likely reflects the fact that only a small fraction of variants in these large regions have true regulatory effects, leading to a lower resolution of annotating functional variants compared with the nucleotide-level annotation of, for example, nonsense or canonical splice site variants. Context-specific genetic effects of tissue-specific and cell type interaction cis-eQTLs are enriched in enhancers and related elements and their variable activity across tissues and cell types.

While cis-eQTLs are enriched for a wide range of functional regions, the vast majority of cis-sQTL are located in transcribed regions, with likely cotranscriptional and/or posttranscriptional regulatory effects. Interestingly, these appear to be less tissue specific, which likely contributes to the higher tissue sharing of cis-sQTLs than cis-eQTLs. The higher tissue sharing of all cotranscriptional or posttranscriptional regulatory effects may facilitate interpretation of potentially disease-related functional effects of (rare) coding variants triggering nonsense-mediated decay or splicing changes, even when the disease-relevant tissues are not available.

About a third of the observed trans-eQTLs are mediated by cis-eQTLs, demonstrating how local genetic regulatory effects can translate to effects at the level of cellular pathways. All types of QTLs that were studied are strong mediators of genetic associations to complex traits, with a higher relative enrichment for cis-sQTLs than cis-eQTLs and with trans-eQTLs having the highest enrichment of all (*35*). With large genome- and phenome-wide studies having uncovered extensive pleiotropy of complex trait associations, the GTEx data provide important

Downloaded from http://science.sciencemag.org/ on October 31, 2020

insights into the molecular underpinnings of this observed pleiotropy: Variants that affect the expression of multiple genes and multiple tissues have a higher degree of complex trait pleiotropy, indicating that some of the pleiotropy arises at the proximal regulatory level. Dissecting this complexity and pinpointing truly causal molecular effects that mediate specific phenotype associations will be a considerable challenge for the future.

This study of the GTEx v8 data has provided insights into genetic regulatory architecture and functional mechanisms. The catalog of QTLs and associated datasets of annotations, cell type enrichments, and GWAS summary statistics requires careful interpretation but provides insights into the biology of gene regulation and functional mechanisms of complex traits. We demonstrate how QTL data can be used to inform on multiple aspects of GWAS interpretation: potential causal variants from fine-mapping, proximal regulatory mechanisms, target genes in cis, and pathway effects in trans, in the context of multiple tissues and cell types. However, our understanding of genetic effects on cellular phenotypes is far from complete. We envision that further investigation into genetic regulatory effects in specific cell types, study of additional tissues and developmental time points not covered by GTEx, incorporation of a diverse set of molecular phenotypes, and continued investment in increasing sample sizes from diverse populations will continue to provide transformative scientific discoveries.

REFERENCES AND NOTES

1. ENCODE Project Consortium, *Nature* **489**, 57–74 (2012).
2. A. Kundaje *et al*., *Nature* **518**, 317–330 (2015).
3. A. Battle *et al*., *Genome Res.* **24**, 14–24 (2014).
4. T. Lappalainen *et al*., *Nature* **501**, 506–511 (2013).
5. M. J. Bonder *et al*., *Nat. Genet.* **49**, 131–138 (2017).
6. J. Lonsdale *et al*., *Nat. Genet.* **45**, 580–585 (2013).
7. L. J. Carithers *et al*., *Biopreserv. Biobank.* **13**, 311–319 (2015).
8. L. A. Siminoff, M. Wilson-Genderson, H. M. Gardiner, M. Mosavel, K. L. Barker, *Biopreserv. Biobank.* **16**, 200–206 (2018).
9. GTEx Consortium, *Science* **348**, 648–660 (2015).
10. GTEx Consortium, *Nature* **550**, 204–213 (2017).
11. See supplementary materials.
12. O. M. de Goede *et al*., bioRxiv 793091 [Preprint]. 4 October 2019. https://doi.org/10.1101/793091.
13. Y. I. Li *et al*., *Nat. Genet.* **50**, 151–158 (2018).
14. R. Jansen *et al*., *Hum. Mol. Genet.* **26**, 1444–1451 (2017).
15. F. Hormozdiari *et al*., *Am. J. Hum. Genet.* **100**, 789–802 (2017).
16. A. Saha, A. Battle, *F1000Res.* **7**, 1860 (2018).
17. S. E. Castel *et al*., *Genome Biol.* 10.1186/s13059-020-02122-z (2020).
18. E. A. Khramtsova, L. K. Davis, B. E. Stranger, *Nat. Rev. Genet.* **20**, 173–190 (2019).
19. B. E. Stranger *et al*., *PLOS Genet.* **8**, e1002639 (2012).
20. T. Raj *et al*., *Science* **344**, 519–523 (2014).
21. P. Mohammadi, S. E. Castel, A. A. Brown, T. Lappalainen, *Genome Res.* **27**, 1872–1884 (2017).
22. M. Oliva *et al*., *Science* **369**, eaba3066 (2020).
23. T. Sun *et al*., *Carcinogenesis* **25**, 2225–2230 (2004).
24. A. Ewart-Toland *et al*., *Carcinogenesis* **26**, 1368–1373 (2005).
25. Y. Ruan *et al*., *J. Pathol.* **225**, 535–543 (2011).
26. H. M. Koh *et al*., *J. Pathol. Transl. Med.* **51**, 32–39 (2017).
27. K. Dhanasekaran *et al*., *FASEB J.* **33**, 219–230 (2019).
28. S. Girardeau-Hubert *et al*., *Sci. Rep.* **9**, 7456 (2019).
29. L. Yin *et al*., *Exp. Dermatol.* **23**, 731–735 (2014).
30. A. A. Brown *et al*., *Nat. Genet.* **49**, 1747–1751 (2017).
31. F. Hormozdiari, E. Kostem, E. Y. Kang, B. Pasaniuc, E. Eskin, *Genetics* **198**, 497–508 (2014).
32. X. Wen, R. Pique-Regi, F. Luca, *PLOS Genet.* **13**, e1006646 (2017).
33. R. Tewhey *et al*., *Cell* **165**, 1519–1529 (2016).
34. J. van Arensbergen *et al*., *Nat. Genet.* **51**, 1160–1169 (2019).
35. Y. I. Li *et al*., *Science* **352**, 600–604 (2016).
36. O. Delaneau *et al*., *Science* **364**, eaat8266 (2019).
37. K. S. Small *et al*., *Nat. Genet.* **43**, 561–564 (2011).
38. F. Yang, J. Wang, B. L. Pierce, L. S. Chen; GTEx Consortium, *Genome Res.* **27**, 1859–1871 (2017).
39. H.-J. Westra *et al*., *Nat. Genet.* **45**, 1238–1243 (2013).
40. B. Liu, M. J. Gloudemans, A. S. Rao, E. Ingelsson, S. B. Montgomery, *Nat. Genet.* **51**, 768–769 (2019).
41. A. N. Barbeira *et al*., bioRxiv 814350 [Preprint]. 23 May 2020. https://doi.org/10.1101/814350.
42. D. L. Nicolae *et al*., *PLOS Genet.* **6**, e1000888 (2010).20369019
43. E. R. Gamazon *et al*., *Nat. Genet.* **50**, 956–967 (2018).
44. H. K. Finucane *et al*., *Nat. Genet.* **47**, 1228–1235 (2015).
45. F. Hormozdiari *et al*., *Nat. Genet.* **50**, 1041–1047 (2018).
46. T. Berisa, J. K. Pickrell, *Bioinformatics* **32**, 283–285 (2016).
47. E. R. Gamazon *et al*., *Nat. Genet.* **47**, 1091–1098 (2015).
48. E. T. Cirulli *et al*., *Nat. Commun.* **11**, 542 (2020).
49. N. M. Ferraro *et al*., *Science* **369**, eaaz5900 (2020).
50. D. M. Jordan, M. Verbanck, R. Do, *Genome Biol.* **20**, 222–18 (2019).
51. S. M. Urbut, G. Wang, P. Carbonetto, M. Stephens, *Nat. Genet.* **51**, 187–195 (2019).
52. P. Mohammadi *et al*., *Science* **366**, 351–356 (2019).
53. D. Aran, Z. Hu, A. J. Butte, *Genome Biol.* **18**, 220 (2017).
54. S. Kim-Hellmuth *et al*., *Science* **369**, eaaz8528 (2020).
55. D. V. Zhernakova *et al*., *Nat. Genet.* **49**, 139–145 (2017).
56. J. E. Peters *et al*., *PLOS Genet.* **12**, e1005908 (2016).
57. R. M. Plenge, E. M. Scolnick, D. Altshuler, *Nat. Rev. Drug Discov.* **12**, 581–594 (2013).
58. F. Aguet, broadinstitute/gtex-pipeline: GTEx v8, Zenodo (2020); http://doi.org/10.5281/zenodo.3727189.
59. F. Aguet, broadinstitute/gtex-v8: main_figures, Zenodo (2020); http://doi.org/10.5281/zenodo.3930961.

ACKNOWLEDGMENTS

We thank the donors and their families for their generous gifts of organ donation for transplantation and tissue donations for the GTEx research project; the Genomics Platform at the Broad Institute for data generation; J. Struewing for support and leadership of the GTEx project; M. Khan and C. Stolte for the illustrations in Fig. 1; and R. Do, D. Jordan, and M. Verbanck for providing GWAS pleiotropy scores. **Funding:** This work was supported by the Common Fund of the Office of the Director, U.S. National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, NIA, NIAID, and NINDS through NIH contracts HHSN261200800001E (Leidos Prime contract with NCI: 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., and A.U.), 10XS170 (NDRI: 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., and L.K.B.), 10XS171 (Roswell Park Cancer Institute: B.A.F., M.T.M., E.K., B.M.G., K.D.R., and J.B.), 10X172 (Science Care Inc.), 12ST1039 (IDOX), 10ST1035 (Van Andel Institute: S.D.J., D.C.R., and D.R.V.), HHSN268201000029C (Broad Institute: F.A., G.G., K.G.A., A.V.S., X.Li., E.T., S.G., A.G., S.A., K.H.H., D.T.N., K.H., S.R.M., and J.L.N.), 5U41HG009494 (F.A., G.G., and K.G.A.), and through NIH grants R01 DA006227-17 (University of Miami Brain Bank: D.C.M. and D.A.D.), Supplement to University of Miami grant DA006227 (D.C.M. and D.A.D.), R01 MH090941 (University of Geneva), R01 MH090951 and R01 MH090937 (University of Chicago), R01 MH090936 (University of North Carolina–Chapel Hill), R01MH101814 (M.M.-A., V.W., S.B.M., R.G., E.T.D., D.G.-M., and A.V.), U01HG007593 (S.B.M.), R01MH101822 (C.D.B.), U01HG007598 (M.O. and B.E.S.), U01MH104393 (A.P.F.), extension H002371 to 5U41HG002371 (W.J.K.), as well as other funding sources: R01MH106842 (T.L., P.M., E.F., and P.J.H.), R01HL142028 (T.L., Si.Ka., and P.J.H.), R01GM122924 (T.L. and S.E.C.), R01MH107666 (H.K.I.), P30DK020595 (H.K.I.), UM1HG008901 (T.L.), R01GM124486 (T.L.), R01HG010067 (Y.Pa.), R01HG002585 (G.Wa. and M.St.), Gordon and Betty Moore Foundation GBMF 4559 (G.Wa. and M.St.), 1K99HG009916-01 (S.E.C.), R01HG006855 (Se.Ka. and R.E.H.), BIO2015-70777-P, Ministerio de Economia y Competitividad and FEDER funds (M.M.-A., V.W., R.G., and D.G.-M.), la Caixa Foundation ID 100010434 under agreement LCF/BQ/SO15/52260001 (D.G.-M.), NIH CTSA grant UL1TR002550-01 (P.M.), Marie-Skłodowska Curie fellowship H2020 Grant 706636 (S.K.-H.), R35HG010718 (E.R.G.), FPU15/03635, Ministerio de Educación, Cultura y Deporte (M.M.-A.), R01MH109905, 1R01HG010480 (A.Ba.), Searle Scholar Program (A.Ba.), R01HG008150 (S.B.M.), 5T32HG000044-22, NHGRI Institutional Training Grant in Genome Science (N.R.G.), EU IMI program (UE7-DIRECT-115317-1) (E.T.D. and A.V.), FNS funded project RNA1 (31003A_149984) (E.T.D. and A.V.), DK110919 (F.H.), F32HG009987 (F.H.), Massachusetts Lions Eye Research Fund Grant (A.R.H.), Wellcome grant WT108749/Z/15/Z (P.F.), and European Molecular Biology Laboratory (P.F. and D.Z.). **Author contributions:** See section 17 of the supplementary materials. **Competing interests:** F.A. is an inventor on a patent application related to TensorQTL. S.E.C. is a cofounder, 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 of consultant to, and holds privately held equity in Scorpion Therapeutics. S.B.M. is on the scientific advisory board of MyOme. D.G.M. is a cofounder 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 member of the scientific advisory boards of Fabric Genomics, Inc., and Eagle Genomes, Ltd. P.G.F. is a partner of Bioinf2Bio. **Data and materials availability:** All GTEx protected data are available through the database of Genotypes and Phenotypes (dbGaP) (accession no. phs000424.v8). Access to the raw sequence data is now provided through the AnVIL platform (https://gtexportal.org/home/protectedDataAccess). Public-access data, including QTL summary statistics and expression levels, are available on the GTEx Portal as downloadable files and through multiple data visualizations and browsable tables (www.gtexportal.org), as well as in the UCSC and Ensembl browsers. All components of the single-tissue cis-QTL pipeline are available at https://github.com/broadinstitute/gtex-pipeline and Zenodo (*58*), and analysis scripts are available at https://github.com/broadinstitute/gtex-v8 and Zenodo (*59*). Residual GTEx biospecimens have been banked and are available as a resource for further studies (access can be requested on the GTEx Portal at www.gtexportal.org/home/samplesPage).

Authors

Lead Analysts†: François Aguet[1]‡, Alvaro N. Barbeira[2], Rodrigo Bonazzola[2], Andrew Brown[3,4], Stephane E. Castel[5,6], Brian Jo[7,8], Silva Kasela[5,6], Sarah Kim-Hellmuth[5,6,9], Yanyu Liang[2], Meritxell Oliva[2,10], Princy Parsana[11]

Analysts†: Elise D. Flynn[5,6], Laure Fresard[12], Eric R. Gamazon[13,14,15,16], Andrew R. Hamel[17,1], Yuan He[18], Farhad Hormozdiari[19,1], Pejman Mohammadi[5,6,20,21], Manuel Muñoz-Aguirre[22,23], YoSon Park[24,25], Ashis Saha[11], Ayellet V. Segrè[1,17], Benjamin J. Strober[18], Xiaoquan Wen[26], Valentin Wucher[22]

Manuscript Working Group†: François Aguet[1], Kristin G. Ardlie[1], Alvaro N. Barbeira[2], Alexis Battle[18,11], Rodrigo Bonazzola[2], Andrew Brown[3,4], Christopher D. Brown[24], Stephane E. Castel[5,6], Nancy Cox[16], Sayantan Das[26], Emmanouil T. Dermitzakis[3,27,28], Barbara E. Engelhardt[7,8], Elise D. Flynn[5,6], Laure Fresard[12], Eric R. Gamazon[13,14,15,16], Diego Garrido-Martín[22], Nicole R. Gay[29], Gad A. Getz[1,30,31], Roderic Guigó[22,32], Andrew R. Hamel[17,1], Robert E. Handsaker[33,33,35], Yuan He[18], Paul J. Hoffman[5], Farhad Hormozdiari[19,1], Hae Kyung Im[2], Brian Jo[7,8], Silva Kasela[5,6], Seva Kashin[33,34,35], Sarah Kim-Hellmuth[5,6,9], Alan Kwong[26], Tuuli Lappalainen[5,6], Xiao Li[1], Yanyu Liang[2], Daniel G. MacArthur[34,36], Pejman Mohammadi[5,6,20,21], Stephen B. Montgomery[12,29], Manuel Muñoz-Aguirre[22,23], Meritxell Oliva[2,10], YoSon Park[24,25], Princy Parsana[11], John M. Rouhana[17,1], Ashis Saha[11], Ayellet V. Segrè[1,17], Matthew Stephens[37], Barbara E. Stranger[2,38], Benjamin J. Strober[18], Ellen Todres[1], Ana Viñuela[39,3,27,28], Gao Wang[37], Xiaoquan Wen[26], Valentin Wucher[22], Yuxin Zou[40]

Analysis Team Leaders†: François Aguet[1], Alexis Battle[18,11], Andrew Brown[3,4], Stephane E. Castel[5,6], Barbara E. Engelhardt[7,8], Farhad Hormozdiari[19,1], Hae Kyung Im[2], Sarah Kim-Hellmuth[5,6,9], Meritxell Oliva[2,10], Barbara E. Stranger[2,38], Xiaoquan Wen[26]

Downloaded from http://science.sciencemag.org/ on October 31, 2020

Senior Leadership†: Kristin G. Ardlie[1], Alexis Battle[18,11], Christopher D. Brown[24], Nancy Cox[16], Emmanouil T. Dermitzakis[3,27,28], Barbara E. Engelhardt[7,8], Gad A. Getz[1,30,31], Roderic Guigó[22,33], Hae Kyung Im[2], Tuuli Lappalainen[5,6], Stephen B. Montgomery[12,29], Barbara E. Stranger[2,38]
Manuscript Writing Group: François Aguet[1], Hae Kyung Im[2], Alexis Battle[18,11], Kristin G. Ardlie[1], Tuuli Lappalainen[5,6]

GTEx Consortium†
Laboratory and Data Analysis Coordinating Center (LDACC): François Aguet[1], Shankara Anand[1], Kristin G. Ardlie[1], Stacey Gabriel[1], Gad Getz[1,30,31], Aaron Graubert[1], Kane Hadley[1], Robert E. Handsaker[33,34,35], Katherine H. Huang[1], Seva Kashin[33,34,35], Xiao Li[1], Daniel G. MacArthur[34,36], Samuel R. Meier[1], Jared L. Nedzel[1], Duyen T. Nguyen[1], Ayellet V. Segrè[1,17], Ellen Todres[1]
Analysis Working Group Funded by GTEx Project Grants: François Aguet[1], Shankara Anand[1], Kristin G. Ardlie[1], Brunilda Balliu[41], Alvaro N. Barbeira[2], Alexis Battle[18,11], Rodrigo Bonazzola[2], Andrew Brown[3,4], Christopher D. Brown[24], Stephane E. Castel[5,6], Donald F. Conrad[42,43], Daniel J. Cotter[29], Nancy Cox[16], Sayantan Das[26], Olivia M. deGoede[29], Emmanouil T. Dermitzakis[3,27,28], Jonah Einson[44,5], Barbara E. Engelhardt[7,8], Eleazar Eskin[45], Tiffany Y. Eulalio[46], Nicole M. Ferraro[46], Elise D. Flynn[5,6], Laure Fresard[12], Eric R. Gamazon[13,14,15,16], Diego Garrido-Martín[22], Nicole R. Gay[29], Gad A. Getz[1,30,31], Michael J. Gloudemans[46], Aaron Graubert[1], Roderic Guigó[22,32], Kane Hadley[1], Andrew R. Hamel[17,1], Robert E. Handsaker[33,34,35], Yuan He[18], Paul J. Hoffman[5], Farhad Hormozdiari[19,1], Lei Hou[47,1], Katherine H. Huang[1], Hae Kyung Im[2], Brian Jo[7,8], Silva Kasela[5,6], Seva Kashin[33,34,35], Manolis Kellis[47,1], Sarah Kim-Hellmuth[5,6,9], Alan Kwong[26], Tuuli Lappalainen[5,6], Xiao Li[1], Xin Li[12], Yanyu Liang[2], Daniel G. MacArthur[34,36], Serghei Mangul[45,48], Samuel R. Meier[1], Pejman Mohammadi[5,6,20,21], Stephen B. Montgomery[12,29], Manuel Muñoz-Aguirre[22,23], Daniel C. Nachun[12], Jared L. Nedzel[1], Duyen T. Nguyen[1], Andrew B. Nobel[49], Meritxell Oliva[2,10], YoSon Park[24,25], Yongjin Park[47,1], Princy Parsana[11], Abhiram S. Rao[50], Ferran Reverter[51], John M. Rouhana[17,1], Chiara Sabatti[52], Ashis Saha[11], Ayellet V. Segrè[1,17], Andrew D. Skol[2,53], Matthew Stephens[37], Barbara E. Stranger[2,38], Benjamin J. Strober[18], Nicole A. Teran[12], Ellen Todres[1], Ana Viñuela[39,3,27,28], Gao Wang[37], Xiaoquan Wen[26], Fred Wright[54], Valentin Wucher[22], Yuxin Zou[40]
Analysis Working Group Not Funded by GTEx Project Grants: Pedro G. Ferreira[55,56,57,58], Gen Li[59], Marta Melé[60], Esti Yeger-Lotem[61,62]
Leidos Biomedical Project Management: Mary E. Barcus[63], Debra Bradbury[63], Tanya Krubit[63], Jeffrey A. McLean[63], Liqun Qi[63], Karna Robinson[63], Nancy V. Roche[63], Anna M. Smith[63], Leslie Sobin[63], David E. Tabor[63], Anita Undale[63]
Biospecimen Collection Source Sites: Jason Bridge[64], Lori E. Brigham[65], Barbara A. Foster[66], Bryan M. Gillard[66], Richard Hasz[67], Marcus Hunter[68], Christopher Johns[69], Mark Johnson[70], Ellen Karasik[66], Gene Kopen[71], William F. Leinweber[71], Alisa McDonald[71], Michael T. Moser[66], Kevin Myer[68], Kimberley D. Ramsey[66], Brian Roe[68], Saboor Shad[71], Jeffrey A. Thomas[71,70], Gary Walters[70], Michael Washington[70], Joseph Wheeler[69]
Biospecimen Core Resource: Scott D. Jewell[72], Daniel C. Rohrer[72], Dana R. Valley[72]
Brain Bank Repository: David A. Davis[73], Deborah C. Mash[73]
Pathology: Mary E. Barcus[63], Philip A. Branton[74], Leslie Sobin[63]
ELSI Study: Laura K. Barker[75], Heather M. Gardiner[75], Maghboeba Mosavel[76], Laura A. Siminoff[75]
Genome Browser Data Integration and Visualization: Paul Flicek[77], Maximilian Haeussler[78], Thomas Juettemann[77], W. James Kent[78], Christopher M. Lee[78], Conner C. Powell[78], Kate R. Rosenbloom[78], Magali Ruffier[77], Dan Sheppard[77], Kieron Taylor[77], Stephen J. Trevanion[77], Daniel R. Zerbino[77]
eGTEx Groups: Nathan S. Abell[29], Joshua Akey[79], Lin Chen[10], Kathryn Demanelis[10], Jennifer A. Doherty[80], Andrew P. Feinberg[81], Kasper D. Hansen[82], Peter F. Hickey[83], Lei Hou[47,1], Farzana Jasmine[10], Lihua Jiang[29], Rajinder Kaul[84,85], Manolis Kellis[47,1], Muhammad G. Kibriya[10], Jin Billy Li[29], Qin Li[29], Shin Lin[86], Sandra E. Linder[29], Stephen B. Montgomery[12,29], Meritxell Oliva[2,10], Yongjin Park[47,1], Brandon L. Pierce[10], Lindsay F. Rizzardi[87], Andrew D. Skol[2,53], Kevin S. Smith[12], Michael Snyder[29], John Stamatoyannopoulos[84,88], Barbara E. Stranger[2,38], Hua Tang[29], Meng Wang[29]
NIH Program Management: Philip A. Branton[74], Latarsha J. Carithers[74,89], Ping Guan[74], Susan E. Koester[90], A. Roger Little[91], Helen M. Moore[74], Concepcion R. Nierras[92], Abhi K. Rao[74], Jimmie B. Vaught[74], Simona Volpi[93]

[1]The Broad Institute of MIT and Harvard, Cambridge, MA, USA. [2]Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL, USA. [3]Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland. [4]Population Health and Genomics, University of Dundee, Dundee, Scotland, UK. [5]New York Genome Center, New York, NY, USA. [6]Department of Systems Biology, Columbia University, New York, NY, USA. [7]Department of Computer Science, Princeton University, Princeton, NJ, USA. [8]Center for Statistics and Machine Learning, Princeton University, Princeton, NJ, USA. [9]Statistical Genetics, Max Planck Institute of Psychiatry, Munich, Germany. [10]Department of Public Health Sciences, University of Chicago, Chicago, IL, USA. [11]Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA. [12]Department of Pathology, Stanford University, Stanford, CA, USA. [13]Data Science Institute, Vanderbilt University, Nashville, TN, USA. [14]Clare Hall, University of Cambridge, Cambridge, UK. [15]MRC Epidemiology Unit, University of Cambridge, Cambridge, UK. [16]Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. [17]Ocular Genomics Institute, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, USA. [18]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA. [19]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [20]Scripps Research Translational Institute, La Jolla, CA, USA. [21]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA. [22]Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Barcelona, Catalonia, Spain. [23]Department of Statistics and Operations Research, Universitat Politècnica de Catalunya (UPC), Barcelona, Catalonia, Spain. [24]Department of Genetics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA. [25]Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA. [26]Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA. [27]Institute for Genetics and Genomics in Geneva (iGE3), University of Geneva, Geneva, Switzerland. [28]Swiss Institute of Bioinformatics, Geneva, Switzerland. [29]Department of Genetics, Stanford University, Stanford, CA, USA. [30]Cancer Center and Department of Pathology, Massachusetts General Hospital, Boston, MA, USA. [31]Harvard Medical School, Boston, MA, USA. [32]Universitat Pompeu Fabra (UPF), Barcelona, Catalonia, Spain. [33]Department of Genetics, Harvard Medical School, Boston, MA, USA. [34]Program in Medical and Population Genetics, The Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA. [35]Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA. [36]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [37]Department of Human Genetics, University of Chicago, Chicago, IL, USA. [38]Center for Genetic Medicine, Department of Pharmacology, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA. [39]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [40]Department of Statistics, University of Chicago, Chicago, IL, USA. [41]Department of Biomathematics, University of California, Los Angeles, Los Angeles, CA, USA. [42]Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA. [43]Division of Genetics, Oregon National Primate Research Center, Oregon Health & Science University, Portland, OR, USA. [44]Department of Biomedical Informatics, Columbia University, New York, NY, USA. [45]Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, USA.[46]Program in Biomedical Informatics, Stanford University School of Medicine, Stanford, CA, USA. [47]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA. [48]Department of Clinical Pharmacy, School of Pharmacy, University of Southern California, Los Angeles, CA, USA. [49]Department of Statistics and Operations Research and Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA. [50]Department of Bioengineering, Stanford University, Stanford, CA, USA. [51]Department of Genetics, Microbiology and Statistics, University of Barcelona, Barcelona, Spain. [52]Departments of Biomedical Data Science and Statistics, Stanford University, Stanford, CA, USA. [53]Department of Pathology and Laboratory Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA. [54]Bioinformatics Research Center and Departments of Statistics and Biological Sciences, North Carolina State University, Raleigh, NC, USA. [55]Department of Computer Sciences, Faculty of Sciences, University of Porto, Porto, Portugal. [56]Instituto de Investigação e Inovação em Saúde, University of Porto, Porto, Portugal. [57]Institute of Molecular Pathology and Immunology, University of Porto, Porto, Portugal. [58]Laboratory of Artificial Intelligence and Decision Support, Institute for Systems and Computer Engineering, Technology and Science, Porto, Portugal. [59]Columbia University Mailman School of Public Health, New York, NY, USA. [60]Life Sciences Department, Barcelona Supercomputing Center, Barcelona, Spain. [61]Department of Clinical Biochemistry and Pharmacology, Ben-Gurion University of the Negev, Beer-Sheva, Israel.[62]National Institute for Biotechnology in the Negev, Beer-Sheva, Israel. [63]Leidos Biomedical, Rockville, MD, USA. [64]UNYTS, Buffalo, NY, USA. [65]Washington Regional Transplant Community, Annandale, VA, USA. [66]Therapeutics, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA. [67]Gift of Life Donor Program, Philadelphia, PA, USA. [68]Life Gift, Houston, TX, USA. [69]Center for Organ Recovery and Education, Pittsburgh, PA, USA. [70]LifeNet Health, Virginia Beach, VA, USA. [71]National Disease Research Interchange, Philadelphia, PA, USA. [72]Van Andel Research Institute, Grand Rapids, MI, USA. [73]Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA. [74]Biorepositories and Biospecimen Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD, USA. [75]College of Public Health, Temple University, Philadelphia, PA, USA. [76]Virginia Commonwealth University, Richmond, VA, USA. [77]European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK. [78]Genomics Institute, University of California Santa Cruz, Santa Cruz, CA, USA. [79]Carl Icahn Laboratory, Princeton University, Princeton, NJ, USA. [80]Department of Population Health Sciences, The University of Utah, Salt Lake City, UT, USA. [81]Departments of Medicine, Biomedical Engineering, and Mental Health, Johns Hopkins University, Baltimore, MD, USA. [82]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA. [83]Department of Medical Biology, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia. [84]Altius Institute for Biomedical Sciences, Seattle, WA, USA. [85]Division of Genetics, University of Washington, Seattle, WA, USA. [86]Department of Cardiology, University of Washington, Seattle, WA, USA. [87]Hudson Alpha Institute for Biotechnology, Huntsville, AL, USA. [88]Genome Sciences, University of Washington, Seattle, WA, USA. [89]National Institute of Dental and Craniofacial Research, Bethesda, MD, USA. [90]Division of Neuroscience and Basic Behavioral Science, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA. [91]National Institute on Drug Abuse, Bethesda, MD, USA. [92]Office of Strategic Coordination, Division of Program Coordination, Planning and Strategic Initiatives, Office of the Director, National Institutes of Health, Rockville, MD, USA. [93]Division of Genomic Medicine, National Human Genome Research Institute, Bethesda, MD, USA.

†Alphabetical order
‡First author

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6509/1318/suppl/DC1
Supplementary Text
Figs. S1 to S52
Tables S1 to S16
References (*60–129*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

30 September 2019; accepted 30 July 2020
10.1126/science.aaz1776

Downloaded from http://science.sciencemag.org/ on October 31, 2020

Science

The GTEx Consortium atlas of genetic regulatory effects across human tissues

The GTEx Consortium

Science **369** (6509), 1318-1330.
DOI: 10.1126/science.aaz1776

A survey of transcription across tissues

Some human genetic variants affect the amount of RNA produced and the splicing of gene transcripts, crucial steps in development and maintaining a healthy individual. However, some of these changes only occur in a small number of tissues within the body. The Genotype-Tissue Expression (GTEx) project has been expanded over time, and, looking at the final data in version 8, Aguet *et al.* present a deep characterization of genetic associations and gene expression and splicing in 838 individuals over 49 tissues (see the Perspective by Wilson). This large study was able to characterize the details underlying many aspects of gene expression and provides a resource with which to better understand the fundamental molecular mechanisms of how genetic variants affect gene regulation and complex traits in humans.

Science, this issue p. 1318; see also p. 1298

ARTICLE TOOLS http://science.sciencemag.org/content/369/6509/1318

SUPPLEMENTARY MATERIALS http://science.sciencemag.org/content/suppl/2020/09/09/369.6509.1318.DC1

RELATED CONTENT
http://science.sciencemag.org/content/sci/369/6509/1316.full
http://science.sciencemag.org/content/sci/369/6509/eaba3066.full
http://science.sciencemag.org/content/sci/369/6509/eaaz8528.full
http://science.sciencemag.org/content/sci/369/6509/eaaz5900.full
http://science.sciencemag.org/content/sci/369/6509/eaaz6876.full

REFERENCES This article cites 126 articles, 13 of which you can access for free
http://science.sciencemag.org/content/369/6509/1318#BIBL

PERMISSIONS http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2020 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works

Downloaded from http://science.sciencemag.org/ on October 31, 2020