Cell type-specific genetic regulation of gene expression across human tissues

Kim-Hellmuth, Sarah; Aguet, François; Oliva, Meritxell; Muñoz-Aguirre, Manuel; Kasela, Silva; Wucher, Valentin

Published in:
Science (New York, N.Y.)

DOI:
10.1126/science.aaz8528

Publication date:
2020

Document Version
Peer reviewed version

Cell type specific genetic regulation of gene expression across human tissues

Authors: Sarah Kim-Hellmuth[1,2,3, 28†*], François Aguet[4*], Meritxell Oliva[5,6], Manuel Muñoz-Aguirre[7,8], Silva Kasela[2,3], Valentin Wucher[7], Stephane E. Castel[2,3], Andrew R. Hamel[4,9], Ana Viñuela[10,11,12,13], Amy L. Roberts[10], Serghei Mangul[14,15], Xiaoquan Wen[16], Gao Wang[17], Alvaro N. Barbeira[5], Diego Garrido-Martín[7], Brian Nadel[18], Yuxin Zou[19], Rodrigo Bonazzola[5], Jie Quan[20], Andrew Brown[11,21], Angel Martinez-Perez[22], José Manuel Soria[22], GTEx Consortium, Gad Getz[4,23,24], Emmanouil T. Dermitzakis[11,12,13], Kerrin S. Small[10], Matthew Stephens[17], Hualin S. Xi[25], Hae Kyung Im[5], Roderic Guigó[7,26], Ayellet V. Segrè[4,9], Barbara E. Stranger[5,27], Kristin G. Ardlie[4], Tuuli Lappalainen[2,3†]

Affiliations:
[1] Statistical Genetics, Max Planck Institute of Psychiatry, Munich, Germany
[2] New York Genome Center, New York, NY, USA
[3] Department of Systems Biology, Columbia University, New York, NY, USA
[4] The Broad Institute of MIT and Harvard, Cambridge, MA, USA
[5] Section of Genetic Medicine, Department of Medicine, The University of Chicago, Chicago, IL, USA
[6] Department of Public Health Sciences, The University of Chicago, Chicago, IL, USA
[7] Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Barcelona, Catalonia, Spain
[8] Department of Statistics and Operations Research, Universitat Politècnica de Catalunya (UPC), Barcelona, Catalonia, Spain
[9] Ocular Genomics Institute, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, USA
[10] Department of Twin Research and Genetic Epidemiology, King's College London, London, UK
[11] Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland
[12] Institute for Genetics and Genomics in Geneva (iGE3), University of Geneva, Geneva, Switzerland
[13] Swiss Institute of Bioinformatics, Geneva, Switzerland
[14] Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, USA
[15] Department of Clinical Pharmacy, School of Pharmacy, University of Southern California, Los Angeles, CA, USA
[16] Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA
[17] Department of Human Genetics, University of Chicago, Chicago, IL, USA
[18] Department of Molecular, Cellular, and Developmental Biology, University of California, Los Angeles, Los Angeles, CA, USA
[19] Department of Statistics, University of Chicago, Chicago, IL, USA
[20] Inflammation & Immunology, Pfizer, Cambridge, MA, USA.
[21] Population Health and Genomics, University of Dundee, Dundee, Scotland, UK
[22] Unit of Genomic of Complex Diseases, Institut d'Investigació Biomèdica Sant Pau (IIB-Sant Pau), Barcelona. Spain
[23] Cancer Center and Department of Pathology, Massachusetts General Hospital, Boston, MA, USA
[24] Harvard Medical School, Boston, MA, USA
[25] Foundational Neuroscience Center, AbbVie, Cambridge, MA, USA
[26] Universitat Pompeu Fabra (UPF), Barcelona, Catalonia, Spain

[27] Center for Genetic Medicine, Department of Pharmacology, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA
[28] current address: Department of Pediatrics, Dr. von Hauner Children's Hospital, University Hospital LMU Munich, Munich, Germany
*These authors contributed equally to this work.
†Correspondence to: skimhellmuth@gmail.com, tlappalainen@nygenome.org

One Sentence Summary:
Estimated cell type abundances from bulk RNA-seq across tissues reveal the cellular specificity of quantitative trait loci.

Abstract:
The Genotype-Tissue Expression (GTEx) project has identified expression and splicing quantitative trait loci in *cis* (QTLs) for the majority of genes across a wide range of human tissues. However, the functional characterization of these QTLs has been limited by the heterogeneous cellular composition of GTEx tissue samples. Here, we map interactions between computational estimates of cell type abundance and genotype to identify cell type interaction QTLs for seven cell types and show that cell type interaction eQTLs provide finer resolution to tissue specificity than bulk tissue *cis*-eQTLs. Analyses of genetic associations with 87 complex traits show a contribution from cell type interaction QTLs and enables the discovery of hundreds of previously unidentified colocalized loci that are masked in bulk tissue.

Main Text:
The Genotype-Tissue Expression (GTEx) project (*1*) and other studies (*2-5*) have shown that genetic regulation of the transcriptome is widespread. The GTEx Consortium, in particular, has built an extensive catalog of expression and splicing quantitative trait loci in *cis* (*cis*-eQTLs and *cis*-sQTLs) across a large range of tissues, showing that these *cis*-eQTLs and *cis*-sQTLs (collectively referred to here as QTLs) are generally either highly tissue-specific or widely shared, even across dissimilar tissues and organs (*1*, *6*). However, the majority of these studies have been performed using heterogeneous bulk tissue samples comprising diverse cell types. This limits the power, interpretation, and downstream applications of QTL studies. Genetic effects that are active only in rare cell types within a sampled tissue may be undetected, a mechanistic interpretation of QTL sharing across tissues and other contexts is complicated without understanding differences in cell type composition, and inference of downstream molecular effects of regulatory variants without the specific cell type context is challenging. Efforts to map eQTLs in individual cell types have been largely restricted to blood, using purified cell types (*7-11*) or single cell sequencing (*12*).

While there are many ongoing efforts to optimize single cell and single nucleus sequencing of human tissues (*13*, *14*), including as part of the Human Cell Atlas (*15*), these methods are not yet scalable to sample sizes and coverage sufficient to achieve comparable power to bulk eQTL studies (*16-18*). However, cell type specific eQTLs can be computationally inferred from bulk tissue measurements, using estimated proportions or enrichments of relevant cell types to test for interactions with genotype. To date, such approaches have only been applied to a limited

range of cell types such as blood cells (*19*, *20*) and adipocytes (*21*). These studies identified thousands of cell type interactions in eQTLs discovered in whole blood samples from large cohorts (5,683 samples (*19*); 2,116 samples, (*20*)), indicating that large numbers of interactions are likely to be identified by expanding this type of analysis to other tissues and cell types.

Identifying cell types *in silico* in bulk tissue

Here, we used computational estimates of cell type enrichment to characterize the cell type specificity of *cis*-eQTLs and *cis*-sQTLs for 43 cell type-tissue combinations, using seven cell types across 35 tissues (Fig. 1A). Estimating the cell type composition of a tissue biospecimen from RNA-seq remains a challenging problem (*22*), and multiple approaches for inferring cell type proportions have been proposed (*23*). We performed extensive benchmarking for multiple cell types across several expression datasets (figs. S1 and S2). The xCell method (*24*), which estimates the enrichment of 64 cell types using reference profiles, was most suitable on the combined basis of correlation with cell counts in blood (fig. S1A), *in silico* simulations (fig. S1B), correlation with expression of marker genes for each cell type (fig. S1C,D), and diversity of reference cell types. Concordance between methods was generally high (fig. S1A and E). Furthermore, the inferred abundances reflected differences in histology (fig. S1C) and tissue pathologies (fig. S2). For each cell type, we selected tissues where the cell type was highly enriched (fig. S3). The xCell scores for these tissue-cell type pairs were highly correlated with the PEER factors used to correct for unobserved confounders in the expression data for QTL mapping (*1*) (fig. S4A), but were generally weakly correlated with known technical confounders (fig. S4B), suggesting that cell type composition accounts for a large fraction of inter-sample variation in gene expression.

Mapping cell type interaction eQTLs and sQTLs

To identify *cis*-eQTLs and *cis*-sQTLs whose effect varies depending on the enrichment of the cell type, we leveraged the variability in cell type composition across GTEx samples to test for an interaction between cell type and genotype using a linear regression model for either gene expression or splicing (*25*) (Figs. 1B,C and S5A,B). Since QTLs identified this way are not necessarily specific to the estimated cell type but may reflect another (anti)correlated cell type, we refer to these eQTLs and sQTLs as cell type interaction eQTLs (ieQTLs) and cell type interaction sQTLs (isQTLs), respectively (or iQTLs in aggregate).

Across cell types and tissues, we detected 3,347 protein coding and lincRNA genes with an ieQTL (ieGenes(*26*)) and 987 genes with an isQTL (isGenes) at 5% FDR per cell type-tissue combination (Figs. 2A, S5C, S6, and table S1). In the following analyses, we use ieQTLs and isQTLs identified with 5% FDR unless indicated otherwise. Notably, while 85% of ieQTLs corresponded to genes with at least one standard *cis*-eQTL (eGenes; we refer to *cis*-eQTLs mapped in bulk tissue as standard eQTLs for simplicity (*26*)) 21% of these ieQTLs were not in LD ($R^2 < 0.2$) with any of the corresponding eGene's conditionally independent eQTLs (*1*) (fig. S7, A and B). For comparison, the proportion of genes with at least one standard eQTL varies as a function of sample size (*1*), with a median of 42% across tissues (48% in transverse colon and

63% in whole blood). This indicates that ieQTL analysis frequently reveals genetic regulatory effects that are not detected by standard eQTL analysis of heterogeneous tissue samples. Unlike standard *cis*-QTL discovery, iQTL discovery was only modestly correlated with sample size (Spearman's ρ = 0.53 and 0.35, for ieQTLs and isQTLs, respectively; fig. S7C,D). The tissues with most iQTLs included blood, as well as transverse colon and breast, which both stratified into at least two distinct groups on the basis of histology (*27*): epithelial vs. adipose tissue (breast) and mucosal vs. muscular tissue (colon) (fig. S1C). This suggests that inter-individual variance (which partially reflects variation in biospecimen collection) in cell type enrichment driven by tissue heterogeneity is a major determinant in discovery power, and benefits iQTL mapping despite being a potential confounding factor for other types of gene expression analyses. Downsampling analyses in whole blood and transverse colon revealed linear relationships between sample size and ieQTL discovery in these tissues, suggesting that significantly larger numbers of ieQTLs may be discovered with larger sample sizes (fig. S7E). ieQTL discovery was largely robust to the choice of deconvolution method, with ~77% of neutrophil ieQTLs detected with xCell also detected using CIBERSORT, and close to complete replication ($\pi1 > 0.99$; fig. S7F).

The QTL effect of ieQTLs and isQTLs can increase or decrease as a function of cell type enrichment (Fig. 1C and fig. S8A). This correlation is usually positive (56%; median across cell type-tissue combinations). As an example, a keratinocyte ieQTL for *CNTN1* in skin had a stronger effect in samples with high enrichment of keratinocytes. However, for a significant number of ieQTLs the effect was negatively correlated (19%) suggesting that the interaction we identified likely captures an eQTL that is only active in at least one other cell type (fig. S8B). For 24% of ieQTLs the correlation was ambiguous. At a more stringent FDR cutoff (FDR < 0.01), the median proportion of ieQTLs with ambiguous cell type correlation decreased to 11% (fig. S8B, right panel), while the proportion of ieQTLs with positive correlation increased to 77%. Moreover, the ieQTLs with ambiguous direction tended to have lower MAF (fig. S8C), suggesting that at less stringent FDR this category might be enriched for false positives.

Altogether, we identified numerous cell type ieQTLs and isQTLs across 43 cell type-tissue combinations, including iQTLs that are not detected by standard eQTLs analysis in bulk tissue. These cell type iQTLs pinpoint the cellular specificity of QTLs that might not necessarily be specific to the tested cell type, but may also capture eQTL effects of (anti)correlated cell types.

Validation and replication of cell type iQTLs

Since few external replication datasets exist, we used allele-specific expression (ASE) data of eQTL heterozygotes (*28*, *29*) to correlate individual-level quantifications of the eQTL effect size (measured as allelic fold-change, aFC) with individual-level cell type enrichments. If the eQTL is active in the cell type of interest, we expect to see low aFC in individuals with low cell type abundance, and higher aFC in individuals with high cell type abundance (fig. S9). The correlation between cell type abundance and aFC across heterozygous individuals can thus be used as a measure of validation for a specific ieQTL.

Using this approach, the median proportion of ieQTLs with a significant ($P < 0.05$) aFC-cell type Pearson correlation was 0.62 (Fig. 2B). For 13 cell type-tissue combinations with > 20 significant ieQTLs, the corresponding π_1 statistic (the proportion of true positives, (*30*)) confirmed the high validation rate (mean $\pi_1 = 0.75$, fig. S10). While this approach does not constitute formal replication in an independent cohort, it is applicable to all tested cell type-tissue combinations and corroborates that ieQTLs are not statistical artefacts of the interaction model.

Next, we performed replication analyses in external cohorts, including whole blood from the GAIT2 study (*31*), purified neutrophils (*9*), adipose and skin tissues from the TwinsUK study for ieQTLs (*5*) and temporal cortex from the Mayo RNA sequencing study for both ieQTLs and isQTLs (*32*). Replication rates ranged from $\pi_1 = 0.32 – 0.67$, with the highest rate observed in purified neutrophils for whole blood (fig. S11). The differences in replication rate likely reflect a combination of lower power to detect cell type ieQTLs/isQTLs compared to standard eQTLs/sQTLs, as well as differences in tissue heterogeneity across studies. Taken together, these results show that ieQTLs and isQTLs can be detected with reasonable robustness for diverse cell types and tissues.

Cell type ieQTLs contribute to tissue specificity

Next, we sought to determine to what extent cell type ieQTLs contribute to the tissue specificity of *cis*-eQTLs. First, we analyzed ieQTL sharing across cell types, observing that ieQTLs for one cell type were generally not ieQTLs for other cell types (e.g. myocyte ieQTLs in muscle tissues were not hepatocyte ieQTLs in liver, etc.; fig. S12A). To determine if a significant cell type interaction effect is associated with the tissue specificity of an eQTL, we tested whether cell type ieQTLs are predictors of tissue sharing. We annotated the top *cis*-eQTLs per gene across tissues with their cell type ieQTL status for the five cell types with at least 20 ieQTLs (adipocytes, epithelial cells, keratinocytes, myocytes, and neutrophils). This annotation was included as a predictor in a logistic regression model of eQTL tissue sharing on the basis of eQTL properties including effect size, minor allele frequency, eGene expression correlation, genomic annotations, and chromatin state (*1*). In all five cell types, ieQTL status was a strong negative predictor of tissue sharing, with the magnitude of the effect similar to that of enhancers, indicating that ieQTLs are an important mechanism for tissue-specific regulation of gene expression (Figs. 3A, S12B). Testing whether cell type isQTLs are predictors of tissue sharing for four cell types with at least 20 isQTLs (adipocytes, epithelial cells, myocytes, and neutrophils) revealed only neutrophil isQTL status as a significant negative predictor (fig. S13). This is likely due to a combination of lower power to detect isQTLs and higher likelihood of splicing-affecting variants having shared effects if a gene is expressed in a tissue or cell type (*1*).

We corroborated the finding for ieQTLs using multi-tissue eQTL mapping with MASH (*1*), testing whether eGenes that are tissue-specific (eQTLs discovered at LSFR < 0.05 only in the tissue type of interest) have a higher proportion of cell type ieQTLs compared to eGenes that are shared across tissues (LSFR < 0.05 in multiple tissues). Indeed, the proportion of cell type ieQTLs across all 43 cell type-tissue combinations was significantly higher in tissue-specific eGenes compared to tissue-shared eGenes ($P = 1.9 \times 10^{-05}$, one-sided Wilcoxon rank sum test, Fig. 3B) further

highlighting the contribution of cell type-specific genetic gene regulation to tissue specificity of eQTLs. For tissues with notably high inter-sample heterogeneity (e.g. breast, transverse colon, and stomach), the above-average enrichment is likely at least partially driven by higher power to detect ieQTLs.

To examine the sharing patterns of cell type ieQTLs across tissues we used two cell types with ieQTLs mapped in >10 tissues (16 tissues for epithelial cells and 13 for neurons). We observed that while standard eQTLs were highly shared across the subsets of 16 and 13 tissues, cell type ieQTLs tended to be highly tissue-specific, reflected by an average of four and five tissues with shared ieQTL effects compared to 11 and 12 for eQTLs in epithelial and brain tissues respectively (Fig. 3C,D, left panels). These findings were robust to power differences in detecting eQTLs vs. ieQTLs, with eQTLs remaining predominantly shared even when limited to 20% of samples (fig. S14). 25.3% of neuron ieQTLs were shared between nine brain tissues, highlighting that tissues of the cerebrum (e.g. cortex, basal ganglia, limbic system) show particularly high levels of sharing compared to cerebellar tissues, the hypothalamus, and the spinal cord (Fig. 3D, left panel). This pattern was absent when analyzing standard eQTLs. Pairwise tissue sharing comparisons further confirmed that cell type ieQTLs showed greater tissue specificity and more diverse tissue sharing patterns than standard eQTLs, which were broadly shared across all tissues (Fig. 3C and D, middle and right panels). These results show that incorporating cell type composition is essential for characterizing the sharing of genetic regulatory effects across tissues.

GWAS and tissue-specific eQTLs and sQTLs

To study the contribution of cell type interaction QTLs to genome-wide association study (GWAS) results for 87 complex traits, we first examined the enrichment of iQTLs of each cell type/tissue combination for trait associations (GWAS $P \leq 0.05$) using QTLEnrich (*33*). We used 23 and 7 cell type/tissue pairs (19 and 7 unique tissues) with >100 ieQTLs or isQTLs, respectively, at a relaxed FDR of 40% to generate robust enrichment estimates of 87 GWAS traits. Across all tested cell type/tissue-trait pairs, the GWAS signal was clearly enriched among ieQTLs and isQTLs (1.3 and 1.4 median fold-enrichments, respectively), similarly to standard eQTLs and sQTLs (Fig. 4A, table S4). The GWAS enrichments were robust to the iQTL FDR cutoffs (fig. S15A and B).

We next analyzed the enrichments of the individual traits for iQTLs of two cell types that we estimated had the largest number of ieQTLs: neutrophil iQTLs in blood and epithelial cell iQTLs in transverse colon. We compared them to the corresponding standard QTLs (Fig. 4B, fig. S15C and D), focusing on traits that had a significant enrichment for either QTL type (Bonferroni-adjusted $P < 0.05$). Interestingly, in blood we observed a significant shift towards higher enrichment for ieQTLs (one-sided, paired Wilcoxon rank sum test; $P = 0.0026$) and especially isQTLs ($P = 2.8 \times 10^{-05}$), which appears to be driven by GWAS for blood cell traits, and also immune traits having a higher enrichment for iQTLs. The higher iQTL signal is absent in colon (ieQTL $P = 1$ and isQTL $P = 0.13$), even though the standard QTL enrichment for blood cell traits appear similar for blood and colon. This pattern suggests that cell type interaction QTLs may have better resolution for indicating relevant tissues and cell types for complex traits compared to tissue QTLs, but further studies are needed to fully test this hypothesis.

Next, we asked whether cell type iQTLs can be linked to loci discovered in genome-wide association studies (GWAS) and used to pinpoint their cellular specificity. To this end, we tested 13,702 ieGenes and 2,938 isGenes (40% FDR) for colocalization with 87 GWAS traits (*1*), using both the cell type ieQTL/isQTL and corresponding standard QTL. 1,370 (10.3%) cell type ieQTLs and 89 (3.7%) isQTLs colocalized with at least one GWAS trait (Figs. 5A and B, and tables S5 and S6). The larger number of colocalizations identified for neutrophil ieQTLs and isQTLs in whole blood relative to other cell type-tissue pairs likely reflects a combination of the larger number of ieQTLs and isQTLs and the abundance of significant GWAS loci for blood-related traits in our set of 87 GWASs (Fig. 5B).

Our analysis revealed a substantial proportion of loci for which only the ieQTL/isQTL colocalizes with the trait (467/1370, 34%; Fig. 5B), or where the joint colocalization of the ieQTL/isQTL and corresponding standard eQTL indicates the cellular specificity of the trait as well as its potential cellular origin (401/1370, 29%; Fig. 5B). For example, a colocalization between the *DHX58* gene in the left ventricle of the heart and an asthma GWAS was only identified through the corresponding myocyte ieQTL (PP4 = 0.64), but not the standard eQTL (PP4 = 0.00; Fig. 5C). Cardiac cells such as cardiomyocytes are not primarily viewed to have a causal role in asthma, but their presence along pulmonary veins and their potential contribution to allergic airway disease have been described (*34*).

An example where both the standard eQTL and the cell type ieQTL colocalize with the trait is given in Fig. 5C for *KREMEN1* in adipocytes in subcutaneous adipose tissue and a birth weight GWAS (PP4 ~0.8); *KREMEN1* has been linked to adipogenesis in mice (*35*). We highlight two analogous examples for isQTLs: the epithelial cell isQTL for *CDHR5* in small intestine colocalized with eosinophil counts whereas the standard sQTL did not (Fig. 5D), and conversely, both the standard sQTL and myocyte isQTL for *ATP5SL* in the left ventricle of the heart colocalized with standing height (Fig. 5D). Additional examples of ieQTLs and isQTLs colocalizing with trait associations are provided in figs. S16 and S17. While the iQTLs do not necessarily pinpoint the specific cell type where the regulatory effect is active, they indicate that cell type specificity plays a role in the GWAS locus. Together, our colocalization results indicate that cell type interaction QTLs yield new potential target genes for GWAS loci that are missed by standard QTLs and provide hypotheses for the cellular specificity of regulatory effects underlying complex traits.

Discussion

By mapping interaction effects between cell type enrichment and genotype on the transcriptome across GTEx tissues, we provide an atlas of thousands of eQTLs and sQTLs that are likely to be cell type specific. Notably, the ieQTLs and isQTLs we report here include several immune and stromal cell types in tissues where cell type specific QTLs have not been characterized in prior studies. Cell type ieQTLs are strongly enriched for tissue- and cellular specificity and provide a finer resolution to tissue-specificity than bulk *cis*-QTLs that are highly shared between tissues. Given the enrichment of GWAS signal in cell type iQTLs for cell types potentially relevant to the traits, and the large fraction of colocalizations with GWAS traits that are only found with cell type

iQTLs, exhaustive characterization of cell type specific QTLs is a highly promising approach towards a mechanistic understanding of these loci, complementing experimental assays of variant function. However, the substantial allelic heterogeneity observed in standard QTLs (*1*) and limited power to deconvolve QTLs that are specific to rare cell types or with weak or opposing effects indicate that many more cell type specific QTLs exist beyond those that can be currently computationally inferred from bulk tissue data. We therefore anticipate that upcoming population-scale single-cell QTL studies will be essential to complement the approaches presented here. However, as those data are still difficult to obtain for many tissues, our demonstration of the insights gained from cell type iQTLs indicates that improving deconvolution approaches and increasing sample sizes will be valuable in this effort and enable discoveries for cell types and tissues not considered in this study.

Methods summary

The GTEx V8 data (*1*) was used for all analyses. Cell type enrichments were computed using xCell (*24*). Interaction QTL mapping was performed using tensorQTL (*36*). Full methods are available in (*26*).

References:

1. GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *bioRxiv*. **7**, 1860–1832 (2019).

2. U. Võsa *et al.*, Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. *bioRxiv*, 1–57 (2018).

3. R. Joehanes *et al.*, Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol.* **18**, 16 (2017).

4. H. Kirsten *et al.*, Dissecting the genetics of the human transcriptome identifies novel trait-related trans-eQTLs and corroborates the regulatory relevance of non-protein coding loci†. *Human Molecular Genetics*. **24**, 4746–4763 (2015).

5. A. Buil *et al.*, Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat. Genet.* **47**, 88–91 (2014).

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

7. B. P. Fairfax *et al.*, Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat. Genet.* **44**, 502–510 (2012).

8. T. Raj *et al.*, Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science*. **344**, 519–523 (2014).

9. V. Naranbhai *et al.*, Genomic modulators of gene expression in human neutrophils. *Nat Commun*. **6**, 7545 (2015).

10. S. Kim-Hellmuth *et al.*, Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat Commun*. **8**, 266 (2017).

11. S. Kasela *et al.*, Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells. *PLoS Genet*. **13**, e1006643–21 (2017).

12. M. G. P. van der Wijst *et al.*, Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat. Genet.* **50**, 493–497 (2018).

13. N. Habib *et al.*, Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Meth*. **14**, 955–958 (2017).

14. M. Slyper *et al.*, A single-cell and single-nucleus RNA-seq toolbox for fresh and frozen human tumors. *bioRxiv*, 1–121 (2019).

15. HCA Consortium, The Human Cell Atlas. *arXiv*, 1–109 (2018).

16. A. K. Sarkar *et al.*, Discovery and characterization of variance QTLs in human induced pluripotent stem cells. *PLoS Genet*. **15**, e1008045–16 (2019).

17. I. Mandric *et al.*, Optimal design of single-cell RNA sequencing experiments for cell-type-specific eQTL analysis. **4**, 85–25 (2019).

18. M. G. P. van der Wijst *et al.*, Single-cell eQTLGen Consortium: a personalized understanding of disease. *arXiv*, 1–26 (2019).

19. H.-J. Westra *et al.*, Cell Specific eQTL Analysis without Sorting Cells. *PLoS Genet*. **11**, e1005223–17 (2015).

20. D. V. Zhernakova *et al.*, Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139–145 (2016).

21. C. A. Glastonbury, A. Couto Alves, J. S. El-Sayed Moustafa, K. S. Small, Cell-Type Heterogeneity in Adipose Tissue Is Associated with Complex Traits and Reveals Disease-Relevant Cell-Specific eQTLs. *Am. J. Hum. Genet.* **104**, 1013–1024 (2019).

22. F. Avila Cobos, J. Vandesompele, P. Mestdagh, K. De Preter, Computational deconvolution of transcriptomics data from mixed cell populations. *Bioinformatics*. **34**, 1969–1979 (2018).

23. G. Sturm *et al.*, Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics*. **35**, i436–i445 (2019).

24. D. Aran, Z. Hu, A. J. Butte, xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).

25. Y. I. Li *et al.*, Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**, 151–158 (2018).

26. See supplementary materials.

27. A. Breschi *et al.*, A limited set of transcriptional programs define major histological types and provide the molecular basis for a cellular taxonomy of the human body. *bioRxiv*. **9**, 75–25 (2019).

28. S. E. Castel *et al.*, A vast resource of allelic expression data spanning human tissues. *bioRxiv*. **7**, 12817–21 (2019).

29. P. Mohammadi, S. E. Castel, A. A. Brown, T. Lappalainen, Quantifying the regulatory effect size of cis-acting genetic variation using allelic fold change. *Genome Res.* **27**, 1872–1884 (2017).

30. J. D. Storey, R. Tibshirani, Statistical significance for genomewide studies. *Proc Natl Acad Sci USA*. **100**, 9440–9445 (2003).

31. L. Vila *et al.*, Heritability of thromboxane A2 and prostaglandin E2 biosynthetic machinery in a Spanish population. *Arterioscler. Thromb. Vasc. Biol.* **30**, 128–134 (2010).

32. M. Allen *et al.*, Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci Data*. **3**, 160089 (2016).

33. E. R. Gamazon *et al.*, Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat. Genet.* **50**, 1–18 (2018).

34. S. S. Folmsbee, C. J. Gottardi, Cardiomyocytes of the Heart and Pulmonary Veins: Novel Contributors to Asthma? *Am. J. Respir. Cell Mol. Biol.* **57**, 512–518 (2017).

35. C. Christodoulides *et al.*, The Wnt antagonist Dickkopf-1 and its receptors are coordinately regulated during early human adipogenesis. *J. Cell. Sci.* **119**, 2613–2620 (2006).

36. A. Taylor-Weiner *et al.*, Scaling computational genomics to millions of individuals with GPUs. *Genome Biol.* **20**, 228–5 (2019).

37. J. Ernst, M. Kellis, Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc*. **12**, 2478–2492 (2017).

38. A. M. Newman *et al.*, Robust enumeration of cell subsets from tissue expression profiles. *Nat Meth*. **12**, 453–457 (2015).

39. Y. Zhang *et al.*, Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron*. **89**, 37–53 (2016).

40. B. Nadel *et al.*, The Gene Expression Deconvolution Interactive Tool (GEDIT): Accurate Cell Type Quantification from Gene Expression Data. *bioRxiv*. **14**, 395–30 (2019).

41. A. I. Su *et al.*, A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA*. **101**, 6062–6067 (2004).

42. W. R. Swindell, A. Johnston, J. J. Voorhees, J. T. Elder, J. E. Gudjonsson, Dissecting the psoriasis transcriptome: inflammatory- and cytokine-driven gene expression in lesions from 163 patients. *BMC Genomics*. **14**, 527–20 (2013).

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

44. H. E. Wheeler *et al.*, Survey of the Heritability and Sparse Architecture of Gene Expression Traits across Human Tissues. *PLoS Genet*. **12**, e1006423–23 (2016).

45. X. Zheng *et al.*, A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*. **28**, 3326–3328 (2012).

46. S. M. Gogarten *et al.*, Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics*. **35**, 5346–5348 (2019).

47. M. Uhlén *et al.*, Proteomics. Tissue-based map of the human proteome. *Science*. **347**, 1260419–1260419 (2015).

48. G. Monaco *et al.*, RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Reports*. **26**, 1627–1640.e7 (2019).

49. V. Garcia Palacios, I. Morita, S. Murota, Expression of adipogenesis markers in a murine stromal cell line treated with 15-deoxy Delta(12,14)-prostaglandin J2, interleukin-11, 9-cis retinoic acid and vitamin K2. *Prostaglandins Leukot. Essent. Fatty Acids*. **65**, 215–221 (2001).

50. Y. Xu *et al.*, Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight*. **1**, e90558 (2016).

51. J. Y. Zheng *et al.*, Regulation of the expression of the prostate-specific antigen by claudin-7. *J. Membr. Biol.* **194**, 187–197 (2003).

52. W. D. Kuhlmann, P. Peschke, Hepatic progenitor cells, stem cells, and AFP expression in models of liver injury. *Int J Exp Pathol*. **87**, 343–359 (2006).

53. S. Joost *et al.*, Single-Cell Transcriptomics Reveals that Differentiation and Spatial Signatures Shape Epidermal and Hair Follicle Heterogeneity. *Cell Systems*. **3**, 221–237.e9 (2016).

54. J. Veevers *et al.*, Cell-Surface Marker Signature for Enrichment of Ventricular Cardiomyocytes Derived from Human Embryonic Stem Cells. *Stem Cell Reports*. **11**, 828–841 (2018).

55. I. Agarkova, D. Auerbach, E. Ehler, J. C. Perriard, A novel marker for vertebrate embryonic heart, the EH-myomesin isoform. *J. Biol. Chem.* **275**, 10256–10264 (2000).

56. P. Mauffrey *et al.*, Progenitors from the central nervous system drive neurogenesis in cancer. *Nature*. **569**, 672–678 (2019).

57. J. R. Davis *et al.*, An Efficient Multiple-Testing Adjustment for eQTL Studies that Accounts for Linkage Disequilibrium between Variants. *Am. J. Hum. Genet.* **98**, 216–224 (2015).

58. S. M. Urbut, G. Wang, P. Carbonetto, M. Stephens, Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet.* **5**, 1–15 (2018).

59. Y.-F. Huang, B. Gulko, A. Siepel, Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data. **49**, 618–624 (2017).

60. D. R. Zerbino, S. P. Wilder, N. Johnson, T. Juettemann, P. R. Flicek, The ensembl regulatory build. *Genome Biol.* **16**, 56 (2015).

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

62. A. N. Barbeira *et al.*, Widespread dose-dependent effects of RNA expression and splicing on complex diseases and traits. *bioRxiv*. **6**, e1000888–89 (2019).

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; M. Khan for the illustrations in Figure 1.
Funding: This work was funded by following funding sources: Marie-Skłodowska Curie fellowship H2020 Grant 706636 (S.K-H.), 1K99HG009916-01 (S.E.C.), R01HG002585 (G.W.,M.S.), BIO2015-70777-P, Ministerio de Economia y Competitividad and FEDER funds (M.M-A.,V.W.,R.G.,D.G-M.), FPU15/03635, Ministerio de Educación, Cultura y Deporte (M.M-A.), "la Caixa" Foundation (ID 100010434) agreement LCF/BQ/SO15/52260001 (D.G.-M.), EU IMI program (UE7-DIRECT-115317-1) (A.V.,E.T.D.), FNS funded project RNA1 (31003A_149984) (A.V.,E.T.D.), Massachusetts Lions Eye Research Fund Grant (A.R.H.), MRC Grants MR/R023131/1 and MR/M004422/1 (K.S.S), Biomedical Big Data Training Grant 5T32LM012424-03 (B.N.). The TwinsUK study was funded by the Wellcome Trust and European Community's Seventh Framework Programme (FP7/2007-2013). The TwinsUK study also receives support from the National Institute for Health Research (NIHR)- funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.
This work was further 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., 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., L.K.B.), 10XS171 (Roswell Park Cancer Institute: B.A.F., M.T.M., E.K., B.M.G., K.D.R., J.B.), 10X172 (Science Care Inc.), 12ST1039 (IDOX), 10ST1035 (Van Andel Institute: S.D.J., D.C.R., 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., J.L.N.), 5U41HG009494 (F.A., G.G., K.G.A.), and through NIH grants R01 DA006227-17 (Univ. of Miami Brain Bank: D.C.M., D.A.D.), Supplement to University of Miami grant DA006227 (D.C.M., D.A.D.), R01 MH090941 (Univ. of Geneva), R01 MH090951 and R01 MH090937 (Univ. of Chicago), R01 MH090936 (Univ. of North Carolina–Chapel Hill), 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.), U01MH104393 (A.P.F.). Extension H002371 to 5U41HG002371 (W.J.K) as well as other funding sources: R01MH106842 (T.L., P.M., E.F., P.J.H.), R01HL142028 (T.L., Si.Ka., P.J.H.), R01GM122924 (T.L., S.E.C.), R01MH107666 (H.K.I.), P30DK020595 (H.K.I.), UM1HG008901 (T.L.), R01GM124486 (T.L.), R01HG010067 (Y.Pa.), R01HG002585 (G.Wa., M.St.), Gordon and Betty Moore Foundation GBMF 4559 (G.Wa., M.St.), R01HG006855 (Se.Ka., R.E.H.), NIH CTSA grant UL1TR002550-01 (P.M.), R35HG010718 (E.R.G.), 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.), F32HG009987 (F.H.).
Author contributions: S.K.-H., F.A. and T.L. conceived the study. S.K.-H. and F.A. led the writing, figure generation and editing of the manuscript and supplement. S.K.-H. coordinated analyses of all contributing authors; S.K.-H. and F.A. generated pipelines and performed iQTL mapping; S.K.-H., F.A., M.O., M.M.-A., V.W., D.G.-M., S.M., B.N., J.Q. performed cell type

benchmarking analyses; S.K. performed ieQTL validation with ASE data using the validation pipeline and ASE data generated by S.E.C.; F.A., A.V., A.L.R. performed replication analyses; S.E.C. performed QTL tissue activity prediction analysis; S.K.-H. and S.E.C. generated tissue sharing (MASH) data; S.K.-H. performed tissue specificity, multi-tissue analysis and colocalization analysis; A.R.H. performed QTLEnrich analysis; G.W. and Y.Z. provided software support for multi-tissue eQTL analysis; X.W. and H.I. provided advice on colocalization analysis; A.B., A.M.-P., J.M.-S. contributed to replication analysis; F.A. and K.G.A. generated and oversaw GTEx v8 data generation, LDACC & pipelines; A.N.B. and R.B. generated GWAS data; K.S.S., M.S., H.S.X., G.G., E.T.D., H.I., R.G., A.V.S., B.E.S., K.G.A., T.L. supervised the work of trainees in his/her lab; M.O. and T.L. contributed to editing of the manuscript; All authors read and approved the final manuscript.

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; J.Q. is an employee of Pfizer Inc.; H.X. is an employee of AbbVie; H.I. has received speaker honoraria from GSK and AbbVie; E.T.D. is chairman and member of the board of Hybridstat LTD; 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; 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. Other GTEx members: E.R.G. is on the Editorial Board of Circulation Research, and does consulting for the City of Hope / Beckman Research Institut; B.E.E. is on the scientific advisory boards of Celsius Therapeutics and Freenome; S.B.M. is on the scientific advisory board of Prime Genomics Inc.; 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-Genzym.

Data and materials availability:

All GTEx open-access data, including summary statistics and visualizations of cell type iQTLs, are available on the GTEx Portal (https://gtexportal.org/home/datasets). All GTEx protected data are available via dbGaP (accession phs000424.v8). Access to the raw sequence data is now provided through the AnVIL platform (https://gtexportal.org/home/protectedDataAccess). 87 harmonized and imputed GWAS summary stats described in Table S3 are available and linked at https://github.com/hakyimlab/gtex-gwas-analysis and https://zenodo.org/record/3629742#.XxYGoy1h0Ux. Original GWAS studies are cited in (*1*). The QTL mapping pipeline is available at https://github.com/broadinstitute/gtex-pipeline and https://doi.org/10.5281/zenodo.3727189, tensorQTL is available at https://github.com/broadinstitute/tensorqtl and https://doi.org/10.5281/zenodo.3726360. Residual GTEx biospecimens have been banked and remain available as a resource for further studies (access can be requested on the GTEx Portal, at https://www.gtexportal.org/home/samplesPage).

Supplementary Materials:

Materials and Methods

Figures S1 – S17

Tables S1 – S6
References (37 - 62)

Fig. 1. Study design of mapping cell type ieQTLs and isQTLs in GTEx v8 project. (**A**) Illustration of 43 cell type-tissue pairs included in the GTEx v8 project. See (*1*) for the full list of tissues included in the GTEx v8 project; two brain regions (frontal cortex and cerebellum) were sampled in replicates. Cell types with median xCell enrichment score > 0.1 within a tissue were used (fig. S2). (**B**) Schematic representation of a cell type interaction eQTL and sQTL. RNA-seq coverage is depicted in gray, blue, and red, representing different genotypes. Differences in coverage between genotypes, corresponding to a QTL effect, are only observed with high cell type enrichment. The scatter plot illustrates the regression model used to identify iQTLs, where the dots represent individual samples. (**C**) Example cell type ieQTL and isQTL. The *CNTN1* eQTL effect in not sun-exposed skin is associated with keratinocyte abundance ($p = 4.1 \times 10^{-19}$; left panel). The *TNFRSF1A* sQTL effect in whole blood is associated with neutrophil abundance but is only detected in samples with lower neutrophil abundances ($p = 6.7 \times 10^{-78}$; right panel). Each data point represents an individual and is colored by genotype. Cell type enrichment scores and gene expression were inverse normal transformed, and intron excision ratios were standardized. The regression lines from the interaction model illustrate how the QTL effect is modulated by cell type enrichment.

Fig. 2. Cell type ieQTL and isQTL discovery. (**A**) Number of cell type ieQTLs (left panel) and isQTLs (right panel) discovered in each cell type-tissue combination at FDR < 5%. Bar labels show the number of ieQTLs and isQTLs, respectively. See Fig. 1A for the legend of tissue colors. (**B**) Proportion of cell type ieQTLs that validated in ASE data. Validation was defined as ieQTLs for which the Pearson correlation between allelic fold-change (aFC) estimates from ASE and cell type estimates was nominally significant ($P < 0.05$). Tissue abbreviations are provided in table S2. Bar labels indicate the number of ieQTLs with validation/number of ieQTLs tested.

Fig. 3. Cell type ieQTLs contribute to *cis*-eQTL tissue specificity. (**A**) Coefficients from logistic regression models of *cis*-eQTL tissue sharing, where epithelial cell ieQTL status is one of the predictors: All significant top *cis*-eQTLs per tissue were annotated based on if they were also a significant ieQTL for a given cell type. The coefficients represent the log(odds ratio) that an eQTL is active in a replication tissue given a predictor. Chromatin states were defined using matched Epigenomics Roadmap tissues and the 15-state ChromHMM (*37*). Genomic annotations, conservation, and overlaps with Ensembl regulatory build TF, CTCF, and DHS peaks are also included. Bars represent the 95% confidence interval. (**B**) Proportion of cell type ieQTL-genes (ieGenes) among tissue-specific and tissue-shared eGenes. An eGene is considered tissue-specific if its eQTL had a MASH local false sign rate (LFSR, equivalent to FDR) < 0.05 only in the cell type ieQTL tissue (or tissue type) otherwise it is considered tissue-shared. Results of all 43 cell type-tissue combinations are shown. See Fig. 1A for the legend of tissue colors. (**C+D**) Tissue activity of cell type ieQTLs and eQTLs, where a cell type ieQTL and eQTL was considered active in a tissue if it had an LFSR < 0.05 (left panel). Pairwise tissue-sharing of ieQTLs (middle panel) or lead standard *cis*-eQTLs (right panel) respectively. The color-coded sharing signal is the proportion of significant QTLs (LFSR < 0.05) that are shared in magnitude (within a factor of 2) and sign between two tissues.

Fig. 4. Cell type iQTLs are enriched for GWAS signals. (A) Distribution of adjusted GWAS fold-enrichment of 23x87 (top panel) and 7x87 (bottom panel) tissue-trait combinations using the most significant iQTL or standard QTL per eGene/sGene. (**B**) Adjusted GWAS fold-enrichments of 87 GWAS traits among iQTLs on the x-axis and standard QTLs on the y-axis. Filled circles indicate significant GWAS enrichment among iQTLs at $P < 0.05$ (Bonferroni-corrected). Colors represent GWAS categories of the 87 GWAS traits (see table S3).

Fig. 5. Cell type iQTLs improve GWAS-QTL matching. (A) Proportion of cell type ieQTLs (left panel) or isQTLs (right panel) with evidence of colocalization using COLOC posterior probabilities (PP4 > 0.5), for ieQTLs and isQTL at FDR < 0.4. Color saturation indicates if a trait colocalized with the cell type iQTL only (dark), the *cis*-QTL only (light) or both QTLs (medium). Bar labels indicate the number of cell type iQTLs with evidence of colocalization (either as iQTL or *cis*-QTL)/number of iQTLs tested. (**B**) Summary of all QTL-trait colocalizations from (A). (**C**) Association p-values in the *DHX58* locus for an asthma GWAS (top), standard heart left ventricle *cis*-eQTL (middle) and myocyte ieQTL (bottom), and in the *KREMEN1* locus for a birth weight GWAS (top), standard subcutaneous adipose *cis*-eQTL (middle), and adipocyte ieQTL (bottom). (**D**) Association p-values in the *CDHR5* locus for an eosinophil count GWAS (top), standard small intestine terminal ileum *cis*-sQTL (middle) and epithelial cell isQTL (bottom), and in the *ATP5SL* locus for a standing height GWAS (top), standard heart left ventricle *cis*-sQTL (middle), and myocyte isQTL (bottom).

GTEx Consortium

Laboratory and Data Analysis Coordinating Center (LDACC): François Aguet[1], Shankara Anand[1], Kristin G Ardlie[1], Stacey Gabriel[1], Gad A Getz[1,2,3], Aaron Graubert[1], Kane Hadley[1], Robert E Handsaker[4,5,6], Katherine H Huang[1], Seva Kashin[4,5,6], Xiao Li[1], Daniel G MacArthur[5,7], Samuel R Meier[1], Jared L Nedzel[1], Duyen T Nguyen[1], Ayellet V Segrè[1,8], Ellen Todres[1]

Analysis Working Group (funded by GTEx project grants):
François Aguet[1], Shankara Anand[1], Kristin G Ardlie[1], Brunilda Balliu[9], Alvaro N Barbeira[10], Alexis Battle[11,12], Rodrigo Bonazzola[10], Andrew Brown[13,14], Christopher D Brown[15], Stephane E Castel[16,17], Donald F Conrad[18,19], Daniel J Cotter[20], Nancy Cox[21], Sayantan Das[22], Olivia M de Goede[20], Emmanouil T Dermitzakis[13,23,24], Jonah Einson[25,16], Barbara E Engelhardt[26,27], Eleazar Eskin[28], Tiffany Y Eulalio[29], Nicole M Ferraro[29], Elise D Flynn[16,17], Laure Fresard[30], Eric R Gamazon[31,32,33,21], Diego Garrido-Martín[34], Nicole R Gay[20], Gad A Getz[1,2,3], Michael J Gloudemans[29], Aaron Graubert[1], Roderic Guigó[34,35], Kane Hadley[1], Andrew R Hame[l8,1], Robert E Handsaker[4,5,6], Yuan He[11], Paul J Hoffman[16], Farhad Hormozdiari[36,1], Lei Hou[37,1], Katherine H Huang[1], Hae Kyung Im[10], Brian Jo[26,27], Silva Kasela[16,17], Seva Kashin[4,5,6], Manolis Kellis[37,1], Sarah Kim-Hellmuth[16,17,38], Alan Kwong[22], Tuuli Lappalainen[16,17], Xiao Li[1], Xin Li[30], Yanyu Liang[10], Daniel G MacArthur[5,7], Serghei Mangul[28,39], Samuel R Meier[1], Pejman Mohammadi[16,17,40,41], Stephen B Montgomery[30,20], Manuel Muñoz-Aguirre[34,42], Daniel C Nachun[30], Jared L Nedzel[1], Duyen T Nguyen[1], Andrew B Nobel[43], Meritxell Oliva[10,44], YoSon Park[15,45], Yongjin Park[37,1], Princy Parsana[12], Abhiram S Rao[46], Ferran Reverter[47], John M Rouhana[8,1], Chiara Sabatti[48], Ashis Saha[12], Ayellet V Segrè[1,8,] Andrew D Skol[10,49], Matthew Stephens[50], Barbara E Stranger[10,51], Benjamin J Strober[11], Nicole A Teran[30], Ellen Todres[1], Ana Viñuela[52,13,23,24], Gao Wang[50], Xiaoquan Wen[22], Fred Wright[53], Valentin Wucher[34], Yuxin Zou[54]

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 & 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[20], Joshua Akey[79], Lin Chen[44], Kathryn Demanelis[44], Jennifer A Doherty[80], Andrew P Feinberg[81], Kasper D Hansen[82], Peter F Hickey[83], Lei Hou[37,1], Farzana

Jasmine[44], Lihua Jiang[20], Rajinder Kaul[84,85], Manolis Kellis[37,1], Muhammad G Kibriya[44], Jin Billy Li[20], Qin Li[20], Shin Lin[86], Sandra E Linder[20], Stephen B Montgomery[30,20], Meritxell Oliva[10,44], Yongjin Park[37,1], Brandon L Pierce[44], Lindsay F Rizzardi[87], Andrew D Skol[10,49], Kevin S Smith[30], Michael Snyder[20], John Stamatoyannopoulos[84,88], Barbara E Stranger[10,51], Hua Tang[20], Meng Wang[20]

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]

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

23. Institute for Genetics and Genomics in Geneva (iGE3), University of Geneva, Geneva, Switzerland
24. Swiss Institute of Bioinformatics, Geneva, Switzerland
25. Department of Biomedical Informatics, Columbia University, New York, NY, USA
26. Department of Computer Science, Princeton University, Princeton, NJ, USA
27. Center for Statistics and Machine Learning, Princeton University, Princeton, NJ, USA
28. Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, USA
29. Program in Biomedical Informatics, Stanford University School of Medicine, Stanford, CA, USA
30. Department of Pathology, Stanford University, Stanford, CA, USA
31. Data Science Institute, Vanderbilt University, Nashville, TN, USA
32. Clare Hall, University of Cambridge, Cambridge, UK
33. MRC Epidemiology Unit, University of Cambridge, Cambridge, UK
34. Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Barcelona, Catalonia, Spain
35. Universitat Pompeu Fabra (UPF), Barcelona, Catalonia, Spain
36. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
37. Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA
38. Statistical Genetics, Max Planck Institute of Psychiatry, Munich, Germany
39. Department of Clinical Pharmacy, School of Pharmacy, University of Southern California, Los Angeles, CA, USA
40. Scripps Research Translational Institute, La Jolla, CA, USA
41. Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA
42. Department of Statistics and Operations Research, Universitat Politècnica de Catalunya (UPC), Barcelona, Catalonia, Spain
43. Department of Statistics and Operations Research and Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA
44. Department of Public Health Sciences, The University of Chicago, Chicago, IL, USA
45. Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA
46. Department of Bioengineering, Stanford University, Stanford, CA, USA
47. Department of Genetics, Microbiology and Statistics, University of Barcelona, Barcelona. Spain.
48. Departments of Biomedical Data Science and Statistics, Stanford University, Stanford, CA, USA
49. Department of Pathology and Laboratory Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA
50. Department of Human Genetics, University of Chicago, Chicago, IL, USA
51. Center for Genetic Medicine, Department of Pharmacology, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA

52. Department of Twin Research and Genetic Epidemiology, King's College London, London, UK
53. Bioinformatics Research Center and Departments of Statistics and Biological Sciences, North Carolina State University, Raleigh, NC, USA
54. Department of Statistics, University of Chicago, Chicago, IL, 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. LifeGift, 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. Temple University, Philadelphia, PA, USA
76. Virginia Commonwealth University, Richmond, VA, USA
77. European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, United Kingdom
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, Utah, 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, University of Washington, Seattle, WA, USA
86. Department of Cardiology, University of Washington, Seattle, WA, USA
87. HudsonAlpha 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