www.sciencemag.org/content/352/6285/600/suppl/DC1

Supplementary Materials for

RNA splicing is a primary link between genetic variation and disease

Yang I. Li, Bryce van de Geijn, Anil Raj, David A. Knowles, Allegra A. Petti, David Golan, Yoav Gilad,* Jonathan K. Pritchard*

*Corresponding author. Email: gilad@uchicago.edu (Y.G.); pritch@stanford.edu (J.K.P.)

Published 29 April 2016, *Science* **352**, 600 (2016)
DOI: 10.1126/science.aad9417

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:
(available at www.sciencemag.org/content/352/6285/600/suppl/DC1)

Materials and Methods

4sU metabolic labelling

We adapted the 4sU labelling method from (*26*). Briefly, cell cultures were grown to log phase in volumes sufficient to yield about 300 ng of 4sU-labeled RNA. Cells were incubated with 4sU for the required length of time (0, 30, or 60 minutes), then washed, pelleted, and frozen. Total RNA was extracted, and 4sU-labeled RNA was separated from total RNA using a bead-based biotin-streptavidin purification protocol.

Cell growth and 4sU-labeling. Cell lines were grown to log phase ($6.5\text{x}10^6$ - $8.5\text{x}10^6$ viable cells/ml, with at least 70% cell viability) in RPMI 1640 (supplemented with 2 mM L-glutamine and 15% fetal bovine serum) using standard procedures (as recommended by Coriell). Culture volumes were sufficient to yield at least $6.4\text{x}10^6$ viable cells, or enough to yield 300 ng of total RNA for the zero time point, and 300 ng of 4sU-labeled RNA for each of the 30' and 60' time points. At the designated "time 0," cultures were divided into 10-ml aliquots, and 50 mM 4sU was added to each aliquot to obtain a final 4sU concentration of $2\text{x}10^{-7}$ uMoles/cell. Aliquots corresponding to the 0' time point were immediately washed in 5 mL PBS, pelleted, and frozen in liquid nitrogen. Aliquots corresponding to subsequent time points were returned to the incubator, and washed and frozen at the appropriate times.

RNA isolation. RNA was isolated using QIAshredder columns (Qiagen) and the RNeasy mini kit (Qiagen), using the protocols provided. RNA concentration was measured using a NanoDrop spectrophotometer, and RNA quality was checked using an Agilent Bioanalyzer.

Biotinylation of 4sU-labeled RNA from the 30' and 60' time points. RNA samples were first adjusted to a concentration of 0.2 ug/ul in 100 ul. 100 ul Biotin labeling buffer (25 mM Tris-HCl 1M pH 7.4, 2.5 mM EDTA 500 mM) and 50 ul of EZ-Link HPDP-Biotin/DMF stock (Thermo Scientific) were added to each sample, and samples were incubated for two hours at room temperature. To remove unbound biotin, 250 ul chloroform:isoamylalcohol (24:1) was added to the biotinylation reaction, and RNA was isolated using Heavy Phase-lock-gel tubes.

RNA Precipitation. Biotinylated RNA was precipitated using ethanol precipitation, and resuspended in 100 ul water. Isolation of 4sU-labeled RNA. Dynabeads MyOne Streptavidin T1 beads (Invitrogen) were washed according to the manufacturer's instructions. 100 ul of washed beads were added to 100 ul of biotinylated RNA. The mixture was incubated at room temperature for 15 minutes, separated with a magnet for 3 minutes, and washed 2-3 times with

1X Binding and Washing Buffer. 4sU-labeled RNA was eluted from the beads using 100 ul of 100mM dithiothreitol (Pierce DTT No-Weigh Format (Thermo Scientific)), and eluted RNA was cleaned using the RNeasy MinElute kit (QIAgen) according to the manufacturer's instructions.

Molecular data mapping

The main molecular phenotypes analyzed in this study were assayed using H3K27ac ChIP-seq, DNase-seq, DNA Methylation 450K arrays, 4sU-sequencing, RNA-sequencing, Ribosome Profiling, and Mass Spectrometry (Table S1). For DNase-seq, Methylation 450K arrays and Mass Spectrometry data, we used previously quantified measurements from (*4, 18, 19*) and used `liftOver` to convert the coordinates from hg18 to hg19 where appropriate.

To map the activity of the other molecular traits to their corresponding genes or genomic regions, we used `bowtie2` (*27*) with option `--very-sensitive` for H3K27ac ChIP-seq data, and `STAR` (*28*) with option `--outSAMstrandField intronMotif` for 4sU-seq, RNA-seq (Pickrell (*17*) and YRI GEUVADIS) and ribo-seq data. We next used the WASP framework (*29*) to re-map reads in order to avoid mapping biased by sequence polymorphisms and remove duplicates for H3K27ac ChIP-seq (but not for gene-level phenotypes). We further used WASP (*4*) to account for read depth heterogeneity and biases for GC-content across libraries. This resulted in normalized read counts for each test window (see next section), which were subsequently used for QTL mapping.

To understand the relationship between our molecular trait measurements **(Fig. 1C)**, we measured the correlation of read counts mapping to different relevant regions of the genome. In particular, we used `featureCounts` (*30*) to count the number of H3K27ac reads mapping to $\pm$1kb of a gene transcription start site (TSS), the number of 4sU-seq, RNA-seq, and ribo-seq reads mapping to the gene body (as defined by its representative exons, see next section). To quantify protein expression, we used iBAC intensity measured at the whole protein level (*19*). We then used Spearman ρ to measure the correlation across genes and molecular phenotypes. We noticed that because of low average H3K27ac read counts mapping to the TSS, the Spearman ρ was unable to detect strong correlation between H3K27ac and gene-level phenotypes, possibly owing to its inability to resolve equal counts. We therefore used Pearson ρ to measure the correlation between H3K27ac at the promoter and gene-level phenotypes.

To test whether the correlation between H3K27ac and 4sU-seq read counts was higher than the one between H3K27ac and RNA-seq read counts, we generated 5,000 random samples of 6,722 genes by drawing with replacement among the 6,722 genes for which we had all three molecular phenotypes measured in individual NA18502. We then

computed their Pearson correlation (27ac-4sU $\mu_\rho = 0.237$ and 27ac-RNA $\mu_\rho = 0.193$) and obtained a mean difference in correlation and standard deviation for this estimate: $\mu_{\Delta\rho^2} = -0.0439$ and $\sigma = 0.0089$. Assuming that our estimates are normally distributed, $\mathcal{Z} = -4.9307$ which corresponds to a p-value = 4.1×10^{-7}. The results are very similar using other individuals.

Peak calling and test windows for molecular traits

We used several strategies to determine appropriate test windows for our molecular traits. For DNaseI and DNA methylation data, we used the same test windows as the original studies.

To determine test windows for H3K27ac, we ran `macs14` with default parameters on each of the 59 H3K27ac `bam` alignment files separately. Overlapping peaks across samples were then merged. MACS windows were then split into segments of 1kb (if they were bigger). We next augmented these peaks with LCL chromHMM annotation windows that were associated with transcription start sites (TssA, TssFlnk), transcription (Tx, TxWk), or enhancers (Enh, EnhG). To do this, we combined all MACS peaks and relevant chromHMM annotations and removed all chromHMM windows that overlapped with MACS peaks. This procedure resulted in 208,512 test windows genome-wide.

To determine test windows for gene-level phenotypes, we downloaded the gencode v19 gene annotations. We then identified all overlapping exons from a gene and considered the longest exons from each overlap as test windows for that gene. In all subsequent analyses, we focused on the top 14,000 most expressed protein-coding genes based on the average RPKM in all 69 RNA-seq samples (Pickrell) (*17*), which corresponds to genes with at least 10.4 RPKM.

Regression between mRNA decay and 4sU-seq, RNA-seq reads counts

To verify that 4sU-seq read counts quantitatively measures gene transcription rates **(Fig. 1B)**, we first computed the average RNA-seq and 4sU-seq read counts (normalized by total sample read counts) over 54 individuals in 8,316 genes for which we had 4sU-seq, RNA-seq, RNA decay rate data (*31*). The ratio of RNA-seq reads to 4sU-seq reads was then regressed against estimated RNA decay rate using the `glm()` function in `R`. The quasi-binomial family was used to account for over-dispersion. The regression between mRNA decay and $\frac{4sU}{RNA + 4sU}$ yielded a p-value = 8.31×10^{-167}. When we performed the same regression between mRNA decay and $\frac{1}{RNA}$ the p-value = 0.0053 (fig. S1), indicating that this effect is not driven by a correlation between RNA and decay.

cis-QTL mapping

As described earlier (Molecular data mapping), we used WASP to adjust differences in sequencing depth and GC content for each of our sample. We then used a normalization and standardization approach developed previously by our group (*4*). Briefly, we first standardized all measurements by gene and then quantile-normalized them to fit a standard normal distribution by individual. We next used principal components analysis (PCA) to regress out unidentified confounders. The numbers of PCs regressed out were chosen to maximize the number of detected QTLs in each data type (we tested 0 to 15 PCs).

To map cis-QTLs for genes, we used all SNPs with MAF ≥ 0.05 and $\pm$100kb of genes, and $\pm$50kb of DNaseI peaks (defined previously in (*4*)), DNA methylation probes and H3K27ac peaks/chromHMM windows. We used the intersection of genotyped position between HapMap 2 and HapMap3 to determine the genotypes of each individual because some of our individuals were genotyped in HapMap2 and some in HapMap3. We had previously imputed all SNPs from the high coverage 1000 Genomes Phase1 data (*5*). Standard linear regression was then used to compute a p-value for each SNP-gene/peak pair.

Effect sizes and correlation of *cis*-QTLs

To compute QTL effect sizes, we used the read depth and GC-corrected count data (H3K27ac ChIP-seq, 4sU-seq, RNA-seq, and ribo-seq) as input to our linear regression and did not regress out any PCs as this can modify effect size estimates. For protein QTL effect sizes, we used the raw (uncorrected) protein data from (*19*). We used the slope of the linear regression as a measure of effect size. As a starting point to compare the correlation of effect sizes across molecular phenotypes (Fig. 2A), we used the 501 gene-snp pairs identified in the GEUVADIS analysis of the YRI samples. Of these, 130 genes were not tested in our study (109 mean expression levels too low, 21 non-coding or with no corresponding IDs), 14 were indels, 33 had MAF < 0.05 and 67 were missing from our imputed SNPs, leaving us with 256 gene-snp pairs. We asked whether the effect sizes of the SNP representing the best SNP-gene association were correlated for H3K27ac read number at TSS, 4sU-seq read depth (at 30 and 60 minutes), RNA-seq read depth (Pickrell (*17*), GEUVADIS), ribo-seq read depth and protein iBAQ intensity **(Fig. 2A)**.

To compute effect sizes for sQTLs, we again used the slope of the linear regression. We used the percent-spliced intron (PSI) of introns as dependent variables and the genotypes [coded as 0 (homozygous reference), 1 (heterozy-

gous), and 2 (homozygous alternate)] as independent variables. As such, the ΔPSI defined as the expected change between the intron splicing levels of the reference versus alternate allele is twice the slope. In fig. S8, we plot the distribution of ΔPSI for our sQTLs. We find that 13.8% of our sQTLs had a ΔPSI $\geq$ 10% and 76.6% over 1%. We should note that while splicing variation with small ΔPSI (e.g. $\leq$ 1%) are expected to have modest impact in general, it is likely that some sQTLs have larger effects in different cell types and/or developmental stages, or that some variants produce aberrant transcripts resulting in toxic peptides.

We next wished to quantify general buffering of regulatory variants by studying average effect sizes of regulatory SNPs across our molecular phenotypes (fig. S4). We decided to start from 565 promoter-H3K4me3 QTLs recently identified in (*8*) at 5% FDR whose H3K4me3 peaks was at most 2.5kb away from the TSS of a gene. We reasoned that by starting from promoter-H3K4me3 QTLs instead of GEUVADIS eQTLs, we can avoid ascertainment biases caused by SNPs with post-transcriptional regulatory effects, thereby allowing us to establish an unbiased accounting of effect size buffering across the regulatory cascade. We therefore computed the effect sizes of the SNP with the strongest association to H3K4me3 levels at promoters on 4sU-seq read depth, RNA-seq (Pickrell (*17*)) read depth, ribo-seq read depth and protein iBAQ intensity. We then polarized the effect size by the direction of effect observed on H3K4me3 levels. Finally, we summarized the effect sizes for each regulatory stage as a boxplot and only SNP effect sizes at the protein level were significantly lower than other regulatory stages (mean ribosome and protein effect sizes were 0.224 and 0.062, respectively, Mann-Whitney U test $p = 0.0016$). We also repeated this same analysis starting from 1,347 eQTLs predicted in GEUVADIS CEU population and again observed that only SNP effect sizes at the protein level were lower than the other regulatory stages (mean ribosome and protein effect sizes of 0.159 and 0.077 respectively, Mann-Whitney U test $p = 2.41 \times 10^{-7}$).

Estimation of QTL sharing between H3K27ac levels at enhancers, H3K27ac levels at promoters, and gene-level phenotypes

To estimate the sharing between QTLs for a molecular trait **A** and another molecular trait **B (Fig. 2B)**, we first identified the best SNP-gene pair for trait **A**. We then used Storey's π_1 method (`qvalue()`) to estimate the sharing between **A** and **B** by considering the association (i.e. p-value) of the best SNP-gene pair from trait **A** in trait **B**. We used several p-value cutoffs on the best SNP-gene association to measure the effect of ascertainment stringency on our estimate of sharing (Table S3). We first found that our estimate of sharing generally increased as our stringency increased (fig. S2). This effect can be caused by (1) a reduction in the power to measure associations when effect sizes are small and (2) a higher false discovery rate (FDR) when the stringency is low. We reasoned however that because FDR is low (<10%)

for associations with p-values $< 10^{-4}$ and the differences in our sharing estimate is high (>30%), these observations are likely due to (1) a reduction in the power to measure associations. Consistent with our speculation, we observed a clear correspondence between SNP effect sizes and p-values of the SNP-gene association for gene expression levels (fig. S2C).

We used a bootstrap approach to compute confidence intervals for our π_1 estimates. Specifically, we resampled p-values with replacement N times (where N is the initial number of samples) and re-computed the π_1 estimate. We also combined our sharing estimates for the 4sU and RNA-seq tests by averaging their estimates and taking the minimum and maximum estimates to compute the lower and upper bound, respectively, of our confidence interval. Owing to the large fluctuations in the π_1 estimates depending on sampling, we report 80% confidence intervals.

To estimate the proportion of QTLs for H3K27ac levels at distal peaks that affect gene expression **(Fig. 2B)**, we took H3K27ac QTL peaks that overlapped a chromHMM-defined enhancer at most 100kb away from the TSS of a gene. We then estimated the sharing between haQTL and QTLs for H3K27ac read counts at the TSS of the closest gene and QTLs for 4sU-seq read count, RNA-seq read count, ribo-seq read counts and/or protein iBAQ intensity of the nearest gene. Overall, we find a modest sharing (20-40%). We reasoned that H3K27ac levels at enhancer may not necessarily affect the expression of the closest gene, but may often affect the expression level of a more distant gene.

We therefore next estimated the sharing between QTLs that affect H3K27ac levels and QTLs for expression levels of **any** nearby genes (fig. S2A). We estimated conservatively that up to 50% of haQTLs do not appear to affect the expression levels of any nearby gene (within 500kb). We were particularly conservative in our estimate because our approach is hindered by our inability to link distal haQTLs to the gene they regulate (when they do regulate a gene). Nevertheless, we used a strategy employed in a previous study (*4*), which estimated that 16% of dsQTLs affect expression levels of a nearby gene. Specifically, for each haQTL we did the following:

i) Classify haQTLs according to their distance to the nearest expressed genes (according to TSS coordinates) into bins $[0, 1\text{kb})$, $[1, 25\text{kb})$, $[25, 100\text{kb}]$.

ii) For each haQTL, take the minimum p-value for association between the QTL and expression of all genes within 1kb, 25kb, and 100kb according to the haQTL's bin.

ii*) We also performed step ii) by computing the association between haQTLs and expression of all genes within 500kb for every bin.

ii) Repeat (ii) for 10,000 permutations of the genotypes.

iii) Estimate a corrected p-value for the most significant eQTL based on the permutations.

iv) Use Storey's π_1 method to estimate the amount of sharing.

Using this procedure (with step ii), we observed a clear relationship between the estimated sharing and the distance between the haQTL peak and the closest TSS (fig. S2a). This observation also holds when using this procedure with step ii* (fig. S2A), so it cannot be fully explained by differences in the number of tests (and the multiple testing correction). When we restricted to the haQTLs with the strongest associations ($p < 10^{-8}$) in each distance category (256 in 1–25kb and 215 in 25–100kb), we estimated that 50.3% of haQTLs are also eQTLs (with step ii). We obtained lower estimates of sharing (40.4%) when we computed the association between haQTLs and expression of all genes within 500kb (using step ii*, fig. S2A). This is due to the larger number of tests and a stricter multiple testing correction. Again, we must stress here that the larger number of tests cannot account for the differences between our estimates of promoter (H3K27ac QTLs proximal to genes <1kb away) percolation versus enhancer (H3K27ac QTLs distal to genes 1kb–100kb away) percolation. Indeed, we observed that even when testing against all genes within 500kb (Right panel of fig. S2A), we estimated a percolation rate of >75% for promoter haQTLs compared to ∼40% for distal haQTLs. In conclusion, we estimate that up to half of these enhancer-like elements do not show evidence for regulating a nearby gene.

We should note here that H3K27ac peaks, although associated with enhancers, are not all expected to be *bona fide* enhancers. For example, it has been suggested that active transposable elements can create novel enhancer-like elements which may subsequently be "wired" into gene regulatory networks (*32*).

As estimates of the numbers of QTLs that affect each regulatory stage, we used the numbers from Table S3 ($p < 10^{-4}$, LR t-test). We should note that one challenge in interpreting these numbers is that they strongly depend on our power to detect association and do not reflect the numbers (even relative) of genetic loci that affect a molecular trait. Nevertheless, we estimated the number of shared QTLs according to Storey's π_1. These numbers are likely to be conservative (see fig. S2B,E), which is why we do not provide an estimate for the number of stage-specific QTLs for RNA-level molecular phenotypes. We did estimate the number of stage-specific QTLs for H3K27ac levels at enhancers, because the ascertainment cutoff had a marginal effect on our estimates of percolation (fig. S2A).

Characterization of the shared effects of eQTLs and sQTLs on chromatin

To estimate the number of eQTLs that may function through DNA-level phenotypes (e.g. transcription factor binding and/or epigenetic modifications) **(Fig. 2C)**, we estimated the number of eQTLs (called using 86 RNA-seq GEUVADIS samples) that were also QTLs for H3K27ac, DNaseI, and/or DNA methylation levels. We also obtained H3K4me1, H3K4me3 QTLs from (*8*) and tested whether our eQTLs were H3K4me1 or H3K4me3-level QTLs.

Specifically, we took the top 2,000 genes with association to gene-level phenotypes, which corresponds roughly to a FDR of 10% (9.82% FDR to be exact). We then took the best SNP for each gene according to two models (Min p-value or max posterior for the hierarchical model) and asked whether they were QTLs (at 10% FDR) for H3K27ac, DNaseI, and/or DNA methylation levels, or were associated at a p-value ($<$0.01) for H3K4me1 or H3K4me3-level. These cutoffs are arbitrary but were determined in relation to our control SNP set. Our control SNP set was derived from SNPs matched for distance from TSS (250bp bins), gene expression (quintile) and minor allele frequency (decile), but with the restriction that they must not have an association p-value < 0.1 with the expression level of any gene. As such, because control SNPs are highly enriched near transcription start sites, we expect a proportion of them to be true DNA-level trait QTLs. To define a similar control set of sQTLs, we sampled one random SNP that lie within an intron cluster (the SNP must therefore be within a gene) with no sQTL association.

Overall, we found that 63.8% of the top 2,000 eQTL SNP-gene pairs were also QTLs for one or more DNA-level traits (versus $\sim$20% for our control SNPs). These results suggest that the vast majority of eQTLs function through DNA-level traits. We note here that it is also possible some genetic variants affect both aspects of chromatin regulation and post-transcriptional regulation separately, or that some eQTLs function through co- or post-transcriptional processes and only indirectly affects DNA-level phenotypes. We do not test these alternative models in the current study, however a recent study (*9*) suggests that these are relatively uncommon ($<$20% of the time). We find only a small excess of sQTLs that are chromatin-QTLs (17.5%) compared to sQTLs control SNPs (13.3%).

Of the 2,000 genes we considered, we expect 10% of the eQTLs to be false positives. We found that 63.8% of eQTLs and $\sim$20% of our control SNPs were also chromatin-QTLs. We may therefore estimate that $(63.8\% - 20\%)/0.90 = 48.7\%$ of all eQTLs are chromatin-QTLs. However, we also expect some SNPs to be chromatin-QTLs by chance. Therefore if we assumed that if $\sim$15$\pm$5% of random SNPs within promoter regions are chromatin-level QTLs, the estimate that $\sim$65% of eQTLs are chromatin-trait QTLs is likely to be within 6% of the true estimate:

$\frac{63.8\%-20\%+[15\%\pm5\%]}{0.90} = 65.3 \pm 5.5\%$. Again, this 6% confidence level is based on the assumption that 15% ($\pm$5%) of SNPs near promoters are chromatin-QTLs. While we found this number useful to obtain an intuition of how accurate our 65% estimate is, we do not think of this as a confidence interval in a statistical sense.

We next wondered whether eQTLs without DNA-level association (unexplained eQTLs or ueQTLs) were enriched in genomic regions informative of regulatory mechanisms **(Fig. 2D)**. We therefore took the 709 ueQTLs and asked whether they were enriched in any particular genomic annotation. Like eQTLs that are chromatin QTLs, we found that ueQTLs were enriched within gene bodies. However, unlike eQTLs, we found no enrichment of ueQTLs near the TSS. This observation is consistent with the notion that they do not function by altering transcription initiation dynamics, which is often reflected in histone modifications.

To further identify genomic annotations enriched for ueQTLs, we used the posterior weights from the hierarchical model (see section entitled "Bayesian hierarchical model to identify causal eQTLs") to assign to each SNP-gene pair a probability that it is the causal eQTL (or ueQTL). We then identified genomic annotations (Table S4) in which these eQTLs (and ueQTLs) lie and summed their probabilities for each annotation independently (note that annotations are not disjoint). For all annotations, we then computed the $\log_2$ differences between the associated ueQTLs and eQTLs probabilities. We computed bootstraps for these estimates by resampling 1,000 times with replacement from the original dataset and computing the 95% confidence interval (fig. S6).

To compute a p-value measuring the difference in contribution of SNPs that lie in regions associated with transcription elongation, gene exons (regions annotated as exonic, and 3'UTR, but not 5' UTR), and gene introns, we first summed the posteriors of SNPs that fell in chromHMM regions associated with transcription elongation (*33*), gene exons and introns, respectively, for each gene as follows:

$$G_{\text{eQTL}} = \{\sum_i P^g_{\text{snp}_i} \cdot \mathbb{1}_{\text{snp}_i \in \text{Txn elongation}} : g \text{ has an eQTL}\} \quad (1)$$

$$G_{\text{ueQTL}} = \{\sum_i P^g_{\text{snp}_i} \cdot \mathbb{1}_{\text{snp}_i \in \text{Txn elongation}} : g \text{ has an ueQTL}\} \quad (2)$$

where $P^g_{\text{snp}_i}$ is the posterior that snp_i is the causal eQTL (or ueQTL) for a gene g. We then compared the distribution G_{eQTL} to G_{ueQTL} using a two-tailed Mann-Whitney U test yielding a p-value = 3.02×10^{-22} for transcription elongation, p = 5.42×10^{-8} for exons and p = 0.006 for introns.

RNA decay and alternative polyadenylation site QTLs

RNA decay. To identify genetic variants associated with RNA decay, we used quantitative measurements previously established in (*31*) and performed linear regression using our panel of imputed individuals. In total, we obtained only 9 RNA decay QTL at 10% FDR. We then asked how many of the regulatory QTLs we identified (at $p < 10^{-4}$) for H3K27ac levels at TSS, transcription, steady-state mRNA, ribosome occupancy, and protein expression levels were RNA decay QTLs using Storey's π_0 method (Table S5).

Interestingly, we estimated that 17.4% and 17.3% of haQTLs at gene promoters and transcription rate QTLs, respectively, were also classified as decay QTLs, versus ∼20% of QTLs for steady-state mRNA and ribosome occupancy levels (although note that our confidence intervals are very large for these estimates). These observations support the notions that (1) a significant fraction of decay QTLs are the result of complex regulatory circuits and/or transcription-coupled decay (*31*) and that (2) genetic variation that directly affect mRNA decay rates do not play a major role in steady-state mRNA level variation.

Alternative polyadenylation sites. To identify alternative polyadenylation site QTLs (apaQTLs), we predicted and quantified alternatively poly-adenylated (APA) sites using a customized version (we lowered the hard-coded coverage threshold from 20 to 3) of DaPars (*34*) and found 11,157 putative APA sites. Using the ratio of distal to proximal polyA site usage as quantitative trait, we identified 585 apaQTLs at 10% FDR.

We then asked how many of the regulatory QTLs we identified for H3K27ac levels at TSS, transcription, steady-state mRNA, ribosome occupancy, and protein expression levels were APA QTLs using Storey's π_0 method, like we did for RNA decay QTLs. Unlike for RNA decay QTLs, we estimated that none of the QTLs for H3k27ac levels at gene TSS were apaQTLs and 7% of transcription (4sU) QTLs were apaQTLs. However, 19%, 42%, and 36% of the genes with a QTL for steady-state, ribosome occupancy, and protein expression levels that also had a predicted APA site were estimated to have a apaQTL (Table S5). Although our estimates have large confidence intervals, these estimates suggest a decoupling between transcription and alternative polyadenylation site usage, as expected. We further found that the top SNP for nearly a quarter of apaQTLs (142 of 585 or 24.2%) were located less than 2kb away from the variable polyadenylation sites (fig. S11). This observation suggests an important contribution of alternative polyadenylation site usage to post-transcriptional gene regulation, as previously proposed (*35*).

Joint model and estimation of QTL sharing

Previous work has estimated the number of shared QTLs across, for example, different tissues (*24*). Rather than hypothesis testing over the 2^P possible hypotheses for P phenotypes, we directly model effect sizes. We assume a prior for the true underlying effect sizes for SNP-gene pair a as follows,

$$\beta_a | z_a \sim z_a N(0, \Sigma) + (1 - z_a)\delta_0, \tag{3}$$

$$z_a \sim \text{Bernoulli}(\pi), \tag{4}$$

where

- β_a is a vector of effect sizes across P phenotypes
- Σ is the $P \times P$ covariance matrix of coefficients for non-null pairs a
- z_a is a mixture indicator: $z_a = 1$ denotes that SNP-gene pair a is regulatory, $z_a = 0$ that it is not.
- δ_0 is a delta-spike at zero in P-dimensional space
- π is the prior probability of a given SNP-gene pair a having non-zero coefficients

We take as the likelihood for the model the fitted estimates from standard linear modeling with associated standard errors, i.e.

$$\beta_{ap} | \hat{\beta}_{ap}, \hat{\sigma^2_{ap}} \sim N(\hat{\beta}_{ap}, \hat{\sigma}^2_{ap}) \tag{5}$$

where p indexes the phenotypes. We use Expectation-Maximization to integrate over β_a and z_a while optimizing with respect to Σ and π.

E-step. Calculate $q(\beta, z) \propto p(\mathcal{D}, \beta, z | \Sigma, \pi)$

$$q(z_a = 1) \propto P(\mathcal{D}, z_a = 1 | \Sigma, \pi) \tag{6}$$

$$q(\beta_a | z_a = 1) = N(B_a^{-1} b_a, B_a^{-1}) \tag{7}$$

where

$$B_a = \Sigma_a^{-1} + \text{diag}(1/\hat{\sigma_a^2}) \tag{8}$$

$$b_{ap} = \hat{\beta}_{ap}/\hat{\sigma}_{ap} \tag{9}$$

The probability required in Equation 6 can be calculated as

$$\log P(\mathcal{D}, z_a = 1|\Sigma, \pi) = \log P(z_a = 1|\pi) + \log \int P(\mathcal{D}|\beta) N(\beta|0, \Sigma) d\beta \tag{10}$$

$$= \log \pi + \log \frac{N(0, 0, \Sigma) \prod_p N(\hat{\beta}_{ap}|0, \hat{\sigma}_{ap}^2)}{N(0|B_a^{-1} b_a, B_a^{-1})} \tag{11}$$

$$\log P(\mathcal{D}, z_a = 0|\Sigma, \pi) = \log P(z_a = 0|\pi) + \log P(\mathcal{D}|\beta = 0) \tag{12}$$

$$= \log(1 - \pi) + \log \prod_p N(\hat{\beta}_{ap}|0, \hat{\sigma}_{ap}^2) \tag{13}$$

$$P(z_a = 1|\mathcal{D}, \Sigma, \pi) = \sigma \left(\log P(\mathcal{D}, z_a = 0|\Sigma, \pi) - \log P(\mathcal{D}, z_a = 0|\Sigma, \pi) \right) \tag{14}$$

$$= \sigma \left(\log \frac{\pi}{1 - \pi} + \log \frac{N(\beta|0, \Sigma)}{N(0|B_a^{-1} b_a, B_a^{-1}} \right) \tag{15}$$

where $\mathcal{D}$ are the data $(\hat{\beta}, \hat{\sigma})$ and $\sigma(x) = 1/(1 + e^{-x})$.

M-step. This involves calculating $\arg\max_{\pi,\Sigma} \mathbb{E}_q \log p(\mathcal{D}, \beta, c|\Sigma, \pi)$, which is achieved as

$$\Sigma := \sum_a q(z_a = 1) \mathbb{E}_{q(\beta_{a:}|z_a=1)}[\beta_{:d} \beta_{:a}^T] / \sum_a q(z_a = 1) \tag{16}$$

$$\pi := \sum_a q(z_a = 1)/A \tag{17}$$

Note 1: We do not explicitly account for linkage disequilibrium (LD), which will result in SNPs in LD with the causal SNP appearing to have significant effects. However, this will not bias estimates of the correlation of effect sizes since this correlation will be maintained at these linked SNPs.

Note 2: We additionally tried two alternative models. The first includes an additional diagonal mixture component which modesl SNP-gene pairs where the sharing across phenotypes does not match the global pattern,

$$\beta_{a:} \sim \pi_2 \text{MVN}(0, \Sigma) + \pi_1 \text{MVN}(0, \text{diag}(\tau)) + \pi_0 \delta_0. \tag{18}$$

We found the diagonal component was given extremely low weight, so the estimated level of global sharing was highly consistent with the original model.

The second allows an arbitrary number of general covariance multivariate Gaussian mixture components (we experimented with 3 and 10):

$$\beta_{a:} \sim \pi_0 \delta_0 + \sum_{c=1}^{C} \pi_c \text{MVN}(0, \Sigma_c) \tag{19}$$

The mixture components were used by the model to fit a heavier tailed distribution, but the estimates of the expected correlation of the effect sizes was again highly consistent with the original model.

Bayesian hierarchical model to identify causal eQTLs

To identify causal eQTLs, we adapted the hierarchical generative model for genetic association with gene expression (*36*) to the setting where only summary statistics of genetic associations with gene expression are available. Under this model, each gene has a latent indicator variable $Q_g \in \{0, 1\}$ that specifies whether the gene has 0 or 1 causal SNP associated with variation in its expression. This causal SNP is hereafter referred to as an expression quantitative trait nucleotide or eQTN. The prior probability for a gene to have an eQTN, $\Pr(Q_g = 1) = \pi$, is assumed to be same for all genes.

For each SNP s tested against a gene g, a latent indicator variable $C_{sg} \in \{0, 1\}$ specifies whether the SNP s is the eQTN for that gene. Conditional on the gene having an eQTN, the prior probability that a particular SNP s is the eQTN depends on whether it falls within a particular genomic annotation (e.g. chromatin mark peaks) and/or sequence-level annotations

$$\Pr(C_{sg} = 1 | Q_g = 1) = \gamma_{sg} \tag{20}$$

where $\gamma_{sg} \propto \exp(\sum_k \alpha_k A_{sk})$. The value of the k^{th} annotation for the s^{th} SNP is encoded as A_{sk}; for our purposes, A_{sk} is a binary-valued variable (for example, is a SNP in an open chromatin region? Is a SNP nonsynonymous or synonymous?). The weight α_k denotes the importance of the k^{th} annotation in specifying the prior probability that a SNP is an eQTN. We impose the constraint $\sum_s \gamma_{sg} = 1$ to ensure that a gene can have at most one eQTN.

The true effect of an eQTN is modeled by a normal distribution centered at zero with variance σ^2 while the true

effect of a SNP that is not an eQTN is fixed exactly to zero:

$$\Pr(\beta_{sg}|C_{sg}) = (1 - C_{sg}) \cdot 0 + C_{sg} \cdot \mathcal{N}(0, \sigma). \tag{21}$$

Given the true effects, the estimated effects are then modeled as $\Pr(\hat{\beta_{sg}}|\beta_{sg}, \hat{\sigma_{sg}}) = \mathcal{N}(\beta_{sg}, \hat{\sigma_{sg}})$, where $\hat{\sigma_{sg}}$ is the standard error of the estimated effect. Integrating out the true effects, we get

$$\Pr(\hat{\beta_{sg}}|C_{sg}) = (1 - C_{sg}) \cdot \mathcal{N}(0, \hat{\sigma_{sg}}) + C_{sg} \cdot \mathcal{N}\left(0, \sqrt[2]{\sigma^2 + \hat{\sigma_{sg}}^2}\right). \tag{22}$$

Using the summary statistics $(\hat{\beta_{sg}}, \hat{\sigma_{sg}})$ from testing each gene with each SNP in a 100kb window around that gene and the annotation values for these SNPs A_{sk}, we estimate the parameters α_k that maximize the overall likelihood of the summary statistics by sharing information across all tested genes. The summary statistics for each gene-SNP pair contribute to the likelihood via a Bayes factor that measures the relative support for the alternative hypothesis (the SNP is an eQTN) compared against the null (the SNP is not associated with gene expression). This model also allows us to compute the posterior probability that a SNP is the eQTN for a particular gene. The software implementing this model can be found at `https://github.com/rajanil/qtlBHM`.

Identification of splicing events and splicing QTLs

We developed a method that clusters introns defined by RNA-seq reads to identify differential splicing events that are variable within and across individuals. In particular, we leveraged the large fraction of RNA-seq reads (86 YRI GEUVADIS samples) across samples which span exon-exon junctions. This allows the identification of splicing events without relying on pre-existing annotations which are typically incomplete, especially in the setting of large genes or individual/population-specific isoforms.

More precisely, we defined "clusters" of introns that represent alternative splicing choices. To do this, we first grouped together overlapping introns (defined by spliced reads). For each of these groups we constructed a graph where nodes are introns and edges represent overlapping introns. The connected components of this graph define our clusters. Singleton nodes (introns) are discarded. For each intron cluster, we next iteratively (1) removed introns that were supported with fewer than 100 reads or fewer than 5% of the total number of intronic read counts for the entire cluster, and (2) re-clustered introns according to the procedure above. As a further filtering step, we

used gene annotations and removed all clusters with intron from 2 or more genes. Of 14,791 clusters, this step removed 231 and the rest were used for subsequent analyses. An implementation of this approach is available at `https://github.com/davidaknowles/leafcutter` and described in (*20*).

To identify genetic variation that affects splicing (sQTLs), we used as quantitative trait the proportion of intron-defining reads to the total number of reads from the intron cluster it belongs to (fig. S7A). As such, this intron ratio describes how often an intron is used relative to other introns in the same cluster. Finally, we tested the association between intron ratio usage and SNPs (with MAF > 0.05) that were within 100kb of the intron cluster.

Estimation of gene-wide and cluster-wide eQTL and sQTL p-values and false discovery rates

To estimate the FDR for eQTLs and sQTLs identified from 86 YRI LCL RNA-seq samples from GEUVADIS, we ran `Matrix eQTL` (*37*) 10,000 times, each time with permuted genotype sample label. We then computed the empirical gene-level p-value for the most significant QTL for each gene as the fraction of permutation p-values at least as significant as the best p-value from non-permutated genotypes, i.e. $p = \frac{\sum_{i=1}^{10^4} \mathbb{1}_{p^i_{\text{perm}} \leq p_{\text{real}}}}{10^4}$, where p^i_{perm} is the p-value obtained from the i^{th} permutation. Using this procedure, we estimated that at 10% and 5% FDR, we identified 2,015 and 1,593 eQTLs, respectively and 2,893 and 1,602 sQTLs, respectively.

Comparisons between eQTLs and sQTLs

Location of eQTLs and sQTLs relative to genes **(Fig. 3A–B)**: For both eQTLs and sQTLs at 5% FDR, we identified their relative location to the gene by dividing each gene into 10 bins of equal sizes and assigned each QTL into its corresponding distance bin if they lie within the gene. When a QTL lies outside the gene, we identify the bin (again of equal size) in which the QTL falls away from the transcription start site or transcription end site.

Distance between the best eQTL and sQTL of a gene **(Fig. 3C)**: For genes whose expression and intron splicing were controlled by an eQTL and sQTL at 10% FDR respectively, we computed the distance between the best eQTL and the best sQTL (according to p-values). We required that the best eQTL (sQTL) must be at least twice (2 times) more significant than the second best eQTL (sQTL). As such, 275 genes satisfied this criterion. We also applied this same approach to compare our eQTL identified from RNA-seq (GEUVADIS) with eQTLs identified from RNA-seq (Pickrell (*17*)) (fig. S12).

Hierarchical model to identify annotations most relevant to gene expression and splicing regulation **(Fig. 3D)**: We obtained LCL-specific chromHMM annotations and gene-level annotations (e.g. intronic variant, splice region variants, UTR variants, etc...) from snpEff (*38*) v4.1. Splice region variants include:

- A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron.
- A variant affecting putative (Lariat) branch point, located in the intron.
- A variant affecting putative (Lariat) branch point from U12 splicing machinery, located in the intron.

We next annotated all tested SNPs according to their overlapping annotations and ran our hierarchical model to identify the annotations that were most likely to harbor a sQTLs and eQTLs. We note that the number of SNPs in each chromHMM annotation were substantially smaller than the number of SNPs in gene-base annotations, which explains the differences in 95% confidence intervals.

Chromatin QTLs that are more likely to be sQTLs **(Fig. 3E)**: Variation in CTCF binding and H3K27ac levels have been shown to correlate with differences in splicing across tissues or experimental conditions (*21, 39*), and splicing QTLs are enriched in regulatory regions including CTCF binding sites (*22*). To identify direct links between chromatin and splicing, we compared the distribution of sQTL p-values of SNPs that were classified as chromatin QTLs (H3K27ac QTLs, CTCF QTLs (*40*), DNA methylation QTLs, and DNaseI QTLs) to those that merely fall within LCLs H3K27ac peaks ($\pm$ 1kb), CTCF peaks ($\pm$1kb), and $\pm$1kb surrounding a DNA methylation probe (450K Illumina Methylation array), and DNaseI peaks, respectively. We find that only H3K27ac and CTCF QTLs were more likely to be sQTLs than SNPs that fall within 1kb of their respective peaks.

To do this, we first identified SNPs that fell into H3K27ac/CTCF peaks ($\pm$ 1kb) and SNPs that were classified as H3K27ac/CTCF QTLs at 10% FDR. To avoid underestimating enrichment p-values due to linkage disequilibrium, we binned each set of SNPs into genomic blocks of 10kb, and then took for each bin a SNP at random. We next asked how many bins had a SNP with an association to splicing below a p-value threshold p = 8.41×10^{-4}, which corresponds to a 10% FDR for the entire set of SNPs tested. We find that $\sim$619 of 11,577 ($\sim$5.3%) haQTLs had an association with splicing, while $\sim$1,105 of 38,642 SNPs (2.9%) SNPs within 1kb of a H3K27ac peak had an association with splicing (χ^2 p=1.1×10^{-34}). For CTCF-QTLs, 30 of 488 were splicing QTLs and 468 of 17,373 CTCF SNPs were sQTLs (χ^2 p=2.2×10^{-5}). These enrichments were robust to different choices of FDR and window sizes.

To identify sQTLs that are associated with chromatin-level phenotypes, we searched for associations between sQTLs and chromatin-level phenotypes that were measured at most 50kb away from each SNP. A p-value was reported if the SNP was a chromatin QTL at 10% FDR. Note here that no Bonferroni correction was applied because QTLs tend to be associated with multiple nearby peaks that show similar variation in terms of genotype. The list of 171 chrom-sQTLs (Table S7) is therefore expected to contain false positives and we advise only to take sQTLs with strong association to chromatin-level phenotypes for follow-up (e.g. $p < 10^{-5}$).

COMT CTCF-QTL **(Fig. 3F)**: We downloaded CTCF-QTL data described from (*40*) and found that a SNP in the *COMT* loci was a CTCF-QTL and affected splicing. Upon further investigation, we found that this SNP was a QTL for DNaseI sensitivity and DNA methylation.

Estimating the relative contribution of haQTLs, eQTLs and sQTLs on complex traits

Genome-wide association studies. We obtained the summary statistics for 6 GWAS: rheumatoid arthritis (*41*), multiple sclerosis (*42*), Alzheimer's disease (*43*), schizophrenia (*44*), height (*45*), and body mass index (*46*). Specifically, we obtained and compiled for each genotyped and imputed SNP, an effect size, standard error and minor allele frequency which were used as input to `fgwas` and `polyTest`.

Annotations. To annotate SNPs, we used snpEff (*38*) to annotate SNPs that lie in genic annotations (5' UTRs, 3' UTRs, introns, synonymous sites, non-synonymous sites, etc...), in conserved regions (phastCons scores > 0.4), chromatin peaks (ENCODE LCL ChIP-seq peaks), and chromHMM annotations. We then encoded whether a SNPs is a QTL for a regulatory phenotype or not as a binary using a p-value cutoff of 10^{-4}, but our observations are consistent with different choices of cutoffs (fig. S16). To further distinguish the effects of genetic variants on complex traits through gene expression level regulation versus splicing regulation **(Fig. 4C)**, we removed any sQTL that had an eQTL association that was more than four order of magnitudes less significant than sQTL association (our observations are also robust to this parameter). We further obtained eQTLs identified from purified Monocytes and T-cells (*47*) and from B-cells (*48*) (fig. S17).

fgwas. **(Fig. 4A)** We ran `fgwas` with default parameters on each annotation separately. Jointly inputing multiple annotations yielded similar results (not shown). We used polyTest (see section entitled "Overview of polyTest") to estimate the relative contribution when accounting for multiple annotations jointly **(Fig. 4B)** (fig. S16). To verify

that our choice of QTL cutoffs (i.e. what we consider a QTL for a particular regulatory phenotype) does not affect our interpretations, we ran fgwas with different QTL association p-value cutoffs (p-values $< 10^{-3}$, $< 10^{-4}$, $< 10^{-5}$, $< 10^{-6}$, $< 10^{-8}$) (fig. S16) representing a wide number of effect sizes and gene numbers with a QTL (e.g. the number of eQTLs ranged from $\sim$500 to over $4,000$). We observed that our estimated effect sizes did not vary significantly and our interpretation remains the same with different choices of cutoffs (fig. S18).

Comparison to the results from Xiong et al. 2015

Xiong et al. (*15*) compared the regulatory scores (predicted using their deep learning method) of 457 SNPs implicated in GWAS to the regulatory scores of a matched set of non-GWAS SNPs; they found no difference in regulatory score distribution between the two sets. Our analyses suggest instead that genetic variation that affects splicing do indeed contribute to complex traits.

There are several differences between their analysis and ours. For example, Xiong et al. use predicted effects of SNPs on splicing to search for GWAS signal while we use direct evidence that a SNP is associated with splicing variation. As such, Xiong et al.'s approach focused on SNPs that are predicted to disrupt splicing with $|\Delta\text{PSI}| \geq 5\%$. Our approach is able to detect variation in splicing that have much smaller effects. Therefore we believe that our approach is more sensitive to detecting association between genetic variants and complex diseases because genetic variants with small effects (even at the molecular level) may contribute substantially to complex traits.

The most critical difference between our two studies is the use of SNPs pooled from hundreds of heterogeneous GWAS to form Xiong et al's set of GWAS SNPs. This is problematic because complex traits can have different etiologies and affect different tissues/cell-types. We considered all SNPs of every GWAS independently and tested statistical enrichment of sQTLs among SNPs with large effect sizes using two polygenic models (fgwas and polyTest). This has several advantages. Firstly, we were able to determine whether sQTLs in a precise cell-type were likely to modulate the risk of a relevant disease. For example, we find that LCL sQTLs were enriched in GWAS signal among immune-related diseases. Secondly, our polygenic model allows us to use the thousands of putative trait-affecting SNPs (some of which have small effects) to support our inference. In contrast only the top 457 SNPs were used in Xiong et al's analysis.

Overview of polyTest

PolyTest is a novel method for joint analysis of entire GWAS and functional annotations. We assume that the effect of the ith SNP, denoted β_i, is drawn from a normal distribution with mean 0, whose variance is determined by the functional annotations associated with the SNP:

$$\hat{\beta}_i|\beta_i \sim N\big(0, \sigma^2(\vec{x}_i)\big),$$

where $\vec{x}_i$ is the vector of annotations associated with SNP i, and $\sigma^2(.)$ is some function mapping the vector of annotations to a positive value. Following the common practice in generalized linear models (*49*), we use $\sigma^2(\vec{x}_i) = e^{\vec{x_i}\bullet\vec{\gamma}}$, where $\vec{\gamma}$ is a vector of coefficients, and $\bullet$ denotes the dot product. In other words, $\gamma_j = 0$ implies that the jth annotation has no effect of the typical effect size, while a non-zero value, implies that the variance is affected by the jth annotation.

Under the assumption of linkage equilibrium, the estimated effect of each SNP is an unbiased estimate of the true effect size, with a standard error that is denoted $\hat{\sigma}_i$ and depends on many factors (such as sample size, allele frequency, imputation quality, etc...) In other words:

$$\hat{\beta}_i \mid \beta_i \sim N(\beta_i, \hat{\sigma}_i^2). \tag{23}$$

Integrating out the true effects, we get:

$$\hat{\beta}_i \sim N\big(0, \hat{\sigma}_i^2 + \sigma^2(\vec{x}_i)\big) = N\big(0, \hat{\sigma}_i^2 + e^{\vec{x}_i\bullet\vec{\gamma}}\big). \tag{24}$$

To incorporate linkage disquilibrium (LD), we follow (*50*) who note that $E[Z_\beta] = RZ_\beta$ where Z_β is the vector of (expected) Z-scores and R is the matrix of SNP by SNP correlations. We can then undo the effect of LD by multiplying both sizes by R^{-1} resulting in unbiased Z-scores, and transforming those to unbiased estimates of effect size (the unbiased estimate, however, typically have a much larger standard error). We use publicly available data to estimate R, and apply the shrinkage method of (*51*) to account for the fact that R is estimated rather than known.

Once the mean and marginal variance of each effect estimate is specified as a function of $\vec{\gamma}$, we maximize the composite log-likelihood function (the product of all the marginal log-likelihoods) using Newton's second-order method. Since classical inference is invalid for maximum composite likelihood estimates, we re-run the estimation for each chromosome separately, resulting in 22 independent samples from each estimate's distribution, which are then weighted by chromosome size to provide an estimate of the standard error of the estimate and compute a p-value.

Comparison to fgwas

Another popular method used in this study is fgwas (*23*). Several major differences exist between fgwas and polyTest: First, fgwas explicitly assumes a sparse genetic architecture of disease, where each segment (500Kb to 1Mb long) harbors at most a single causal SNP, while polyTest assumes a highly polygenic model of disease. Second, the sparsity assumption of fgwas is used as a justification for not modelling the tagging due to LD. However, if two annotations are often close together (e.g. exons and introns, 5'UTR and promoters etc.), both might appear to be significant even if only one of them is indeed significant when LD is not accounted for. Lastly, polyTest adds $\log(f(1-f))$ as an addtional annotation, thus allowing any dependence of the form $Var(\beta) \propto [f(1-f)]^{\alpha}$ between the variance of effect sizes and allele frequency f (as is often observed), while fgwas assumes a constant variance of effect sizes.

Raw and processed data availability

The newly generated 4sU-seq (30m and 60m) data are available on GEO (accession ID GSE75220). Other data used can be found in Table S8, at `http://mitra.stanford.edu/kundaje/portal/chromovar3d/qtls.html` (Histone QTLs) and `http://www.ebi.ac.uk/birney-srv/CTCF-QTL/` (CTCF-QTLs).

Supplementary Figures

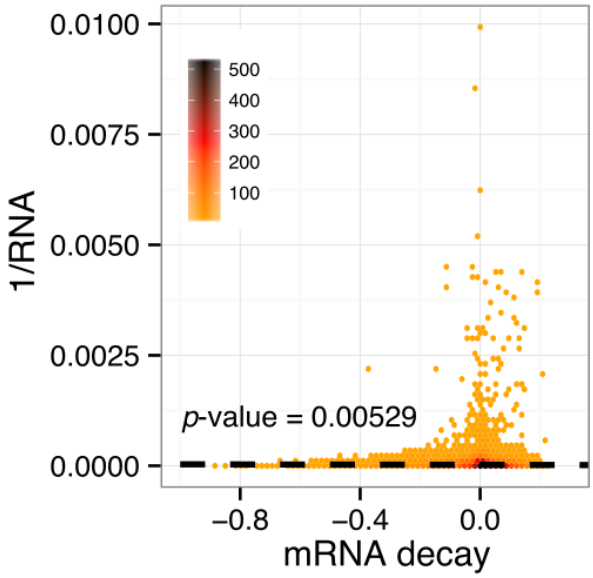

Figure S1: Weak correlation between mRNA expression level and RNA decay.

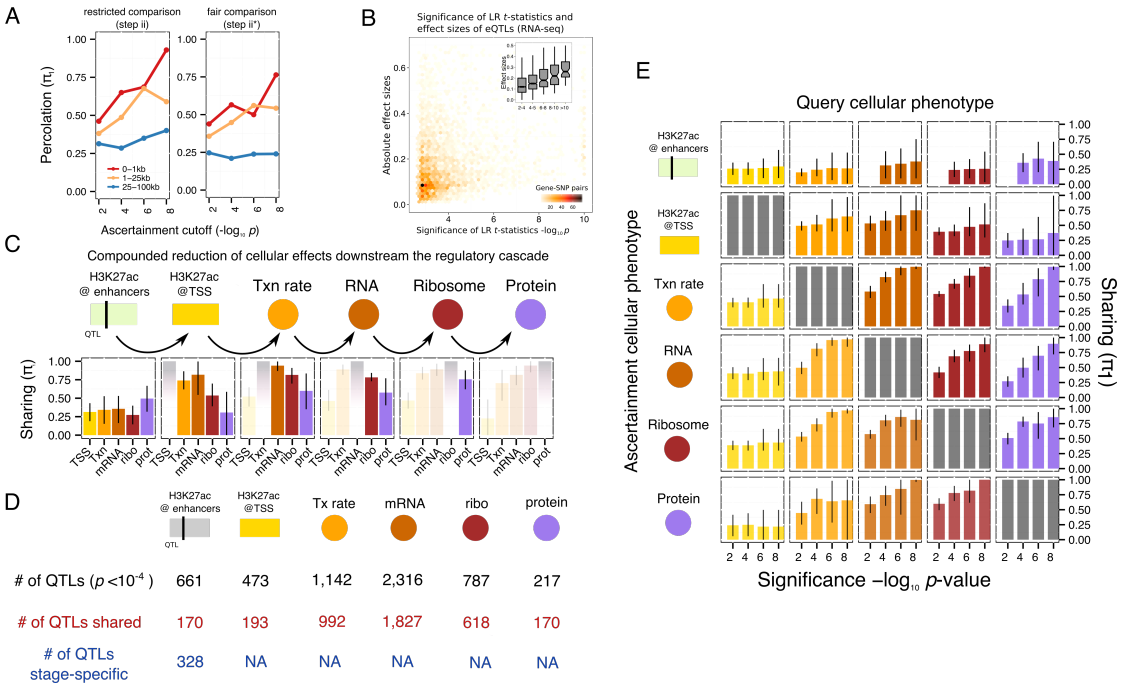

Figure S2: Sharing of QTLs across the regulatory cascade. A) Left panel: estimates of sharing between H3K27ac QTLs and eQTLs (total gene expression) for genes whose TSS are less than 1kb, 25kb, and 100kb away from peaks. For each of these categories (1kb, 25kb, 100kb), only genes within 1kb, 25kb, and 100kb away, respectively were considered (restricted comparison). Right panel: the same estimates for the percolation of effects when considering all genes that are 500kb for every peak (fair comparison). B) eQTLs with strong associations have larger effect sizes (linear regression slopes) on average than eQTLs with weaker associations. C) Compounded reduction of QTL sharing across the regulatory cascade suggest a small amount of buffering between transcription, translation, and protein. D) Estimates of QTL numbers (see Table S3) and sharing across gene phenotypes. The number of stage-specific QTLs were estimated according to the section entitled "Estimation of QTL sharing between H3K27ac levels at enhancers, H3K27ac levels at promoters, and gene-level phenotypes" of the Supplementary Materials. E) Estimates of QTL sharing for all regulatory stage pairs, using ascertainment from one to the other stage. Bars represent 80% confidence intervals.

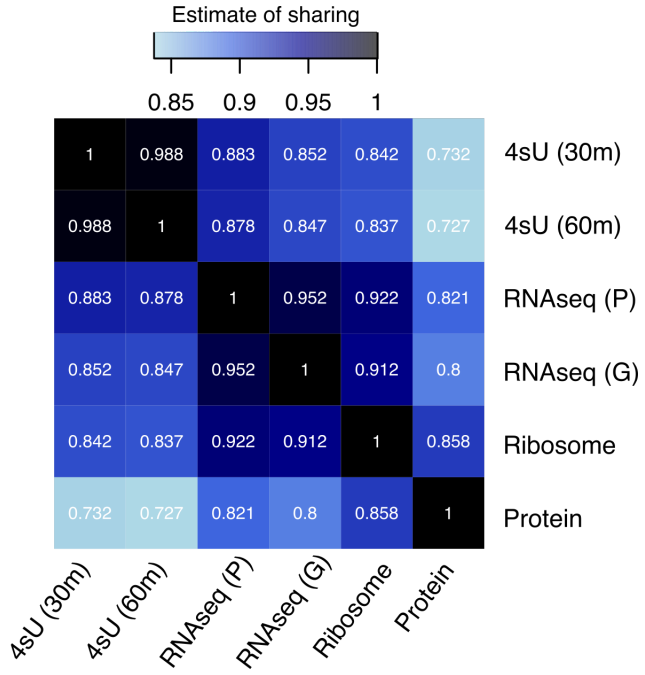

Figure S3: Estimates of sharing across regulatory QTLs using a joint modeling approach (see section entitled "Joint model and estimation of QTL sharing" for full model.

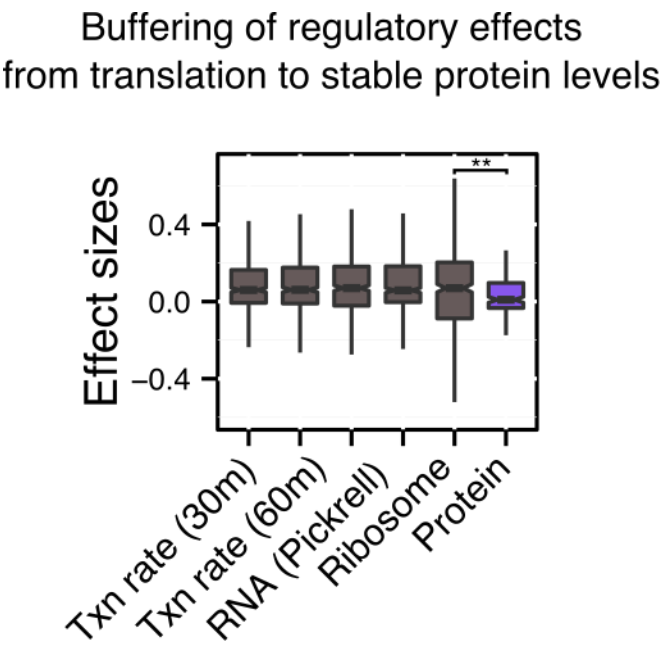

Figure S4: Effect sizes are similar from transcription to translation rates but appear to be partially buffered at the protein expression level. Ascertainment was made from 565 H3K4me3 QTLs recently identified in (*8*) at 5% FDR. Protein QTL effect sizes are on average smaller than those of transcription rates, RNA, and ribosome QTLs (Mann-Whitney U, $p = 0.002$)

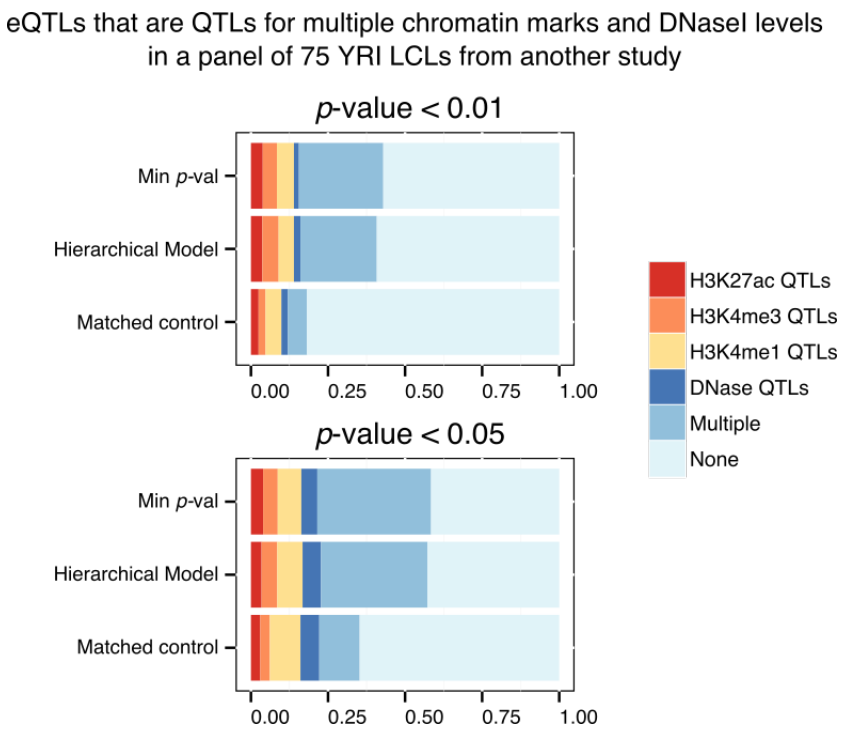

Figure S5: Estimates of eQTLs that are also chromatin-level QTLs using QTL data from (*8*) alone. The estimated proportion of eQTLs that also affect chromatin-level traits using chrom-QTLs from (*8*) is lower than the one estimated using our chromatin QTLs (for both $p < 0.01$ and $p < 0.05$), and the proportion of control eQTLs that are chromatin QTLs is higher (for $p < 0.05$). This suggests that additional chromatin marks representing gene body (H3K4me3) and enhancers (H3K4me1) do not account for eQTLs that were found to be unexplained in our study.

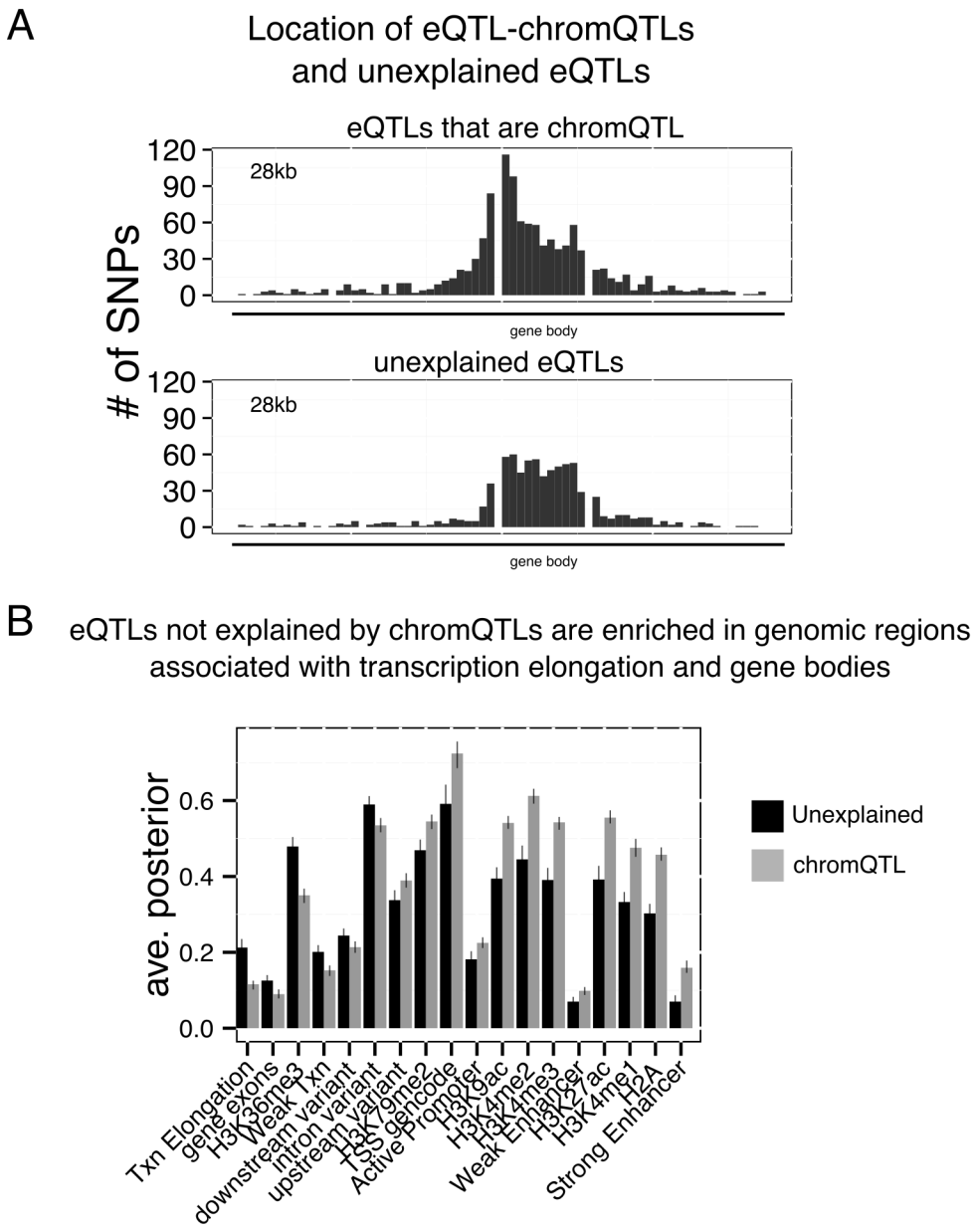

Figure S6: A) eQTLs that are not chromatin-level QTLs are enriched within the gene body, but not near TSS. B) Enrichment of eQTLs that are not chromatin-level QTLs in regions associated to transcription elongation, weak transcription and H3K36me3 compared to eQTLs that also appear to affect chromatin-traits. Bars represent 95% confidence intervals obtained from bootstraps.

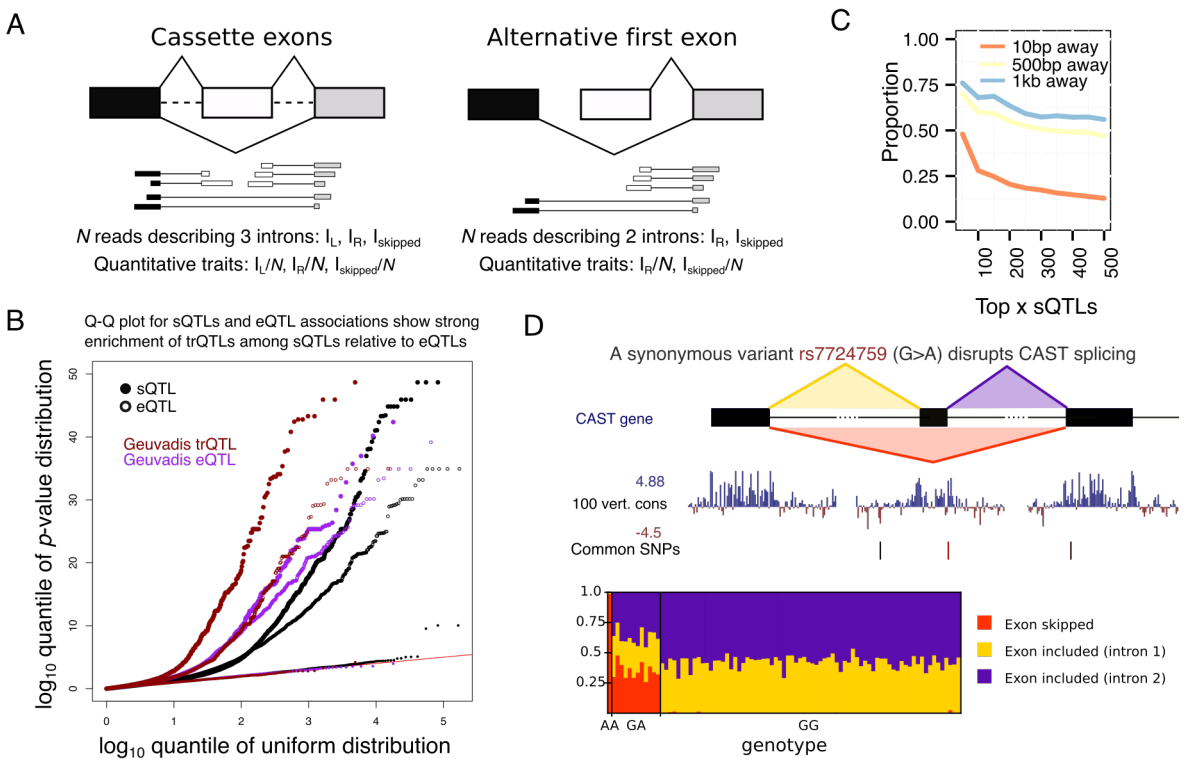

Figure S7: Identifying splicing quantitative trait loci. A) To identify genetic variants that affect splicing (sQTLs), we quantified the ratio of intron usage within intron clusters. Each intron is thus assigned a ratio between 0 and 1, which is independent from total gene expression levels. B) sQTLs are highly enriched for transcript ratio QTLs (trQTLs) previously identified in (*6*), which indicates that our approaches, although very different, can detect the same splicing signals. C) Of the 100 most significant associations, we find that 28 lie in close proximity (at most 10bp away) to the splice sites of the exons they affect and 60 within 500bp. This suggests that genetic variants often affect transcript abundance independently from transcription rate. D) Example of a synonymous variant located 1 bp away from the 5' splice site of an intron in *CAST* which disrupts splicing of an internal intron. This example was previously identified and experimentally validated (*52*).

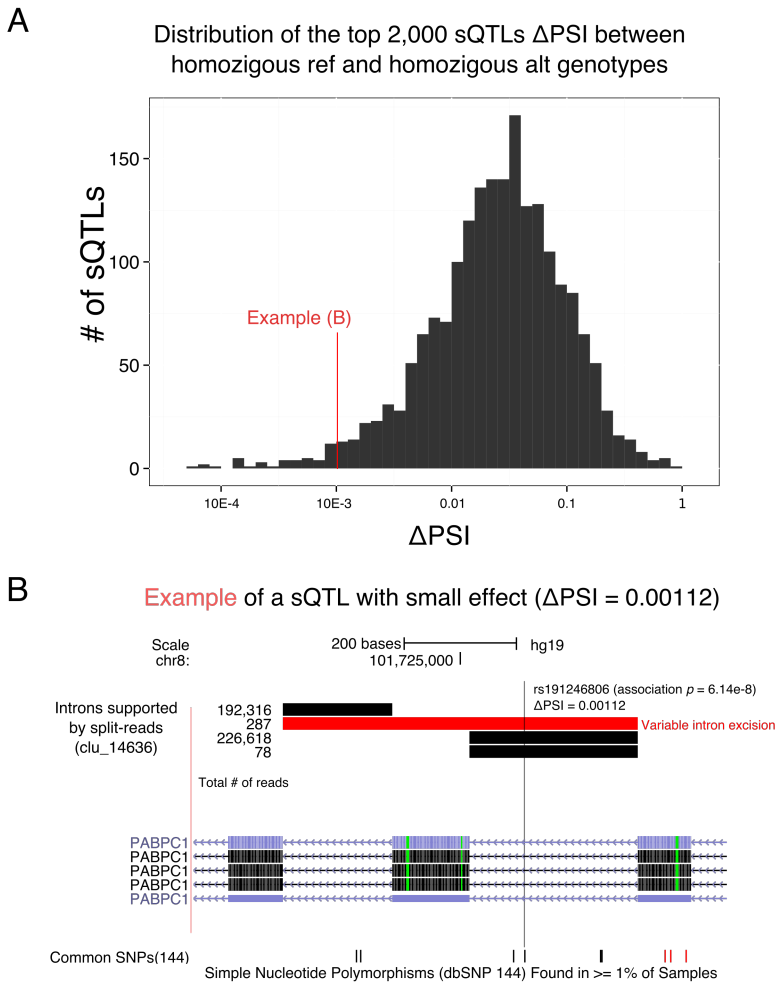

Figure S8: A) Distribution of $|\Delta\text{PSI}|$ for the top 2,000 sQTLs detected in this study. ΔPSI was computed as twice the coefficient (2β) of the linear fit ($Y = \beta X + \epsilon$) between genotypes (coded as $X = 0, 1$, and 2) and raw intron excision levels (Y, no transformation or covariates). B) Example of a sQTL with a small (0.00112) ΔPSI. A genetic variant (rs191246806) controls the excision level of an intronic stretch that contains an otherwise constitutively expressed exon.

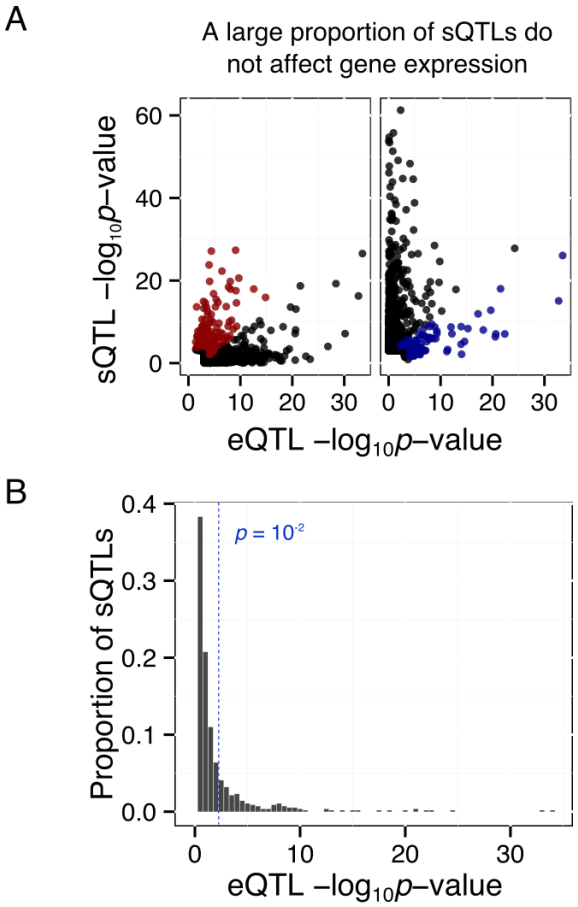

Figure S9: A) Left panel: each point represents a SNP with the most significant association to the expression levels of a gene. A large minority of genes have eQTLs that affect splicing and may be driven by variants that affect splicing (red points). Right panel: each point represents a SNP with the most significant association to the splicing ratio of an intron ration. A large majority of genes have sQTLs that do not appear to affect gene expression. Some variants may affect splicing by altering transcription initiation (blue points). B) Histogram of the strength of association between sQTLs and total gene expression: the top sQTL for >74% intron clusters do not appear to affect the expression levels of corresponding gene ($p > 0.01$).

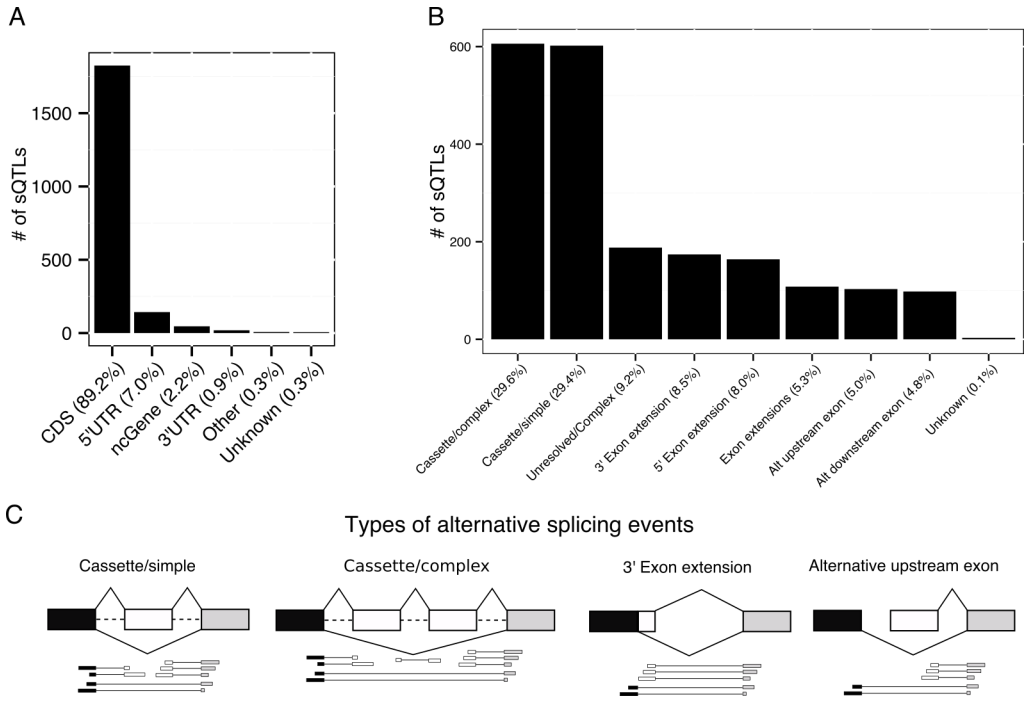

Figure S10: Effects of sQTLs on mRNA isoforms. A) Nearly 90% of sQTLs affect an intron that flanks a coding gene. B) Using gene annotation, we were able to classify about 60% of events as the skipping of one exon (cassette/simple) or more (cassette/complex), ∼20% alternative acceptor or donor sites, and ∼10% alternative 5' or 3' exon. The rest were events that are more complex. C) Examples of how reads spanning intron-exon boundaries can be used to determine alternative splicing type.

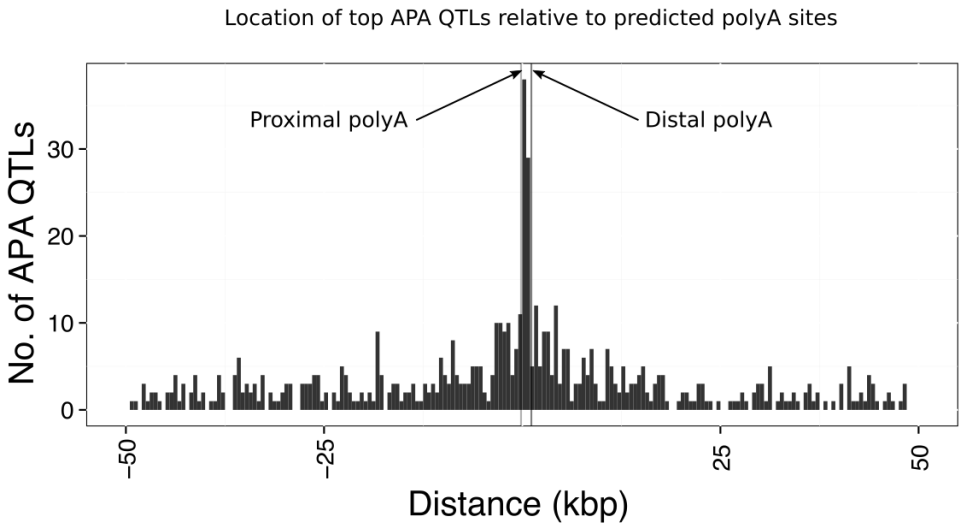

Figure S11: Position distribution of 585 APA QTLs identified at 10% FDR.

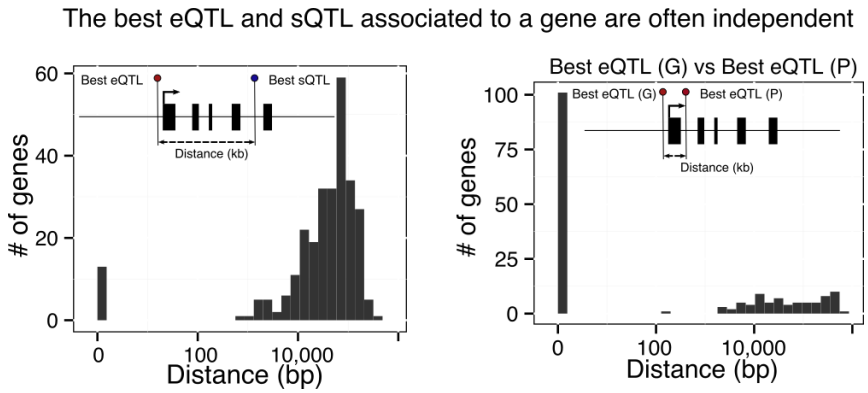

Figure S12: Left: Genes with top eQTL and sQTL with p-values at least 2 times smaller compared to the next best associations to total expression and intron ratios, respectively. Right: Genes with top eQTL ascertained in GEUVADIS RNA-seq data and eQTL ascertained in Pickrell RNA-seq data also with p-values at least 2 times smaller compared to the next best associations.

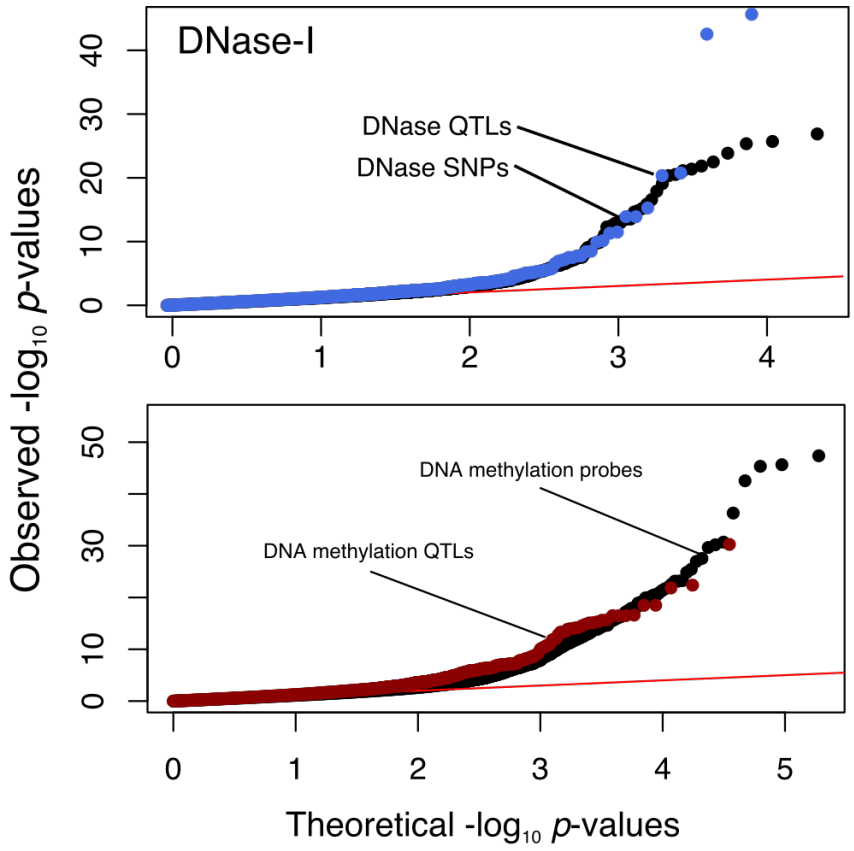

Figure S13: QTLs for DNaseI and DNA methylation levels are not more likely to be sQTLs than SNPs that fall in DNaseI peaks or $\pm 1kb$ away from DNA methylation probes, respectively.

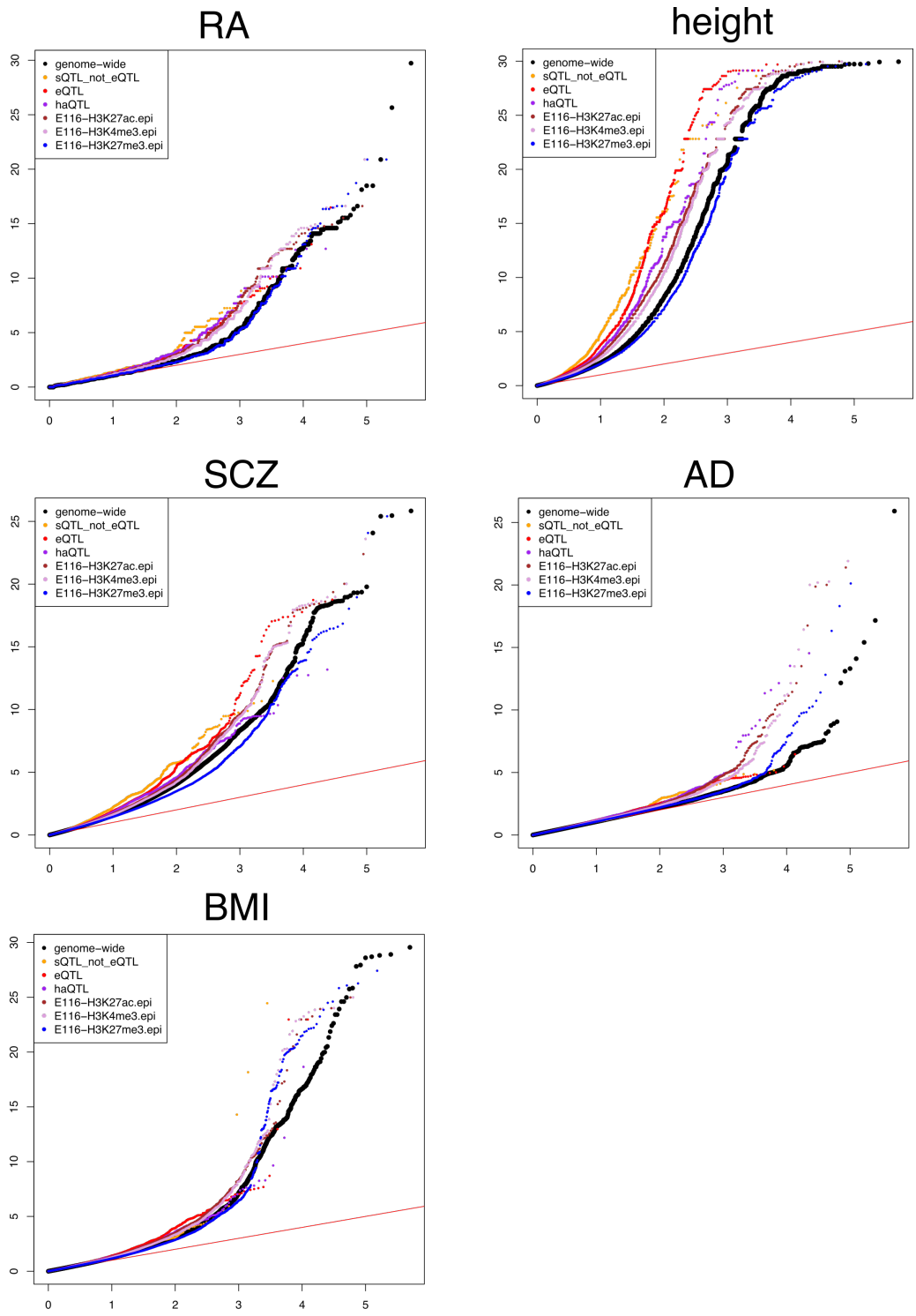

Figure S14: Q-Q plot for rheumatoid arthritis (RA), height, schizophrenia (SCZ), Alzheimer's disease (AD), and body mass index (BMI). All hits in the extended MHC region (26Mb–34Mb on hg19 chromosome 16) and hits with association p-value $< 10^{-30}$ were removed.

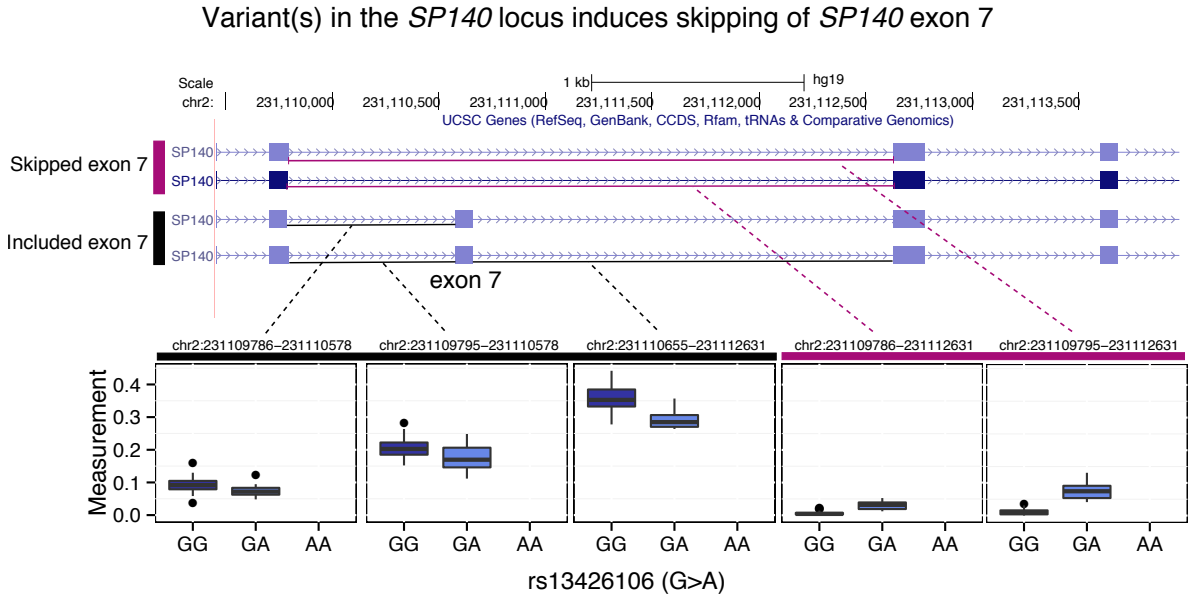

Figure S15: A sQTL in *SP140* is associated with an exon skipping event linked to multiple sclerosis (*53*). Our intron-based analysis of splicing revealed an association between genotype and a complex splicing event. The splicing of two different mRNA isoforms are affected. These two isoforms differ by 9bp at the 3' of the first depicted exon. The alternate allele, A, promotes skipping of exon 7 in both isoforms.

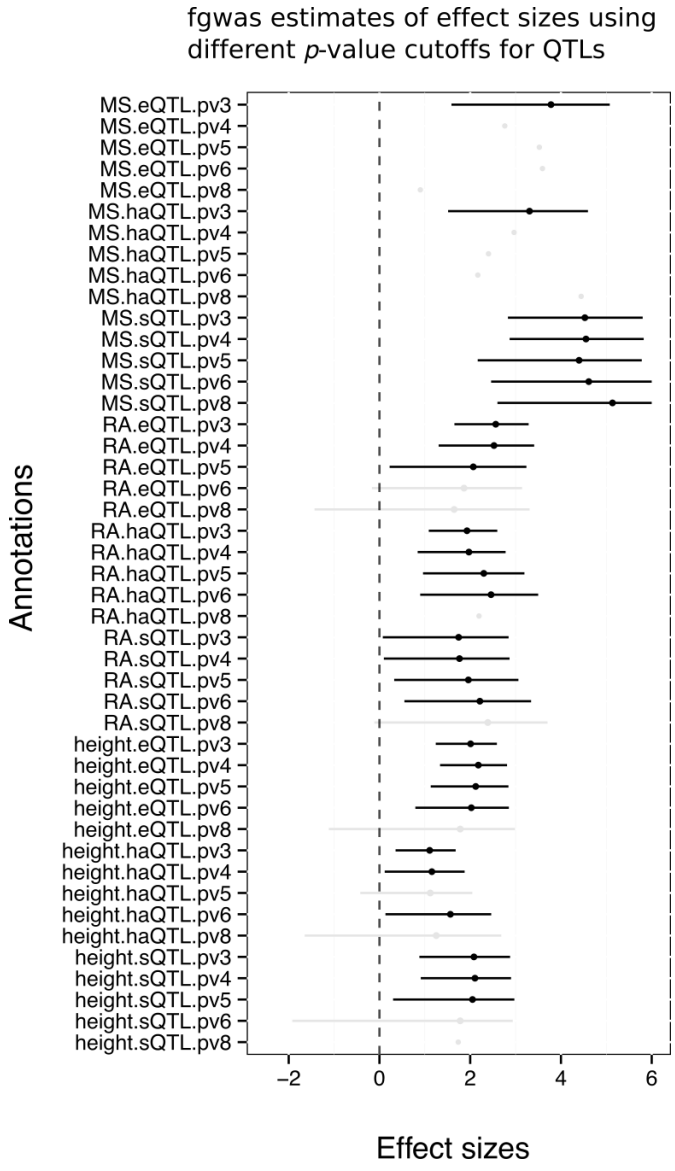

Figure S16: Robustness of QTL cutoffs in fgwas analysis. Various cutoffs on the minimum p-value of the association between genotype to molecular trait yield similar estimates. The 95% confidence interval overlap 0 for grayed out annotations.

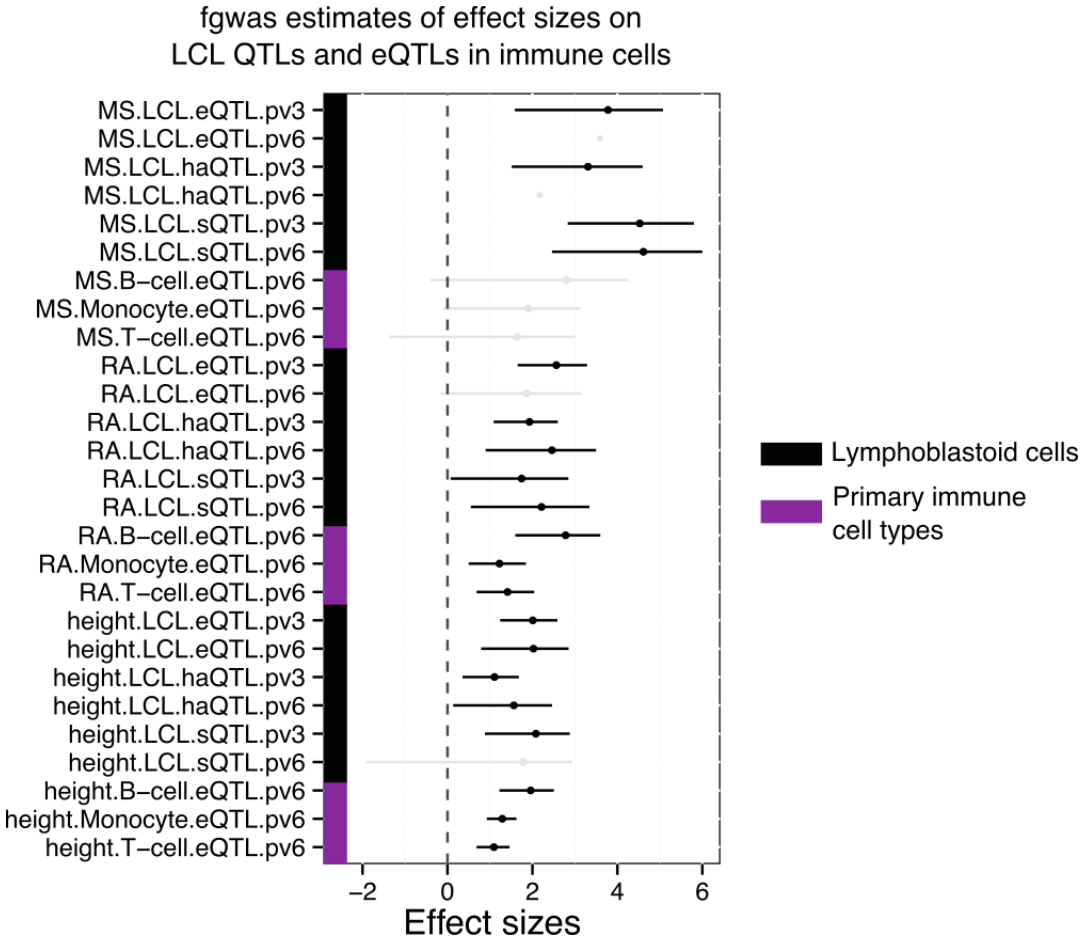

Figure S17: Comparison of fgwas estimated effect sizes between QTLs in LCLs and eQTLs from purified immune cells (T-cells and monocytes eQTLs were obtained from (*47*) and B-cells eQTLs from (*48*)).

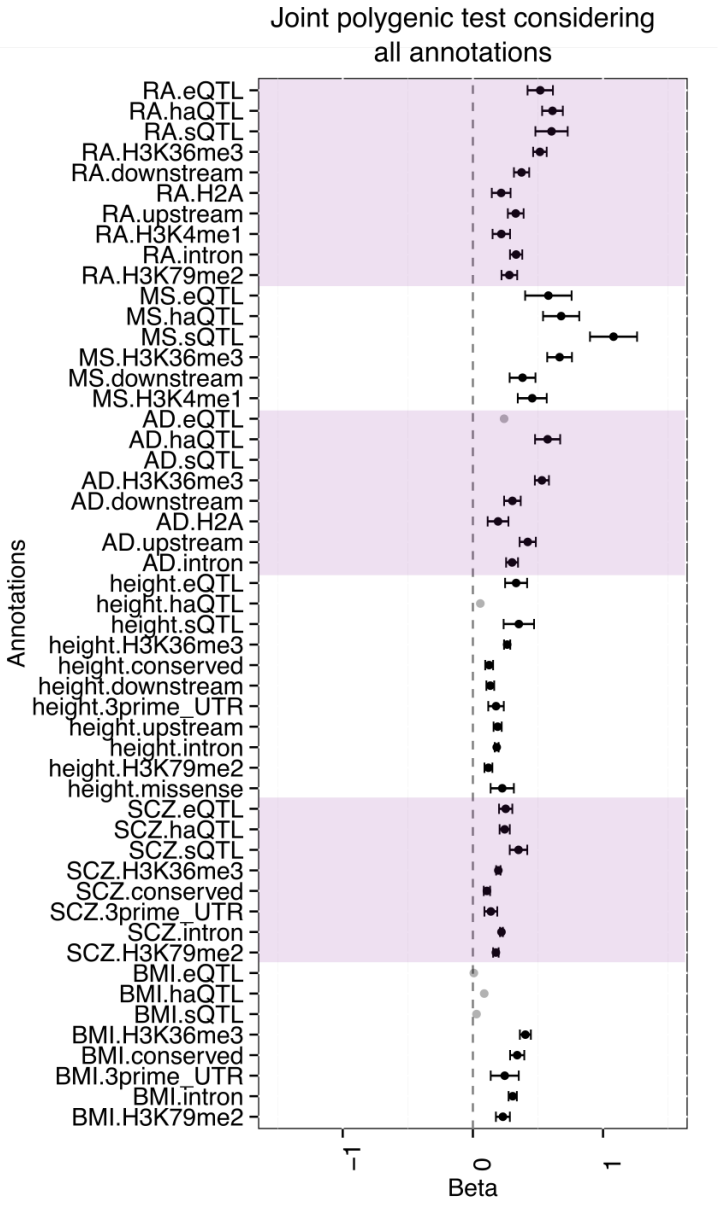

Figure S18: Joint analysis of all annotations including the QTL annotations using polyTest. All QTL estimates are shown (sQTL for AD not significant), other annotations are shown only when significant.

Supplementary Tables

data type	measurement	sample size	source	QTL mapping pipeline
H3K27ac	chromatin modification	59	del Rosario et al., 2015	WASP+LM
DNaseI	open chromatin	70	Degner et al., 2012	Degner+liftOver+LM
Methylation	methylation levels	64	Banovich et al., 2014	Banovich+liftOver+LM
4su (30min)	transcription rate	65	This study (GSE75220)	WASP+LM
4su (60min)	transcription rate	64	This study (GSE75220)	WASP+LM
RNA-seq (P)	steady-state mRNA levels	69	Pickrell et al., 2010	WASP+LM
RNA-seq (G)	steady-state mRNA levels	86	Lappalainen et al., 2013	WASP+LM
RNA decay	RNA decay rates	68	Pai et al., 2012	Pai+LM
riboprofiling	ribosome occupancy	70	Battle et al., 2015	WASP+LM
protein level	steady-state protein levels	64	Battle et al., 2015	Battle+LM

Table S1: Table of all datasets processed in this study. WASP (*4*) was used to account for read depth heterogeneity and GC-content biases across libraries. LM: linear model.

Data type	No. of PCs regressed
H3K27ac	6
4sU (30m)	13
4sU (60m)	11
RNA-seq (Pickrell (*17*))	14
RNA-seq (GEUVADIS)	15
ribo-seq	9
Intronic splicing ratios (GEUVADIS)	3

Table S2: Number of PCs that maximizes the number of QTLs for each data type.

Data type	T = 2	4	6	8
H3K27ac (enhancer)	2,149	661	291	133
H3K27ac (TSS)	1,249	473	131	47
4sU (30m)	3,632	1,142	299	110
4sU (60m)	3,512	1,116	277	119
RNA-seq (Pickrell (*17*))	4,496	1,324	394	180
RNA-seq (GEUVADIS)	6,954	2,316	983	508
ribo-seq	2,117	787	208	91
protein	586	217	68	21

Table S3: Number of QTLs with association $-\log_{10}$p-values $> T$ that were used to estimate sharing. H3K27ac enhancer and TSS refer to H3K27ac level QTLs tested at enhancers (chromHMM-defined (*33*) enhancer $<$ 100kb away from the nearest gene) and transcription start sites of a gene ($\pm$1kb away from gene TSS), respectively.

Annotation	$\log_2$ sum posterior differences	Bootstrap 95% CI
Strong Enhancer	**-1.18**	**(-1.47,-0.86)**
H2A	**-0.60**	**(-0.71,-0.46)**
H3K4me1	**-0.52**	**(-0.63,-0.38)**
H3K27ac	**-0.50**	**(-0.60,-0.38)**
Weak Enhancer	**-0.49**	**(-0.77,-0.18)**
H3K4me3	**-0.47**	**(-0.58,-0.36)**
H3K4me2	**-0.46**	**(-0.56,-0.34)**
H3K9ac	**-0.46**	**(-0.56,-0.35)**
Active Promoter	**-0.31**	**(-0.52,-0.13)**
Weak Promoter	-0.30	(-0.63,0.03)
TSS gencode	**-0.29**	**(-0.42,-0.15)**
H3K79me2	**-0.22**	**(-0.32,-0.12)**
upstream gene variant	**-0.21**	**(-0.34,-0.08)**
H3K9me3	-0.06	(-0.32,0.20)
H4K20me1	-0.03	(-0.21,0.14)
intron variant	**0.14**	**(0.06,0.21)**
Heterochrom	0.14	(-0.12,0.37)
downstream gene variant	**0.19**	**(0.04,0.35)**
Weak Txn	**0.40**	**(0.22,0.59)**
H3K36me3	**0.45**	**(0.35,0.55)**
gene exons	**0.48**	**(0.22,0.71)**
Txn Elongation	**0.88**	**(0.70,1.08)**

Table S4: Genomic annotation differences between chrom-eQTLs and unexplained eQTLs. Significant annotations are in bold font.

Phenotype	No. of genes tested	Decay π_1	80% CI
H3K27acShyam @ TSS	301	0.174	(0.030, 0.327)
Transcription (4sU)	759	0.173	(0.078, 0.277)
RNA-seq (Geuvadis)	1545	0.200	(0.130, 0.269)
Ribosome	562	0.207	(0.094, 0.324)
Protein	182	0.0	(0.0, 0.154)

Table S5: Proportion of regulatory QTLs that are decay QTLs estimated using Storey's π_1. 80%-confidence interval estimated using 500 bootstraps.

Phenotype	No. of genes tested	APA π_1	80% CI
H3K27acShyam @ TSS	198	0.0	(0.0, 0.063)
Transcription (4sU)	449	0.070	(0.0, 0.196)
RNA-seq (Geuvadis)	1040	0.188	(0.102, 0.269)
Ribosome	328	0.423	(0.307, 0.548)
Protein	104	0.362	(0.140, 0.585)

Table S6: Proportion of regulatory QTLs that are apaQTLs estimated using Storey's π_1. 80%-confidence interval estimated using 500 bootstraps.

cluster	sQTL p-value	intron coord	snp id	eQTL p-value	ctcfQTL p-value	haQTL p-value	dsQTL p-value	meQTL p-value
clu5430	2.21×10^{-4}	chr19:1011611:1011740	rs10405583	2.19×10^{-3}	ns	ns	2.49×10^{-3}	3.21×10^{-3}
clu1743	6.99×10^{-8}	chr11:68665480:68668003	rs544370	2.98×10^{-21}	ns	2.06×10^{-7}	ns	1.36×10^{-6}
clu4984	3.83×10^{-5}	chr14:24572464:24573051	rs144856253	7.36×10^{-1}	ns	1.23×10^{-3}	4.13×10^{-3}	ns
clu677	4.29×10^{-5}	chr12:57867959:57868235	rs4760273	4.96×10^{-1}	ns	ns	2.13×10^{-3}	1.43×10^{-3}
clu11546	3.45×10^{-7}	chr2:113955211:113955327	rs148543122	1.98×10^{-1}	ns	1.95×10^{-3}	ns	2.35×10^{-3}
clu2617	5.28×10^{-7}	chr10:127473829:127483449	rs10901449	1.54×10^{-2}	ns	6.02×10^{-4}	ns	8.11×10^{-4}
clu9550	2.04×10^{-4}	chr5:169188629:169191038	rs7721990	3.62×10^{-3}	ns	3.83×10^{-4}	2.16×10^{-3}	ns
clu5714	2.17×10^{-5}	chr19:10675710:10676582	rs3745244	1.26×10^{-1}	ns	7.21×10^{-4}	4.20×10^{-3}	ns
clu7346	4.36×10^{-4}	chr20:44424082:44429713	rs6032531	1.35×10^{-1}	2.18×10^{-12}	6.47×10^{-4}	ns	ns
clu14580	4.84×10^{-4}	chr8:145065075:145065368	rs7462703	7.96×10^{-1}	ns	ns	2.71×10^{-3}	4.10×10^{-3}
clu13108	4.84×10^{-4}	chr1:155238150:155238499	rs1057941	4.59×10^{-1}	ns	2.42×10^{-6}	3.02×10^{-3}	ns
clu3566	3.29×10^{-24}	chr17:74553939:74557370	rs4647885	1.60×10^{-3}	ns	2.31×10^{-3}	ns	4.91×10^{-3}
clu4373	8.10×10^{-20}	chr16:89165171:89167070	rs8060043	1.64×10^{-7}	ns	1.64×10^{-4}	ns	1.99×10^{-4}
clu7010	4.03×10^{-15}	chr22:50316091:50316260	rs6520064	3.17×10^{-1}	1.77×10^{-5}	1.00×10^{-3}	ns	ns
clu7713	2.74×10^{-6}	chr7:6049129:6057445	rs34097030	2.26×10^{-2}	ns	ns	6.80×10^{-4}	6.49×10^{-4}
clu5946	9.91×10^{-6}	chr19:19743075:19744612	rs873870	8.67×10^{-1}	ns	ns	1.81×10^{-3}	3.97×10^{-3}
clu6361	1.66×10^{-16}	chr19:54704756:54705028	rs3852889	5.76×10^{-1}	ns	8.71×10^{-4}	ns	2.61×10^{-3}
clu5321	1.29×10^{-16}	chr14:92583986:92587540	rs2896197	8.47×10^{-3}	ns	1.26×10^{-4}	ns	4.90×10^{-3}
clu7661	3.24×10^{-5}	chr7:643253:647502	rs111526379	1.75×10^{-2}	ns	1.75×10^{-7}	ns	1.58×10^{-3}
clu2223	5.34×10^{-4}	chr10:51592619:51606988	rs2012677	4.39×10^{-1}	ns	3.14×10^{-6}	3.43×10^{-3}	ns

Table S7: List of 20 random (of 171) sQTLs that are associated with two of CTCF, H3K27ac, methylation, or DNase levels at a nearby locus. Full table available in Data Table S1. The association p-values were computed as the minimum p-value of the SNP associations (linear regression t-test) with nearby tested windows for each chromatin-level phenotypes separately (ns: not significant or applicable).

Data	Accession
H3K27ac	GSE58852 (GEO)
DNA methylation	GSE57483 (GEO)
DNase-seq	GSE31388 (GEO)
4sU-seq	GSE75220 (GEO)
RNA-seq (Pickrell)	GSE19480 (GEO)
RNA-seq (GEUVADIS)	E-GEUV-3 (ArrayExpress)
RNA decay	GSE37451 (GEO)
ribo-seq	GSE61742 (GEO)
protein	PXD001406 (ProteomeXchange)

Table S8: Location of datasets used to call QTLs in this study.

Supplementary Data Set

Supplementary Data Table S1. List of 171 sQTLs that are associated with chromatin-level variation.

cluster: Cluster identification number

sQTL p-value: p-value (t-test) of the association with intron excision level

intron coord: coordinates of the intron

snp id: Identification of the SNP

eQTL p-value: association with total gene expression level

ctcfQTL p-value: minimum p-value (t-test) of the association with a nearby CTCF peak

haQTL p-value: minimum p-value (t-test) of the association with a nearby H3K27ac peak

dsQTL p-value: minimum p-value (t-test) of the association with a nearby DNase peak

meQTL p-value: minimum p-value (t-test) of the association with a nearby methylation probe

References and Notes

1. D. L. Nicolae, E. Gamazon, W. Zhang, S. Duan, M. E. Dolan, N. J. Cox, Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. *PLOS Genet.* **6**, e1000888 (2010). Medline doi:10.1371/journal.pgen.1000888

2. The GTEx Consortium, The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015). Medline doi:10.1126/science.1262110

3. M. Kasowski, F. Grubert, C. Heffelfinger, M. Hariharan, A. Asabere, S. M. Waszak, L. Habegger, J. Rozowsky, M. Shi, A. E. Urban, M. Y. Hong, K. J. Karczewski, W. Huber, S. M. Weissman, M. B. Gerstein, J. O. Korbel, M. Snyder, Variation in transcription factor binding among humans. *Science* **328**, 232–235 (2010). Medline doi:10.1126/science.1183621

4. J. F. Degner, A. A. Pai, R. Pique-Regi, J. B. Veyrieras, D. J. Gaffney, J. K. Pickrell, S. De Leon, K. Michelini, N. Lewellen, G. E. Crawford, M. Stephens, Y. Gilad, J. K. Pritchard, DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* **482**, 390–394 (2012). Medline doi:10.1038/nature10808

5. G. McVicker, B. van de Geijn, J. F. Degner, C. E. Cain, N. E. Banovich, A. Raj, N. Lewellen, M. Myrthil, Y. Gilad, J. K. Pritchard, Identification of genetic variants that affect histone modifications in human cells. *Science* **342**, 747–749 (2013). Medline doi:10.1126/science.1242429

6. T. Lappalainen, M. Sammeth, M. R. Friedländer, P. A. ’t Hoen, J. Monlong, M. A. Rivas, M. Gonzàlez-Porta, N. Kurbatova, T. Griebel, P. G. Ferreira, M. Barann, T. Wieland, L. Greger, M. van Iterson, J. Almlöf, P. Ribeca, I. Pulyakhina, D. Esser, T. Giger, A. Tikhonov, M. Sultan, G. Bertier, D. G. MacArthur, M. Lek, E. Lizano, H. P. Buermans, I. Padioleau, T. Schwarzmayr, O. Karlberg, H. Ongen, H. Kilpinen, S. Beltran, M. Gut, K. Kahlem, V. Amstislavskiy, O. Stegle, M. Pirinen, S. B. Montgomery, P. Donnelly, M. I. McCarthy, P. Flicek, T. M. Strom, H. Lehrach, S. Schreiber, R. Sudbrak, A. Carracedo, S. E. Antonarakis, R. Häsler, A. C. Syvänen, G. J. van Ommen, A. Brazma, T. Meitinger, P. Rosenstiel, R. Guigó, I. G. Gut, X. Estivill, E. T. Dermitzakis; The Geuvadis Consortium, Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013). Medline doi:10.1038/nature12531

7. R. C. del Rosario, J. Poschmann, S. L. Rouam, E. Png, C. C. Khor, M. L. Hibberd, S. Prabhakar, Sensitive detection of chromatin-altering polymorphisms reveals autoimmune disease mechanisms. *Nat. Methods* **12**, 458–464 (2015). Medline doi:10.1038/nmeth.3326

8. F. Grubert, J. B. Zaugg, M. Kasowski, O. Ursu, D. V. Spacek, A. R. Martin, P. Greenside, R. Srivas, D. H. Phanstiel, A. Pekowska, N. Heidari, G. Euskirchen, W. Huber, J. K. Pritchard, C. D. Bustamante, L. M. Steinmetz, A. Kundaje, M. Snyder, Genetic control of chromatin states in humans involves local and distal chromosomal interactions. *Cell* **162**, 1051–1065 (2015). Medline doi:10.1016/j.cell.2015.07.048

9. S. M. Waszak, O. Delaneau, A. R. Gschwind, H. Kilpinen, S. K. Raghav, R. M. Witwicki, A. Orioli, M. Wiederkehr, N. I. Panousis, A. Yurovsky, L. Romano-Palumbo, A. Planchon, D. Bielser, I. Padioleau, G. Udin, S. Thurnheer, D. Hacker, N. Hernandez, A. Reymond, B. Deplancke, E. T. Dermitzakis, Population variation and genetic control of modular chromatin architecture in humans. *Cell* **162**, 1039–1050 (2015). Medline doi:10.1016/j.cell.2015.08.001

10. H. K. Finucane, B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, P. R. Loh, V. Anttila, H. Xu, C. Zang, K. Farh, S. Ripke, F. R. Day, S. Purcell, E. Stahl, S. Lindstrom, J. R. Perry, Y. Okada, S. Raychaudhuri, M. J. Daly, N. Patterson, B. M. Neale, A. L. Price; ReproGen Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; RACI Consortium, Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015). Medline doi:10.1038/ng.3404

11. A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y. C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shoresh, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K. H. Farh, S. Feizi, R. Karlic, A. R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L. H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, M. Kellis; Roadmap Epigenomics Consortium, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015). Medline doi:10.1038/nature14248

12. M. Claussnitzer, S. N. Dankel, K. H. Kim, G. Quon, W. Meuleman, C. Haugen, V. Glunk, I. S. Sousa, J. L. Beaudry, V. Puviindran, N. A. Abdennur, J. Liu, P. A. Svensson, Y. H. Hsu, D. J. Drucker, G. Mellgren, C. C. Hui, H. Hauner, M. Kellis, FTO obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* **373**, 895–907 (2015). Medline doi:10.1056/NEJMoa1502214

13. M. A. Garcia-Blanco, A. P. Baraniak, E. L. Lasda, Alternative splicing in disease and therapy. *Nat. Biotechnol.* **22**, 535–546 (2004). Medline doi:10.1038/nbt964

14. H. B. Fraser, X. Xie, Common polymorphic transcript variation in human disease. *Genome Res.* **19**, 567–575 (2009). Medline doi:10.1101/gr.083477.108

15. H. Y. Xiong, B. Alipanahi, L. J. Lee, H. Bretschneider, D. Merico, R. K. Yuen, Y. Hua, S. Gueroussov, H. S. Najafabadi, T. R. Hughes, Q. Morris, Y. Barash, A. R. Krainer, N. Jojic, S. W. Scherer, B. J. Blencowe, B. J. Frey, The human splicing

code reveals new insights into the genetic determinants of disease. *Science* **347**, 1254806 (2015). Medline doi:10.1126/science.1254806

16. See the supplementary materials and methods on *Science* Online.

17. J. K. Pickrell, J. C. Marioni, A. A. Pai, J. F. Degner, B. E. Engelhardt, E. Nkadori, J. B. Veyrieras, M. Stephens, Y. Gilad, J. K. Pritchard, Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768–772 (2010). Medline doi:10.1038/nature08872

18. N. E. Banovich, X. Lan, G. McVicker, B. van de Geijn, J. F. Degner, J. D. Blischak, J. Roux, J. K. Pritchard, Y. Gilad, Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. *PLOS Genet.* **10**, e1004663 (2014). Medline doi:10.1371/journal.pgen.1004663

19. A. Battle, Z. Khan, S. H. Wang, A. Mitrano, M. J. Ford, J. K. Pritchard, Y. Gilad, Impact of regulatory variation from RNA to protein. *Science* **347**, 664–667 (2015). Medline doi:10.1126/science.1260793

20. Y. I. Li, D. A. Knowles, J. K. Pritchard, **http://biorxiv.org/content/early/2016/03/16/044107** (2016).

21. S. Shukla, E. Kavak, M. Gregory, M. Imashimizu, B. Shutinoski, M. Kashlev, P. Oberdoerffer, R. Sandberg, S. Oberdoerffer, CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74–79 (2011). Medline doi:10.1038/nature10442

22. M. Gutierrez-Arcelus, H. Ongen, T. Lappalainen, S. B. Montgomery, A. Buil, A. Yurovsky, J. Bryois, I. Padioleau, L. Romano, A. Planchon, E. Falconnet, D. Bielser, M. Gagnebin, T. Giger, C. Borel, A. Letourneau, P. Makrythanasis, M. Guipponi, C. Gehrig, S. E. Antonarakis, E. T. Dermitzakis, Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. *PLOS Genet.* **11**, e1004958 (2015). Medline doi:10.1371/journal.pgen.1004958

23. J. K. Pickrell, Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94**, 559–573 (2014). Medline doi:10.1016/j.ajhg.2014.03.004

24. T. Flutre, X. Wen, J. Pritchard, M. Stephens, A statistical framework for joint eQTL analysis in multiple tissues. *PLOS Genet.* **9**, e1003486 (2013). Medline doi:10.1371/journal.pgen.1003486

25. E. Ira, M. Zanoni, M. Ruggeri, P. Dazzan, S. Tosato, COMT, neuropsychological function and brain structure in schizophrenia: A systematic review and neurobiological interpretation. *J. Psychiatry Neurosci.* **38**, 366–380 (2013). Medline doi:10.1503/jpn.120178

26. M. Rabani, J. Z. Levin, L. Fan, X. Adiconis, R. Raychowdhury, M. Garber, A. Gnirke, C. Nusbaum, N. Hacohen, N. Friedman, I. Amit, A. Regev, Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat. Biotechnol.* **29**, 436–442 (2011). Medline doi:10.1038/nbt.1861

27. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009). Medline doi:10.1186/gb-2009-10-3-r25

28. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). Medline doi:10.1093/bioinformatics/bts635

29. B. van de Geijn, G. McVicker, Y. Gilad, J. K. Pritchard, WASP: Allele-specific software for robust molecular quantitative trait locus discovery. *Nat. Methods* **12**, 1061–1063 (2015). Medline doi:10.1038/nmeth.3582

30. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014). Medline doi:10.1093/bioinformatics/btt656

31. A. A. Pai, C. E. Cain, O. Mizrahi-Man, S. De Leon, N. Lewellen, J. B. Veyrieras, J. F. Degner, D. J. Gaffney, J. K. Pickrell, M. Stephens, J. K. Pritchard, Y. Gilad, The contribution of RNA decay quantitative trait loci to inter-individual variation in steady-state gene expression levels. *PLOS Genet.* **8**, e1003000 (2012). Medline doi:10.1371/journal.pgen.1003000

32. J. Ernst, M. Kellis, ChromHMM: Automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012). Medline doi:10.1038/nmeth.1906

33. R. C. del Rosario, N. A. Rayan, S. Prabhakar, Noncoding origins of anthropoid traits and a new null model of transposon functionalization. *Genome Res.* **24**, 1469–1484 (2014). Medline doi:10.1101/gr.168963.113

34. Z. Xia, L. A. Donehower, T. A. Cooper, J. R. Neilson, D. A. Wheeler, E. J. Wagner, W. Li, Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3′-UTR landscape across seven tumour types. *Nat. Commun.* **5**, 5274 (2014). Medline doi:10.1038/ncomms6274

35. O. K. Yoon, T. Y. Hsu, J. H. Im, R. B. Brem, Genetics and regulatory impact of alternative polyadenylation in human B-lymphoblastoid cells. *PLOS Genet.* **8**, e1002882 (2012). Medline doi:10.1371/journal.pgen.1002882

36. J. B. Veyrieras, S. Kudaravalli, S. Y. Kim, E. T. Dermitzakis, Y. Gilad, M. Stephens, J. K. Pritchard, High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLOS Genet.* **4**, e1000214 (2008). Medline doi:10.1371/journal.pgen.1000214

37. A. A. Shabalin, Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358 (2012). Medline doi:10.1093/bioinformatics/bts163

38. P. Cingolani, A. Platts, L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, D. M. Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* (*Austin*) **6**, 80–92 (2012). Medline doi:10.4161/fly.19695

39. J. Curado, C. Iannone, H. Tilgner, J. Valcárcel, R. Guigó, Promoter-like epigenetic signatures in exons displaying cell type-specific splicing. *Genome Biol.* **16**, 236 (2015). Medline doi:10.1186/s13059-015-0797-8

40. Z. Ding, Y. Ni, S. W. Timmer, B. K. Lee, A. Battenhouse, S. Louzada, F. Yang, I. Dunham, G. E. Crawford, J. D. Lieb, R. Durbin, V. R. Iyer, E. Birney, Quantitative genetics of CTCF binding reveal local sequence effects and different modes of X-chromosome association. *PLOS Genet.* **10**, e1004798 (2014). Medline doi:10.1371/journal.pgen.1004798

41. Y. Okada, D. Wu, G. Trynka, T. Raj, C. Terao, K. Ikari, Y. Kochi, K. Ohmura, A. Suzuki, S. Yoshida, R. R. Graham, A. Manoharan, W. Ortmann, T. Bhangale, J. C. Denny, R. J. Carroll, A. E. Eyler, J. D. Greenberg, J. M. Kremer, D. A. Pappas, L. Jiang, J. Yin, L. Ye, D. F. Su, J. Yang, G. Xie, E. Keystone, H. J. Westra, T. Esko, A. Metspalu, X. Zhou, N. Gupta, D. Mirel, E. A. Stahl, D. Diogo, J. Cui, K. Liao, M. H. Guo, K. Myouzen, T. Kawaguchi, M. J. Coenen, P. L. van Riel, M. A. van de Laar, H. J. Guchelaar, T. W. Huizinga, P. Dieudé, X. Mariette, S. L. Bridges Jr., A. Zhernakova, R. E. Toes, P. P. Tak, C. Miceli-Richard, S. Y. Bang, H. S. Lee, J. Martin, M. A. Gonzalez-Gay, L. Rodriguez-Rodriguez, S. Rantapää-Dahlqvist, L. Arlestig, H. K. Choi, Y. Kamatani, P. Galan, M. Lathrop, S. Eyre, J. Bowes, A. Barton, N. de Vries, L. W. Moreland, L. A. Criswell, E. W. Karlson, A. Taniguchi, R. Yamada, M. Kubo, J. S. Liu, S. C. Bae, J. Worthington, L. Padyukov, L. Klareskog, P. K. Gregersen, S. Raychaudhuri, B. E. Stranger, P. L. De Jager, L. Franke, P. M. Visscher, M. A. Brown, H. Yamanaka, T. Mimori, A. Takahashi, H. Xu, T. W. Behrens, K. A. Siminovitch, S. Momohara, F. Matsuda, K. Yamamoto, R. M. Plenge; RACI Cnsortium; GARNET Consortium, Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014). Medline doi:10.1038/nature12873

42. S. Sawcer, G. Hellenthal, M. Pirinen, C. C. Spencer, N. A. Patsopoulos, L. Moutsianas, A. Dilthey, Z. Su, C. Freeman, S. E. Hunt, S. Edkins, E. Gray, D. R. Booth, S. C. Potter, A. Goris, G. Band, A. B. Oturai, A. Strange, J. Saarela, C. Bellenguez, B. Fontaine, M. Gillman, B. Hemmer, R. Gwilliam, F. Zipp, A. Jayakumar, R. Martin, S. Leslie, S. Hawkins, E. Giannoulatou, S. D'alfonso, H. Blackburn, F. Martinelli Boneschi, J. Liddle, H. F. Harbo, M. L. Perez, A. Spurkland, M. J. Waller, M. P. Mycko, M. Ricketts, M. Comabella, N. Hammond, I. Kockum, O. T. McCann, M. Ban, P. Whittaker, A. Kemppinen, P. Weston, C. Hawkins, S. Widaa, J. Zajicek, S. Dronov, N. Robertson, S. J. Bumpstead, L. F. Barcellos, R. Ravindrarajah, R. Abraham, L. Alfredsson, K. Ardlie, C. Aubin, A. Baker, K. Baker, S. E. Baranzini, L. Bergamaschi, R. Bergamaschi, A. Bernstein, A. Berthele, M. Boggild, J. P. Bradfield, D. Brassat, S. A. Broadley, D. Buck, H. Butzkueven, R. Capra, W. M. Carroll, P. Cavalla, E. G. Celius, S. Cepok, R. Chiavacci, F. Clerget-Darpoux, K. Clysters, G. Comi, M. Cossburn, I. Cournu-Rebeix, M. B. Cox, W. Cozen, B. A. Cree, A. H. Cross, D. Cusi, M. J. Daly, E. Davis, P. I. de Bakker, M. Debouverie, M. B. D'hooghe, K. Dixon, R. Dobosi, B. Dubois, D. Ellinghaus, I. Elovaara, F. Esposito, C. Fontenille, S. Foote, A. Franke, D. Galimberti, A. Ghezzi, J. Glessner, R. Gomez, O. Gout, C. Graham, S. F. Grant, F. R. Guerini, H. Hakonarson, P. Hall, A. Hamsten, H. P. Hartung, R. N.

Heard, S. Heath, J. Hobart, M. Hoshi, C. Infante-Duarte, G. Ingram, W. Ingram, T. Islam, M. Jagodic, M. Kabesch, A. G. Kermode, T. J. Kilpatrick, C. Kim, N. Klopp, K. Koivisto, M. Larsson, M. Lathrop, J. S. Lechner-Scott, M. A. Leone, V. Leppä, U. Liljedahl, I. L. Bomfim, R. R. Lincoln, J. Link, J. Liu, A. R. Lorentzen, S. Lupoli, F. Macciardi, T. Mack, M. Marriott, V. Martinelli, D. Mason, J. L. McCauley, F. Mentch, I. L. Mero, T. Mihalova, X. Montalban, J. Mottershead, K. M. Myhr, P. Naldi, W. Ollier, A. Page, A. Palotie, J. Pelletier, L. Piccio, T. Pickersgill, F. Piehl, S. Pobywajlo, H. L. Quach, P. P. Ramsay, M. Reunanen, R. Reynolds, J. D. Rioux, M. Rodegher, S. Roesner, J. P. Rubio, I. M. Rückert, M. Salvetti, E. Salvi, A. Santaniello, C. A. Schaefer, S. Schreiber, C. Schulze, R. J. Scott, F. Sellebjerg, K. W. Selmaj, D. Sexton, L. Shen, B. Simms-Acuna, S. Skidmore, P. M. Sleiman, C. Smestad, P. S. Sørensen, H. B. Søndergaard, J. Stankovich, R. C. Strange, A. M. Sulonen, E. Sundqvist, A. C. Syvänen, F. Taddeo, B. Taylor, J. M. Blackwell, P. Tienari, E. Bramon, A. Tourbah, M. A. Brown, E. Tronczynska, J. P. Casas, N. Tubridy, A. Corvin, J. Vickery, J. Jankowski, P. Villoslada, H. S. Markus, K. Wang, C. G. Mathew, J. Wason, C. N. Palmer, H. E. Wichmann, R. Plomin, E. Willoughby, A. Rautanen, J. Winkelmann, M. Wittig, R. C. Trembath, J. Yaouanq, A. C. Viswanathan, H. Zhang, N. W. Wood, R. Zuvich, P. Deloukas, C. Langford, A. Duncanson, J. R. Oksenberg, M. A. Pericak-Vance, J. L. Haines, T. Olsson, J. Hillert, A. J. Ivinson, P. L. De Jager, L. Peltonen, G. J. Stewart, D. A. Hafler, S. L. Hauser, G. McVean, P. Donnelly, A. Compston; International Multiple Sclerosis Genetics Consortium; Wellcome Trust Case Control Consortium 2, Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011). Medline doi:10.1038/nature10251

43. A. C. Naj, G. Jun, G. W. Beecham, L. S. Wang, B. N. Vardarajan, J. Buros, P. J. Gallins, J. D. Buxbaum, G. P. Jarvik, P. K. Crane, E. B. Larson, T. D. Bird, B. F. Boeve, N. R. Graff-Radford, P. L. De Jager, D. Evans, J. A. Schneider, M. M. Carrasquillo, N. Ertekin-Taner, S. G. Younkin, C. Cruchaga, J. S. Kauwe, P. Nowotny, P. Kramer, J. Hardy, M. J. Huentelman, A. J. Myers, M. M. Barmada, F. Y. Demirci, C. T. Baldwin, R. C. Green, E. Rogaeva, P. St George-Hyslop, S. E. Arnold, R. Barber, T. Beach, E. H. Bigio, J. D. Bowen, A. Boxer, J. R. Burke, N. J. Cairns, C. S. Carlson, R. M. Carney, S. L. Carroll, H. C. Chui, D. G. Clark, J. Corneveaux, C. W. Cotman, J. L. Cummings, C. DeCarli, S. T. DeKosky, R. Diaz-Arrastia, M. Dick, D. W. Dickson, W. G. Ellis, K. M. Faber, K. B. Fallon, M. R. Farlow, S. Ferris, M. P. Frosch, D. R. Galasko, M. Ganguli, M. Gearing, D. H. Geschwind, B. Ghetti, J. R. Gilbert, S. Gilman, B. Giordani, J. D. Glass, J. H. Growdon, R. L. Hamilton, L. E. Harrell, E. Head, L. S. Honig, C. M. Hulette, B. T. Hyman, G. A. Jicha, L. W. Jin, N. Johnson, J. Karlawish, A. Karydas, J. A. Kaye, R. Kim, E. H. Koo, N. W. Kowall, J. J. Lah, A. I. Levey, A. P. Lieberman, O. L. Lopez, W. J. Mack, D. C. Marson, F. Martiniuk, D. C. Mash, E. Masliah, W. C. McCormick, S. M. McCurry, A. N. McDavid, A. C. McKee, M. Mesulam, B. L. Miller, C. A. Miller, J. W. Miller, J. E. Parisi, D. P. Perl, E. Peskind, R. C. Petersen, W. W. Poon, J. F. Quinn, R. A. Rajbhandary, M. Raskind, B. Reisberg, J. M. Ringman, E. D. Roberson, R. N. Rosenberg, M. Sano, L. S. Schneider, W. Seeley, M. L. Shelanski, M. A. Slifer, C. D. Smith, J. A. Sonnen, S. Spina, R. A.

Stern, R. E. Tanzi, J. Q. Trojanowski, J. C. Troncoso, V. M. Van Deerlin, H. V. Vinters, J. P. Vonsattel, S. Weintraub, K. A. Welsh-Bohmer, J. Williamson, R. L. Woltjer, L. B. Cantwell, B. A. Dombroski, D. Beekly, K. L. Lunetta, E. R. Martin, M. I. Kamboh, A. J. Saykin, E. M. Reiman, D. A. Bennett, J. C. Morris, T. J. Montine, A. M. Goate, D. Blacker, D. W. Tsuang, H. Hakonarson, W. A. Kukull, T. M. Foroud, J. L. Haines, R. Mayeux, M. A. Pericak-Vance, L. A. Farrer, G. D. Schellenberg, Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat. Genet.* **43**, 436–441 (2011). Medline

44. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014). Medline doi:10.1038/nature13595

45. A. R. Wood, T. Esko, J. Yang, S. Vedantam, T. H. Pers, S. Gustafsson, A. Y. Chu, K. Estrada, J. Luan, Z. Kutalik, N. Amin, M. L. Buchkovich, D. C. Croteau-Chonka, F. R. Day, Y. Duan, T. Fall, R. Fehrmann, T. Ferreira, A. U. Jackson, J. Karjalainen, K. S. Lo, A. E. Locke, R. Mägi, E. Mihailov, E. Porcu, J. C. Randall, A. Scherag, A. A. Vinkhuyzen, H. J. Westra, T. W. Winkler, T. Workalemahu, J. H. Zhao, D. Absher, E. Albrecht, D. Anderson, J. Baron, M. Beekman, A. Demirkan, G. B. Ehret, B. Feenstra, M. F. Feitosa, K. Fischer, R. M. Fraser, A. Goel, J. Gong, A. E. Justice, S. Kanoni, M. E. Kleber, K. Kristiansson, U. Lim, V. Lotay, J. C. Lui, M. Mangino, I. Mateo Leach, C. Medina-Gomez, M. A. Nalls, D. R. Nyholt, C. D. Palmer, D. Pasko, S. Pechlivanis, I. Prokopenko, J. S. Ried, S. Ripke, D. Shungin, A. Stancáková, R. J. Strawbridge, Y. J. Sung, T. Tanaka, A. Teumer, S. Trompet, S. W. van der Laan, J. van Setten, J. V. Van Vliet-Ostaptchouk, Z. Wang, L. Yengo, W. Zhang, U. Afzal, J. Arnlöv, G. M. Arscott, S. Bandinelli, A. Barrett, C. Bellis, A. J. Bennett, C. Berne, M. Blüher, J. L. Bolton, Y. Böttcher, H. A. Boyd, M. Bruinenberg, B. M. Buckley, S. Buyske, I. H. Caspersen, P. S. Chines, R. Clarke, S. Claudi-Boehm, M. Cooper, E. W. Daw, P. A. De Jong, J. Deelen, G. Delgado, J. C. Denny, R. Dhonukshe-Rutten, M. Dimitriou, A. S. Doney, M. Dörr, N. Eklund, E. Eury, L. Folkersen, M. E. Garcia, F. Geller, V. Giedraitis, A. S. Go, H. Grallert, T. B. Grammer, J. Gräßler, H. Grönberg, L. C. de Groot, C. J. Groves, J. Haessler, P. Hall, T. Haller, G. Hallmans, A. Hannemann, C. A. Hartman, M. Hassinen, C. Hayward, N. L. Heard-Costa, Q. Helmer, G. Hemani, A. K. Henders, H. L. Hillege, M. A. Hlatky, W. Hoffmann, P. Hoffmann, O. Holmen, J. J. Houwing-Duistermaat, T. Illig, A. Isaacs, A. L. James, J. Jeff, B. Johansen, Å. Johansson, J. Jolley, T. Juliusdottir, J. Junttila, A. N. Kho, L. Kinnunen, N. Klopp, T. Kocher, W. Kratzer, P. Lichtner, L. Lind, J. Lindström, S. Lobbens, M. Lorentzon, Y. Lu, V. Lyssenko, P. K. Magnusson, A. Mahajan, M. Maillard, W. L. McArdle, C. A. McKenzie, S. McLachlan, P. J. McLaren, C. Menni, S. Merger, L. Milani, A. Moayyeri, K. L. Monda, M. A. Morken, G. Müller, M. Müller-Nurasyid, A. W. Musk, N. Narisu, M. Nauck, I. M. Nolte, M. M. Nöthen, L. Oozageer, S. Pilz, N. W. Rayner, F. Renstrom, N. R. Robertson, L. M. Rose, R. Roussel, S. Sanna, H. Scharnagl, S. Scholtens, F. R. Schumacher, H. Schunkert, R. A. Scott, J. Sehmi, T. Seufferlein, J. Shi, K. Silventoinen, J. H. Smit, A. V. Smith, J. Smolonska, A. V. Stanton, K. Stirrups, D. J. Stott, H. M. Stringham, J. Sundström, M. A. Swertz, A. C.

Syvänen, B. O. Tayo, G. Thorleifsson, J. P. Tyrer, S. van Dijk, N. M. van Schoor, N. van der Velde, D. van Heemst, F. V. van Oort, S. H. Vermeulen, N. Verweij, J. M. Vonk, L. L. Waite, M. Waldenberger, R. Wennauer, L. R. Wilkens, C. Willenborg, T. Wilsgaard, M. K. Wojczynski, A. Wong, A. F. Wright, Q. Zhang, D. Arveiler, S. J. Bakker, J. Beilby, R. N. Bergman, S. Bergmann, R. Biffar, J. Blangero, D. I. Boomsma, S. R. Bornstein, P. Bovet, P. Brambilla, M. J. Brown, H. Campbell, M. J. Caulfield, A. Chakravarti, R. Collins, F. S. Collins, D. C. Crawford, L. A. Cupples, J. Danesh, U. de Faire, H. M. den Ruijter, R. Erbel, J. Erdmann, J. G. Eriksson, M. Farrall, E. Ferrannini, J. Ferrières, I. Ford, N. G. Forouhi, T. Forrester, R. T. Gansevoort, P. V. Gejman, C. Gieger, A. Golay, O. Gottesman, V. Gudnason, U. Gyllensten, D. W. Haas, A. S. Hall, T. B. Harris, A. T. Hattersley, A. C. Heath, C. Hengstenberg, A. A. Hicks, L. A. Hindorff, A. D. Hingorani, A. Hofman, G. K. Hovingh, S. E. Humphries, S. C. Hunt, E. Hypponen, K. B. Jacobs, M. R. Jarvelin, P. Jousilahti, A. M. Jula, J. Kaprio, J. J. Kastelein, M. Kayser, F. Kee, S. M. Keinanen-Kiukaanniemi, L. A. Kiemeney, J. S. Kooner, C. Kooperberg, S. Koskinen, P. Kovacs, A. T. Kraja, M. Kumari, J. Kuusisto, T. A. Lakka, C. Langenberg, L. Le Marchand, T. Lehtimäki, S. Lupoli, P. A. Madden, S. Männistö, P. Manunta, A. Marette, T. C. Matise, B. McKnight, T. Meitinger, F. L. Moll, G. W. Montgomery, A. D. Morris, A. P. Morris, J. C. Murray, M. Nelis, C. Ohlsson, A. J. Oldehinkel, K. K. Ong, W. H. Ouwehand, G. Pasterkamp, A. Peters, P. P. Pramstaller, J. F. Price, L. Qi, O. T. Raitakari, T. Rankinen, D. C. Rao, T. K. Rice, M. Ritchie, I. Rudan, V. Salomaa, N. J. Samani, J. Saramies, M. A. Sarzynski, P. E. Schwarz, S. Sebert, P. Sever, A. R. Shuldiner, J. Sinisalo, V. Steinthorsdottir, R. P. Stolk, J. C. Tardif, A. Tönjes, A. Tremblay, E. Tremoli, J. Virtamo, M. C. Vohl, P. Amouyel, F. W. Asselbergs, T. L. Assimes, M. Bochud, B. O. Boehm, E. Boerwinkle, E. P. Bottinger, C. Bouchard, S. Cauchi, J. C. Chambers, S. J. Chanock, R. S. Cooper, P. I. de Bakker, G. Dedoussis, L. Ferrucci, P. W. Franks, P. Froguel, L. C. Groop, C. A. Haiman, A. Hamsten, M. G. Hayes, J. Hui, D. J. Hunter, K. Hveem, J. W. Jukema, R. C. Kaplan, M. Kivimaki, D. Kuh, M. Laakso, Y. Liu, N. G. Martin, W. März, M. Melbye, S. Moebus, P. B. Munroe, I. Njølstad, B. A. Oostra, C. N. Palmer, N. L. Pedersen, M. Perola, L. Pérusse, U. Peters, J. E. Powell, C. Power, T. Quertermous, R. Rauramaa, E. Reinmaa, P. M. Ridker, F. Rivadeneira, J. I. Rotter, T. E. Saaristo, D. Saleheen, D. Schlessinger, P. E. Slagboom, H. Snieder, T. D. Spector, K. Strauch, M. Stumvoll, J. Tuomilehto, M. Uusitupa, P. van der Harst, H. Völzke, M. Walker, N. J. Wareham, H. Watkins, H. E. Wichmann, J. F. Wilson, P. Zanen, P. Deloukas, I. M. Heid, C. M. Lindgren, K. L. Mohlke, E. K. Speliotes, U. Thorsteinsdottir, I. Barroso, C. S. Fox, K. E. North, D. P. Strachan, J. S. Beckmann, S. I. Berndt, M. Boehnke, I. B. Borecki, M. I. McCarthy, A. Metspalu, K. Stefansson, A. G. Uitterlinden, C. M. van Duijn, L. Franke, C. J. Willer, A. L. Price, G. Lettre, R. J. Loos, M. N. Weedon, E. Ingelsson, J. R. O'Connell, G. R. Abecasis, D. I. Chasman, M. E. Goddard, P. M. Visscher, J. N. Hirschhorn, T. M. Frayling; Electronic Medical Records and Genomics (eMEMERGEGE) Consortium; MIGen Consortium; PAGEGE Consortium; LifeLines Cohort Study, Defining the role of common variation in the genomic

and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014). Medline

46. A. E. Locke, B. Kahali, S. I. Berndt, A. E. Justice, T. H. Pers, F. R. Day, C. Powell, S. Vedantam, M. L. Buchkovich, J. Yang, D. C. Croteau-Chonka, T. Esko, T. Fall, T. Ferreira, S. Gustafsson, Z. Kutalik, J. Luan, R. Mägi, J. C. Randall, T. W. Winkler, A. R. Wood, T. Workalemahu, J. D. Faul, J. A. Smith, J. Hua Zhao, W. Zhao, J. Chen, R. Fehrmann, Å. K. Hedman, J. Karjalainen, E. M. Schmidt, D. Absher, N. Amin, D. Anderson, M. Beekman, J. L. Bolton, J. L. Bragg-Gresham, S. Buyske, A. Demirkan, G. Deng, G. B. Ehret, B. Feenstra, M. F. Feitosa, K. Fischer, A. Goel, J. Gong, A. U. Jackson, S. Kanoni, M. E. Kleber, K. Kristiansson, U. Lim, V. Lotay, M. Mangino, I. Mateo Leach, C. Medina-Gomez, S. E. Medland, M. A. Nalls, C. D. Palmer, D. Pasko, S. Pechlivanis, M. J. Peters, I. Prokopenko, D. Shungin, A. Stančáková, R. J. Strawbridge, Y. Ju Sung, T. Tanaka, A. Teumer, S. Trompet, S. W. van der Laan, J. van Setten, J. V. Van Vliet-Ostaptchouk, Z. Wang, L. Yengo, W. Zhang, A. Isaacs, E. Albrecht, J. Ärnlöv, G. M. Arscott, A. P. Attwood, S. Bandinelli, A. Barrett, I. N. Bas, C. Bellis, A. J. Bennett, C. Berne, R. Blagieva, M. Blüher, S. Böhringer, L. L. Bonnycastle, Y. Böttcher, H. A. Boyd, M. Bruinenberg, I. H. Caspersen, Y. D. Ida Chen, R. Clarke, E. W. Daw, A. J. de Craen, G. Delgado, M. Dimitriou, A. S. Doney, N. Eklund, K. Estrada, E. Eury, L. Folkersen, R. M. Fraser, M. E. Garcia, F. Geller, V. Giedraitis, B. Gigante, A. S. Go, A. Golay, A. H. Goodall, S. D. Gordon, M. Gorski, H. J. Grabe, H. Grallert, T. B. Grammer, J. Gräßler, H. Grönberg, C. J. Groves, G. Gusto, J. Haessler, P. Hall, T. Haller, G. Hallmans, C. A. Hartman, M. Hassinen, C. Hayward, N. L. Heard-Costa, Q. Helmer, C. Hengstenberg, O. Holmen, J. J. Hottenga, A. L. James, J. M. Jeff, Å. Johansson, J. Jolley, T. Juliusdottir, L. Kinnunen, W. Koenig, M. Koskenvuo, W. Kratzer, J. Laitinen, C. Lamina, K. Leander, N. R. Lee, P. Lichtner, L. Lind, J. Lindström, K. Sin Lo, S. Lobbens, R. Lorbeer, Y. Lu, F. Mach, P. K. Magnusson, A. Mahajan, W. L. McArdle, S. McLachlan, C. Menni, S. Merger, E. Mihailov, L. Milani, A. Moayyeri, K. L. Monda, M. A. Morken, A. Mulas, G. Müller, M. Müller-Nurasyid, A. W. Musk, R. Nagaraja, M. M. Nöthen, I. M. Nolte, S. Pilz, N. W. Rayner, F. Renstrom, R. Rettig, J. S. Ried, S. Ripke, N. R. Robertson, L. M. Rose, S. Sanna, H. Scharnagl, S. Scholtens, F. R. Schumacher, W. R. Scott, T. Seufferlein, J. Shi, A. Vernon Smith, J. Smolonska, A. V. Stanton, V. Steinthorsdottir, K. Stirrups, H. M. Stringham, J. Sundström, M. A. Swertz, A. J. Swift, A. C. Syvänen, S. T. Tan, B. O. Tayo, B. Thorand, G. Thorleifsson, J. P. Tyrer, H. W. Uh, L. Vandenput, F. C. Verhulst, S. H. Vermeulen, N. Verweij, J. M. Vonk, L. L. Waite, H. R. Warren, D. Waterworth, M. N. Weedon, L. R. Wilkens, C. Willenborg, T. Wilsgaard, M. K. Wojczynski, A. Wong, A. F. Wright, Q. Zhang, E. P. Brennan, M. Choi, Z. Dastani, A. W. Drong, P. Eriksson, A. Franco-Cereceda, J. R. Gådin, A. G. Gharavi, M. E. Goddard, R. E. Handsaker, J. Huang, F. Karpe, S. Kathiresan, S. Keildson, K. Kiryluk, M. Kubo, J. Y. Lee, L. Liang, R. P. Lifton, B. Ma, S. A. McCarroll, A. J. McKnight, J. L. Min, M. F. Moffatt, G. W. Montgomery, J. M. Murabito, G. Nicholson, D. R. Nyholt, Y. Okada, J. R. Perry, R. Dorajoo, E. Reinmaa, R. M. Salem, N. Sandholm, R. A. Scott, L. Stolk, A. Takahashi, T. Tanaka, F. M. Van't Hooft, A.

A. Vinkhuyzen, H. J. Westra, W. Zheng, K. T. Zondervan, A. C. Heath, D. Arveiler, S. J. Bakker, J. Beilby, R. N. Bergman, J. Blangero, P. Bovet, H. Campbell, M. J. Caulfield, G. Cesana, A. Chakravarti, D. I. Chasman, P. S. Chines, F. S. Collins, D. C. Crawford, L. A. Cupples, D. Cusi, J. Danesh, U. de Faire, H. M. den Ruijter, A. F. Dominiczak, R. Erbel, J. Erdmann, J. G. Eriksson, M. Farrall, S. B. Felix, E. Ferrannini, J. Ferrières, I. Ford, N. G. Forouhi, T. Forrester, O. H. Franco, R. T. Gansevoort, P. V. Gejman, C. Gieger, O. Gottesman, V. Gudnason, U. Gyllensten, A. S. Hall, T. B. Harris, A. T. Hattersley, A. A. Hicks, L. A. Hindorff, A. D. Hingorani, A. Hofman, G. Homuth, G. K. Hovingh, S. E. Humphries, S. C. Hunt, E. Hyppönen, T. Illig, K. B. Jacobs, M. R. Jarvelin, K. H. Jöckel, B. Johansen, P. Jousilahti, J. W. Jukema, A. M. Jula, J. Kaprio, J. J. Kastelein, S. M. Keinanen-Kiukaanniemi, L. A. Kiemeney, P. Knekt, J. S. Kooner, C. Kooperberg, P. Kovacs, A. T. Kraja, M. Kumari, J. Kuusisto, T. A. Lakka, C. Langenberg, L. Le Marchand, T. Lehtimäki, V. Lyssenko, S. Männistö, A. Marette, T. C. Matise, C. A. McKenzie, B. McKnight, F. L. Moll, A. D. Morris, A. P. Morris, J. C. Murray, M. Nelis, C. Ohlsson, A. J. Oldehinkel, K. K. Ong, P. A. Madden, G. Pasterkamp, J. F. Peden, A. Peters, D. S. Postma, P. P. Pramstaller, J. F. Price, L. Qi, O. T. Raitakari, T. Rankinen, D. C. Rao, T. K. Rice, P. M. Ridker, J. D. Rioux, M. D. Ritchie, I. Rudan, V. Salomaa, N. J. Samani, J. Saramies, M. A. Sarzynski, H. Schunkert, P. E. Schwarz, P. Sever, A. R. Shuldiner, J. Sinisalo, R. P. Stolk, K. Strauch, A. Tönjes, D. A. Trégouët, A. Tremblay, E. Tremoli, J. Virtamo, M. C. Vohl, U. Völker, G. Waeber, G. Willemsen, J. C. Witteman, M. C. Zillikens, L. S. Adair, P. Amouyel, F. W. Asselbergs, T. L. Assimes, M. Bochud, B. O. Boehm, E. Boerwinkle, S. R. Bornstein, E. P. Bottinger, C. Bouchard, S. Cauchi, J. C. Chambers, S. J. Chanock, R. S. Cooper, P. I. de Bakker, G. Dedoussis, L. Ferrucci, P. W. Franks, P. Froguel, L. C. Groop, C. A. Haiman, A. Hamsten, J. Hui, D. J. Hunter, K. Hveem, R. C. Kaplan, M. Kivimaki, D. Kuh, M. Laakso, Y. Liu, N. G. Martin, W. März, M. Melbye, A. Metspalu, S. Moebus, P. B. Munroe, I. Njølstad, B. A. Oostra, C. N. Palmer, N. L. Pedersen, M. Perola, L. Pérusse, U. Peters, C. Power, T. Quertermous, R. Rauramaa, F. Rivadeneira, T. E. Saaristo, D. Saleheen, N. Sattar, E. E. Schadt, D. Schlessinger, P. E. Slagboom, H. Snieder, T. D. Spector, U. Thorsteinsdottir, M. Stumvoll, J. Tuomilehto, A. G. Uitterlinden, M. Uusitupa, P. van der Harst, M. Walker, H. Wallaschofski, N. J. Wareham, H. Watkins, D. R. Weir, H. E. Wichmann, J. F. Wilson, P. Zanen, I. B. Borecki, P. Deloukas, C. S. Fox, I. M. Heid, J. R. O'Connell, D. P. Strachan, K. Stefansson, C. M. van Duijn, G. R. Abecasis, L. Franke, T. M. Frayling, M. I. McCarthy, P. M. Visscher, A. Scherag, C. J. Willer, M. Boehnke, K. L. Mohlke, C. M. Lindgren, J. S. Beckmann, I. Barroso, K. E. North, E. Ingelsson, J. N. Hirschhorn, R. J. Loos, E. K. Speliotes; LifeLines Cohort Study; ADIPOGen Consortium; AGEN-BMI Working Group; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GLGC; ICBP; MAGIC Investigators; MuTHER Consortium; MIGen Consortium; PAGE Consortium; ReproGen Consortium; GENIE Consortium; International Endogene Consortium, Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015). Medline

47. T. Raj, K. Rothamel, S. Mostafavi, C. Ye, M. N. Lee, J. M. Replogle, T. Feng, M. Lee, N. Asinovski, I. Frohlich, S. Imboywa, A. Von Korff, Y. Okada, N. A. Patsopoulos, S. Davis, C. McCabe, H. I. Paik, G. P. Srivastava, S. Raychaudhuri, D. A. Hafler, D. Koller, A. Regev, N. Hacohen, D. Mathis, C. Benoist, B. E. Stranger, P. L. De Jager, Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* **344**, 519–523 (2014). Medline doi:10.1126/science.1249547

48. B. P. Fairfax, S. Makino, J. Radhakrishnan, K. Plant, S. Leslie, A. Dilthey, P. Ellis, C. Langford, F. O. Vannberg, J. C. Knight, Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat. Genet.* **44**, 502–510 (2012). Medline doi:10.1038/ng.2205

49. P. McCullagh, J. A. Nelder, P. McCullagh, *Generalized Linear Models*, vol. 2 (Chapman and Hall, London, 1989).

50. G. Kichaev, W. Y. Yang, S. Lindstrom, F. Hormozdiari, E. Eskin, A. L. Price, P. Kraft, B. Pasaniuc, Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLOS Genet.* **10**, e1004722 (2014). Medline doi:10.1371/journal.pgen.1004722

51. J. Schäfer, K. Strimmer, A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.* **4** (2005). doi:10.2202/1544-6115.1175

52. T. Kwan, D. Benovoy, C. Dias, S. Gurd, D. Serre, H. Zuzan, T. A. Clark, A. Schweitzer, M. K. Staples, H. Wang, J. E. Blume, T. J. Hudson, R. Sladek, J. Majewski, Heritability of alternative splicing in the human genome. *Genome Res.* **17**, 1210–1218 (2007). Medline doi:10.1101/gr.6281007

53. F. Matesanz *et al*., *Hum. Mol. Genet*. **24**, 5619 (2015).