

1

## 2 Local genetic effects on gene expression across 44 human tissues

3

4 François Aguet<sup>1\*</sup>, Andrew A. Brown<sup>2,3,4\*</sup>, Stephane E. Castel<sup>5,6\*</sup>, Joe R. Davis<sup>7,8\*</sup>, Pejman  
5 Mohammadi<sup>5,6\*</sup>, Ayellet V. Segre<sup>1\*</sup>, Zachary Zappala<sup>7,8\*</sup>, Nathan S. Abell<sup>7,8</sup>, Laure Frésard<sup>8</sup>, Eric  
6 R. Gamazon<sup>9</sup>, Ellen Gelfand<sup>1</sup>, Michael J. Gloudemans<sup>8,10</sup>, Yuan He<sup>11</sup>, Farhad Hormozdiari<sup>12</sup>,  
7 Xiao Li<sup>1</sup>, Xin Li<sup>8</sup>, Boxiang Liu<sup>8,13</sup>, Diego Garrido-Martín<sup>14,15</sup>, Halit Ongen<sup>2,3,4</sup>, John J.  
8 Palowitch<sup>16</sup>, YoSon Park<sup>17</sup>, Christine B. Peterson<sup>18,19</sup>, Gerald Quon<sup>1,20</sup>, Stephan Ripke<sup>21,22</sup>,  
9 Andrey A. Shabalin<sup>23</sup>, Tyler C. Shimko<sup>7,8</sup>, Benjamin J. Strober<sup>11</sup>, Timothy J. Sullivan<sup>1</sup>, Nicole  
10 A. Teran<sup>7,8</sup>, Emily K. Tsang<sup>8,10</sup>, Hailei Zhang<sup>1</sup>, Yi-Hui Zhou<sup>24</sup>, Alexis Battle<sup>25</sup>, Carlos D.  
11 Bustamante<sup>7,26</sup>, Nancy J. Cox<sup>9</sup>, Barbara E. Engelhardt<sup>27</sup>, Eleazar Eskin<sup>12,28</sup>, Gad Getz<sup>1,29</sup>,  
12 Manolis Kellis<sup>1,20</sup>, Gen Li<sup>30</sup>, Daniel G. MacArthur<sup>1,20</sup>, Andrew B. Nobel<sup>16</sup>, Chiara Sabatti<sup>18,26</sup>,  
13 Xiaoquan Wen<sup>31</sup>, Fred A. Wright<sup>24,32</sup>, GTEx Consortium, Tuuli Lappalainen<sup>5,6</sup>, Kristin G.  
14 Ardlie<sup>1</sup>, Emmanouil T. Dermitzakis<sup>2,3,4†</sup>, Christopher D. Brown<sup>17†</sup>, Stephen B. Montgomery<sup>7,8†</sup>

15 **1** The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, 02142, USA

16 **2** Department of Genetic Medicine and Development, University of Geneva Medical School, 1211  
17 Geneva, Switzerland

18 **3** Institute for Genetics and Genomics in Geneva (iG3), University of Geneva, 1211 Geneva,  
19 Switzerland

20 **4** Swiss Institute of Bioinformatics, 1211 Geneva, Switzerland

21 **5** New York Genome Center, New York, NY, 10013, USA

22 **6** Department of Systems Biology, Columbia University, New York, NY, 10032, USA

23 **7** Department of Genetics, Stanford University, Stanford, CA, 94305, USA

24 **8** Department of Pathology, Stanford University, Stanford, CA, 94305, USA

25 **9** Division of Genetic Medicine, Department of Medicine, Vanderbilt University, Nashville, TN,  
26 37232, USA

27 **10** Biomedical Informatics Program, Stanford University, Stanford, CA, 94305, USA

28 **11** Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, 21218, USA

29 **12** Department of Human Genetics, University of California, Los Angeles, CA, 90095, USA

30 **13** Department of Biology, Stanford University, Stanford, CA, 94305, USA

31 **14** Bioinformatics and Genomics, Centre for Genomic Regulation (CRG), Barcelona Institute of  
32 Science and Technology, Barcelona 08003, Spain.

33 **15** Department of Experimental and Health Sciences, Universitat Pompeu Fabra (UPF),  
34 Barcelona 08002, Spain

35 **16** Department of Statistics and Operations Research, University of North Carolina, Chapel Hill,  
36 NC, 27599, USA

37 **17** Department of Genetics, University of Pennsylvania, Perelman School of Medicine,  
38 Philadelphia, PA, 19104, USA

39 **18** Department of Biomedical Data Science, Stanford University, Stanford, CA, 94305, USA

40 **19** Present address: Department of Biostatistics, The University of Texas MD Anderson Cancer  
41 Center, Houston, TX, 77030, USA

42 **20** Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, 02139, USA

43 **21** Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA,  
44 02114, USA

45 **22** Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge,  
46 02142, MA, USA

47 **23** Center for Biomarker Research and Personalized Medicine, Virginia Commonwealth  
48 University, Richmond, VA 23298, USA

49 **24** Bioinformatics Research Center and Department of Biological Sciences, North Carolina State  
50 University, Raleigh, NC, 27695, USA

51 **25** Department of Computer Science, Johns Hopkins University, Baltimore, MD, 21218, USA

52 **26** Department of Statistics, Stanford University, Stanford, CA, 94305, USA

53 **27** Department of Computer Science, Center for Statistics and Machine Learning, Princeton  
54 University, Princeton, NJ, 08540, USA

55 **28** Department of Computer Science, University of California, Los Angeles, CA, 90095, USA

56 **29** Massachusetts General Hospital Cancer Center and Department of Pathology, Massachusetts  
57 General Hospital and Harvard Medical School, Boston, MA, 02114, USA

58 **30** Department of Biostatistics, Mailman School of Public Health, Columbia University, New  
59 York, NY, 10032, USA

60 **31** Department of Biostatistics, University of Michigan, Ann Arbor, MI, 48109, USA

61 **32** Department of Statistics, North Carolina State University, Raleigh NC, 27695, USA

62 \* Co-first authors, listed alphabetically

63 † Co-corresponding authors

## 64 **Abstract**

65 Expression quantitative trait locus (eQTL) mapping provides a powerful means to identify func-  
66 tional variants influencing gene expression and disease pathogenesis. We report the identification  
67 of cis-eQTLs from 7,051 post-mortem samples representing 44 tissues and 449 individuals as part  
68 of the Genotype-Tissue Expression (GTEx) project. We find a cis-eQTL for 88% of all annotated  
69 protein-coding genes, with one-third having multiple independent effects. We identify numerous  
70 tissue-specific cis-eQTLs, highlighting the unique functional impact of regulatory variation in di-  
71 verse tissues. By integrating large-scale functional genomics data and state-of-the-art fine-mapping  
72 algorithms, we identify multiple features predictive of tissue-specific and shared regulatory effects.  
73 We improve estimates of cis-eQTL sharing and effect sizes using allele specific expression across tis-  
74 sues. Finally, we demonstrate the utility of this large compendium of cis-eQTLs for understanding  
75 the tissue-specific etiology of complex traits, including coronary artery disease. The GTEx project  
76 provides an exceptional resource that has improved our understanding of gene regulation across  
77 tissues and the role of regulatory variation in human genetic diseases.

## 78 **Introduction**

79 Genome-wide association studies (GWAS) have identified a wealth of genetic variants associated  
80 with complex traits and disease risk. However, characterizing the molecular and cellular mechanisms  
81 through which these variants act remains a major challenge that limits our understanding of disease  
82 pathogenesis and the development of therapeutic interventions. Expression quantitative trait locus  
83 (eQTL) studies provide a systematic approach to characterize the molecular consequences of genetic  
84 variation across tissues and cell types<sup>1-4</sup>. Multiple studies have identified eQTLs for thousands of  
85 genes<sup>5-7</sup>, providing novel insights into gene regulation and enabling the interpretation of GWAS  
86 signals<sup>8-12</sup>. These studies have largely been performed in a few easily accessible cell types and cell  
87 lines, precluding interpretation of the systemic and tissue-specific consequences of genetic variation.  
88 To overcome these limitations, the Genotype Tissue Expression (GTEx) project was designed to  
89 identify and characterize eQTLs across a broad range of tissues. During the pilot phase, which  
90 focused on nine tissues, the GTEx project highlighted patterns of eQTL tissue-specificity and

91 demonstrated the value of multi-tissue study designs for identifying causal genes and tissues for  
92 trait-associated variants<sup>1</sup>. These results indicated that the identification of eQTLs across an even  
93 broader range of tissues would drastically improve characterization of the gene- and tissue-specific  
94 consequences of genetic variants.

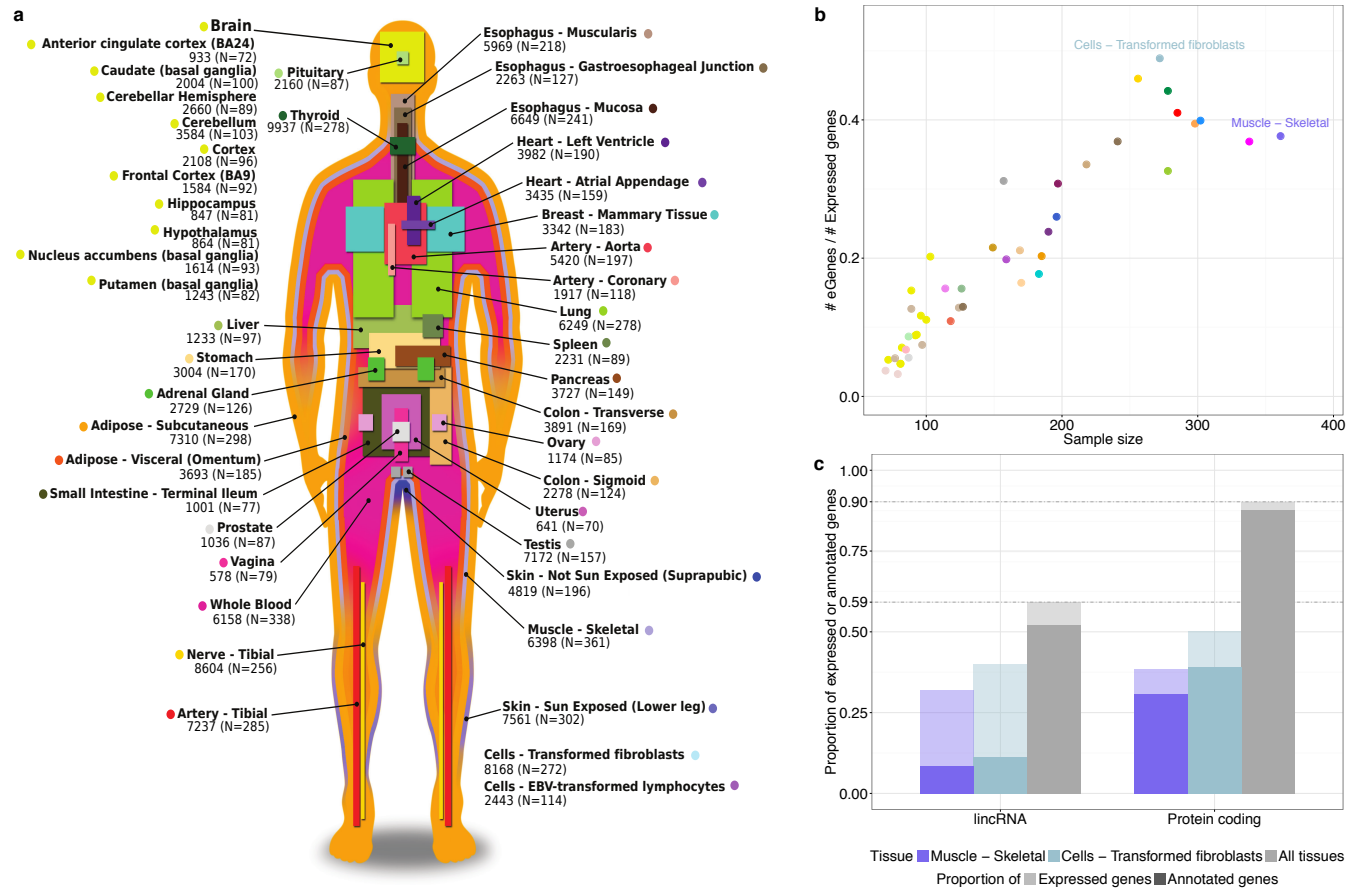
95 Here, we report on the discovery of cis-eQTLs across an expanded collection of 44 tissues in  
96 the GTEx V6p study. This dataset consists of 7,051 transcriptomes from 449 individuals and  
97 44 tissues (median 16 tissues per individual, 127 samples per tissue), including multiple tissues  
98 that are difficult to sample such as 10 distinct brain regions. With this dataset, we identified cis-  
99 eQTLs within each tissue and characterized the sharing of eQTLs across tissues. We next assessed  
100 the relationship between tissue-specific and shared eQTLs with different functional annotations,  
101 including promoters, enhancers and Hi-C contacts, and with allele-specific expression (ASE). Fi-  
102 nally, we demonstrated the utility of this multi-tissue resource for the interpretation of genetic  
103 variation associated with complex disease. We provide openly available summary statistics of cis-  
104 eQTLs for all 44 tissues on the GTEx Portal (<http://gtexportal.org>) and all raw data in dbGaP  
105 (phs000424.v6.p1).

## 106 Single-tissue cis-eQTL discovery

107 cis-eQTLs, or associations between local genetic variation and gene expression ( $\leq 1$  Mb from  
108 the transcription start site, TSS), were identified using genotype and RNA-seq data generated  
109 from 44 tissues ( $N = 70$ –361 samples per tissue) using a linear model (FastQTL)<sup>13</sup> (Fig. 1a,b).  
110 Within each tissue, we identified a median of 2,866 genes with cis-eQTLs at a 5% FDR (hereafter  
111 referred to as eGenes). In total, we found 159,760 cis-eQTLs for 20,175 genes, representing 82.6%  
112 of all genes tested in GTEx and 78.3% of all annotated autosomal lincRNA and protein coding  
113 genes<sup>14</sup>. For autosomal protein-coding genes alone, we identified 16,605 eGenes representing 90.2%  
114 of all expressed protein-coding genes in GTEx and 88% of all annotated protein-coding genes (Fig.  
115 1c). For genes without an eQTL in any tissue, we observed less selective constraint as well as  
116 enrichment of functions related to transcriptional regulation, environmental response, and cellular  
117 differentiation, indicating that biological context influences the discovery of eQTLs for these genes  
118 (Extended Data Fig. 1). eGene discovery increased linearly with sample size with no evidence of  
119 saturation at the full sample size for each tissue, suggesting that all genes may ultimately be shown  
120 to be influenced by regulatory variation (Extended Data Fig. 2).

121 We also identified conditionally independent regulatory variants for each eGene (secondary  
122 cis-eQTLs) using forward-backward stepwise regression separately in each tissue. This approach  
123 revealed an additional 22,099 cis-eQTLs across the 44 tissues, with 36.7% of protein-coding genes  
124 and 12.5% of lincRNAs having multiple, conditionally independent cis-eQTLs in at least one tissue  
125 (Extended Data Fig. 3).

126 The large sampling of tissues allowed us to develop a comprehensive view of the sharing of  
127 cis-eQTLs across tissues in the human body. We tested the replication of cis-eQTLs using the  $\pi_1$   
128 statistic<sup>15</sup> for all tissue pairs (Fig. 2a). We observed patterns of sharing that reflected previously  
129 identified relationships between tissues<sup>1</sup>. For example, we found a high degree of sharing between  
130 brain tissues (mean  $\pi_1$  of 0.864), arterial tissues (mean  $\pi_1$  of 0.854), and skeletal muscle and heart  
131 tissues (mean  $\pi_1$  of 0.819). The mean  $\pi_1$  sharing across all tissue pairs was 0.727 ranging from  
132 0.354 to 0.981. Since individuals in the GTEx dataset contribute samples for multiple tissues, we  
133 investigated the effect of this grouping on sharing estimates by calculating  $\pi_1$  for tissues subsampled  
134 to have complete sharing among individuals (Extended Data Fig. 4). These sharing estimates  
135 correlated with estimates from variable levels of individual overlap between tissues (Spearman  $\rho =$   
136 0.53,  $P < 2.2 \times 10^{-16}$ ). Furthermore, in the full dataset for each tissue, we observed that even for

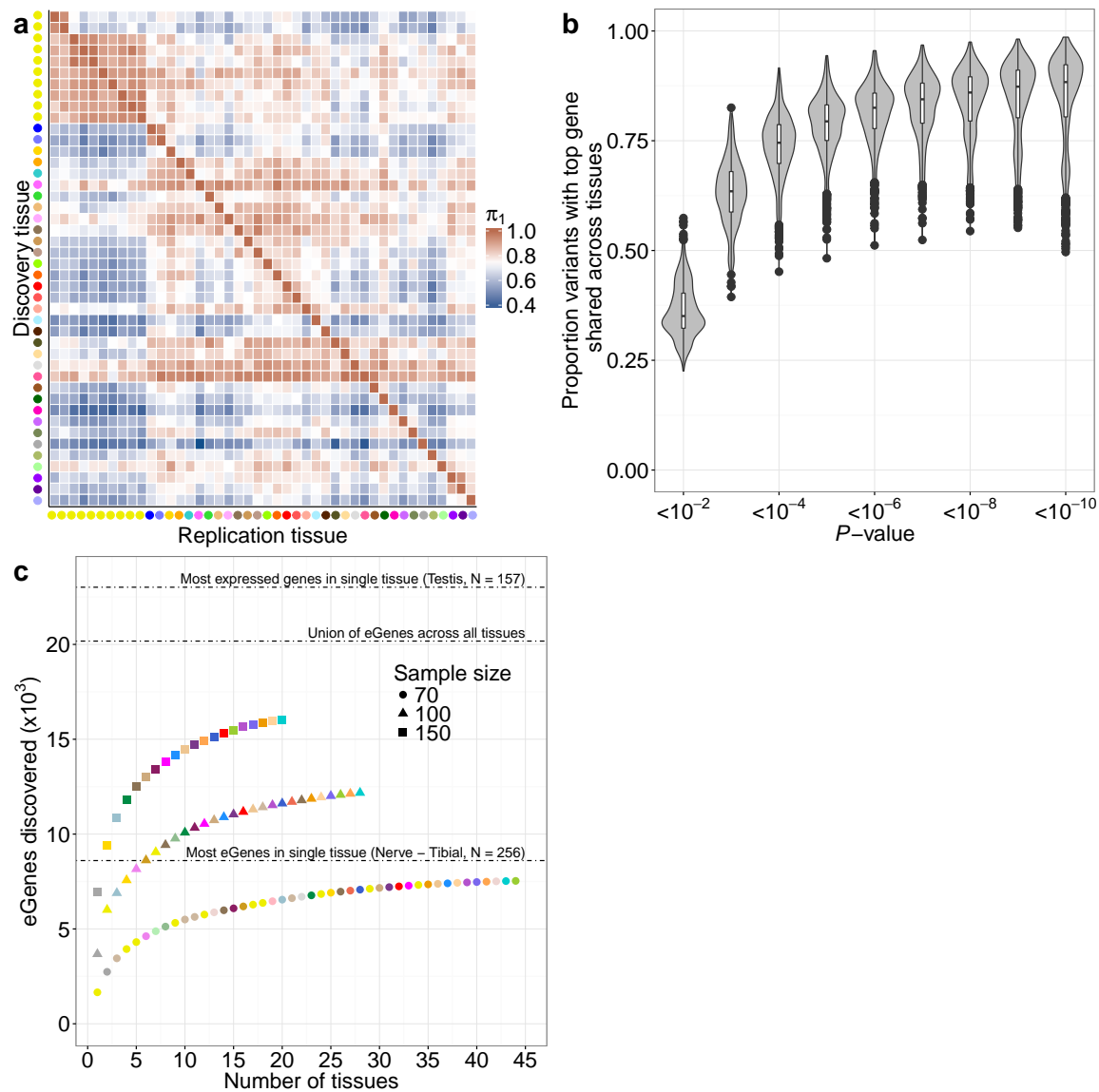


**Figure 1. Sample size and eGene discovery in the GTEx V6p study.** (a) Illustration of the 44 tissues and cell lines included in the GTEx V6p project with the associated number of eGenes and sample sizes. (b) The proportion of expressed genes discovered as eGenes versus sample size. Cells - Transformed fibroblasts are highlighted as the tissue with the highest proportion. Muscle - Skeletal has the largest sample size. (c) Fraction of genes that are eGenes across all tissues by transcript class. As in (b), Cells - Transformed fibroblasts and Muscle - Skeletal are shown as a reference. Annotated genes are all known human genes for each transcript class as curated in GENCODE v19.

137 very strong shared associations ( $P < 10^{-10}$  in each tissue), roughly 10% exhibited different single  
 138 top gene associations across tissues, indicating that the interpretation of the regulatory effect of  
 139 these variants can still be tissue-dependent (Fig. 2b).

140 To quantify the impact of sample size and number of tissues studied on cis-eQTL discovery, we  
 141 first compared eGene discovery across a range of sample sizes and tissues (Fig. 2c). The discovery  
 142 of new eGenes was most influenced by sample size. However, a diverse sampling of tissues also  
 143 improved eGene discovery. At its full sample size of 256 individuals, tibial nerve had the most  
 144 eGenes of any tissue at 8,604, yet 9,394 unique eGenes were found for the top two tissues at a  
 145 subsample size of 150 individuals. Cerebellum, Testis, Nerve - Tibial, and Thyroid were among  
 146 the most effective tissues in increasing the total number of unique eGene discoveries. We next  
 147 tested how sample size influenced patterns of cis-eQTL sharing across tissues. We observed that

148 cis-eQTLs discovered in GTEx tissues with large sample sizes were less likely to be shared in other  
 149 tissues, indicating that weaker associations identified in deeply sampled tissues remain difficult to  
 150 replicate due to their smaller and possibly tissue-specific effects (Fig. 3; Extended Data Fig. 5).



**Figure 2. Single-tissue eQTL discovery across tissues.** (a) Replication of eQTLs between tissues. Pairwise  $\pi_1$  statistics are reported for single-tissue eQTL discoveries in each tissue. Higher  $\pi_1$  values indicate an increased replication of eQTLs. Tissues are grouped using hierarchical clustering on rows and columns separately with a distance metric of  $1 - \rho$ , where  $\rho$  is the Spearman correlation of  $\pi_1$  values.  $\pi_1$  is only calculated when the gene is expressed and testable in the replication tissue. (b) Proportion of variants with top associated protein-coding gene preserved between tissues shown for varying nominal association thresholds. (c) eGene discovery as a function of sample size and number of tissues assayed. Each tissue was subsampled to 70, 100, and 150 individuals and a greedy algorithm was used to assess sequential combinations of tissues that maximize the total number of unique eGenes discovered.



## 151 Multi-tissue cis-eQTL discovery

152 Multi-tissue cis-eQTL analyses have been shown to increase power while explicitly modeling sharing  
153 patterns across tissues<sup>16–18</sup>. We performed a meta-analysis across all 44 tissues using METASOFT<sup>19</sup>  
154 and identified between 4,538 and 9,327 eGenes (m-value  $\geq 0.9$ ) per tissue. On average, each cis-  
155 eQTL effect was shared across 15 tissues. The advantage of meta-analysis was most apparent  
156 for individual tissues with smaller sample sizes (Fig. 3a), most notably for the 10 sampled brain  
157 regions. For example, in the hippocampus ( $N = 81$ ), the number of single-tissue eGenes is 847  
158 whereas the number of eGenes detected through meta-analysis is 4,636. eGenes identified by meta-  
159 analysis were more likely to be significant in single tissue analyses at larger sample sizes (Extended  
160 Data Fig. 6). Our meta-analysis approach demonstrates that sharing of cis-eQTL effects across  
161 multiple tissues can improve discovery in specialized or difficult-to-access tissues.

162 To ensure these findings did not depend on the modeling assumptions of METASOFT, we an-  
163 alyzed the FastQTL  $P$ -values for all genes and all tissues with TreeQTL, a hierarchical multiple  
164 comparison procedure, that controls the FDR of eGene discoveries across tissues<sup>20</sup>. This procedure  
165 identified 19,610 eGenes, 565 fewer eGenes than with the single-tissue analysis. While more conser-  
166 vative overall than the tissue-by-tissue analysis, we observed an increase in the number of eGenes  
167 detected in the tissues with the smallest sample sizes, as well as an increase in the average number  
168 of tissues in which an eGene is detected (from 7.9 for single-tissue analysis to 8.5; Extended Data  
169 Fig. 7).

170 Modeling of cis-eQTL sharing across tissues using METASOFT showed a bimodal pattern with  
171 increasing tissue-specificity for tissues with larger sample sizes (Fig. 3b). Increased tissue-specificity  
172 likely emerges from differences in discovery power and effect sizes across tissues. It also suggests  
173 that deep sampling diminishes the gains of meta-analysis, instead benefiting identification of more  
174 tissue-specific effects. The bimodal pattern of sharing was further supported by three different  
175 methods: simple overlap of the single-tissue results, the hierarchical procedure of TreeQTL, and an  
176 empirical Bayes model<sup>18</sup> (Extended Data Fig. 8).

## 177 Genomic features of cis-eQTLs

178 To characterize the genomic properties of cis-eQTLs, we annotated the associated variants (hereafter  
179 referred to as eVariants) with chromatin state predictions from 128 cell types sampled by the  
180 Roadmap Epigenomics Consortium, including 26 tissues that match GTEx tissues<sup>21</sup>. eVariants were  
181 enriched in predicted promoter and enhancer states across a broad range of tissues and exhibited  
182 significantly greater enrichment in promoters and enhancers from their matched tissues (linear  
183 model controlling for discovery cell type,  $P < 5.7 \times 10^{-10}$ ), illustrating consistent patterns of cell  
184 type specificity for both cis-regulatory elements (CREs) and cis-eQTLs (Fig. 3c, e). Furthermore,  
185 cis-eQTLs were more likely to be active across pairs of tissues if the eVariant overlapped the same  
186 chromatin state in both tissues (paired Wilcoxon signed rank test,  $P < 2.2 \times 10^{-16}$ , Fig. 3d).

187 Compared to primary eVariants, secondary eVariants were located on average further away from  
188 the TSS (median distance 50.1 kb from the TSS versus 28.9 kb, Wilcoxon rank sum test,  $P < 2.2$   
189  $\times 10^{-16}$ ; Extended Data Fig. 9a) and exhibited less tissue sharing than primary eQTLs (Wilcoxon  
190 rank sum test,  $P < 2.2 \times 10^{-16}$ ; Extended Data Fig. 9b). Both primary and secondary eVariants  
191 were enriched for promoter Hi-C contacts compared to background variant-TSS pairs (Wilcoxon  
192 rank sum test,  $P < 2.2 \times 10^{-16}$ ; Extended Data Fig. 9c). This observation suggests that, despite  
193 their genomic distance from the TSS, many primary and secondary eVariants remain in close  
194 physical contact with their target gene promoters via chromatin looping interactions. Although  
195 primary eVariants are significantly more enriched in promoters than enhancers (Wilcoxon rank

196 sum test,  $P < 2.2 \times 10^{-16}$ ), secondary eVariants show greater enrichment in enhancers, consistent  
197 with their increasing distance from the TSS and tissue-specific activity (Wilcoxon rank sum test,  
198  $P < 2.2 \times 10^{-16}$ ; Fig. 3e; Extended Data Fig. 9c). This result underscores the importance  
199 of analyzing eQTLs beyond the primary association to discover regulatory variants in enhancers,  
200 which are known to be particularly relevant for disease associations<sup>22-24</sup>.

201 Integration of genomic annotations in eQTL testing has been demonstrated to improve power<sup>6, 25-27</sup>.  
202 We applied a Bayesian hierarchical model incorporating variant-level genomic annotations for eQTL  
203 discovery in 26 tissues with cell-type matched annotations from the Epigenomics Roadmap<sup>28</sup> (Wen,  
204 X. submitted). Distance to the TSS and promoter and enhancer annotations improved our ability  
205 to discover eQTLs (Extended Data Fig. 10a). Using these annotations increased the total number  
206 of eGene discoveries by an average of 43% (1,200 genes) across tissues (Extended Data Fig. 10b).

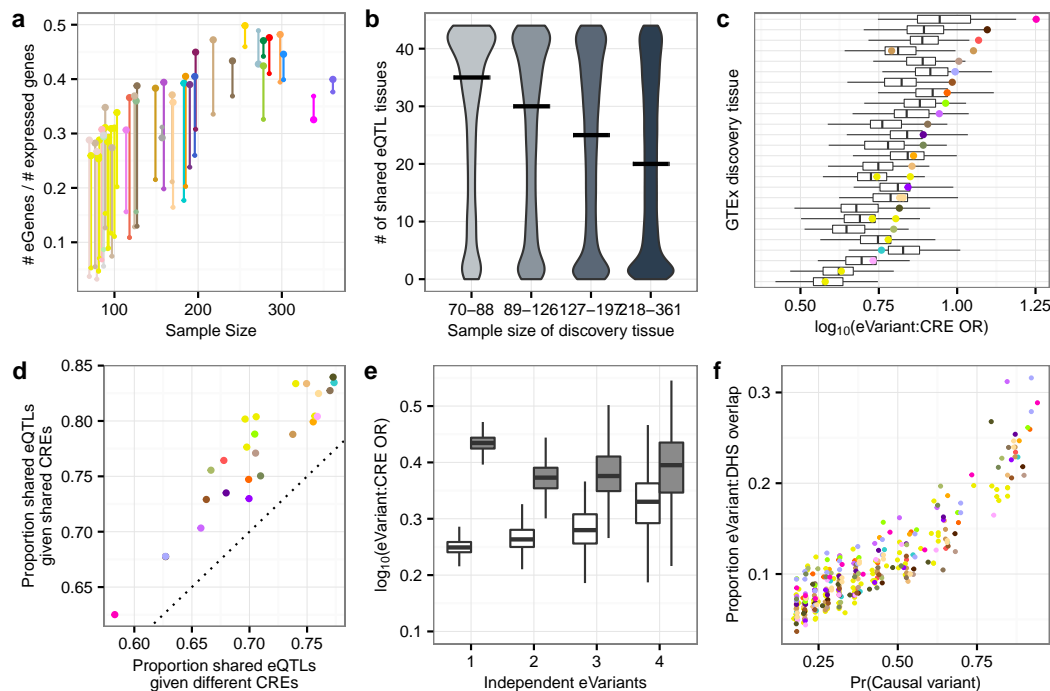
## 207 Fine-mapping eQTL variants

208 To identify likely causal variants underlying eQTLs, we applied two computational fine-mapping  
209 strategies. First, we identified 90% credible sets for each eGene in each tissue using CAVIAR<sup>29</sup>, a  
210 probabilistic method that utilizes the observed marginal test statistics and LD structure to detect  
211 variant sets that may harbor more than one causal variant<sup>29</sup>. Across all tissues, the mean credible  
212 set size was 29 variants (per tissue means ranged from 25 to 31). Credible set size decreased with  
213 increasing discovery tissue sample size. The addition of 100 samples reduced credible set size by an  
214 average of one variant indicating that large sample sizes are required to identify causal variants using  
215 association strength alone (Extended Data Fig. 11a). As expected, credible sets overlapped across  
216 tissues more extensively for tissue-shared eQTLs compared to tissue-specific eQTLs (Extended  
217 Data Fig. 11b).

218 We estimated the probability that each eVariant is a causal variant using CaVEMaN, a non-  
219 parametric sampling-based approach that accounts for noise in expression measurements and linkage  
220 structure (Brown et al. in preparation). Across tissues, we estimated that between 3.5%-11.7%  
221 of primary eVariants are causal (probability  $\geq 0.8$ ; Extended Data Fig. 12). For predicted causal  
222 variants, the same variant is predicted as causal for 13.3% to 32.6% of variants at the same proba-  
223 bility threshold in separate tissues where an eGene is also identified. However, the replication rate  
224  $\pi_1$  was considerably higher (59.6%-93.5%), demonstrating the difficulties in fine mapping variants  
225 even when the LD structure is expected to be preserved across tissues. Consistent with predicted  
226 causal variants being functional regulatory variants (as opposed to LD proxies), 24.3% of eVariants  
227 with causal probabilities in the top 10th percentile ( $P > 0.77$ ) overlapped open chromatin regions  
228 compared to 11.2% of all eVariants and 6.6% of eVariants in the lowest 10th percentile ( $0.027 < P$   
229  $< 0.19$ ; Fig. 3f).

## 230 cis-eQTL effect sizes

231 To determine the effect sizes of eQTLs discovered in GTEx, we used an additive model of eQTL  
232 alleles on total gene expression, allowing for biologically meaningful interpretation of effect sizes as  
233 an allelic fold change between the two eQTL alleles (see Methods; Mohammadi et al. in prepara-  
234 tion). 17.4% of eGenes had eQTLs with median effect sizes of  $\geq 2$ -fold across tissues (Fig. 4a).  
235 As expected, mean effect sizes per tissue were influenced by sample size (Extended Data Fig. 13).  
236 When stratifying each gene by the number of tissues that it is expressed in, we observed a decrease  
237 in the average effect size per gene indicating that genes expressed in multiple tissues are less likely  
238 to have eQTLs with large regulatory effects (Spearman  $\rho = -0.29$ ,  $P < 2.2 \times 10^{-16}$ , Fig. 4b).  
239 Supporting this observation, tissue-shared eQTLs had significantly smaller effect sizes than tissue-



**Figure 3. Multi-tissue eQTL discovery and genomic context.** (a) The proportion of expressed genes for which eGenes are discovered in single tissues (5% FDR; small dots) and the multi-tissue meta-analysis (m-value  $\geq 0.9$ ; large dots), stratified by the sample size of individual tissues. In the meta-analysis, eQTL discoveries are made using METASOFT to identify tissues where the posterior probability a given eQTL effect exists (i.e. the tissue's m-value) is  $\geq 0.9$ . (b) The number of tissues in which a given eQTL is shared as a function of tissue sample size. For each tissue, we calculated the degree of sharing (i.e. the number of tissues with m-value  $\geq 0.9$ ) for all eQTLs identified in that tissue at a 5% FDR. Tissues were then binned into quartiles based on sample size. The median number of shared tissues is plotted for each quartile as a horizontal black line. (c) Enrichment of eVariants in cis-regulatory elements (CREs) across 128 NIH Epigenomics Roadmaps cell types is depicted for each GTEx discovery tissue. Stronger enrichment was observed in matched tissues (colored dots) compared to unmatched tissues (boxplots). (d) Proportion of eQTLs that are shared between two tissues (m-value in both tissues  $\geq 0.9$ ) if the eVariant overlaps the same Roadmap annotation in both tissues (y-axis) or different annotations (x-axis). Points represent the mean of pairwise comparisons between all tissues, colored by the discovery tissue. (e) Enrichment of eVariants in tissue-matched enhancers (white) and promoters (grey) for the first four conditionally independent eQTLs discovered for each eGene (x-axis, sorted by discovery order). (f) Proportion of eVariants overlapping tissue-matched DNase I hypersensitive sites as a function of the probability that a variant is causal. Points are colored by the eQTL discovery tissue.

240 shared eQTLs matched for significance level (Wilcoxon rank sum test,  $P < 2.2 \times 10^{-16}$ , Fig. 4c;  
 241 Extended Data Fig. 13).

242 We assessed whether variants with distinct functional annotations had different average effects  
 243 on gene expression using the large number of eQTLs we discovered. eVariants at canonical splice  
 244 sites exhibited the strongest effects, followed by variants in noncoding transcripts (Fig. 4d). Vari-



ants in the 3' UTR had the weakest effect, significantly weaker than those in 5' UTRs (Wilcoxon rank sum test,  $P < 4.81 \times 10^{-10}$ ). Missense variants had a significantly stronger effect on gene expression than synonymous variants (Wilcoxon rank sum test,  $P < 8.65 \times 10^{-5}$ ). Analysis of eQTL effect sizes around the TSS demonstrated that upstream variants had a stronger effect on gene expression than downstream variants (Wilcoxon rank sum test,  $P < 1.94 \times 10^{-15}$ ; Fig. 4e), an effect that seems to persist through the gene body and beyond. These results suggest that eVariants likely to affect transcription have stronger effects on gene expression levels than variants likely to impact post-transcriptional regulation of mRNA levels.

### Allele-specific expression (ASE)

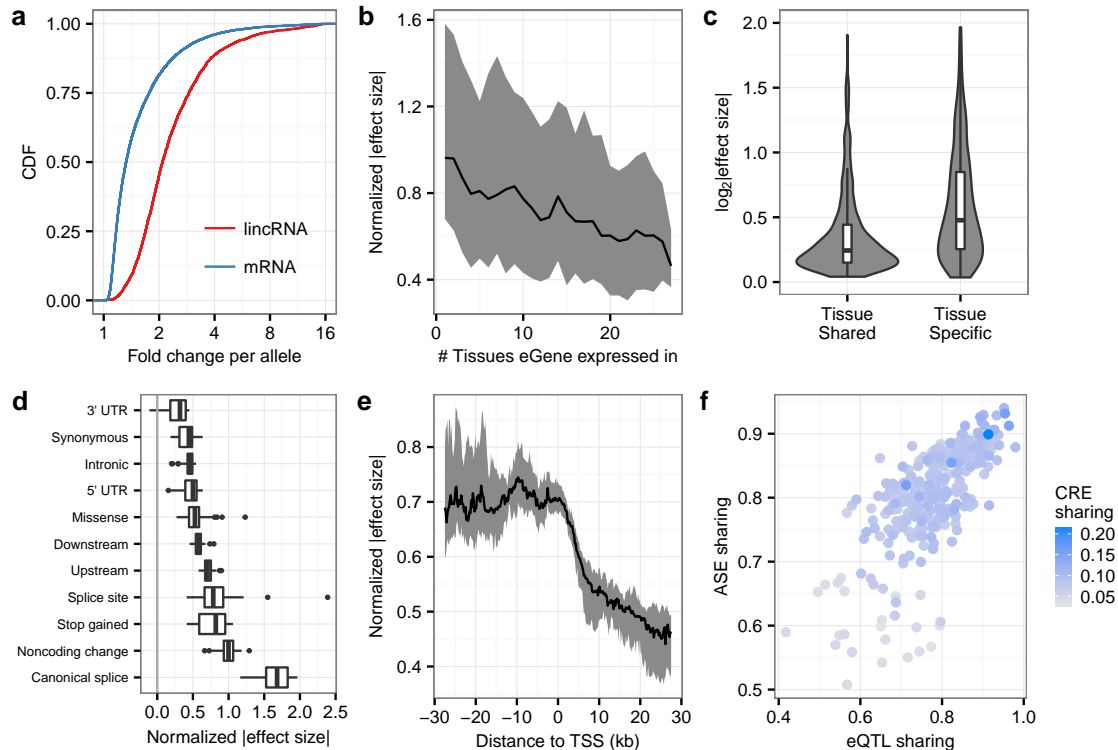
The impact of a regulatory variant on expression may be estimated from either total expression or allele specific expression (ASE) estimates. We measured ASE<sup>30</sup> at over 135 million sites across tissues and individuals, with a median of over 10,000 genes quantified per donor (Extended Data Fig. 14a-f). In total, 63.5% of all protein-coding genes could be tested for ASE in at least one individual and tissue with 62.6% having ASE data from multiple individuals in at least one tissue. 87.9% of testable genes had significant allelic imbalance in at least one individual (binomial test, FDR < 0.05), demonstrating an abundance of cis-linked regulatory effects. Across individuals, a median of 1,963 genes had significant allelic imbalance in at least one tissue, with a median of 570 genes where the individual was not heterozygous for a top eQTL. We independently estimated the effects of the primary eVariant for each eGene in each tissue using both allele-specific and total gene expression measurements (see Methods). Effect size estimates from both approaches are highly consistent with an average Spearman correlation of 0.84 (std. dev. = 2%; Extended Data Fig. 15) and an average ratio of ASE effect size to eQTL effect size of 98.5% (std. dev. = 1%). This observation confirms that cis-eQTLs and ASE capture the same biological phenomenon.

We modeled allelic expression in genes across different tissues of each individual in order to capture tissue-specificity of regulatory variant function. Over 17% of genes exhibit allelic expression patterns that differed across tissues in at least one individual. Patterns of ASE sharing in these genes were used to cluster tissues independently of total gene expression levels, which may be more susceptible to shared environmental influences, and without the strong dependency with sample size that complicates analyses of eQTL sharing (Extended Data Fig. 14g; Fig. 2a). Indeed, pairwise ASE sharing was highly correlated with pairwise eQTL sharing (Spearman  $\rho = 0.70$ ,  $P < 2.2 \times 10^{-16}$ ). Moreover, both pairwise ASE and eQTL sharing are correlated with pairwise tissue sharing of eVariant CRE annotation (Spearman  $\rho > 0.29$ ,  $P < 2.6 \times 10^{-7}$ ; Fig. 4f).

### eQTLs and GWAS

The expanded GTEx resource provides a unique opportunity to interpret GWAS associations for a wide range of complex traits and diseases. The increased diversity of tissue sampling has resulted in more identified tissue-specific eQTLs. Indeed, the degree of tissue sharing of an eQTL is associated with several indicators of phenotypic impact. eGenes shared across many tissues harbor fewer protein-coding loss-of-function (LoF) variants curated in the ExAC database<sup>31</sup> (Fig. 5a), consistent with purifying selection removing large effect regulatory variants that involve many tissues. Tissue-shared eGenes were also less likely to be annotated disease genes compared to tissue-specific eGenes (Fisher's exact test, nominal  $P < 10^{-6}$  for GWAS, OMIM, and LoF intolerant gene sets; Fig. 5a, Extended Data Fig. 16), highlighting that the cell-type specific mechanisms underlying complex genetic diseases may be elucidated only through broad tissue sampling.

This broad sampling affects the interpretation of eQTL data in the context of GWAS variants.



**Figure 4. ASE and the epigenomic context of cis-eQTLs across tissues.** (a) For each autosomal protein-coding and lincRNA eGene, the median effect size was computed across all tissues with eVariants for that eGene. The empirical CDF of these median effect sizes is depicted. (b) Median (line) and interquartile range (ribbon) of absolute eQTL effect size, corrected for median expression level across tissues and the minor allele frequency of the eVariant, as a function of the number of tissues the eGene is expressed in. (c) Comparison of effect sizes between q-value-matched tissue-shared eQTLs (m-value > 0.9 in at least 35 tissues) and tissue-specific eQTLs (m-value  $\geq$  0.9 in only the discovery tissue). (d) Normalized absolute eQTL effect size for each top eVariant, for each eVariant annotation. Normalized effect sizes were estimated by correcting for eVariant minor allele frequency and cross tissue effect size differences. (e) Normalized (as in d) eQTL effect size depicted in 200bp bins, relative to the eGene TSS. Bin medians and interquartile ranges plotted as lines and ribbons, respectively. (f) Pairwise tissue sharing of ASE effects for genes with bimodal ASE effects (proportion with same ASE mode; y-axis) is correlated with pairwise eQTL sharing ( $\pi_1$ , x-axis), and the fraction of eVariants overlapping the same Roadmap annotation in both tissues.

289 We observed that 92.7% of all common variants assayed by GTEx are nominally associated with  
 290 the expression of one or more genes in one or more tissues ( $P < 0.05$ ) and nearly 50% are significant  
 291 when performing a Bonferroni correction based on the number of tissues tested (Fig. 5b). Given  
 292 the ubiquity of eQTL associations, caution is warranted when using eQTL data to interpret the  
 293 function of candidate variants without assessing whether GWAS and eQTL association signals are  
 294 likely driven by the same causal variant by colocalization approaches that examine local LD and  
 295 trait summary statistics<sup>22,32-34</sup>.

296 To illustrate the utility of GTEx for the interpretation of disease-associated variation, we ap-

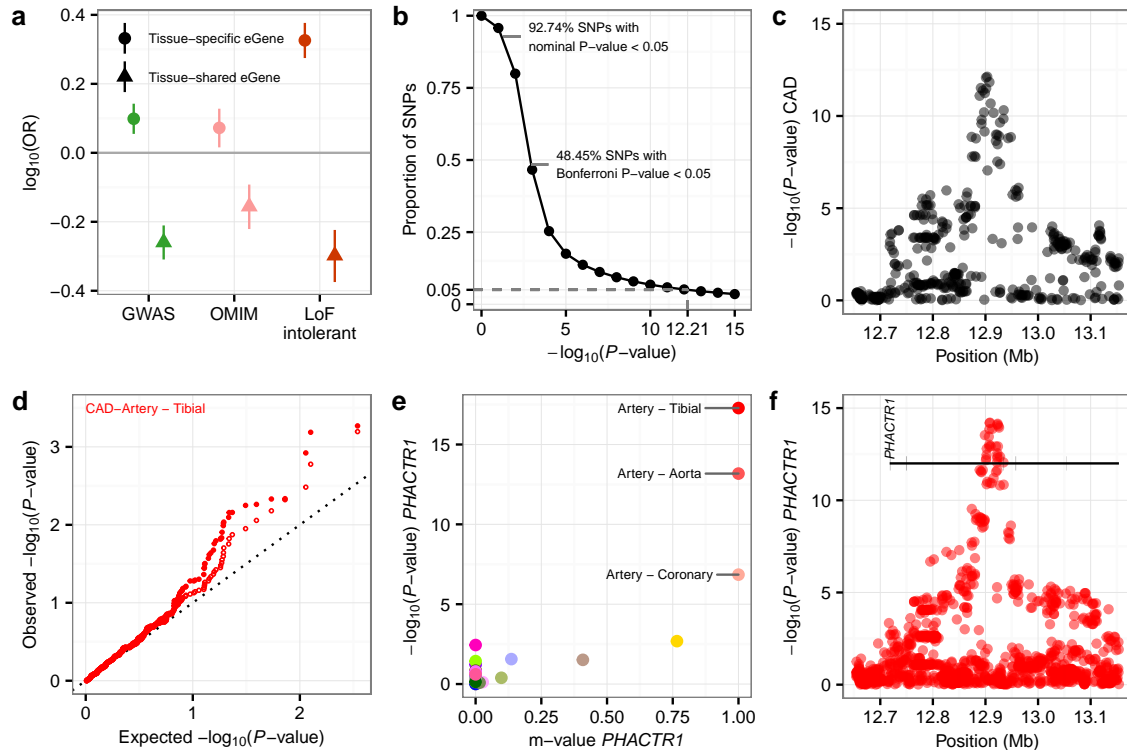
297 plied GTEx to the *PHACTR1* locus which is associated with a range of complex traits, including  
298 myocardial infarction (MI)<sup>35</sup>, coronary artery disease (CAD)<sup>36–38</sup>, cervical artery dissection<sup>39</sup>, and  
299 migraines<sup>40</sup> (Fig. 5c). Notably, the CAD and MI risk allele (G) at rs9349379 is protective for cer-  
300 vical artery dissection and migraines. Initial targeted analyses<sup>41</sup> demonstrated that the CAD risk  
301 allele (G) at rs9349379 is associated with decreased expression of *PHACTR1* in coronary arteries.

302 To investigate the mechanism and tissue of action of this pleiotropic SNP, we characterized  
303 the effect of rs9349379 across the 44 GTEx tissues. rs9349379G was strongly associated with  
304 decreased *PHACTR1* expression (Fig. 5e; meta-analysis  $P < 2.2 \times 10^{-16}$ ), with a tissue-specific  
305 eQTL effect observed only in aorta, coronary, and tibial arteries (Fig. 5e; m-value  $\geq 0.9$ ), where  
306 the risk allele expression is 72%, 57% and 65% of the protective allele expression, respectively.  
307 *PHACTR1*, *TBC1D7*, and the nearby noncoding RNA, *RP1-257A7.5*, were the only genes within  
308 1 Mb associated with genotype at rs9349379 in any tissue. Notably, the tissue specificity of the  
309 eQTL effect was not mirrored in the tissue specificity of *PHACTR1* gene expression (Extended  
310 Data Fig. 17). Colocalization analysis in arterial tissues indicated that rs9349379 is likely the  
311 variant responsible for both the GWAS and the eQTL signal in the locus (Fig. 5f; RTC = 1,  
312 eCAVIAR = 0.95)<sup>34,42</sup>. Applying the PrediXcan method<sup>12</sup> to the BioVU repository<sup>43</sup>, we found  
313 that genetically predicted decreased *PHACTR1* expression in coronary and aorta arteries was  
314 associated with tachycardia (meta-analysis  $P < 10^{-6}$ ), whereas genetically predicted increased  
315 *PHACTR1* expression was associated with migraines ( $P = 1.2 \times 10^{-7}$ ). *PHACTR1* is the sole  
316 gene in the locus that was implicated by PrediXcan in BioVU, using arterial tissues, for either  
317 trait. These results suggest that the pleiotropic effects of rs9349379 are driven by a consistent,  
318 tissue-specific molecular phenotype that causes diverse downstream consequences.

## 319 Discussion

320 The most immediate effects of functional genetic variation are on molecular phenotypes. Combining  
321 trait and disease associated variants with molecular QTL data has been a successful strategy for  
322 resolving causal genes and tissues<sup>44</sup>. In particular, these approaches have provided key information  
323 on human-specific traits and therapeutic interventions<sup>11,45,46</sup>. While the pilot phase of the GTEx  
324 project identified cis-eQTLs in nine tissues, the GTEx V6p collection has been expanded to 44  
325 tissues providing a wealth of additional cis-eQTL discoveries. These data facilitate both systematic  
326 and targeted interpretation of the functional consequences of genetic variants across a range of  
327 biological contexts.

328 We found a pervasive effect of common regulatory variation on the vast majority of human  
329 genes with a sizable proportion of genes having multiple independent loci associated with their  
330 expression levels. By combining cis-eQTL data across tissues, we demonstrated that GTEx V6p  
331 data may be used to enable cis-eQTL discovery in tissues with limited sample sizes. Many of the  
332 largest, primary effects are shared across tissues. Additionally, we observed that both secondary  
333 cis-eQTLs and cis-eQTLs from deeply sampled tissues exhibit more tissue-specificity. cis-eQTLs  
334 are enriched in both tissue-specific enhancers and promoters and patterns of regulatory element  
335 overlap are predictive of tissue sharing for cis-eQTLs. Secondary cis-eQTLs were as enriched as  
336 primary cis-eQTLs for Hi-C contacts suggesting a direct effect on gene expression facilitated by  
337 chromosome looping and local nuclear organization. Furthermore, we demonstrated that tissue-  
338 specific genes and eQTLs have larger effect sizes, and we have presented a large resource of allelic  
339 expression data that demonstrates correlated estimates of tissue-sharing and effect size estimates  
340 with eQTLs. Overall, these observations illustrate the systemic effects of regulatory variants and  
341 inform eQTL study design by highlighting the unique contributions of tissue-specific eQTLs that



**Figure 5. Intersection of cis-eQTLs with GWAS.** (a) Enrichment of tissue-specific and tissue-shared eGenes in disease and loss of function mutation intolerant genes. eGenes were defined in each tissue using METASOFT ( $m\text{-value} \geq 0.9$ ). Tissue-specific and shared eGenes were defined as eGenes in the bottom and top 10% of the distribution of proportion of tissues with an eQTL effect, respectively. Points represent the log odds ratio for enrichment of the eGene category in each gene list. Bars represent 95% confidence intervals (Fisher's exact test). (b) Proportion of eQTLs discovered as a function of  $P$ -value cutoffs. Nearly 93% of all SNPs passed a nominal significance threshold of 0.05. More than 48% of all SNPs passed a Bonferroni threshold defined as the nominal threshold divided by the number of tissues (44). To control the type I error rate at 5%, a stringent cutoff of  $10^{-12}$  is needed. (c) CAD association significance (y-axis) for all SNPs within 250 kb of the sentinel SNP (x-axis). (d) Quantile-quantile plot for CAD GWAS associations. Observed GWAS  $P$ -values (y-axis) plotted as a function of expected  $P$ -values (x-axis), for the top 1,000 eQTLs (closed circles) and MAF and distance matched SNPs (open circles). (e) For each tissue, the METASOFT  $m$ -value (x-axis) of the lead CAD SNP is plotted against the single-tissue eQTL association significance (y-axis). Points are colored by tissue. (f) eQTL association significance (y-axis) for *PHACTR1* for all SNPs within 250 kb of the sentinel CAD SNP (x-axis).

342 can only be identified through broad tissue sampling.

343 The wealth of cis-eQTLs identified in this study bears important implications for GWAS inter-  
 344 pretation. We demonstrated that 92.7% of variants tested in our study have a nominally significant  
 345 association with expression ( $P$ -value < 0.05) and that approximately 10% of eVariants may change  
 346 their top associated gene when tested in another tissue. Given the abundance of associations for  
 347 any variant, care must be taken in using cis-eQTL data to propose novel biological mechanisms

348 for disease-associated variants. GWAS variants are enriched among tissue-specific cis-eQTLs, high-  
349 lighting the necessity for sampling of diverse tissues. The wealth of associations in GTEx may  
350 further aid in selecting candidate causal tissues where multiple GWAS signals for specific traits  
351 are enriched. Within these targeted tissues, colocalization strategies that combine locus-specific  
352 trait and expression association information are required to understand the underlying biological  
353 mechanism. We showed that a combination of these approaches may be used to interpret a GWAS  
354 signal of relevance to coronary artery disease, and we revealed novel tissue-specific biology identified  
355 through the analysis of GTEx V6p data.

356 Together, cis-eQTL data in GTEx V6p provide the most comprehensive characterization of the  
357 local effects of regulatory variation to date. We expect that these data will be of considerable  
358 utility for the interpretation of gene regulatory mechanisms, human evolution, and complex trait  
359 and disease biology.

## 360 Online Methods

### 361 Sample procurement

362 The GTEx V6p eQTL analysis freeze represents 44 distinct tissue sites collected from 449 post-  
363 mortem donors representing a total of 7,051 tissues. All human subjects were deceased donors.  
364 Informed consent was obtained for all donors via next-of-kin consent to permit the collection and  
365 banking of de-identified tissue samples for scientific research. Complete descriptions of the donor  
366 enrollment and consent process, as well as biospecimen procurement, methods, sample fixation  
367 and histopathological review procedures were previously described<sup>1,47</sup>. Briefly, whole blood was  
368 collected from each donor, along with fresh skin samples, for DNA genotyping, RNA expression  
369 and culturing of lymphoblastoid and fibroblast cells, and shipped overnight to the GTEx Labo-  
370 ratory Data Analysis and Coordination Center (LDACC) at the Broad Institute. Two adjacent  
371 aliquots were then prepared from each sampled tissue and preserved in PAXgene tissue kits. One  
372 of each paired sample was embedded in paraffin (PFPE) for histopathological review, the second  
373 was shipped to the LDACC for processing and molecular analysis. Brains were collected from  
374 approximately 1/3rd of the donors, and were shipped on ice to the brain bank at the University of  
375 Miami, where 11 brain sub-regions were sampled and flash frozen. These samples were also shipped  
376 to the LDACC at the Broad Institute for processing and analysis.

377 All DNA genotyping was performed on blood-derived DNA samples, unless unavailable, in which  
378 case a tissue-derived DNA sample was substituted. RNA was extracted from all tissues, but quality  
379 varied<sup>1</sup>. RNA sequencing was performed on all samples with a RIN score of 5.7 or higher and with  
380 at least 500ng of total RNA. Nucleic acid isolation protocols, and sample QC metrics applied, are  
381 as described in<sup>1</sup>.

### 382 Data production

383 RNA was isolated from a total of 9,547 postmortem samples from 54 tissue types from up to  
384 550 individuals. 44 tissues were sampled from at least 70 individuals: 31 solid-organ tissues, 10  
385 brain subregions with two duplicate regions (cortex and cerebellum), whole blood, and two cell  
386 lines derived from donor blood and skin samples. Each tissue had a different number of unique  
387 samples. Non-strand specific, polyA+ selected RNA-seq libraries were generated using the Illu-  
388 mina TruSeq protocol. Libraries were sequenced to a median depth of 78 million 76-bp paired  
389 end reads. RNA-seq reads were aligned to the human genome (hg19/GRCh37) using TopHat<sup>48</sup>  
390 (v1.4) based on GENCODE v19 annotations<sup>14</sup>. This annotation is available on the GTEx Por-



391 tal (gencode.v19.genes.v6p\_model.patched\_contigs.gtf.gz). Gene-level expression was estimated as  
392 reads per kilobase of transcript per million mapped reads (RPKM) using RNA-SeQC on uniquely  
393 mapped, properly paired reads fully contained with exon boundaries and with alignment distances  
394  $\leq 6$ . Samples with less than 10 million mapped reads or with outlier expression measurements  
395 based on the D-statistic were removed<sup>49</sup>.

396 DNA isolated from blood was used for genotyping. 450 individuals were genotyped using  
397 Illumina Human Omni 2.5M and 5M Beadchips. Genotypes were phased and imputed with  
398 SHAPEIT<sup>50</sup> and IMPUTE<sup>51</sup>, respectively, using multi-ethnic panel reference from 1000 Genomes  
399 Project Phase 1 v3<sup>52</sup>. Variants were excluded from analysis if they: (1) had a call rate  $< 95\%$ ; (2)  
400 had minor allele frequencies  $< 1\%$ ; (3) deviated from Hardy-Weinberg Equilibrium ( $P < 10^{-6}$ ); or  
401 (4) had an imputation info score less than 0.4.

## 402 cis-eQTL mapping

403 We conducted cis-eQTL mapping within the 44 tissues with at least 70 samples each. Only genes  
404 with  $\geq 10$  individuals with expression estimates  $> 0.1$  RPKM and an aligned read count  $\geq 6$   
405 within each tissue were considered significantly expressed and used for cis-eQTL mapping. Within  
406 each tissue, the distribution of RPKMs in each sample was quantile-transformed using the average  
407 empirical distribution observed across all samples. Expression measurements for each gene in each  
408 tissue were subsequently transformed to the quantiles of the standard normal distribution. The  
409 effects of unobserved confounding variables on gene expression were quantified with PEER<sup>53</sup>, run  
410 independently for each tissue. 15 PEER factors were identified for tissues with less than 150  
411 samples; 30 for tissues with sample sizes between 150 and 250; and 35 for tissues with more than  
412 250 tissues.

413 Within each tissue, cis-eQTLs were identified by linear regression, as implemented in FastQTL<sup>13</sup>,  
414 adjusting for PEER factors, gender, genotyping platform, and three genotype-based PCs. We  
415 restricted our search to variants within 1 Mb of the transcription start site of each gene and, in  
416 the tissue of analysis, minor allele frequencies  $\geq 0.01$  with the minor allele observed in at least  
417 10 samples. Nominal  $P$ -values for each variant-gene pair were estimated using a two-tailed t-test.  
418 Significance of the most highly associated variant per gene was estimated by adaptive permutation  
419 with the setting "--permutate 1000 10000". These empirical  $P$ -values were subsequently corrected  
420 for multiple testing across genes using Storey's q-value method<sup>15</sup>.

421 To identify the list of all significant variant-gene pairs associated with eGenes, a genome-wide  
422 empirical  $P$ -value threshold,  $p_t$ , was defined as the empirical  $P$ -value of the gene closest to the 0.05  
423 FDR threshold.  $p_t$  was then used to calculate a nominal  $P$ -value threshold for each gene based on  
424 the beta distribution model (from FastQTL) of the minimum  $P$ -value distribution  $f(p_{\min})$  obtained  
425 from the permutations for the gene. Specifically, the nominal threshold was calculated as  $F^{-1}(p_t)$ ,  
426 where  $F^{-1}$  is the inverse cumulative distribution. For each gene, variants with a nominal  $P$ -value  
427 below the gene-level threshold were considered significant and included in the final list of variant-  
428 gene pairs.

## 429 Multi-tissue cis-eQTL mapping

430 To increase sensitivity of cis-eQTL detection, in particular of cis-eQTLs with smaller effect sizes, we  
431 ran METASOFT<sup>54</sup>, a meta-analysis method, on all variant-gene pairs that were significant (FDR  
432  $< 5\%$ ) in at least one of the 44 tissues based on the single-tissue results from FastQTL. The goal  
433 of this analysis was to gain power to discover additional tissues for a cis-eQTL. A random effects  
434 model in METASOFT (called RE2), designed to find loci with effects that may have heterogeneity

435 between datasets/tissues (and assumes estimates are independent and consistent in effect direction)  
436 was used<sup>19</sup>. The posterior probability that an eQTL effect exists in a given tissue, or m-value<sup>54</sup>,  
437 was calculated for each variant-gene pair and tissue tested. A significance cutoff of m-value  $\geq 0.9$   
438 was used to discover high-confidence cis-eQTLs.

439 We applied a separate hierarchical multiple testing correction method to identify multi-tissue  
440 eGenes. First, we constructed a  $P$ -value for each eGene across tissues using the Simes combination  
441 rule<sup>55</sup> on the tissue-specific beta-approximation  $P$ -values provided by FastQTL. Storey's q-value  
442 method<sup>15</sup> was then used to identify eGenes that are active in any tissue. To identify the specific  
443 tissues in which these eGenes are regulated, we applied the Benjamini and Bogomolov procedure<sup>56</sup>  
444 at the 0.05 level. This approach not only allowed us to control the FDR for the discovery of eGenes  
445 across tissues and the expected average proportion of false tissue discoveries across these eGenes,  
446 but also to gain power to detect eGenes in tissues with smaller sample sizes when there is evidence  
447 from other tissues supporting their regulation.

## 448 Independent cis-eQTL mapping

### 449 *Single-tissue analysis*

450 Multiple independent signals for a given expression phenotype were identified by forward stepwise  
451 regression followed by a backwards selection step. The gene-level significance threshold was set to  
452 be the maximum beta-adjusted  $P$ -value (correcting for multiple-testing across the variants) over  
453 all eGenes in a given tissue. At each iteration, we performed a scan for cis-eQTLs using FastQTL,  
454 correcting for all previously discovered variants and all standard GTEx covariates. If the beta  
455 adjusted  $P$ -value for the lead variant was not significant at the gene-level threshold, the forward  
456 stage was complete and the procedure moved on to the backward stage. If this  $P$ -value was sig-  
457 nificant, the lead variant was added to the list of discovered cis-eQTLs as an independent signal  
458 and the forward step moves on to the next iteration. The backwards stage consisted of testing  
459 each variant separately, controlling for all other discovered variants. To do this, for each eVariant,  
460 we scanned for cis-eQTLs controlling for standard covariates and all other eVariants. If no variant  
461 was significant at the gene-level threshold the variant in question was dropped, otherwise the lead  
462 variant from this scan, which controls for all other signals found in the forward stage, was chosen  
463 as the variant that represents the signal best in the full model.

### 465 *Multi-tissue analysis*

466 We ran a modified version of forward stepwise regression to select an ordered list of independent  
467 variants associated with a given gene across all tissues types. In each step  $k$ , we identify variants  
468 associated with expression of each gene across tissues, and refer to these as the tier  $k$  variants. In  
469 each tier  $k$ , for each tissue, Matrix-eQTL was run independently for each gene that had a variant  
470 added to the model at every previous step  $1..k-1$  (all genes are assessed in tier 1). In each tier, any  
471 significant variants identified in tiers  $1..k-1$  are included as covariates. Significant tier  $k$  variants  
472 were assessed as follows. For each tissue, we obtained gene-level  $P$ -values for tier  $k$  via eigenMT<sup>57</sup>.  
473 Genome-wide significance of multiple independent variants per gene (in each tissue independently)  
474 was assessed via Benjamini-Hochberg (FDR  $< 0.05$ ) for all gene-level  $P$ -values tested in tier  $k$   
475 combined with all those tested in previous tiers<sup>58</sup>. To identify the cross-tissue tier  $k$  variant for  
476 a given gene, we selected the variant (out of all variants genome-wide significant for the gene in  
477 at least one tissue) with the smallest geometric mean  $P$ -value (across tissues). If no variant was  
478 genome-wide significant, no cross-tissue tier  $k$  variant was selected for that gene, and that gene will  
479 be estimated to have  $k-1$  total independent cross-tissue variants. If a particular tissue's tier  $j$   
480 genome-wide significant variant for a particular gene differed from the cross-tissue tier  $j$  variant for

481 the same gene, the  $P$ -value of that tissue's tier  $j$  genome-wide significant variant was used in the  
482 Benjamini-Hochberg procedure. If a particular gene's cross-tissue variant for tier  $k$  does not meet  
483 genome-wide significance in all tissues in the tier  $(k + 1)$  step due to increased multiple testing,  
484 that gene will be conservatively considered to have  $(k - 1)$  independent cross-tissue variants.

## 485 Allele-specific expression

### 486 *Data generation*

487 For each sample, allele-specific RNA-seq read counts were generated at all heterozygous SNPs with  
488 the GATK ASEReadCounter tool using default settings<sup>30</sup>. Only uniquely mapping reads with a  
489 base quality  $\geq 10$  at the SNP were counted, and only those SNPs with coverage of at least 8  
490 reads were reported. Unless otherwise mentioned, SNPs that met any of the following criteria  
491 were flagged and removed from downstream analyses: (1) UCSC 50mer mappability of  $< 1$ , (2)  
492 simulation-based evidence of mapping bias<sup>59</sup>, (3) heterozygous genotype not supported by RNA-  
493 seq data across all samples for that subject (test adapted from Castel et al.<sup>30</sup>). Phasing between  
494 variants was determined using population phasing, and for some analyses was used to aggregate  
495 allelic counts across variants. Full ASE data is available through dbGAP.

### 496 *Modeling patterns of ASE sharing across tissues*

497 We used a beta-binomial mixture to model ASE across tissues, with each component corresponding  
498 to a distinct mode of allelic imbalance. The model was learned independently for each heterozygous  
499 coding SNP in each individual. Optimization was performed using five independent initial param-  
500 eters values. The number of components in the mixture model,  $K$ , was selected using Bayesian  
501 Information Criterion (BIC). Variance of the BIC was estimated by bootstrapping and the most  
502 parsimonious model within one standard deviation of the global minimum BIC model was chosen  
503 as the optimal model.

504  
505 Individuals with RNA-seq data from at least 20 tissues were included in the analysis ( $N = 131$ ).  
506 The most highly expressed, coding, heterozygous SNP in each gene was selected. Genes with at least  
507 30 reads in at least two tissues and at least one tissue with allelic imbalance (defined as  $P < 10^{-3}$   
508 under a binomial null model) were included in the analysis. In total 207,943 SNPs spanning 13,030  
509 genes were modeled using 1, 2, 3, and 4 modes of allelic imbalance. 2% of SNPs exhibited more  
510 than one pattern of allelic imbalance across tissues ( $K=2$ : 4219 cases,  $K=3$ : 64 cases, and  $K=4$ : 4  
511 cases). These multimodal cases involved 2,226 genes across individuals. SNPs with bimodal ( $K=2$ )  
512 pattern of allelic expression were used to derive estimates of ASE tissue sharing. Tissues with less  
513 than 100 cases were excluded from analysis. Tissue similarity was measured as the proportion of  
514 times two tissues exhibit the same mode of allelic imbalance.

## 515 Effect size estimation

### 516 *cis-eQTL effect size*

517 cis-eQTL effect size was defined as the ratio between the expression of the haplotype carrying the  
518 alternative eVariant allele to the one carrying the reference allele in  $\log_2$  scale and was calculated  
519 using the method presented in (Mohammadi et al. in preparation). In short, the model assumes  
520 an additive model of expression in which the total expression of a gene in a given genotype group  
521 is the sum of the expression of the two haplotypes:  $e(\text{genotype}) = 2e_r, e_r + e_a, 2e_a$ , for reference  
522 homozygotes, heterozygotes, and alternate homozygotes, respectively, where  $e_r$  is expression of the  
523 haplotype carrying the reference allele and  $e_a$ , expression of the haplotype carrying the alternative  
524 allele is:  $e_a = ke_r$  where  $0 < k < \infty$ .

525 cis-eQTL effect size is represented in  $\log_2$  scale as  $s = \log_2 k$ , and is capped at 100-fold to  
526 avoid outliers ( $|s| < \log_2 100$ ). Expression counts were retrieved for all top eGenes in all tissues  
527 and PEER corrected. Data was log-transformed with one pseudo-count to stabilize the variance.  
528 The model was fit using non-linear least squares to derive maximum likelihood estimates of the  
529 model parameters  $k$  and  $e_r$ . A similar maximum likelihood approach with additive effects and  
530 multiplicative errors (prior to log transformation)<sup>60</sup> was compared in several tissues to the effect  
531 size estimates reported here, exhibiting rank correlation 0.98. Confidence intervals for the effect  
532 sizes were derived using bias corrected and accelerated (BCa) bootstrap with 100 samples.

533 For all analyses in a given tissue only the top eVariant per eGene was used. Only those eQTLs  
534 whose 95% confidence interval of the effect size estimate did not overlap zero were used for down-  
535 stream analysis. To control for differences in power due to eVariant allele frequency, the effect of  
536 MAF on eQTL effect size was estimated using LOWESS regression (Matlab function `malowess`:  
537 `span=0.2, robust=true`), and was subtracted from the effect sizes on a per tissue basis.

538

### 539 *ASE effect size*

540 For each sample, haplotypic expression at all eGenes was calculated by summing counts from all  
541 phased, heterozygous SNPs. For a given cis-eQTL variant, assume  $x_i$  is the number of RNA-seq  
542 reads aligned to one haplotype, and  $y_i$  is the total number of reads aligned to either haplotype  
543 in the  $i$ th individual. Regulatory effect size of the cis-eQTL was calculated as median log-ratio:  
544  $s(x, y) = \text{median}[\log_2(x_i) \log_2(y_i - x_i)]$ . Effect sizes were calculated for cis-eQTLs for which 10  
545 or more individuals with  $y_i \geq 10$ , and the effect sizes were constrained to be less than 100 fold  
546 ( $|s(x, y)| < \log_2 100$ ). Confidence intervals for the effect sizes were derived using BCa bootstrap  
547 with 100 samples.

### 548 **cis-eQTL fine-mapping**

#### 549 *CaVEMaN*

550 We utilized CaVEMaN (Causal Variant Evidence Mapping with Non-parametric resampling) to  
551 estimate the probability that an eVariant was a causal variant (Brown et al., in preparation). We  
552 used a non-GTEX reference cis-eQTL dataset from subcutaneous adipose tissue, lymphoblastoid  
553 cell lines, skin and whole blood, to simulate causal variants with characteristics matching genuine  
554 cis-eQTLs<sup>61</sup> (effect size, residual variance, minor allele frequency, and distance to the TSS). For  
555 each simulation, we calculated the proportion of times the simulated causal variant was among the  
556  $i$ th most significant eVariants and denoted this proportion as  $p_i$ . For each lead eVariant in GTEX,  
557 we generated a single-signal expression phenotype by controlling for all covariates fitted in the cis-  
558 eQTL mapping and all other eVariants for the gene except the eVariant whose signal we wished  
559 to preserve. These data were sampled with replacement 10,000 times and cis-eQTL mapping was  
560 performed on each resample. The proportion of times a given eVariant was ranked  $i$  was calculated,  
561 denoted  $F_i$ . The CaVEMaN score is then defined as  $\sum_{i=1}^{10} p_i \cdot F_i$ . To calibrate CaVEMaN scores,  
562 across all genes and tissues simulated (removing blood as an outlier) we divided the CaVEMaN  
563 scores of the peak variants into twenty quantiles. Within each quantile, we calculated the propor-  
564 tion of times the lead variant was the causal variant and then drew a monotonically increasing  
565 smooth spline from the origin, through the 20 quantiles, to the point (1, 1) using the `gsl` interpolate  
566 functions with the `steffen` method (`gsl-2.1`, <https://www.gnu.org/software/gsl/>). This function  
567 provides our mapping of CaVEMaN score of the lead SNP onto the probability it is the causal  
568 variant, calibrated using the simulations.

569

#### 570 *CAVIAR*

571 CAVIAR (CAusal Variants Identification in Associated Regions)<sup>29</sup> uses LD structure to model the  
572 observed marginal test statistics for each eGene as following a multivariate normal distribution  
573 (MVN). Applying this model, CAVIAR can define a credible set containing all causal variants  
574 with probability  $\rho$ . To define these credible sets in each tissue, we used a threshold of  $\rho = 90\%$ .  
575 We utilized eCAVIAR (eQTL and GWAS CAVIAR) to colocalize GWAS and eQTL studies for  
576 detection of the target genes and relevant tissues<sup>42</sup>. eCAVIAR computes a posterior probability  
577 of a variant identified as causal in both GWAS and eQTL studies. We used a cut-off of 1% for  
578 colocalization posterior probability based on observations from previous simulations<sup>42</sup>.

579 To test for a significant relationship between tissue sample size and the size of the 90% credible  
580 set, we compared credible set sizes for the top 100 single-tissue cis-eQTLs across tissues (Extended  
581 Data Fig. 11a). We further combined cis-eQTL sharing results from METASOFT with CAVIAR's  
582 90% credible sets to test if tissues with shared cis-eQTLs could be used to fine-map the causal  
583 variant. Here, from the initial METASOFT results, we identified the top shared cis-eQTL for each  
584 eGene by selecting the cis-eQTL with the smallest RE2  $P$ -value. For eGenes that had a shared  
585 cis-eQTL or a tissue-specific cis-eQTL, we compared the intersection of the 90% credible sets within  
586 and between each group (Extended Data Fig. 11b).

## 587 **Overlap of tissue-specific and tissue-shared eGenes with disease genes**

588 For each gene tested for multi-tissue eQTLs using METASOFT, we calculated the proportion  
589 of tissues for which the gene had a strong eQTL effect (i.e. the proportion of tissues with  $m$ -  
590 value  $\geq 0.9$ ). We defined tissue-specific eGenes as genes in the bottom 10% of the empirical  
591 distribution of this proportion. Similarly, we defined tissue-shared eGenes as genes in the top 10%  
592 of this distribution. We examined the enrichment of tissue-specific and tissue-shared eGenes in  
593 six different gene lists: the NHGRI-EBI GWAS Catalog<sup>62</sup>, the Online Mendelian Inheritance in  
594 Man (OMIM) database<sup>63</sup>, the Orphanet database, the ClinVar database<sup>64</sup>, the list of genes with  
595 clinically actionable variants reported by the American College of Medical Genetics (ACMG)<sup>65</sup>,  
596 and the list of LoF intolerant genes from ExAC<sup>31</sup>. For the GWAS catalog, we restricted to only  
597 genes with reported associations. LoF intolerant genes were defined as those with a pLI score  $\geq$   
598 0.9 in ExAC<sup>31</sup>. We calculated odds ratios and 95% confidence intervals using Fisher's exact test  
599 for both tissue-specific and tissue-shared eGenes in each gene list. For the tissue-specific eGenes,  
600 we used as a background the remaining set of genes tested in METASOFT that were not classified  
601 as tissue-specific eGenes. Similarly, for tissue-shared eGenes, we used as a background the set of  
602 genes not classified as tissue-shared eGenes.

## 603 **GWAS analysis**

604 We have previously described the Regulatory Trait Concordance (RTC) score to assess whether a  
605 GWAS variant is tagging the same functional variant as a regulatory variant<sup>34</sup>. Briefly, for a cis-  
606 eQTL and GWAS variant located in the same region between recombination hotspots, we correct  
607 the eQTL phenotype (i.e., gene expression) for all the  $N$  variants within the region using linear  
608 regression, creating  $N$  pseudo-phenotypes from the residuals of the linear regression. We then test  
609 for eQTL association between the cis-eQTL variant and the  $N$  pseudo-phenotypes. These  $P$ -values  
610 are subsequently sorted (descending) and ranked, and the rank of the  $P$ -value arising from the  
611 cis-eQTL and GWAS variant corrected phenotype association is found and the score is defined as  
612  $(N - \text{GWASrank}) / N$ . The RTC score ranges from 0 to 1 with 1 indicating higher likelihood of  
613 shared functional effect.

614



## 615 *CAD GWAS*

616 Data on coronary artery disease and myocardial infarction have been contributed by CARDIo-  
617 GRAMplusC4D investigators and have been downloaded from [www.cardiogramplusc4d.org](http://www.cardiogramplusc4d.org).

## 618 **Data availability**

619 Genotype data from the GTEx V6p release are available in dbGaP (study accession phs000424.v6.p1;  
620 [www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000424.v6.p1](http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000424.v6.p1)). The  
621 VCFs for the imputed array data are in phg000520.v2.GTEx\_MidPoint\_Imputation.genotype-calls-  
622 vcf.c1.GRU.tar (the archive contains a VCF for chromosomes 1-22 and a VCF for chromosome X).  
623 Allelic expression data is also available in dbGap. Expression data (read counts and RPKM) and  
624 eQTL input files (normalized expression data and covariates for 44 the tissues) from the GTEx V6p  
625 release are available from the GTEx Portal (<http://gtexportal.org>). eQTL results are available  
626 from the GTEx Portal. In addition to results tables for the 44 tissues in this study (eGenes, signif-  
627 icant variant-gene pairs, and all variant-gene pairs tested), the portal provides multiple interactive  
628 visualization and data exploration features for eQTLs, including:

- 629 • eQTL box plot: displays variant-gene associations
- 630 • Gene eQTL Visualizer: displays all significant associations for a gene across tissues and linkage  
631 disequilibrium information
- 632 • Multi-tissue eQTL plot: displays multi-tissue posterior probabilities from meta-analysis against  
633 single-tissue association results
- 634 • IGV browser: displays eQTL across tissues and GWAS Catalog results for a selected genomic  
635 region

## 636 **References**

- 637 1. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue  
638 gene regulation in humans. *Science* **348**, 648–660 (2015).
- 639 2. Nica, A. C. *et al.* The architecture of gene regulatory variation across multiple human tissues:  
640 the MuTHER study. *PLoS Genetics* **7**, e1002003 (2011).
- 641 3. Dimas, A. S. *et al.* Common regulatory variation impacts gene expression in a cell type-  
642 dependent manner. *Science* **325**, 1246–1250 (2009).
- 643 4. Huang, G.-J. *et al.* High resolution mapping of expression QTLs in heterogeneous stock mice  
644 in multiple tissues. *Genome Research* **19**, 1133–1140 (2009).
- 645 5. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation  
646 in humans. *Nature* **501**, 506–511 (2013).
- 647 6. Battle, A. *et al.* Characterizing the genetic basis of transcriptome diversity through RNA-  
648 sequencing of 922 individuals. *Genome Research* **24**, 14–24 (2014).
- 649 7. Zhernakova, D. *et al.* Hypothesis-free identification of modulators of genetic risk factors.  
650 *bioRxiv* 033217 (2015).

- 651 8. Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and disease.  
652 *Nature reviews. Genetics* **16**, 197–212 (2015).
- 653 9. Westra, H.-J. & Franke, L. From genome to function by studying eQTLs. *Biochimica et*  
654 *biophysica acta* **1842**, 1896–1902 (2014).
- 655 10. Montgomery, S. B. & Dermitzakis, E. T. From expression QTLs to personalized transcrip-  
656 tomics. *Nature reviews. Genetics* **12**, 277–282 (2011).
- 657 11. Gibson, G., Powell, J. E. & Marigorta, U. M. Expression quantitative trait locus analysis  
658 for translational medicine. *Genome Medicine* **7**, 60 (2015).
- 659 12. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference  
660 transcriptome data. *Nature Genetics* **47**, 1091–1098 (2015).
- 661 13. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient  
662 QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).
- 663 14. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE  
664 Project. *Genome Research* **22**, 1760–1774 (2012).
- 665 15. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proceedings*  
666 *of the National Academy of Sciences* **100**, 9440–9445 (2003).
- 667 16. Flutre, T., Wen, X., Pritchard, J. & Stephens, M. A statistical framework for joint eQTL  
668 analysis in multiple tissues. *PLoS Genetics* **9**, e1003486 (2013).
- 669 17. Sul, J. H., Han, B., Ye, C., Choi, T. & Eskin, E. Effectively identifying eQTLs from multiple  
670 tissues by combining mixed model and meta-analytic approaches. *PLoS Genetics* **9**, e1003491  
671 (2013).
- 672 18. Li, G., Shabalin, A. A., Rusyn, I., Wright, F. A. & Nobel, A. B. An Empirical Bayes  
673 Approach for Multiple Tissue eQTL Analysis (2013). 1311.2948.
- 674 19. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis  
675 of genome-wide association studies. *American Journal of Human Genetics* **88**, 586–598  
676 (2011).
- 677 20. Peterson, C. B., Bogomolov, M., Benjamini, Y. & Sabatti, C. Treeqtl: hierarchical error  
678 control for eqtl findings. *Bioinformatics* **32**, 2556–2558 (2016).
- 679 21. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human  
680 epigenomes. *Nature* **518**, 317–330 (2015).
- 681 22. Farh, K. K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease  
682 variants. *Nature* **518**, 337–343 (2015).
- 683 23. Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans.  
684 *N. Engl. J. Med.* **373**, 895–907 (2015).
- 685 24. Harismendy, O. *et al.* 9p21 DNA variants associated with coronary artery disease impair  
686 interferon- $\gamma$  signalling response. *Nature* **470**, 264–268 (2011).

- 687 25. Brown, C. D., Mangravite, L. M. & Engelhardt, B. E. Integrative modeling of eQTLs and  
688 cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. *PLoS*  
689 *Genetics* **9**, e1003649 (2013).
- 690 26. Das, A. *et al.* Bayesian integration of genetics and epigenetics detects causal regulatory SNPs  
691 underlying expression variability. *Nature Communications* **6**, 8555 (2015).
- 692 27. Wang, D., Rendon, A. & Wernisch, L. Transcription factor and chromatin features predict  
693 genes associated with eQTLs. *Nucleic Acids Research* **41**, 1450–1463 (2013).
- 694 28. Wen, X. Molecular QTL discovery incorporating genomic annotations using Bayesian false  
695 discovery rate control. *Annals of Applied Statistics* (in press).
- 696 29. Hormozdiari, F., Kostem, E., Kang, E. Y., Pasaniuc, B. & Eskin, E. Identifying causal  
697 variants at loci with multiple signals of association. *Genetics* **198**, 497–508 (2014).
- 698 30. Castel, S. E., Levy Moonshine, A., Mohammadi, P., Banks, E. & Lappalainen, T. Tools  
699 and best practices for data processing in allelic expression analysis. *Genome Biology* **16**, 195  
700 (2015).
- 701 31. Exome Aggregation Consortium *et al.* Analysis of protein-coding genetic variation in 60,706  
702 humans. *bioRxiv* (2015).
- 703 32. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association  
704 studies using summary statistics. *PLoS Genetics* **10**, e1004383 (2014).
- 705 33. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex  
706 trait gene targets. *Nature Genetics* **48**, 481–487 (2016).
- 707 34. Nica, A. C. *et al.* Candidate causal regulatory effects by integration of expression QTLs with  
708 complex trait genetic associations. *PLoS Genetics* **6**, e1000895 (2010).
- 709 35. Myocardial Infarction Genetics Consortium *et al.* Genome-wide association of early-onset  
710 myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nature*  
711 *Genetics* **41**, 334–341 (2009).
- 712 36. CARDIoGRAMplusC4D Consortium *et al.* Large-scale association analysis identifies new  
713 risk loci for coronary artery disease. *Nature Genetics* **45**, 25–33 (2013).
- 714 37. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for  
715 coronary artery disease. *Nature Genetics* **43**, 333–338 (2011).
- 716 38. Lu, X. *et al.* Genome-wide association study in Han Chinese identifies four new susceptibility  
717 loci for coronary artery disease. *Nature Genetics* **44**, 890–894 (2012).
- 718 39. Debette, S. *et al.* Common variation in PHACTR1 is associated with susceptibility to cervical  
719 artery dissection. *Nature Genetics* **47**, 78–83 (2015).
- 720 40. Anttila, V. *et al.* Genome-wide meta-analysis identifies new susceptibility loci for migraine.  
721 *Nature Genetics* **45**, 912–917 (2013).
- 722 41. Beaudoin, M. *et al.* Myocardial Infarction-Associated SNP at 6p24 Interferes With MEF2  
723 Binding and Associates With PHACTR1 Expression Levels in Human Coronary Arteries.  
724 *Arteriosclerosis, Thrombosis, and Vascular Biology* **35**, 1472–1479 (2015).

- 725 42. Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes.  
726 *bioRxiv* 065037 (2016).
- 727 43. Denny, J. C. *et al.* Systematic comparison of phenome-wide association study of electronic  
728 medical record data and genome-wide association study data. *Nature Biotechnology* **31**,  
729 1102–1110 (2013).
- 730 44. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.  
731 *Nucleic Acids Research* **42**, D1001–6 (2014).
- 732 45. Obeidat, M. *et al.* Molecular mechanisms underlying variations in lung function: a systems  
733 genetics analysis. *The Lancet. Respiratory Medicine* **3**, 782–795 (2015).
- 734 46. Folkersen, L. *et al.* Applying genetics in inflammatory disease drug discovery. *Drug discovery*  
735 *today* **20**, 1176–1181 (2015).
- 736 47. Carithers, L. J. *et al.* A Novel Approach to High-Quality Postmortem Tissue Procurement:  
737 The GTEx Project. *Biopreservation and Biobanking* **13**, 311–319 (2015).
- 738 48. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-  
739 Seq. *Bioinformatics* **25**, 1105–1111 (2009).
- 740 49. Wright, F. A. *et al.* Heritability and genomics of gene expression in peripheral blood. *Nature*  
741 *Genetics* **46**, 430–437 (2014).
- 742 50. O’Connell, J. *et al.* A general approach for haplotype phasing across the full spectrum of  
743 relatedness. *PLoS Genetics* **10**, e1004234 (2014).
- 744 51. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes.  
745 *G3* **1**, 457–470 (2011).
- 746 52. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092  
747 human genomes. *Nature* **491**, 56–65 (2012).
- 748 53. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of  
749 expression residuals (PEER) to obtain increased power and interpretability of gene expression  
750 analyses. *Nature Protocols* **7**, 500–507 (2012).
- 751 54. Han, B. & Eskin, E. Interpreting meta-analyses of genome-wide association studies. *PLoS*  
752 *Genetics* **8**, e1002555 (2012).
- 753 55. Simes, R. J. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*  
754 **73**, 751–754 (1986).
- 755 56. Benjamini, Y. & Bogomolov, M. Selective inference on multiple families of hypotheses.  
756 *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**, 297–318  
757 (2014).
- 758 57. Davis, J. R. *et al.* An Efficient Multiple-Testing Adjustment for eQTL Studies that Accounts  
759 for Linkage Disequilibrium between Variants. *American Journal of Human Genetics* **98**, 216–  
760 224 (2016).

- 761 58. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and  
762 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series*  
763 *B Methodological* **57**, 289–300 (1995).
- 764 59. Panousis, N. I., Gutierrez-Arcelus, M., Dermitzakis, E. T. & Lappalainen, T. Allelic mapping  
765 bias in RNA-sequencing is not a major confounder in eQTL studies. *Genome Biology* **15**,  
766 467 (2014).
- 767 60. Palowitch, J., Shabalín, A., Zhou, Y., Nobel, A. B. & Wright, F. A. Estimation of in-  
768 terpretable eqtl effect sizes using a log of linear model. *arXiv preprint arXiv:1605.08799*  
769 (2016).
- 770 61. Buil, A. *et al.* Gene-gene and gene-environment interactions detected by transcriptome  
771 sequence analysis in twins. *Nature Genetics* **47**, 88–91 (2015).
- 772 62. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.  
773 *Nucleic Acids Res.* **42**, D1001–1006 (2014).
- 774 63. Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online  
775 Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disor-  
776 ders. *Nucleic Acids Res.* **33**, D514–517 (2005).
- 777 64. Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants.  
778 *Nucleic Acids Res.* **44**, D862–868 (2016).
- 779 65. Green, R. C. *et al.* ACMG recommendations for reporting of incidental findings in clinical  
780 exome and genome sequencing. *Genet. Med.* **15**, 565–574 (2013).

## 781 Acknowledgments

782 The Genotype-Tissue Expression (GTEx) project was supported by the Common Fund of the Of-  
783 fice of the Director of the National Institutes of Health ([commonfund.nih.gov/GTEx](http://commonfund.nih.gov/GTEx)). Additional  
784 funds were provided by the National Cancer Institute (NCI), National Human Genome Research  
785 Institute (NHGRI), National Heart, Lung, and Blood Institute (NHLBI), National Institute on  
786 Drug Abuse (NIDA), National Institute of Mental Health (NIMH), and National Institute of Neu-  
787 rological Disorders and Stroke (NINDS). Donors were enrolled at Biospecimen Source Sites funded  
788 by NCISAIC-Frederick, Inc. (SAIC-F) subcontracts to the National Disease Research Interchange  
789 (10XS170) and Roswell Park Cancer Institute (10XS171). The Laboratory, Data Analysis, and Co-  
790 ordinating Center (LDACC) was funded through a contract (HHSN268201000029C) to The Broad  
791 Institute, Inc. Biorepository operations were funded through an SAIC-F subcontract to Van An-  
792 del Institute (10ST1035). Additional data repository and project management were provided by  
793 SAIC-F (HHSN261200800001E). The Brain Bank was supported by a supplement to University of  
794 Miami grant DA006227.

795 We thank Anshul Kundaje and Oana Ursu for input on the HI-C analysis. J.R.D. is supported by  
796 a Lucille P. Markey Biomedical Research Stanford Graduate Fellowship. J.R.D., Z.Z., and N.A.T.  
797 acknowledge the Stanford Genome Training Program (SGTP; NIH/NHGRI T32HG000044). Z.Z. is  
798 also supported by the National Science Foundation (NSF) GRFP (DGE-114747). L.F. is supported  
799 by the Stanford Center for Computational, Evolutionary, and Human Genomics (CEHG). D.G.M. is  
800 supported by a "la Caixa"-Severo Ochoa pre-doctoral fellowship. D.M. is supported by NIH grants  
801 U54DK105566 and R01GM104371. B.J.S is supported by NIH training grant T32 GM007057.



802 E.K.T is supported by a Hewlett-Packard Stanford Graduate Fellowship and a doctoral scholarship  
803 from the Natural Science and Engineering Council of Canada. T.S. is supported by a National Sci-  
804 ence Foundation Graduate Research Fellowship (DGE-1656518). A.B. is supported by the Searle  
805 Scholars Program and NIH grant 1R01MH109905. A.B., C.D.Bu. and S.B.M are supported by  
806 NIH grant R01HG008150 (NHGRI; Non-Coding Variants Program). S.B.M. and C.D.Bu. are sup-  
807 ported by NHGRI grants U01HG007436 and U01HG009080. A.B., C.D.Bu., E.T.D., S.E.C., T.L,  
808 and S.B.M. are supported by NIH grants R01MH101814 (NIH Common Fund; GTEEx Program).  
809 C.D.Br., Y.P., and B.E.E. are supported by NIH grant R01MH101822. G.L., A.B.N, J.J.P., A.S., Y-  
810 H.Z., and F.A.W. are supported by NIH grants R01MH101819, R01HG009125, and R21HG007840.  
811 T.L. and P.M. are supported by NIH grant R01MH106842. T.L. is supported by the NIH grant  
812 UM1HG008901. T.L. and S.E.C. are supported by the NIH contract HHSN2682010000029C. C.S.  
813 and C.B.P. are supported by NIH grant R01 MH101782.

## 814 Contributions

815 F.A., A.V.S., N.J.C., G.G., K.G.A., and E.T.D. contributed to study design. F.A., A.A.B., S.E.C.,  
816 J.R.D., P.M., A.V.S., Z.Z., N.S.A., L.F., E.R.G., E.G., M.J.G., Y.H., F.H., X.L., X.Li., B.L.,  
817 D.G-M., H.O., J.J.P., Y.P., C.B.P., G.Q., S.R., A.A.S., T.C.S., B.J.S., T.J.S., N.A.T., E.K.T.,  
818 H.Z., Y-H.Z., A.B., C.D.Bu., B.E.E., E.E., M.K., G.L., D.G.M., A.B.N., C.S., X.W., F.A.W.,  
819 T.L., K.G.A., E.T.D., C.D.Br., and S.B.M. contributed analysis. C.D.Br. and S.B.M. wrote the  
820 manuscript. F.A., A.A.B., S.E.C., J.R.D., P.M., A.V.S, Z.Z., C.B.P., B.J.S., A.B., C.S., T.L.,  
821 K.G.A., E.T.D., C.D.Br., and S.B.M. contributed text to the manuscript. All authors reviewed  
822 and revised the manuscript.

## 823 Competing financial interests

824 C.D.Bu is on the scientific advisory boards (SABs) of Ancestry.com, Personalis, Liberty Biosecurity,  
825 and Etalon DX. C.D.Bu. is also a founder and chair of the SAB of IdentifyGenomics. None of  
826 these entities played a role in the design, interpretation, or presentation of these results.

## 827 Corresponding authors

828 Emmanouil T. Dermitzakis (emmanouil.dermitzakis@unige.ch), Christopher D. Brown (chrbro@upenn.edu),  
829 Stephen B. Montgomery (smontgom@stanford.edu)