

## 1 Distant regulatory effects of genetic variation in multiple human tissues

2 Brian Jo<sup>1&</sup>, Yuan He<sup>2&</sup>, Benjamin J. Strober<sup>2&</sup>, Princy Parsana<sup>3&</sup>, François Aguet<sup>4</sup>, Andrew A. Brown<sup>5,6,7</sup>, Stephane  
3 E. Castel<sup>8,9</sup>, Eric R. Gamazon<sup>10,11</sup>, Ariel Gewirtz<sup>1</sup>, Genna Gliner<sup>12</sup>, Buhm Han<sup>13</sup>, Amy Z. He<sup>3</sup>, Eun Yong Kang<sup>14</sup>, Ian  
4 C. McDowell<sup>15</sup>, Xiao Li<sup>4</sup>, Pejman Mohammadi<sup>8,9</sup>, Christine B. Peterson<sup>16</sup>, Gerald Quon<sup>4,17</sup>, Ashis Saha<sup>3</sup>, Ayellet V.  
5 Segrè<sup>4</sup>, Jae Hoon Sul<sup>18</sup>, Timothy J. Sullivan<sup>4</sup>, Kristin G. Ardlie<sup>4</sup>, Christopher D. Brown<sup>19</sup>, Donald F. Conrad<sup>20</sup>,  
6 Nancy J. Cox<sup>10</sup>, Emmanouil T. Dermitzakis<sup>5,6,7</sup>, Eleazar Eskin<sup>14,21</sup>, Manolis Kellis<sup>4,17</sup>, Tuuli Lappalainen<sup>8,9</sup>, Chiara  
7 Sabatti<sup>22</sup>, GTEx Consortium, Barbara E. Engelhardt<sup>23\*</sup>, Alexis Battle<sup>3\*</sup>

8 <sup>1</sup> Lewis Sigler Institute, Princeton University, Princeton, NJ 08540

9 <sup>2</sup> Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, 21218

10 <sup>3</sup> Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218

11 <sup>4</sup> The Broad Institute of Massachusetts Institute of Technology and Harvard University Cambridge, Massachusetts  
12 02142

13 <sup>5</sup> Department of Genetic Medicine and Development, University of Geneva Medical School, 1211 Geneva,  
14 Switzerland

15 <sup>6</sup> Institute for Genetics and Genomics in Geneva (iG3), University of Geneva, 1211 Geneva, Switzerland

16 <sup>7</sup> Swiss Institute of Bioinformatics, 1211 Geneva, Switzerland

17 <sup>8</sup> New York Genome Center, 101 Avenue of the Americas, New York, NY, 10013

18 <sup>9</sup> Department of Systems Biology, Columbia University Medical Center, New York, NY 10032

19 <sup>10</sup> Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN  
20 37232

21 <sup>11</sup> Department of Clinical Epidemiology, Biostatistics and Bioinformatics and Department of Psychiatry, Academic  
22 Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

23 <sup>12</sup> Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08540

24 <sup>13</sup> Department of Convergence Medicine, University of Ulsan College of Medicine, Asian Medical Center, Korea

25 <sup>14</sup> Department of Computer Science, University of California, Los Angeles, CA 90095

26 <sup>15</sup> Computational Biology & Bioinformatics Graduate Program, Duke University, Durham, NC 27708

27 <sup>16</sup> Department of Biostatistics, The University of Texas MD Anderson Cancer Center, 1400 Pressler Street, Houston,  
28 TX 77030

29 <sup>17</sup> Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, 02139

30 <sup>18</sup> Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, CA 90095

31 <sup>19</sup> University of Pennsylvania, Perelman School of Medicine, Department of Genetics, Philadelphia, PA, 19104

32 <sup>20</sup> Department of Genetics, Department of Pathology & Immunology, Washington University School of Medicine,  
33 St. Louis, Missouri

34 <sup>21</sup> Department of Human Genetics, University of California, Los Angeles, CA 90095

35 <sup>22</sup> Departments of Biomedical Data Science and Statistics, HRP Redwood building, Stanford, CA 94305-5404

36 <sup>23</sup> Princeton University, Department of Computer Science, Center for Statistics and Machine Learning, 35 Olden  
37 Street, Princeton, NJ 08540

38 & Equal contribution

39 \* Corresponding authors: [ajbattle@cs.jhu.edu](mailto:ajbattle@cs.jhu.edu), [bee@princeton.edu](mailto:bee@princeton.edu)

## 40 Abstract

41 Understanding the genetics of gene regulation provides information on the cellular mechanisms  
42 through which genetic variation influences complex traits. Expression quantitative trait loci, or  
43 eQTLs, are enriched for polymorphisms that have been found to be associated with disease risk.  
44 While most analyses of human data has focused on regulation of expression by nearby variants  
45 (cis-eQTLs), distal or trans-eQTLs may have broader effects on the transcriptome and important  
46 phenotypic consequences, necessitating a comprehensive study of the effects of genetic variants  
47 on distal gene transcription levels. In this work, we identify trans-eQTLs in the Genotype Tissue  
48 Expression (GTEx) project data<sup>1</sup>, consisting of 449 individuals with RNA-sequencing data  
49 across 44 tissue types. We find 81 genes with a trans-eQTL in at least one tissue, and we  
50 demonstrate that trans-eQTLs are more likely than cis-eQTLs to have effects specific to a single  
51 tissue. We evaluate the genomic and functional properties of trans-eQTL variants, identifying  
52 strong enrichment in enhancer elements and Piwi-interacting RNA clusters. Finally, we describe  
53 three tissue-specific regulatory loci underlying relevant disease associations: 9q22 in thyroid that  
54 has a role in thyroid cancer, 5q31 in skeletal muscle, and a previously reported master regulator  
55 near *KLF14* in adipose. These analyses provide a comprehensive characterization of trans-eQTLs  
56 across human tissues, which contribute to an improved understanding of the tissue-specific  
57 cellular mechanisms of regulatory genetic variation.

## 58 Introduction

59 Variation in the human genome influences complex disease risk through changes at a cellular  
60 level. Many disease-associated variants are also associated with gene expression levels through  
61 which they mediate disease risk. The majority of expression quantitative trait locus (eQTL)  
62 studies<sup>1-6</sup> thus far have focused on local- or cis-eQTLs because of the relative simplicity of  
63 association mapping in human for both statistical and biological reasons<sup>7,8</sup>. Trans-eQTLs, or  
64 genetic variants that affect gene expression levels of distant target genes, have received much  
65 less attention in comparison to cis-eQTLs, in part due to the considerable multiple hypotheses  
66 testing burden<sup>9</sup>. Far fewer replicable, strong effect trans-eQTLs have been discovered in human  
67 data as compared to cis-eQTLs, unlike in model organisms such as *Saccharomyces cerevisiae* or  
68 *Arabidopsis thaliana*<sup>7,10,11</sup>. However, a handful of replicable trans-eQTLs have now been  
69 identified in human tissues<sup>3,12,13</sup>. Additionally, recent work has suggested that trans-eQTLs  
70 contribute substantially to the genetic regulation of complex diseases<sup>12</sup>, motivating a careful  
71 examination of the role of trans-eQTLs across human tissues in disease etiology.

72 Here, we identify trans-eQTLs in the Genotype-Tissue Expression (GTEx) v6 data, which  
73 include 449 individuals with imputed genotypes and RNA-seq data across 44 tissues for a total  
74 of 7,051 samples. We evaluate the tissue-specificity of trans-eQTLs, and we demonstrate  
75 replication of trans-eQTLs in a large independent RNA-seq study<sup>14</sup>. We show enrichment of  
76 trans-eQTLs for tests restricted to genetic variants associated with expression of nearby genes

77 and trait-associated variants. We then explore properties of genetic variants with significant  
78 associations with distal gene expression levels including functional enrichment in cis regulatory  
79 elements and Piwi-interacting RNA clusters. We discuss three examples of trans-eQTLs in the  
80 GTEx data: the broad regulatory role of the 9q22 locus near thyroid-specific transcription factor  
81 *FOXE1*; a trait-associated regulatory locus in skeletal muscle acting through interferon  
82 regulatory factor *IRF-1*; and replication of a previously-identified master regulator in adipose  
83 tissue near *KLF14* with broad but differential effects in subcutaneous and visceral adipose.

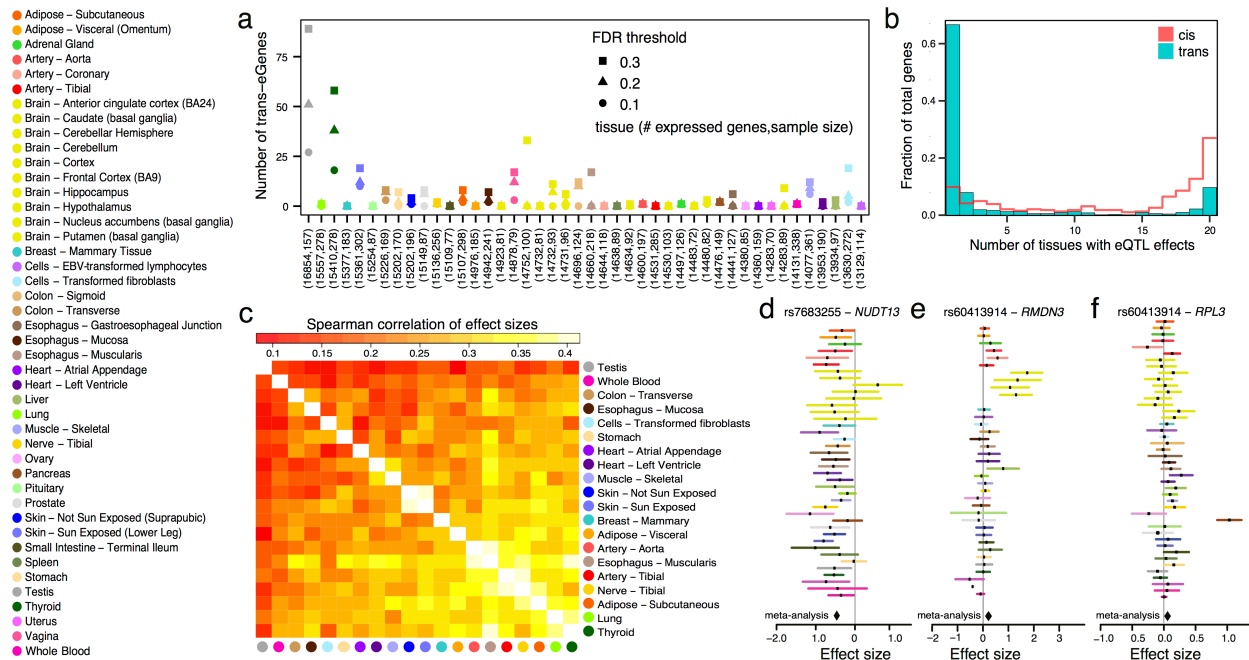
#### 84 **Detection of trans-eQTLs across 44 tissues**

85 We performed trans-eQTL association mapping in each of the 44 GTEx tissues independently.  
86 We applied a linear model controlling for ancestry, sex, genotyping platform, and unobserved  
87 factors in the expression data for each tissue that may reflect batch or other technical  
88 confounders<sup>15,16</sup> (see Online Methods). We tested for association between every protein coding  
89 gene or long non-coding RNA and all autosomal variants (minor allele frequency, MAF > 0.05),  
90 where the gene-variant pair was located on different chromosomes. We developed a  
91 standardized pipeline for filtering detectable false positives from trans-eQTL identification in  
92 RNA-seq data. For example, one major source of artifacts arises from mapping errors in  
93 sequencing data, for which true cis-eQTL variants appear to regulate distal genes due to  
94 sequence similarity between distant regions of the genome<sup>3</sup>. To correct for this, we eliminated  
95 from consideration genes with poor mappability, variants in repetitive elements, and trans-eQTL  
96 associations between pairs of genomic loci that show evidence of cross-mapping (see Online  
97 Methods).

98 Applying this approach, we found a total of 590 trans-eQTLs (false discovery rate, FDR  $\leq$  0.1,  
99 Benjamini-Hochberg, assessed in each tissue separately) including 81 unique genes (*trans-*  
100 *eGenes*, or genes with one or more trans-eQTLs; Fig. 1a) and 532 unique variants (*trans-*  
101 *eVariants*, or variants that are associated with transcription levels of one or more distal genes) in  
102 18 of the 44 GTEx tissues (Table 1). Tissues with larger sample sizes and greater numbers of  
103 expressed genes were more likely to yield trans associations, indicating that low statistical power  
104 limits our analysis. The tissue with the most trans-eGenes was testis (157 individuals; 28 eGenes;  
105 193 trans-eQTLs), reflecting the unusually large number of expressed genes (16,854 genes) and  
106 consistent with the large number of cis-eQTLs detected in this tissue [Aguet et al, GTEx cis-  
107 eQTL manuscript, in submission].

108 We next performed an association test with a restricted subset of variants to control for linkage  
109 disequilibrium (LD), because many of the 532 trans-eVariants are well-correlated variants at the  
110 same genetic locus. To do this, we pruned the set of genotyped and imputed variants to have  
111 local genotype  $R^2 < 0.5$  by random selection, agnostic to gene expression levels or functional  
112 annotation of variants<sup>17</sup>. This LD-pruning led to a set of variants that included approximately  
113 10% of the original variant set. While this may result in false negatives by eliminating some of

114 the strongest associations, it also has the potential to reduce false positives that are not supported  
 115 by associations with well-correlated variants in the same LD block. Performing association  
 116 mapping in this reduced set, we found 41 trans-eQTLs reflecting 40 unique eVariants across 34  
 117 eGenes. These LD-pruned trans-eQTLs spanned 18 tissues, as with the genome-wide set, but  
 118 included a number of tissues that were not observed in the genome-wide analysis such as breast,  
 119 lung, and liver (Table 1).



**Figure 1. Trans-eQTLs across 44 diverse tissues in the GTEx data.** (a) The number of trans-eGenes in all the tissues at three FDR thresholds, ordered with decreasing number of expressed genes. The x-axis labels include (number of expressed gene, number of samples) for each tissue. (b) Distribution of the number of tissues having MetaTissue m-value greater than 0.5 for the top variant for each trans-eGene at  $FDR \leq 0.5$  and each randomly selected cis-eGenes (also  $FDR \leq 0.5$ ). cis-eGenes were matched for discovery tissue distribution to the trans-eGenes. Shown for genes with meta-analysis p-value  $\leq 0.01$ . (c) Hierarchical agglomerative clustering of trans-eGenes ( $FDR \leq 0.5$ ) using a distance metric of  $(1 - \text{Spearman correlation})$  of MetaTissue effect sizes over all genes observed in both tissues. (d) An example of a trans-eQTL (rs7683255 - *NUDT13*) identified in skin - sun-exposed ( $FDR \leq 0.1$ ,  $P \leq 1.1 \times 10^{-10}$ ) that has a global effect across tissues. The lines represent 95% confidence interval of the effect size. (e) An example of a trans-eQTL (rs60413914 - *RMDN3*) identified in brain putamen ( $FDR \leq 0.1$ ,  $P \leq 1.2 \times 10^{-13}$ ) that has an effect in all five brain tissues tested but shows little effect in other tissues. (f) An example of a trans-eQTL (rs758335 - *RPL3*) identified in pancreas ( $FDR \leq 0.1$ ,  $P \leq 2.2 \times 10^{-16}$ ) that has a tissue-specific effect.

120 We also investigated long range eQTLs where the variant lies on the same chromosome as the  
 121 target gene but is not local. We performed association mapping between each gene and variant  
 122 on the same chromosome and we identified 291 intra-chromosomal distal eQTLs ( $\geq 5$  Mb  
 123 between gene and variant;  $FDR \leq 0.1$ ), including 46 eGenes and 247 eVariants (Extended Data  
 124 Table 1). Further, investigated whether intra-chromosomal distal QTLs acted in cis or trans using  
 125 a statistical model to quantify evidence for allele-specific expression (ASE), as cis regulation  
 126 would induce allelic imbalance in gene expression levels for cis-eVariant heterozygotes<sup>3,18</sup>

127 (Mohammadi et al., [GTEx companion paper in preparation]; see Online Methods). Applying  
128 this model to a larger set of 23,953 candidate eQTLs based on a p-value threshold of  $1.0 \times 10^{-5}$ ,  
129 we identified seven distal eQTLs with significant evidence of cis regulation ( $FDR \leq 0.1$ ;  
130 Extended Data Table 2). The support for cis effects overall dropped significantly below  
131 expectation after 3 Mb (Extended Data Fig. 1), possibly suggesting that the majority of distal  
132 intra-chromosomal eQTLs act in trans or represent false positives. However, ASE was observed  
133 for intra-chromosomal gene-variant pairs up to 170 Mb apart, demonstrating that cis regulation  
134 can indeed occur over long genomic distances. While observing ASE provides evidence of cis  
135 regulation, its absence does not guarantee trans regulation, since phasing and power affect  
136 detection (Extended Data Fig. 1). For the remaining analyses, we focus on inter-chromosomal  
137 associations to avoid confounding characterization of cis- and trans-eQTLs.

138 Next, we investigated the level of tissue specificity of the detected trans-eQTLs. We performed a  
139 meta-analysis across the 20 tissues with the greatest number of samples using MetaTissue<sup>19</sup>. We  
140 selected variants for cross-tissue evaluation from the single tissue trans-eQTLs discovered at a  
141 relaxed FDR of 0.5, giving 798 trans-eGenes across the 20 tissues. We estimated that the level of  
142 tissue specificity for each most significant trans-eVariant for each eGene by quantifying the  
143 number of tissues likely to show effects of the eVariant based on MetaTissue m-values (i.e., the  
144 probability that the eQTL effect exists in the tissue). Overall, we observed greater tissue  
145 specificity for trans-eQTLs than a set of cis-eQTLs randomly selected at the same FDR (Fig. 1b);  
146 this observation was robust to choices of m-value threshold and selection of cis-eQTLs  
147 (Extended Data Fig. 2). Extensive tissue-specificity was also observed based on a hierarchical  
148 approach for FDR control, where we found no trans-eQTLs shared across more than a single  
149 tissue (Extended Data Table 3)<sup>20</sup>. Our estimate of greater tissue specificity for trans-eQTLs  
150 agreed with the minimal sharing of trans effects reported in previous eQTL studies with fewer  
151 tissues<sup>21–23</sup>.

152 Although there was greater tissue specificity of trans-eQTLs, we observed trans-eQTL sharing  
153 between pairs of tissues based on MetaTissue effect size estimates that reflected known tissue  
154 relatedness, and were in concordance with patterns of cis-eQTL sharing (Fig. 1c; see Online  
155 Methods; Extended Data Fig. 3). We observed a number of tissue-shared trans-eQTLs, including  
156 rs7683255, which showed moderate trans association with *NUDT13* across most tested GTEx  
157 tissues with consistent direction of effect while only being identified as significant ( $FDR \leq 0.1$ ;  $P$   
158  $\leq 1.1 \times 10^{-10}$ ; Fig. 1d) in skin – sun-exposed. We found examples of trans-eQTLs shared across  
159 a subset of related tissues, such as an association between rs60413914 and *RMDN3*, which was  
160 genome-wide significant in brain – putamen ( $FDR \leq 0.1$ ;  $P \leq 1.2 \times 10^{-13}$ ; Fig. 1e) and had  
161 moderate effects in all tested brain regions but no strong effect in other tissues. *RMDN3* is  
162 widely expressed, with higher average expression levels in brain tissues than outside of the brain  
163 (Extended Data Fig. 4). We observed tissue specific trans-eQTLs, such as rs758335 and *RPL3*,  
164 which is only observed in pancreas ( $FDR \leq 0.1$ ;  $P \leq 2.2 \times 10^{-16}$ ; Fig. 1f).

Tissue	No. of samples	Genome wide		LD pruned		Cis eVariants		Trait associated variants		Any approach	
		gene	var	gene	var	gene	var	gene	var	gene	var
Muscle - Skeletal	361	6	41	0	0	3	4	2	2	6	42
Whole Blood	338	1	2	1	1	1	1	0	0	2	3
Skin - Sun Exposed (Lower Leg)	302	9	21	5	5	1	1	1	1	12	24
Adipose - Subcutaneous	298	2	7	1	1	1	1	0	0	3	8
Lung	278	0	0	1	1	0	0	1	1	2	2
Thyroid	278	19	189	4	5	3	2	2	1	20	190
Cells - Transformed fibroblasts	272	2	11	0	0	2	3	1	1	3	12
Nerve - Tibial	256	1	1	0	0	1	1	0	0	2	2
Esophagus - Mucosa	241	2	10	3	3	2	2	0	0	5	13
Esophagus - Muscularis	218	0	0	1	1	1	1	1	1	3	3
Artery - Aorta	197	1	1	1	1	0	0	0	0	1	1
Skin - Not Sun Exposed (Suprapubic)	196	1	1	1	1	0	0	0	0	1	1
Heart - Left Ventricle	190	0	0	4	4	0	0	0	0	4	4
Breast - Mammary Tissue	183	0	0	2	3	0	0	1	1	3	4
Stomach	170	0	0	0	0	1	1	0	0	1	1
Colon - Transverse	169	3	11	0	0	0	0	0	0	3	11
Heart - Atrial Appendage	159	0	0	2	3	0	0	0	0	2	3
Testis	157	28	193	3	4	2	2	4	4	31	197
Pancreas	149	2	12	0	0	1	2	1	1	3	13
Adrenal Gland	126	1	1	0	0	0	0	0	0	1	1
Cells - EBV-transformed lymphocytes	114	0	0	0	0	0	0	2	2	2	2
Brain - Cerebellum	103	0	0	0	0	2	2	0	0	2	2
Brain - Caudate (basal ganglia)	100	0	0	3	3	0	0	0	0	3	3
Liver	97	0	0	1	1	0	0	0	0	1	1
Brain - Nucleus accumbens (basal ganglia)	93	0	0	0	0	1	1	0	0	1	1
Brain - Cerebellar Hemisphere	89	0	0	0	0	1	1	0	0	1	1
Brain - Putamen (basal ganglia)	82	1	9	1	2	0	0	0	0	1	9
Vagina	79	3	22	0	0	0	0	0	0	3	22
Small Intestine - Terminal Ileum	77	0	0	1	1	0	0	2	2	3	3
Uterus	70	0	0	0	0	0	0	1	1	1	1
<b>Total (union)</b>		<b>81</b>	<b>532</b>	<b>34</b>	<b>40</b>	<b>23</b>	<b>25</b>	<b>19</b>	<b>18</b>	<b>124</b>	<b>580</b>

**Table 1. Trans-eVariant and eGene discoveries for genome-wide and restricted approaches in the GTEx data.** Each tissue with non-zero values in one or more of the restricted approaches is included in the rows with the total on the final row; the columns include the number of samples for that tissue, followed by the number of unique trans-eGenes and trans-eVariants identified in the genome-wide tests, and tests restricted to the LD-pruned, cis-eQTL, and trait associated variants, followed by the number of unique trans-eGenes and trans-eVariants identified by any of the four approaches.

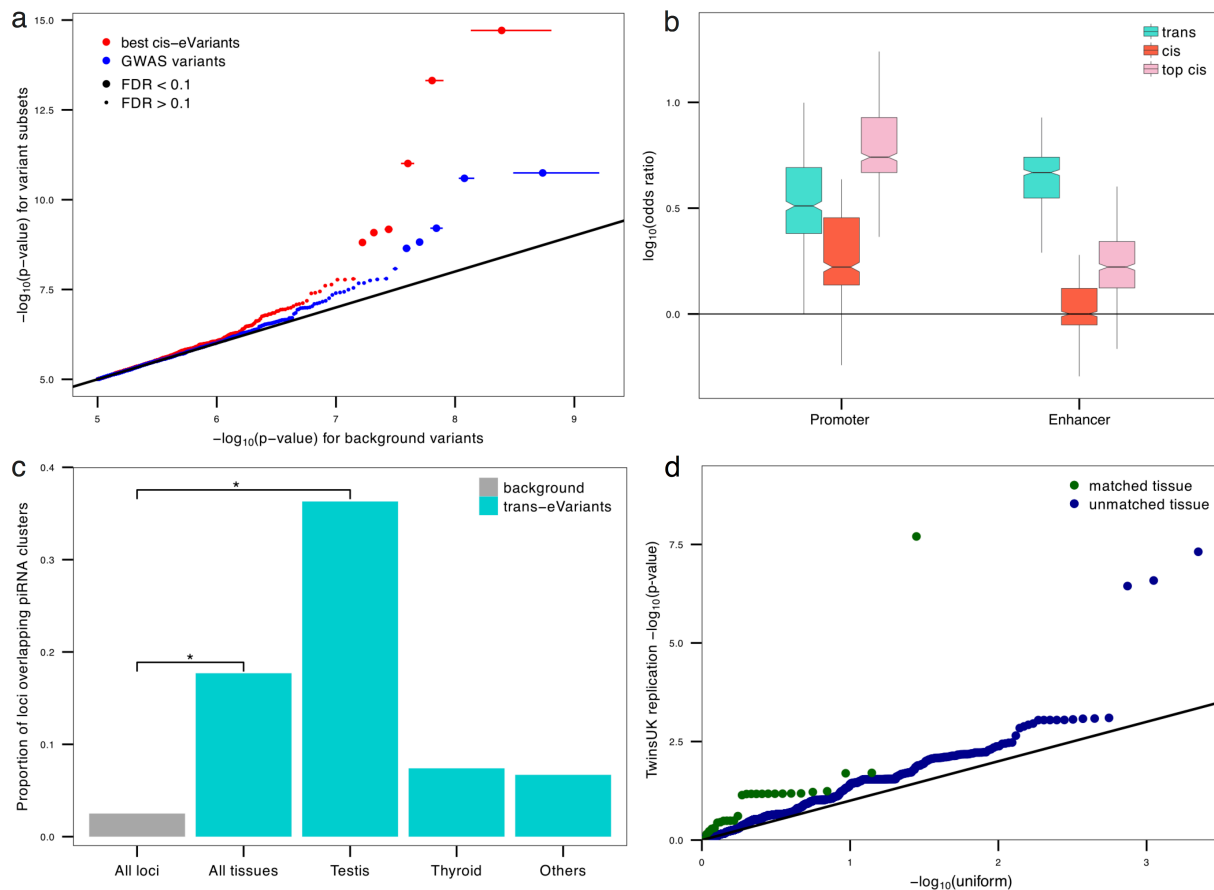
## 165 **Characterization and functional analysis of trans-eQTL variants**

166 To better understand their cellular mechanisms, we characterized the functional properties of  
167 trans-eVariants. Of the 590 trans-eVariants from the genome-wide analysis, 312 were also  
168 identified to have a cis association ( $FDR \leq 0.05$ ), significantly more than expected by chance  
169 (Fisher's exact test;  $P \leq 2.2 \times 10^{-16}$ ). This pattern would suggest a mechanism for trans  
170 association in which the eVariant directly regulates expression of a nearby gene, whose protein  
171 product then affects other genes downstream. We performed an association test, restricting the  
172 variants to the set of cis-eVariants (top variant per cis-eGene) and testing for trans association  
173 with all genes on any other chromosome than the variant's own. Cis-eVariants were significantly  
174 more likely to have low trans-eQTL association p-values than random variants matched for MAF  
175 (Chi-squared test;  $P \leq 2.2 \times 10^{-16}$ ; Fig. 2a). We identified a total of 23 trans-eGenes ( $FDR \leq 0.1$ )  
176 among this subset of tests, 14 of which were not discovered in the genome-wide analysis.  
177 Variants with both cis and trans associations did not show stronger effect sizes in cis (Wilcoxon  
178 rank sum test,  $P \leq 0.22$ ), and the direction of effect was not significantly matched (binomial test;  
179  $P \leq 0.18$ ; Extended Data Fig. 5); however, the small number of trans-eQTLs discovered after  
180 restricting to cis-eVariants limits the interpretability of these results. Trans-eVariants that have  
181 no cis association may alter protein function, may reflect false negatives in the cis association  
182 test, or may arise from unmeasured regulatory mechanisms. We observed a depletion of protein-  
183 coding loci among our eVariants (odds ratio = 0.39; Fisher's exact test,  $P \leq 0.03$ ) suggesting that  
184 modification of protein function is not the dominant mechanism for trans-eQTL effects.

185 It has been also reported that genetic variants associated with complex traits in genome-wide  
186 association studies (GWAS) are enriched for trans-eQTLs<sup>12,24,25</sup>. We evaluated this in the GTEx  
187 data by performing association testing by restricting to variants that have been associated with a  
188 complex trait in a GWAS<sup>26</sup> ( $P \leq 2.0 \times 10^{-5}$ ). Across the 44 tissues, we found 21 trans-eQTL  
189 associations, involving 18 unique variants and 19 unique genes ( $FDR \leq 0.1$ ; Fig. 2a; Table 1).  
190 As with the cis-eQTL restricted analysis, we observed lower trans-eQTL p-values among trait-  
191 associated variants than in a control set of variants matched on MAF and distance to the nearest  
192 gene transcription start site (TSS; Chi-squared test,  $P \leq 1.9 \times 10^{-4}$ ).

193 We investigated whether trans-eVariants were each associated with numerous target genes,  
194 which may reflect broad effects of regulatory loci, as have been reported in model organisms<sup>27-</sup>  
195 <sup>29</sup>. Disambiguating true broad regulatory effects from artifacts remains an important challenge<sup>30</sup>  
196 – PEER and other methods designed for artifact correction<sup>31,32</sup> generally identify and remove  
197 patterns of broad correlation between genes, regardless of whether the source is biological or  
198 technical. We conservatively removed a large number of latent factors (either 15, 30, or 35 PEER  
199 factors<sup>16</sup>, capturing 59-78% of total variance in gene expression depending on tissue sample size;  
200 Extended Data Fig. 6), which reduces false positives<sup>33</sup> but may also remove variance in gene  
201 expression levels arising from broad trans effects. Indeed, we observed loci with numerous  
202 associations in uncorrected data (Extended Data Fig. 7) that disappeared once controlling for

203 unobserved factors estimated by PEER. Despite this, we observed evidence of eVariants with  
 204 multiple targets even after correction. At genome-wide significance, three loci (60 Kb windows,  
 205 potentially containing multiple variants) were associated with two distal eGenes each.  
 206 Additionally, for each eVariant, we evaluated the distribution of association statistics with all  
 207 genes expressed in the corresponding tissue and calculated  $(1 - \pi_0)$ , the estimated total fraction  
 208 of genes associated with the variant (Extended Data Fig. 8)<sup>34</sup>. This suggests that much larger  
 209 numbers of likely target genes for trans-eVariants than for either cis-eVariants or randomly  
 210 selected variants, with significantly higher values of  $(1 - \pi_0)$  (Wilcoxon rank sum test,  $P \leq 3.4 \times$   
 211  $10^{-4}$  and  $P \leq 2.2 \times 10^{-16}$ , respectively).



**Figure 2. Functional characterization of GTEx trans-eVariants.** (a) Partial quantile-quantile (QQ) plot showing enrichment of low trans-eQTL p-values of association for cis-eVariants and trait-associated variants in skeletal muscle ( $n = 361$ ). (b) Cis-regulatory element enrichment analysis of trans-eVariants ( $\text{FDR} \leq 0.1$ ), cis-eVariants ( $\text{FDR} \leq 0.1$ ), and the top most significant cis-eVariants. Boxes show promoter and enhancer element enrichment in any of the GTEx discovery tissue's matched cell type specific Roadmap or ENCODE annotations compared to 500 randomly selected background variants (matched for distance to TSS and MAF). (c) Proportion of loci overlapping with piRNA clusters, including randomly sampled loci, trans-eVariants across all tissues, testis trans-eVariants, thyroid trans-eVariants, and trans-eVariants from all tissues other than testis and thyroid. (d) Replication of trans-eVariants from GTEx in the TwinsUK data (y-axis) across matched tissues (green) and unmatched tissues (blue), versus the expected p-values from the quantiles of a uniform distribution (x-axis).



212 We studied possible molecular mechanisms underlying the trans-eQTLs. Using matched tissue-  
213 specific annotations from the Roadmap Epigenomics project<sup>35,36</sup>, we compared enrichment of  
214 trans-eQTLs in promoter and enhancer regions of the genome to randomly selected variants  
215 matched by distance to nearest TSS, MAF, and chromosome. Trans-eVariants (FDR  $\leq 0.1$ ) were  
216 enriched in cell-type matched enhancers (Fisher's exact,  $P \leq 6.6 \times 10^{-4}$ ) and moderately enriched  
217 for promoters ( $P \leq 0.13$ ), with greater enrichment in enhancers (Fig. 2b). We observed greater  
218 enrichment for trans-eVariants than for cis-eVariants called at the same FDR (promoter  
219 Wilcoxon rank sum test,  $P \leq 2.2 \times 10^{-16}$ ; enhancer,  $P \leq 2.2 \times 10^{-16}$ ). Stronger effect sizes are  
220 needed to detect trans-eVariants at the same FDR, but even comparing to a matched number of  
221 the strongest cis-eVariants, we observed greater enrichment among trans-eVariants for enhancer  
222 element overlap. These results indicate that trans-eVariants were more enriched for enhancer  
223 regions than cis-eVariants, consistent with greater tissue specificity of enhancer activity and  
224 greater tissue-specificity of trans-eVariants (Fig. 1b).

225 Observing the large number of trans-eQTLs detected in testis, we investigated possible  
226 mechanisms for this tissue in more detail. Piwi-interacting RNAs (piRNAs) are small 24-31bp  
227 RNAs that bind to Piwi-class proteins and silence mobile elements by RNA degradation and by  
228 methylation of their DNA source. PiRNAs are strongly expressed in testis and may regulate gene  
229 expression<sup>37,38</sup>. We tested for enrichment of trans-eVariants in piRNA clusters identified in  
230 testis<sup>39</sup>. We found that 36.3% of testis trans-eVariants directly overlap piRNA clusters,  
231 representing a significant enrichment beyond the 2.5% of the genome covered by these regions  
232 (permutation,  $P \leq 1.0 \times 10^{-4}$ ). In aggregate, eVariants from all tissues demonstrated an enriched  
233 overlap of 17.7% with piRNA clusters (permutation,  $P \leq 7.0 \times 10^{-4}$ ) but this enrichment appeared  
234 to be almost entirely driven by testis eQTLs (Fig. 2c).

### 235 **Replication of trans-eQTLs.**

236 Trans-eQTLs have not replicated consistently in human studies as compared to cis-eQTLs<sup>13,40-42</sup>,  
237 due in part to insufficient statistical power and a limited number of studies with comparable  
238 tissue and cohort, but also reflecting potential false positive associations. First, we attempted to  
239 replicate two trans-eQTL associations from lymphoblastoid cell lines (LCLs) identified in the  
240 trait-associated variant restricted analysis. We tested these trans-eQTLs in the GEUVADIS data  
241 ( $n=462$ )<sup>6</sup>, but did not find signal of association for either eQTL ( $P \leq 0.93$ , rs3125734;  $P \leq 0.64$ ,  
242 rs10520789). We then tested the union of the GTEx trans-eQTLs across the four sets of tests  
243 (genome-wide, LD pruned, cis-eVariants, and GWAS hits; FDR  $\leq 0.1$ ) for replication in the  
244 TwinsUK eQTL data<sup>14</sup>, which includes four shared tissues with GTEx—whole blood,  
245 subcutaneous adipose, LCLs, and photo-protected infra-umbilical skin—for  $n=856$  donors of  
246 European ancestry<sup>14</sup>. We found a substantial enrichment of low p-value associations among the  
247 gene-variant pairs in the TwinsUK data for GTEx trans-eQTLs (Wilcoxon rank sum test;  $P \leq 4.8$   
248  $\times 10^{-15}$ ; Fig. 2d); furthermore, this enrichment of association p-value was significantly higher in  
249 matched tissue types than in unmatched tissue types (Wilcoxon rank sum test;  $P \leq 2.4 \times 10^{-4}$ ).

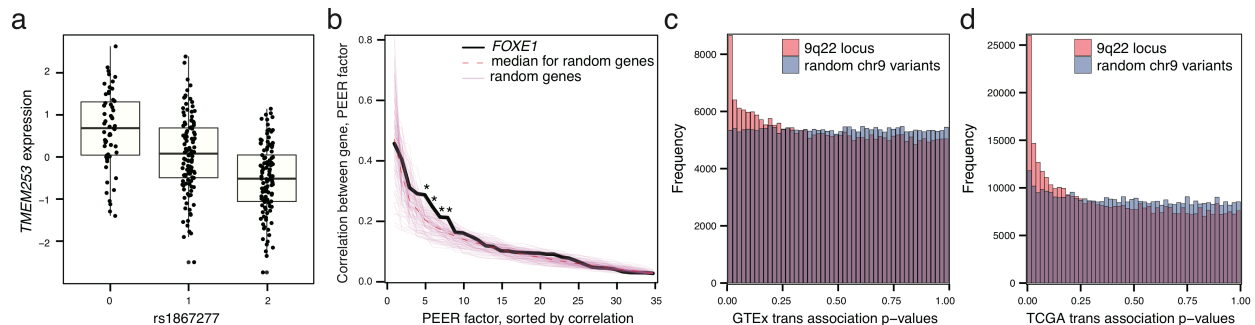
250 In related work in the TwinsUK cohort<sup>23</sup>, with RNA-seq analysis of n=845 individuals in adipose  
251 – subcutaneous, LCLs, and skin, we replicated two strong tissue-specific trans-eQTLs. In GTEx  
252 adipose – subcutaneous, we found two linked variants rs13234269 and rs35722851, which were  
253 not in our trans-eQTL list due to strict repeat element filtering that we relaxed for replication of  
254 these trans effects, that were associated in cis with *KLF14* that showed enrichment for genome-  
255 wide trans effects (discussed in detail below). These variants were in strong LD ( $R^2 \geq 0.98$ ) with  
256 master regulator rs4731702 that was identified in both the TwinsUK study<sup>23</sup> and MuTHER<sup>2,43</sup>  
257 study. In skin – sun-exposed in GTEx, rs289750 was associated in cis with *NLRC5* and in trans  
258 with *TAPI*, while the TwinsUK study found rs289749 (located 469 bp away from rs289750;  $R^2$   
259 = 0.918) associated with the same genes in cis and trans.

### 260 **Broad regulatory locus 9q22 in thyroid tissue**

261 We found two genome-wide significant trans-eVariants in the 9q22 locus for thyroid tissue  
262 (rs7037324 and rs1867277, with correlation coefficient  $R^2 = 0.74$ ; thyroid n = 278) associated  
263 with *TMEM253* (chromosome 14; Fig. 3a) and *ARFGEF3* (chromosome 6). These two trans-  
264 eGenes were also identified as significant in both the cis-eQTL and the GWAS restricted tests.  
265 The cis target gene was *C9orf156*, and the supporting GWAS trait was thyroid cancer<sup>44</sup>  
266 (rs7037324; odds ratio, OR = 1.54;  $P \leq 2.2 \times 10^{-16}$ ). The 9q22 locus has also been linked with  
267 multiple thyroid specific diseases including goiter, hypothyroidism, and thyroid cancer<sup>45-47</sup> and  
268 contains the gene *FOXE1*, a thyroid-specific transcription factor (Extended Data Fig. 9). Loss-  
269 of-function mutations in *FOXE1* manifests as ectopic thyroid tissue or cleft palate in developing  
270 mice<sup>48</sup>, and congenital cleft lip and cleft palate have also shown association with 9q22 variants  
271 in human studies<sup>45</sup>. *FOXE1* was weakly associated in cis with variants rs7037324 and rs1867277  
272 ( $P \leq 5.2 \times 10^{-3}$  and 0.0191, respectively), but only before PEER correction of expression data.  
273 Despite this moderate cis association, based on colocalization analysis<sup>49</sup>, we estimated the  
274 posterior probability that a shared causal variant at this locus drives both cis and trans  
275 associations to be greater than 0.99 for both candidate cis-eGenes (*FOXE1* and *C9orf156*) with  
276 both trans-eGenes (*TMEM253* and *ARFGEF3*). Further, *FOXE1* transcription was strongly  
277 correlated with several of the PEER factors estimated from the thyroid gene expression data (Fig.  
278 3b), suggesting a broad effect of this thyroid-specific regulatory gene and explaining the lack of  
279 cis association signal after controlling for all 35 PEER factors. We evaluated the trans-eVariants  
280 for association across all genes in uncorrected data and found substantial enrichment for low p-  
281 values across many genes (subcutaneous  $(1 - \pi_0) = 0.10$  and visceral  $(1 - \pi_0) = 0.04$ ; Fig. 3c)  
282 indicating a broad regulatory effect.

283 We replicated the effects of this locus in 496 primary thyroid cancer RNA-seq samples from The  
284 Cancer Genome Atlas (TCGA)<sup>50</sup>. We tested 19,153 genes for association with 23 variants in  
285 chromosome 9 locus 100600000 - 100670000, which is the region containing the two eVariants.  
286 Correcting for cross-chromosomal association tests across the 23 variants, we found 1173 unique  
287 trans-eGenes ( $FDR \leq 0.1$ ), substantially more than randomly selected chromosome 9 variants

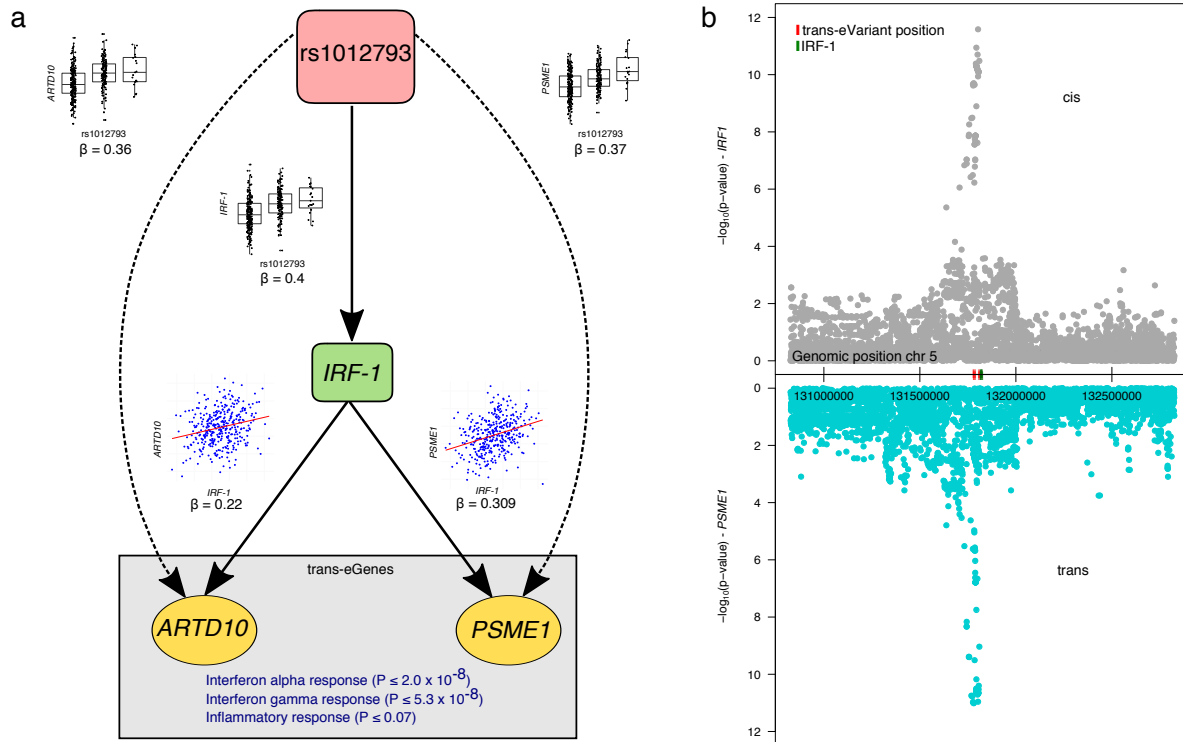
288 (Fig. 3d, Extended Data Fig. 10). Despite the substantial changes to gene expression levels in  
 289 cancer tissue, we replicated both trans-eQTL associations from GTEx in TCGA data, *TMEM253*  
 290 (GTEx  $P \leq 1.2 \times 10^{-4}$ , FDR  $\leq 0.034$ ) and *ARFGEF3* (GTEx  $P \leq 1.1 \times 10^{-5}$ , FDR  $\leq 0.0097$ ).  
 291 Among 15 variants associated with *TMEM253*, rs10115216 was also associated in cis with  
 292 *FOXE1* ( $P \leq 9.3 \times 10^{-3}$ , FDR  $\leq 0.043$ ) and rs6586 in cis with *C9orf156* (FDR  $\leq 3.0 \times 10^{-13}$ ).  
 293 These results demonstrate replication of both the broad impact of the 9q22 locus and particular  
 294 target genes in thyroid tumor tissue.  
 295



**Figure 3. Trans-eQTLs in 9q22 locus in thyroid act as master regulators.** (a) Association of rs1867277 with corrected *TMEM253* expression levels ( $P \leq 2.2 \times 10^{-16}$ ). (b) Correlation between *FOXE1* expression levels and thyroid PEER factors, compared to 100 random genes. For every gene, absolute correlation was sorted in decreasing order. The correlation of *FOXE1* with the 5th, 6th, 7th, and 8th PEER factors was significantly higher than the correlation of random genes at those rank ordered PEER factors (empirical  $P \leq 0.05$ ). (c) P-value histogram of associations between 19 variants in the 9q22 locus and all genes in GTEx thyroid gene expression levels, compared to 19 random variants from the same chromosome. (d) P-value histogram of associations between 23 variants in the 9q22 locus and all genes in TCGA thyroid tumor expression data, compared to 23 random variants from the same chromosome.

### 296 Trait-associated variants in skeletal muscle near interferon regulatory factor *IRF-1*

297 In skeletal muscle, two linked variants in the 5q31 locus (rs2706381 and rs1012793,  $R^2 = 0.84$ )  
 298 were associated in trans with the expression of immune response genes *PSME1* ( $P \leq 9.8 \times 10^{-12}$ ),  
 299 and *ARTD10* ( $P \leq 8.3 \times 10^{-10}$ ). A third variant on the same locus ( $R^2 = 0.50$ ), rs12659708, also  
 300 showed significant association with *ARTD10* ( $P \leq 4.8 \times 10^{-14}$ ) and moderate association with  
 301 *PSME1* ( $P \leq 1.6 \times 10^{-7}$ ). These variants were moderately associated with numerous genes in  
 302 skeletal muscle (47 trans-eGenes at FDR = 0.2, assessed only among the three variants; Extended  
 303 Data Fig. 11). Potential targets (trans-eQTL  $P \leq 0.001$ ) were enriched (right-tailed Fisher's exact  
 304 test) in multiple immune pathways from MsigDB<sup>51</sup> including *interferon alpha response* ( $P \leq 2.0$   
 305  $\times 10^{-8}$ ), *interferon gamma response* ( $P \leq 5.3 \times 10^{-8}$ ) and nominally significant for *inflammatory*  
 306 *response* ( $P \leq 0.07$ ; Extended Data Table 4). The two linked variants rs2706381 and rs1012793  
 307 were also significantly associated with circulating fibrinogen levels in a GWAS<sup>52</sup>. Fibrinogen  
 308 mediates inflammatory disorders including muscle injury and Duchene muscular dystrophy  
 309 (DMD), multiple sclerosis, and rheumatoid arthritis<sup>53-56</sup>, and has been shown to drive fibrosis in  
 310 DMD, where it promotes expression of *IL-1 $\beta$*  and *TGF- $\beta$* <sup>57</sup>.

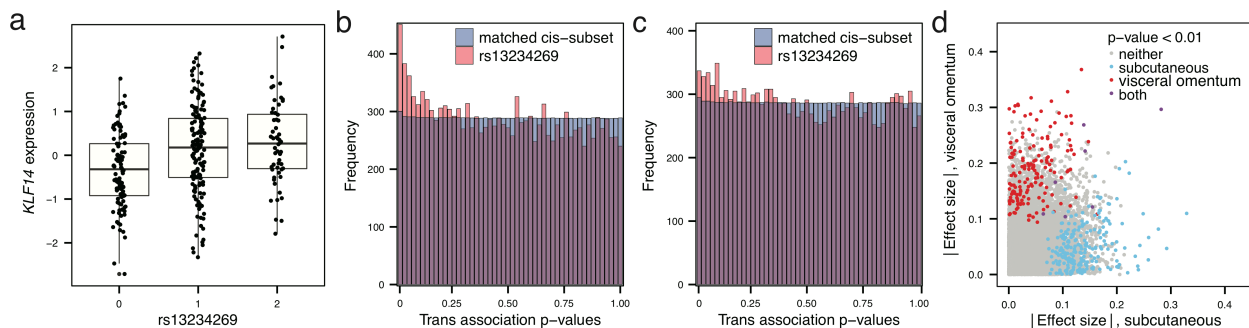


**Figure 4. Skeletal muscle master regulatory network through *IRF-1*.** (a) Network showing cis and trans regulatory effects of rs1012793 mediated through *IRF-1* (*Interferon regulatory factor 1*). Rs1012793 affects expression of *IRF-1* in cis and *PSME1* and *ARTD10* in trans (box plots). *IRF-1* is significantly co-expressed with the trans-eGenes (scatter plots). (b) Cis and trans association significance of variants within 1 Mb of *IRF-1* transcription start site in chromosome 5 locus with cis-eGene *IRF-1* (gray) and trans-eGene *PSME1* (teal) demonstrating concordant signal across the locus.

311 To explore cellular mechanisms underlying these effects, we evaluated cis regulatory  
 312 associations for each variant. Rs1012793 and rs12659708 appeared as cis-eVariants associated  
 313 with *IRF-1*, and rs1012793 was associated with *SLC22A4* (FDR ≤ 0.05). However, the directions  
 314 of effect between cis and trans targets were only consistent for *IRF-1* (Fig. 4a). The association  
 315 statistics in this region were also highly concordant for *IRF-1* (cis-eGene) and *PSME1* (trans-  
 316 eGene), (Fig. 4b), quantified using colocalization analysis<sup>49</sup>, which produced posterior  
 317 probabilities greater than 0.97 that the same causal variant regulates *IRF-1* and each of *PSME1*  
 318 and *ARTD10*. The cis-eGene *IRF-1* is a transcription factor known to facilitate regulation of  
 319 interferon induced immune responses<sup>58–61</sup>, and *PSME1* and *ARTD10* are interferon response  
 320 genes upregulated in inflammation and antigen presentation<sup>58,62–64</sup>. Both trans-eGenes *PSME1*  
 321 and *ARTD10* were also identified as potential *IRF-1* targets in primary human monocytes<sup>65</sup>.  
 322 Together, these results suggest cis regulatory loci affecting *IRF-1* are regulators of the *IFN*  
 323 responsive inflammatory processes involving genes including *PSME1* and *ARTD10*, with  
 324 implications for complex traits affecting muscle tissue.

## 325 Replication of a trans-eQTL master regulator via *KLF14* in adipose tissues

326 The MuTHER study<sup>2,43</sup> (n=776) identified a master trans regulator in adipose – subcutaneous  
327 tissue with the maternally expressed cis target gene *KLF14*, which encodes a transcription factor,  
328 Kruppel-like factor 14<sup>43</sup>. Cis-eQTL rs4731702, targeting *KLF14*, showed enriched association  
329 with genes that are relevant in metabolic phenotypes, such as cholesterol levels<sup>66,67</sup>. In the GTEx  
330 data, rs4731702 was not quite statistically significant as a cis-eQTL in adipose – subcutaneous in  
331 the GTEx data ( $P \leq 8.1 \times 10^{-5}$ , where the FDR  $\leq 0.05$  significance threshold is  $P \leq 5.7 \times 10^{-5}$ ).  
332 Adipose – visceral did not have any significant cis-eQTLs at this locus. However, we identified  
333 two variants, rs13234269 (Fig. 5a) and rs35722851, that are cis-eQTLs for *KLF14* in adipose –  
334 subcutaneous ( $P \leq 2.2 \times 10^{-5}$  and  $4.7 \times 10^{-5}$ , respectively) and in strong LD with rs4731702 ( $R^2 =$   
335 0.98 and 0.99, respectively) [Aguet et al, GTEx cis-eQTL manuscript, in submission]. We used  
336 variant rs13234269 for further testing, which was not in our trans-eQTL list due to strict repeat  
337 element filtering but we included here for replication analysis. We tested the association between  
338 this locus and all expressed genes in two GTEx adipose tissues: subcutaneous (14,461 genes) and  
339 visceral (14,342 genes). Although we found no individually significant trans-eGenes, we found  
340 an enrichment of association with distal gene expression, which was more pronounced in adipose  
341 – subcutaneous ( $1 - \pi_0 = 0.11$  for adipose – subcutaneous,  $1 - \pi_0 = 0.04$  for adipose – visceral,  
342 Figs. 5b, 5c, and Extended Data Table 5), replicating the results of the MuTHER study.  
343 However, the absolute value effect sizes of rs13234269 across 14,105 genes shared in the two  
344 adipose tissues showed poor correlation across the two tissues ( $R^2 = 0.11$ ; Fig. 5d).



**Figure 5: Master regulator in two adipose tissues with sex-specific effects.** (a) Association of rs13234269 with *KLF14* gene expression levels in adipose – subcutaneous in the GTEx data. (b) P-value histogram of associations with all genes for rs13234269 in adipose – subcutaneous as compared to the p-value histogram of associations with all genes of 7,608 variants matched in MAF and distance to TSS of the closest gene with the best cis-eQTLs in adipose – subcutaneous. (c) P-value histogram of associations with all genes for rs13234269 in adipose – visceral. (d) Absolute value effect sizes for trans-association between rs13234269 and 14,105 genes in adipose – subcutaneous (x-axis) and adipose – visceral (y-axis), with colors indicating the tissue for which the association has  $P \leq 0.01$ , and the regression line in blue with  $R^2 = 0.11$ .

345 *KLF14* is a maternally expressed transcription factor in an imprinted locus, and the MuTHER  
346 study included only females. In GTEx data, both tissues included moderate evidence of sex-  
347 differential expression of *KLF14* ( $P \leq 4.3 \times 10^{-3}$  in adipose – subcutaneous;  $P \leq 2.1 \times 10^{-3}$  in  
348 adipose – visceral) when correcting for all covariates other than sex. However, when considering  
349 female and male samples together in the GTEx adipose – subcutaneous data, the effect of

350 rs13234269 on *KLF14* was the same in males and females in adipose – subcutaneous, (gene-by-  
351 sex interaction,  $P \leq 0.44$ ; Extended Data Fig. 12), but we observed a mild gene-by-sex  
352 interaction with *KLF14* in adipose – visceral ( $P \leq 2.7 \times 10^{-3}$ ; Extended Data Fig. 12). This  
353 suggests a role for trans regulation in metabolic diseases, of which many show evidence of  
354 sexual dimorphism<sup>68–70</sup>.

## 355 Discussion

356 Here, we presented an analysis of the trans regulation of gene expression by genetic variation,  
357 measuring association in expression data from 449 individuals and 44 human tissues in the GTEx  
358 project data. We identified 81 trans-eGenes from 18 tissues, and observed an enrichment for  
359 coincident cis regulatory effects and GWAS associations. We observed that trans-eQTL effects  
360 are moderately shared across tissues, but exhibit much greater tissue-specificity than cis-eQTLs.  
361 This increased tissue-specificity was also reflected in greater enrichment in overlap with  
362 enhancer elements. Testis trans-eVariants were highly enriched in Piwi-interacting RNA clusters,  
363 suggesting a possible general mechanism for these trans-eQTLs across tissues; it remains to  
364 directly assess the mediation of regulatory effects by Piwi-interacting RNAs and to determine the  
365 tissue specificity of the piRNA clusters.

366 Trans-eQTL detection remains limited by power and relative effect size, and also by challenges  
367 in disentangling broad regulatory effects from artifacts in gene expression data<sup>3,8,22</sup>. While it is  
368 essential to aggressively control for these unobserved confounders in order to avoid false  
369 positives, this may obscure the effects of the most broad trans-eQTLs and master regulatory  
370 elements, as evidenced by analysis of the thyroid *FOXO1* 9q22 locus. However, in the GTEx  
371 trans-eQTL data, we observed evidence of trans-eVariants associated with multiple genes, and  
372 evaluated three examples in detail. We showed that variants near thyroid-specific transcription  
373 factor *FOXO1* are moderately associated with numerous genes in thyroid, an effect we were able  
374 to reproduce in TCGA thyroid cancer samples. We then explored cis and trans effects of a  
375 regulatory region in skeletal muscle that appears to act through *IRF-1*. Finally, we examined  
376 previously reported master regulatory effects of *KLF14* in the two GTEx adipose tissues. Each  
377 of these three regulatory loci also contained variants associated with tissue-relevant complex  
378 traits.

379 Trans-eQTLs from diverse human tissues will serve as an important resource for characterizing  
380 GWAS variants according to their cellular mechanisms and consequences. Combining GWAS  
381 variants with genome-wide eQTLs will allow us to identify both the proximal and distal  
382 regulatory effects underlying human disease phenotypes, including tissue-specific regulatory  
383 pathways. This study represents the largest multi-tissue study of trans-eQTLs to date, allowing a  
384 more complete characterization of distal regulatory effects and a greater understanding of the  
385 genome-wide, tissue-specific consequences of genetic variation on gene expression relevant to  
386 complex human traits.

## 387 Online Methods

388 **RNA-seq data from GTEx.** The GTEx v6p analysis freeze (phs000424.v6.p1, available in  
389 dbGaP) includes RNA that was isolated from 8,555 postmortem samples from 53 tissue types  
390 across 544 individuals. All human subjects were deceased donors. Informed consent was  
391 obtained for all donors via next-of-kin consent to permit the collection and banking of de-  
392 identified tissue samples for scientific research. A total of 44 tissues were sampled from at least  
393 70 individuals: 31 solid-organ tissues, ten brain subregions with two duplicate regions (cortex  
394 and cerebellum), whole blood, and two cell lines derived from donor blood and skin samples  
395 (Table 1). Each tissue had a different number of unique samples. Non-strand specific, polyA+  
396 selected RNA-seq libraries were generated using the Illumina TruSeq protocol. Libraries were  
397 sequenced to a median depth of 78M 76-bp paired end reads. RNA-seq reads were aligned to the  
398 human genome (hg19/GRCh37) using Tophat (v1.4.1)<sup>71</sup> based on GENCODE v19 annotations.  
399 Gene-level expression was estimated as reads per kilobase of transcript per million mapped reads  
400 (RPKM) with RNA-SeQC using uniquely mapped, properly paired reads<sup>72</sup>.

401 Only genes with  $\geq 10$  individuals with expression estimates  $> 0.1$  RPKM and an aligned read  
402 count  $\geq 6$  within each tissue were considered significantly expressed and used for eQTL  
403 mapping. Within each tissue, the distribution of RPKMs in each sample was transformed to the  
404 average empirical distribution across all samples. Expression measurements for each gene in  
405 each tissue were subsequently transformed to the quantiles of the standard normal distribution.

406 To increase the sensitivity of our analyses, we regressed out both known covariates (three  
407 genotype principal components, sex, and DNA sequencing platform) and PEER factors<sup>16</sup>  
408 calculated independently for each tissue. A total of 15 PEER factors were included for tissues  
409 with fewer than 150 samples; 30 for tissues with sample sizes between 150 and 250; and 35 for  
410 tissues with more than 250 samples.

411 **Genotypes from GTEx.** The initial number of GTEx donors genotyped on Illumina's Omni  
412 arrays in the second phase of GTEx (GTEx\_phs000424, release v6) was 455 before sample  
413 quality control (296 declared as males and 159 as females). These samples included 272 donors  
414 genotyped on Illumina's HumanOmni2.5-Quad Array (2,378,075 variants), and 183 on  
415 Illumina's HumanOmni5-Quad Array (4,276,680 variants), after excluding 2 Klinefelter donors  
416 and 5 duplicates. DNA isolated from blood samples was the primary source of DNA used for  
417 genotyping ( $>360$ ng DNA), performed at the Broad Institute of Harvard and MIT. Genotypes  
418 were called using Illumina's GeneTrain calling algorithm (Autocall). The genotyping call rates  
419 per individual exceeded 98% for all samples. All genotypes and analyses were aligned to  
420 chromosome positions from human genome build 37 (hg19).

421 To merge the genotypes from Illumina's Omni 5M and Omni 2.5M arrays we extracted the  
422 genotype calls of an overlapping subset of  $\sim 2.2$  million variants between the two platforms from

423 all samples, using VCFtools (<http://vcftools.sourceforge.net/>). This enabled imputing the same  
424 set of variants into all samples, a reasonable solution given that the concordance between hard  
425 genotype calls and imputed genotypes is high.

426 Multiple sample and variant quality control (QC) steps were performed before running  
427 imputation to ensure high confidence variants and to remove outlier or related samples from  
428 eQTL analysis. We used the toolset PLINK<sup>17</sup> to perform appropriate genotype QC filters  
429 (Extended Data Table 6). This resulted in 1,883,274 autosomal variants genotyped across 450  
430 GTEx donors.

431 **Imputation of autosomal genotypes.** To increase power and resolution for discovering new  
432 eQTLs in the different GTEx tissues collected from the donors, we imputed variants from 1000  
433 Genomes Project into the QC filtered Omni 5M+2.5M merged genotype data for 451 GTEx  
434 donors. The reference panel version used was the 1000 Genomes Phase 1 integrated variant set  
435 release March 2012 (release v3), updated on 24 August 2012, downloaded from the IMPUTE2  
436 website: [https://mathgen.stats.ox.ac.uk/impute/data\\_download\\_1000G\\_phase1\\_integrated.html](https://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_integrated.html).  
437 This v3 version includes single nucleotide polymorphisms (SNPs) and indels and is limited to  
438 variants with more than one minor allele copy ("minor allele count greater than 1") across all  
439 1,092 individuals.

440 We filtered out variants with incompatible alleles between the Omni 5M or 2.5M arrays and the  
441 1000 Genomes reference data, and variants with a frequency difference larger than 0.15 between  
442 GTEx and 1000 Genomes samples, computed using samples of European descent, which  
443 constitute the majority of samples in GTEx. Variants were aligned between GTEx samples and  
444 1000 Genomes Project by chromosome position (genome build 37), removing variants that did  
445 not align.

446 The imputation of autosomes was run using the Ricopili pipeline  
447 (<https://sites.google.com/a/broadinstitute.org/ricopili/>). Prephasing was performed on all samples  
448 together using SHAPEIT v2.r644  
449 ([https://mathgen.stats.ox.ac.uk/genetics\\_software/shapeit/shapeit.html](https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html)). Imputation was  
450 performed using IMPUTE2 2.2.7\_beta with the default effective population size of 20,000 on 3  
451 Mb segments across each chromosome, which were subsequently merged. This yielded  
452 14,390,153 variants across 451 samples. After imputation was completed, a chromosome 17  
453 trisomy individual (GTEx-UPIC) was discovered and its genotypes was removed from the  
454 analysis freeze VCF, resulting in genotype data for 450 donors.

455 The following QC filters were applied to the genotyped and imputed array VCF for eQTL  
456 analysis: INFO < 0.4, minor allele frequency (MAF) < 1%, Hardy-Weinberg Equilibrium (HWE)  
457  $P \leq 1.0 \times 10^{-6}$ . We calculated missing rate for best-guessed genotypes, and the HWE test was



458 performed using the software tool SNPTEST<sup>73</sup> using only samples from European descent.  
459 Indels with length >51 base pairs were removed. About 13% of variants were hard call genotypes  
460 and 87% of variants were imputed. About 91% of the total numbers of variants were SNPs, and  
461 8.9% were indels. The REF and ALT alleles in the imputed VCF were checked for alignment to  
462 the human reference genome hg19, and the REF and ALT sequences were added for both SNPs  
463 and indels.

464 The final genotyped and imputed array VCF (file format v4.1) for autosomal variants contains  
465 genotype posterior probabilities for each of the three possible genotypes for 11,552,519 variants  
466 across 450 GTEx donors. The dosages of the alternative alleles relative to the human reference  
467 genome hg19 were used as the genotype measure for eQTL analysis. To assess the accuracy of  
468 imputation of autosomal chromosomes, we compared the alternative allele dosages between  
469 imputed and genotyped calls, using the Omni 2.5S set of variants for 183 GTEx samples from  
470 the pilot phase, for which we have both direct calls on the Omni 5M array and imputed calls  
471 from the merged set of 450 samples. Imputation accuracy was assessed using the coefficient of  
472 determination ( $R^2$ ) computed for each of the 2.5S variants separately across 183 samples and  
473 between the alternative allele dosage of the post-QC'd imputed calls and the directly genotyped  
474 calls. The imputation accuracy observed was very high for common variants (mean  $R^2 = 0.931$ -  
475  $0.969$ ; median  $R^2 = 0.985$ - $0.989$ ), and, as expected, somewhat lower, for low frequency variants  
476 (mean  $R^2 = 0.722$ - $0.906$ ; median  $R^2 = 0.804$ - $0.976$ ; Extended Data Table 7).

477 We computed the principal components (PCs) of the genotyped and imputed variants for 451  
478 GTEx samples using EIGENSTRAT<sup>74</sup> as implemented in Ricopili  
479 (<https://sites.google.com/a/broadinstitute.org/ricopili/pca>). This was done using a genome-wide  
480 set of linkage disequilibrium (LD)-pruned variants ( $R^2 > 0.2$ , plink --indep-pairwise 200 100 0.2)  
481 generated from best-guessed genotype calls after imputation (posterior probability > 0.9).  
482 Variant filters were applied, including the exclusion of variants not present in all samples, strand  
483 ambiguous SNPs (AT, CG), variants in the MHC region, variants with MAF < 5% or HWE  $P <$   
484  $1.0 \times 10^{-4}$ , and variant missing rate > 2%. For eQTL analysis, the first three genotype PCs were  
485 used as covariates, as they captured the largest proportions of genotype variance of the top  
486 genotype PCs (See Supplemental material in [Aguet et al, GTEx cis-eQTL manuscript, in  
487 submission]).

488 **Trans-eQTL association testing.** Matrix eQTL<sup>15</sup> was used to test all autosomal variants (MAF  
489 > 0.05) with all gene transcripts, restricted to lying on different chromosomes, in each tissue  
490 independently using an additive linear model. We included the three genotype PCs, genotyping  
491 platform, sex, and PEER factors estimated from expression data in Matrix eQTL when  
492 performing association testing. The correlation between variant and gene expression levels was  
493 evaluated using the estimated t-statistic from this model, and corresponding FDR was estimated

494 using Benjamini-Hochberg FDR correction<sup>15,75</sup> separately within each tissue and also using  
495 permutation analysis.

496 **Trans-eQTL quality control.** Mappability of every k-mer of the reference human genome  
497 (hg19) computed by the ENCODE project<sup>35</sup> has been downloaded from the UCSC genome  
498 browser (accession: wgEncodeEH000318, wgEncodeEH00032)<sup>76</sup>. We have computed exon- and  
499 untranslated region (UTR)-mappability of a gene as the average mappability of all k-mers in  
500 exonic regions and UTRs, respectively. We have chosen  $k = 75$  for exonic regions, as it is the  
501 closest to GTEx read length among all possible values of  $k$ . However, as UTRs are generally  
502 small regions, and 36 is the smallest among all possible values of  $k$ , we have chosen  $k = 36$  for  
503 UTRs. Finally, mappability of a gene is computed as the weighted average of its exon-  
504 mappability and UTR-mappability, weights being proportional to the total length of exonic  
505 regions and UTRs, respectively. We excluded from association testing any gene with  
506 mappability  $< 0.8$ .

507 The set of genetic variants tested have also been reduced by first filtering out all variants with  
508 MAF  $< 0.05$  in individuals sampled for the tissue being tested (reducing the variant set to  
509 6,226,121), and then filtering out all variant that are annotated by RepeatMasker to belong to a  
510 repeat region [<http://www.repeatmasker.org/>], release library version 20140131 for hg19. This  
511 filtering reduced the number of variants tested by roughly 53.6%, from 6,226,121 variants to  
512 2,889,379.

513 Next, we aligned every 75-mer in exonic regions and 36-mers in UTRs of every gene with  
514 mappability below 1.0 to the reference human genome (hg19) using Bowtie (v 1.1.2)<sup>77</sup>. If any of  
515 the alignments started within an exon or a UTR of another gene, then that pair of genes are cross-  
516 mappable. We excluded from consideration any variant-gene pair where the variant is within  
517 100 Kb of a gene that cross-maps with the potential trans-eQTL target gene.

518 Population structure is another source of potential false positives, and we control for three  
519 genotype principal components (PCs). While this should capture most broad effects of ancestry,  
520 we additionally check for residual evidence of strong correlation with a larger set of 20 genotype  
521 PCs (Extended Data Table 8). We observe a modest increase in correlation among trans-  
522 eVariants (Extended Data Fig. 13). While we opted not to apply further filtering, we have  
523 flagged any trans-eVariant with maximum correlation greater than 99% of the levels observed  
524 among random variants for use in future downstream analyses that may depend on ancestry.

525 **Linkage disequilibrium, cis-eQTL, and GWAS restricted trans-eQTL tests.** We performed  
526 restricted trans-eQTL association tests by filtering the set of variants considered in three ways.  
527 First, we filtered the final VCF files using linkage disequilibrium LD-pruning ( $R^2 > 0.5$ , plink  
528 parameters --indep 50 5 2), removing approximately 90% of variants. Next, from the original

529 VCF file, we performed association mapping using only the most significant GTEx cis-eQTL per  
530 eGene per tissue [Aguet et al, GTEx cis-eQTL manuscript, in submission]. From the original  
531 VCF file, we performed association mapping using only variants that had been found to have a  
532 trait association in a genome-wide association study<sup>26</sup> ( $P \leq 2.0 \times 10^{-5}$ ). The three association  
533 mapping analyses and FDR estimation were performed in each tissue separately.

534 **Intra-chromosomal long-range eQTL detection.** Phased allelic expression data were collected  
535 for all LD pruned eQTL ( $FDR \leq 0.1$ ) and only those eQTL with data in at least 10 eVariant  
536 homozygotes and heterozygotes were used. To remove cases where strong allelic imbalance was  
537 seen in eQTL homozygotes, the top 5% of eQTL sorted by homozygote allelic imbalance were  
538 filtered. To minimize the number of phasing errors that occur at long, chromosome wide  
539 distances, we developed a model that predicts the probability of phasing error as a function of the  
540 minor allele frequency of both the eVariant and a coding variant where ASE is assessed, as well  
541 as the distance between them. We used this model to filter cases where the predicted probability  
542 of correct phasing was  $< 99\%$ . A beta-binomial mixture model was then used to determine if the  
543 allelic data supported the presence of a cis-eQTL. To identify long-range cis-eQTL, from eQTL  
544 with TSS distance  $> 5$  Mb the top eQTL per gene was selected, and multiple testing correction  
545 was performed using the Benjamini-Hochberg FDR method on a per tissue basis. We next  
546 quantified the proportion of eQTL with significant (nominal  $P \leq 0.01$ ) ASE supported evidence  
547 of cis regulation as a function of distance to eGene TSS. Although we attempted to reduce  
548 phasing error, we were unable to accurately estimate the remaining error, so we compared the  
549 observed proportion of cis-eQTL to what would be expected under the worst case scenario of  
550 phasing error. Performance under the worst case scenario was determined by introducing phasing  
551 error between eVariants and ASE data at a rate of 50% to LD pruned eQTL ( $FDR \leq 0.1$ ) within  
552 100 Kb of the TSS, which were assumed to act in cis, and then determining the number of  
553 significant (nominal  $P \leq 0.01$ ) ASE supported cis-eQTL that could be identified as a function of  
554 eQTL effect size.

555 **Cross-tissue trans-eQTLs.** We used MetaTissue to quantify the tissue-specificity trans-  
556 eQTLs<sup>19</sup>. We ran MetaTissue with the heuristic option on to increase detection of cross-tissue  
557 differences. As MetaTissue, with the heuristic option on, does not permit analysis across all 44  
558 tissues, we restricted to the 20 tissues with the largest sample sizes. We restricted to the best  
559 variant per trans-eGene ( $FDR \leq 0.5$  in 20 tissue subset; 798 eGenes) and the best variant per  
560 randomly selected cis-eGene ( $FDR \leq 0.5$  in 20 tissue subset). We also analyzed the top cis-  
561 eGenes by p-value in a separate comparison. The distribution of cis-eGene discovery tissues was  
562 matched to that observed in trans. As input to MetaTissue, we used the same genotype and  
563 expression matrices as were used in the tissue-specific Matrix eQTL association analysis. As  
564 MetaTissue does not handle tissue specific covariates and allows for only one genotype file, we  
565 controlled for general covariates (gender, genotype PCs, and DNA platform) in genotype. For  
566 each tissue type, we controlled for all covariates (tissue-specific and general) in the gene

567 expression levels and projected the expression levels of each gene to the quantiles of a standard  
568 normal.

569 **Tissue clustering from effect size in trans-eQTLs.** Hierarchical agglomerative clustering was  
570 performed on trans-eGenes ( $FDR \leq 0.5$ ) using distance metric (1 – Spearman correlation) of  
571 MetaTissue effect sizes across all observed genes between tissue pairs.

572 **Hierarchical FDR control for multi-tissue eVariant discovery.** We applied a hierarchical  
573 FDR control approach to identify significant trans-eVariants across all variants, genes, and  
574 tissues together as a second assessment of tissue-specificity of trans-eQTLs<sup>20</sup>. As input, we  
575 considered 305,822 variants from the LD-pruned set that had a nominal trans association  $P \leq 1.0$   
576  $\times 10^{-7}$  with at least one gene. Let  $H_{ijk}$  denote the null hypothesis of no association between  
577 variant  $i$  and the expression of gene  $j$  in tissue  $k$ ,  $H_{ij\cdot}$  denote the null hypothesis of no association  
578 between variant  $i$  and gene  $j$  in any tissue, and  $H_{i\cdot\cdot}$  denote the null hypothesis of no association  
579 between variant  $i$  and any gene in any tissue. We consider  $i$  to be an eVariant if we reject  $H_{i\cdot\cdot}$ , and  
580 a variant-gene pair to be discovered if we reject  $H_{ij\cdot}$ .

581 To evaluate  $H_{ijk}$ ,  $H_{ij\cdot}$  and  $H_{i\cdot\cdot}$ , we performed a hierarchical testing procedure<sup>20,78</sup>. P-values were  
582 defined starting from the leaf hypotheses  $H_{ijk}$ , where we used the association p-value  $p_{ijk}$   
583 calculated by Matrix eQTL. P-values  $p_{ij\cdot}$  corresponding the variant x gene null hypotheses  $H_{ij\cdot}$   
584 across tissues were then calculated using Simes<sup>79</sup>, and p-values  $p_{i\cdot\cdot}$  corresponding to the variant-  
585 level null hypothesis  $H_{i\cdot\cdot}$  were also calculated using Simes. We then applied the Benjamini-  
586 Hochberg (BH) procedure on  $p_{i\cdot\cdot}$  to identify eVariants at  $FDR \leq 0.1$ . Next, we applied BH with  
587 an adjusted threshold to account for variant selection to the collections of  $p_{ij\cdot}$  for each discovered  
588 eVariant  $i$  to identify which genes it controls. Finally, we applied BH with a threshold adjusted to  
589 account for the two previous selection steps to each of the collections of  $p_{ijk}$  corresponding to  
590 each discovered eVariant-eGene pair to identify the tissues in which this regulation is present.  
591 This three-level procedure controls the FDR of eVariants, the average expected proportion of  
592 false variant-gene associations across eVariants<sup>78</sup>, and the expected weighted average of false  
593 tissue discoveries for the selected variant-gene pairs (weighted by the size of the eVariant and  
594 eGene sets) to the target  $FDR \leq 0.1$ .

595 **Cis regulatory element enrichment analysis.** We annotated discovered trans-eVariants using  
596 chromatin state predictions from 127 cell types or cell lines sampled by the Roadmap  
597 Epigenomics project<sup>33</sup>. Each cell type or cell line has the genome segmented by a 15-state hidden  
598 Markov model (HMM) in 400 bp windows. Several of these states are labeled as types of  
599 'enhancers', 'promoters,' and 'repressed regions.' For the standard 15-state Roadmap  
600 segmentations, regulatory elements are labeled independently for each cell type. Our analysis  
601 was restricted to GTEx tissues that are composed of at least one Roadmap Epigenomics cell type  
602 (26 tissues); which included 84 eVariants and 24 eGenes ( $FDR \leq 0.1$ ). We matched these

603 variants to randomly selected variants based on chromosome, distance to nearest TSS, and MAF.  
604 We quantified enrichment of the trans variants relative to random variants in both enhancer and  
605 promoter elements in the GTEx discovery tissue's matched Roadmap cell type (Extended Data  
606 Table 9). We then performed the same analysis with randomly matched cis-eGenes. Matching  
607 performed as follows: for each of the 24 trans-eGenes  $g$ , each having  $N_g$  associated eVariants  
608 ( $FDR \leq 0.1$ ), we randomly selected a cis-eGene that also had at least  $N_g$  associated variants ( $FDR$   
609  $\leq 0.1$ ). We then selected the top  $N_g$  variants associated with this gene based on p-value to use in  
610 the enrichment analysis. Selecting 24 random cis-eGenes for enrichments yields unstable  
611 enrichment, so we ran cis-eGene selection and enrichment 70 times with different selections. We  
612 rank ordered the 70 trials for both promoters and enhancers based on average odds ratio  
613 enrichment relative to background. We then used the trial that was closest to median rank for  
614 plotting both promoters and enhancers.

615 **piRNA cluster enrichment analysis.** We obtained a list of 6,250 piRNA clusters that were  
616 experimentally determined from RNA sequencing of human testis<sup>36</sup>. When considering all  
617 unique trans-eVariants identified in all tissues, we identified an enrichment of trans-eQTLs  
618 overlapping a piRNA cluster (17.8%) compared to the null expectation if trans-eVariants were  
619 randomly distributed compared to piRNA clusters (2.5%). To further establish the statistical  
620 significance of this observation, we generated a null distribution of piRNA-eVariant overlap by  
621 permutation. Using bedtools<sup>80</sup>, we permuted the location of piRNA clusters on the human  
622 genome 10,000 times, requiring the piRNA clusters be excluded from centromeres and sex  
623 chromosomes. We also evaluated the proportion of trans-eVariants located within 10 Kb of a  
624 piRNA cluster, and estimated the significance of this enrichment using the same permutation  
625 scheme.

626 **TCGA thyroid RNA-seq analysis.** To replicate trans-eVariants in thyroid, we used Thyroid  
627 Carcinoma (THCA) RNA-seq and genotype array data from The Cancer Genome Atlas (TCGA).  
628 Filtering out tumor normal and metastatic samples, we restricted our analysis to 496 primary  
629 tumor samples<sup>45</sup>. Next, after log transforming RNA-seq RSEM measurements<sup>81</sup>, we quantile  
630 normalized the data to the empirical distribution such that each sample has the same distribution.  
631 Next, we ensured that expression of each gene follows a Gaussian distribution by projecting each  
632 gene expression levels to the quantiles of a standard normal. To account for noise and  
633 confounding factors in RNA-seq measurements, we corrected the data by controlling for the first  
634 five gene expression principal components using a linear model. After this, using a linear model,  
635 we tested the effect of each variant in chr 9 position 100600000 - 100670000 on expression  
636 levels of all distal genes. We used the Benjamini-Hochberg method to correct for multiple  
637 hypotheses testing. Genes with  $FDR \leq 0.1$  were called as trans-eGenes.

638 **Colocalization analysis.** To quantify the probability that cis- and trans-eGenes share the same  
639 causal genetic locus in thyroid and muscle, we used Coloc<sup>49</sup> with p-value summary statistics as  
640 input.

#### 641 **Data availability**

642 Genotype data from the GTEx v6 release are available in dbGaP (study accession  
643 phs000424.v6.p1; [http://www.ncbi.nlm.nih.gov/projects/gap/cgi-](http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000424.v6.p1)  
644 [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 VCF files for imputed array data are in the  
645 archive phg000520.v2.GTEx MidPoint Imputation.genotype-calls-vcf.c1.GRU.tar (the archive  
646 contains a VCF for chromosomes 1-22 and a VCF for chromosome X). Allelic expression data is  
647 also available in dbGap. Expression data (read counts and RPKM) and eQTL input files  
648 (normalized expression data and covariates for 44 the tissues) from the GTEx v6p release are  
649 available from the GTEx Portal (<http://gtexportal.org>). eQTL results are available from the GTEx  
650 Portal.

#### 651 **References**

- 652 1. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in  
653 humans. *Science* **348**, 648–660 (2015).
- 654 2. Nica, A. C. *et al.* The Architecture of Gene Regulatory Variation across Multiple Human  
655 Tissues: The MuTHER Study. *PLOS Genet* **7**, e1002003 (2011).
- 656 3. Battle, A. *et al.* Characterizing the genetic basis of transcriptome diversity through RNA-  
657 sequencing of 922 individuals. *Genome Res.* **24**, 14–24 (2014).
- 658 4. Dimas, A. S. *et al.* Common regulatory variation impacts gene expression in a cell type-  
659 dependent manner. *Science* **325**, 1246–1250 (2009).
- 660 5. Huang, G.-J. *et al.* High resolution mapping of expression QTLs in heterogeneous stock mice  
661 in multiple tissues. *Genome Res.* **19**, 1133–1140 (2009).
- 662 6. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation  
663 in humans. *Nature* **501**, 506–511 (2013).
- 664 7. Stranger, B. E. *et al.* Population genomics of human gene expression. *Nat. Genet.* **39**, 1217–  
665 1224 (2007).
- 666 8. Montgomery, S. B. & Dermitzakis, E. T. From expression QTLs to personalized  
667 transcriptomics. *Nat. Rev. Genet.* **12**, 277–282 (2011).
- 668 9. Rockman, M. V. & Kruglyak, L. Genetics of global gene expression. *Nat. Rev. Genet.* **7**,  
669 862–872 (2006).
- 670 10. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional  
671 regulation in budding yeast. *Science* **296**, 752–755 (2002).
- 672 11. Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and disease.  
673 *Nat. Rev. Genet.* **16**, 197–212 (2015).
- 674 12. Westra, H.-J. *et al.* Systematic identification of trans eQTLs as putative drivers of known  
675 disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).

- 676 13. Innocenti, F. *et al.* Identification, Replication, and Functional Fine-Mapping of Expression  
677 Quantitative Trait Loci in Primary Human Liver Tissue. *PLOS Genet* **7**, e1002078 (2011).
- 678 14. Glass, D. *et al.* Gene expression changes with age in skin, adipose tissue, blood and brain.  
679 *Genome Biol.* **14**, R75 (2013).
- 680 15. Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations.  
681 *Bioinformatics* **28**, 1353–1358 (2012).
- 682 16. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of  
683 expression residuals (PEER) to obtain increased power and interpretability of gene  
684 expression analyses. *Nat. Protoc.* **7**, 500–507 (2012).
- 685 17. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based  
686 linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- 687 18. Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E. & Lappalainen, T. Tools and  
688 best practices for data processing in allelic expression analysis. *Genome Biol.* **16**, 195  
689 (2015).
- 690 19. Sul, J. H., Han, B., Ye, C., Choi, T. & Eskin, E. Effectively Identifying eQTLs from Multiple  
691 Tissues by Combining Mixed Model and Meta-analytic Approaches. *PLOS Genet* **9**,  
692 e1003491 (2013).
- 693 20. Peterson, C. B., Bogomolov, M., Benjamini, Y. & Sabatti, C. TreeQTL: hierarchical error  
694 control for eQTL findings. *Bioinforma. Oxf. Engl.* **32**, 2556–2558 (2016).
- 695 21. Buil, A. *et al.* *Quantifying the degree of sharing of genetic and non-genetic causes of gene*  
696 *expression variability across four tissues.* (2016).
- 697 22. Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in  
698 twins. *Nat. Genet.* **44**, 1084–1089 (2012).
- 699 23. Hore, V. *et al.* Tensor decomposition for multiple-tissue gene expression experiments. *Nat.*  
700 *Genet.* **48**, 1094–1100 (2016).
- 701 24. Bryois, J. *et al.* Cis and Trans Effects of Human Genomic Variants on Gene Expression.  
702 *PLoS Genet.* **10**, (2014).
- 703 25. Huan, T. *et al.* A Meta-analysis of Gene Expression Signatures of Blood Pressure and  
704 Hypertension. *PLOS Genet* **11**, e1005035 (2015).
- 705 26. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.  
706 *Nucleic Acids Res.* **42**, D1001–1006 (2014).
- 707 27. Pierce, B. L. *et al.* Mediation Analysis Demonstrates That Trans -eQTLs Are Often  
708 Explained by Cis -Mediation: A Genome-Wide Analysis among 1,800 South Asians. *PLOS*  
709 *Genet* **10**, e1004818 (2014).
- 710 28. Fairfax, B. P. *et al.* Genetics of gene expression in primary immune cells identifies cell type-  
711 specific master regulators and roles of HLA alleles. *Nat. Genet.* **44**, 502–510 (2012).
- 712 29. Zhang, X., Cal, A. J. & Borevitz, J. O. Genetic architecture of regulatory variation in  
713 *Arabidopsis thaliana.* *Genome Res.* **21**, 725–733 (2011).
- 714 30. Rakitsch, B. & Stegle, O. Modelling local gene networks increases power to detect trans-  
715 acting genetic effects on gene expression. *Genome Biol.* **17**, 33 (2016).

- 716 31. Mostafavi, S. *et al.* Normalizing RNA-Sequencing Data by Modeling Hidden Covariates  
717 with Prior Knowledge. *PLOS ONE* **8**, e68141 (2013).
- 718 32. Leek, J. T. & Storey, J. D. Capturing Heterogeneity in Gene Expression Studies by Surrogate  
719 Variable Analysis. *PLOS Genet* **3**, e161 (2007).
- 720 33. Mangravite, L. M. *et al.* A statin-dependent QTL for GATM expression is associated with  
721 statin-induced myopathy. *Nature* **502**, 377–380 (2013).
- 722 34. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl.*  
723 *Acad. Sci. U. S. A.* **100**, 9440–9445 (2003).
- 724 35. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the  
725 human genome. *Nature* **489**, 57–74 (2012).
- 726 36. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human  
727 epigenomes. *Nature* **518**, 317–330 (2015).
- 728 37. Gou, L.-T. *et al.* Pachytene piRNAs instruct massive mRNA elimination during late  
729 spermiogenesis. *Cell Res.* **24**, 680–700 (2014).
- 730 38. Watanabe, T. & Lin, H. Posttranscriptional regulation of gene expression by Piwi proteins  
731 and piRNAs. *Mol. Cell* **56**, 18–27 (2014).
- 732 39. Ha, H. *et al.* A comprehensive analysis of piRNAs from adult human testis and their  
733 relationship with genes and mobile elements. *BMC Genomics* **15**, 545 (2014).
- 734 40. Zeller, T. *et al.* Genetics and Beyond – The Transcriptome of Human Monocytes and  
735 Disease Susceptibility. *PLOS ONE* **5**, e10693 (2010).
- 736 41. Kirsten, H. *et al.* Dissecting the genetics of the human transcriptome identifies novel trait-  
737 related trans-eQTLs and corroborates the regulatory relevance of non-protein coding loci.  
738 *Hum. Mol. Genet.* ddv194 (2015). doi:10.1093/hmg/ddv194
- 739 42. Brown, C. D., Mangravite, L. M. & Engelhardt, B. E. Integrative Modeling of eQTLs and  
740 Cis-Regulatory Elements Suggests Mechanisms Underlying Cell Type Specificity of eQTLs.  
741 *PLOS Genet* **9**, e1003649 (2013).
- 742 43. Small, K. S. *et al.* Identification of an imprinted master trans regulator at the KLF14 locus  
743 related to multiple metabolic phenotypes. *Nat. Genet.* **43**, 561–564 (2011).
- 744 44. Mancikova, V. *et al.* Thyroid cancer GWAS identifies 10q26.12 and 6q14.1 as novel  
745 susceptibility loci and reveals genetic heterogeneity among populations. *Int. J. Cancer* **137**,  
746 1870–1878 (2015).
- 747 45. Lidral, A. C. *et al.* A single nucleotide polymorphism associated with isolated cleft lip and  
748 palate, thyroid cancer and hypothyroidism alters the activity of an oral epithelium and  
749 thyroid enhancer near FOXE1. *Hum. Mol. Genet.* **24**, 3895–3907 (2015).
- 750 46. Eriksson, N. *et al.* Novel associations for hypothyroidism include known autoimmune risk  
751 loci. *PloS One* **7**, e34442 (2012).
- 752 47. Denny, J. C. *et al.* Variants near FOXE1 are associated with hypothyroidism and other  
753 thyroid conditions: using electronic medical records for genome- and phenome-wide studies.  
754 *Am. J. Hum. Genet.* **89**, 529–542 (2011).



- 755 48. De Felice, M. *et al.* A mouse model for hereditary thyroid dysgenesis and cleft palate. *Nat.*  
756 *Genet.* **19**, 395–398 (1998).
- 757 49. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic  
758 Association Studies Using Summary Statistics. *PLOS Genet* **10**, e1004383 (2014).
- 759 50. Agrawal, N. *et al.* Integrated Genomic Characterization of Papillary Thyroid Carcinoma.  
760 *Cell* **159**, 676–690 (2014).
- 761 51. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set  
762 collection. *Cell Syst.* **1**, 417–425 (2015).
- 763 52. Dehghan, A. *et al.* Association of novel genetic Loci with circulating fibrinogen levels: a  
764 genome-wide association study in 6 population-based cohorts. *Circ. Cardiovasc. Genet.* **2**,  
765 125–133 (2009).
- 766 53. Davalos, D. & Akassoglou, K. Fibrinogen as a key regulator of inflammation in disease.  
767 *Semin. Immunopathol.* **34**, 43–62 (2012).
- 768 54. Mann, C. J. *et al.* Aberrant repair and fibrosis development in skeletal muscle. *Skelet. Muscle*  
769 **1**, 21 (2011).
- 770 55. Suelves, M. *et al.* Plasmin activity is required for myogenesis in vitro and skeletal muscle  
771 regeneration in vivo. *Blood* **99**, 2835–2844 (2002).
- 772 56. Suelves, M. *et al.* uPA deficiency exacerbates muscular dystrophy in MDX mice. *J. Cell*  
773 *Biol.* **178**, 1039–1051 (2007).
- 774 57. Vidal, B. *et al.* Fibrinogen drives dystrophic muscle fibrosis via a TGFbeta/alternative  
775 macrophage activation pathway. *Genes Dev.* **22**, 1747–1752 (2008).
- 776 58. Taniguchi, T., Ogasawara, K., Takaoka, A. & Tanaka, N. IRF family of transcription factors  
777 as regulators of host defense. *Annu. Rev. Immunol.* **19**, 623–655 (2001).
- 778 59. White, L. C. *et al.* Regulation of LMP2 and TAP1 genes by IRF-1 explains the paucity of  
779 CD8+ T cells in IRF-1<sup>-/-</sup> mice. *Immunity* **5**, 365–376 (1996).
- 780 60. Penninger, J. M. *et al.* The interferon regulatory transcription factor IRF-1 controls positive  
781 and negative selection of CD8+ thymocytes. *Immunity* **7**, 243–254 (1997).
- 782 61. Kim, P. K. M. *et al.* IRF-1 expression induces apoptosis and inhibits tumor growth in mouse  
783 mammary cancer cells in vitro and in vivo. *Oncogene* **23**, 1125–1135 (2004).
- 784 62. Salazar, J. C. *et al.* Activation of Human Monocytes by Live *Borrelia burgdorferi* Generates  
785 TLR2-Dependent and -Independent Responses Which Include Induction of IFN-β. *PLOS*  
786 *Pathog* **5**, e1000444 (2009).
- 787 63. Bosinger, S. E. *et al.* Global genomic analysis reveals rapid control of a robust innate  
788 response in SIV-infected sooty mangabeys. *J. Clin. Invest.* (2009). doi:10.1172/JCI40115
- 789 64. Boehm, U., Klamp, T., Groot, M. & Howard, J. C. Cellular responses to interferon-gamma.  
790 *Annu. Rev. Immunol.* **15**, 749–795 (1997).
- 791 65. Shi, L., Perin, J. C., Leipzig, J., Zhang, Z. & Sullivan, K. E. Genome-wide analysis of  
792 interferon regulatory factor I binding in primary human monocytes. *Gene* **487**, 21–28 (2011).
- 793 66. Voight, B. F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale  
794 association analysis. *Nat. Genet.* **42**, 579–589 (2010).

- 795 67. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood  
796 lipids. *Nature* **466**, 707–713 (2010).
- 797 68. Varlamov, O., Bethea, C. L. & Roberts, C. T. Sex-Specific Differences in Lipid and Glucose  
798 Metabolism. *Front. Endocrinol.* **5**, (2015).
- 799 69. Van, P. L., Bakalov, V. K. & Bondy, C. A. Monosomy for the X-chromosome is associated  
800 with an atherogenic lipid profile. *J. Clin. Endocrinol. Metab.* **91**, 2867–2870 (2006).
- 801 70. Carr, M. C. The emergence of the metabolic syndrome with menopause. *J. Clin. Endocrinol.*  
802 *Metab.* **88**, 2404–2411 (2003).
- 803 71. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-  
804 Seq. *Bioinforma. Oxf. Engl.* **25**, 1105–1111 (2009).
- 805 72. DeLuca, D. S. *et al.* RNA-SeQC: RNA-seq metrics for quality control and process  
806 optimization. *Bioinformatics* **28**, 1530–1532 (2012).
- 807 73. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat.*  
808 *Rev. Genet.* **11**, 499–511 (2010).
- 809 74. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide  
810 association studies. *Nat. Genet.* **38**, 904–909 (2006).
- 811 75. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and  
812 Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300  
813 (1995).
- 814 76. Rosenbloom, K. R. *et al.* ENCODE data in the UCSC Genome Browser: year 5 update.  
815 *Nucleic Acids Res.* **41**, D56–63 (2013).
- 816 77. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient  
817 alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
- 818 78. Benjamini, Y. & Bogomolov, M. Selective inference on multiple families of hypotheses. *J.*  
819 *R. Stat. Soc. Ser. B Stat. Methodol.* **76**, 297–318 (2014).
- 820 79. Simes, R. J. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*  
821 **73**, 751–754 (1986).
- 822 80. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic  
823 features. *Bioinformatics* **26**, 841–842 (2010).
- 824 81. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or  
825 without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

## 826 **Acknowledgements**

827 The Genotype-Tissue Expression (GTEx) project was supported by the Common Fund of the Office of the Director  
828 of the National Institutes of Health ([commonfund.nih.gov/GTEx](http://commonfund.nih.gov/GTEx)). Additional funds were provided by the National  
829 Cancer Institute (NCI), National Human Genome Research Institute (NHGRI), National Heart, Lung, and Blood  
830 Institute (NHLBI), National Institute on Drug Abuse (NIDA), National Institute of Mental Health (NIMH), and  
831 National Institute of Neurological Disorders and Stroke (NINDS). Donors were enrolled at Biospecimen Source  
832 Sites funded by NCISAIC-Frederick, Inc. (SAIC-F) subcontracts to the National Disease Research Interchange  
833 (10XS170) and Roswell Park Cancer Institute (10XS171). The Laboratory, Data Analysis, and Coordinating Center  
834 (LDACC) was funded through a contract (HHSN268201000029C) to The Broad Institute, Inc. Biorepository  
835 operations were funded through an SAIC-F subcontract to Van Andel Institute (10ST1035). Additional data

836 repository and project management were provided by SAIC-F (HHSN261200800001E). The Brain Bank was  
837 supported by a supplement to University of Miami grant DA006227. A.B., E.T.D., T.L., and S.E.C. are supported by  
838 NIH grant R01MH101814 (NIH Common Fund; GTEEx Program). TwinsUK is funded by the Wellcome Trust,  
839 Medical Research Council, European Union, the National Institute for Health Research (NIHR)-funded  
840 BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS  
841 Foundation Trust in partnership with King's College London. A.B. is supported by the Searle Scholars Program,  
842 NIH grant 1R01MH109905, and NIH grant R01HG008150 (NHGRI; Non-Coding Variants Program). B.E.E. is  
843 supported by NIH grant R00 HG006265, NIH R01 MH101822, NIH U01 HG007900, and a Sloan Faculty  
844 Fellowship. C.D.B. is supported by NIH grant R01 MH101822. E.T.D. is supported by the NIH-NIMH, European  
845 Research Council (ERC), Swiss National Science Foundation and Louis Jeantet Foundation. B.J. is supported by  
846 NIH grant 2T32HG003284-11. T.L. and P.M. are supported by the NIH grant R01MH106842. T.L. is supported by  
847 the NIH grant UM1HG008901. T.L. and SEC are supported by the NIH contract HHSN2682010000029C. C.B.P.  
848 and C.S. are supported by NIH grant R01 MH101782. D.F.C. is supported by NIH grant R01MH101810. E.R.G and  
849 N.J.C are supported by NIH grants R01 MH101820 and R01 MH090937A. The authors would like to thank  
850 Abhinav Nellore and Christopher Wilks for assistance with TCGA data, Kerrin Small for discussions, and Jeffrey T.  
851 Leek for suggestions on the manuscript.

852 **Author contributions**

853 B.J., Y.H., B.J.S., P.P., A.B., and B.E.E. designed the study, performed the analysis, and wrote the manuscript.  
854 G.G., C.P., S.E.C., A.S., A.A.B, A.H., P.M., G.Q., and D.F.C. contributed analysis. All authors provided a critical  
855 review of the analyses and the manuscript.