Genetic variants associated mRNA stability in lung

Jian-Rong Li[1], Mabel Tang[2], Yafang Li[1], Christopher I Amos[1,3,4], Chao Cheng[1,3,4*]

[1] Department of Medicine, Baylor College of Medicine, Houston, Texas, United States of America

[2] Department of BioSciences, Biochemistry and Cell Biology, Rice University, Houston, Texas, United States of America

[3] Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, Texas, United States of America

[4] Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, Texas, United States of America

* Corresponding author

E-mail: Chao.Cheng@bcm.edu (CC)

Abstract

Expression quantitative trait loci (eQTLs) analyses have been widely used to identify genetic variants associated with gene expression levels to understand what molecular mechanisms underlie genetic traits. The resultant eQTLs might affect the expression of associated genes through transcriptional or post-transcriptional regulation. In this study, we attempt to distinguish these two types of regulation by identifying genetic variants associated with mRNA stability of genes (stQTLs). Specifically, we computationally inferred mRNA stability of genes based on RNA-seq data and performed association analysis to identify stQTLs. Using the Genotype-Tissue Expression (GTEx) lung RNA-Seq data, we identified a total of 142,801 stQTLs for 3,942 genes and 186,132 eQTLs for 4,751 genes from 15,122,700 genetic variants for 13,476 genes, respectively. Interesting, our results indicated that stQTLs were enriched in the CDS and 3'UTR regions, while eQTLs are enriched in the CDS, 3'UTR, 5'UTR, and upstream regions. We also found that stQTLs are more likely than eQTLs to overlap with RNA binding protein (RBP) and microRNA (miRNA) binding sites. Our analyses demonstrate that simultaneous identification of stQTLs and eQTLs can provide more mechanistic insight on the association between genetic variants and gene expression levels.

Author Summary

In the past decade, many studies have identified genetic variants associated with gene expression level (eQTLs) in different phenotypes, including tissues and diseases. Gene expression is the result of cooperation between transcriptional regulation, such as transcriptional activity, and post-transcriptional regulation, such as mRNA stability. Here, we present a computational framework that take advantage of recently developed methods to estimate mRNA stability from RNA-Seq, which is widely used to estimate gene expression, and then to identify genetic variants associated with mRNA

stability (stQTLs) in lung tissue. Compared to eQTLs, we found that genetic variants that affects mRNA stability are more significantly located in the CDS and 3'UTR regions, which are known to interact with RNA-binding proteins (RBPs) or microRNAs to regulate stability. In addition, stQTLs are significantly more likely to overlap the binding sites of RBPs. We show that the six RBPs that most significantly bind to stQTLs are all known to regulate mRNA stability. This pipeline of simultaneously identifying eQTLs and stQTLs using only RNA-Seq data can provide higher resolution than traditional eQTLs study to better understand the molecular mechanisms of genetic variants on the regulation of gene expression.

Introduction

Quantitative trait loci (QTLs) are genomic loci that explain variation of a quantitative trait [1]. The most well investigated QTLs are eQTLs, which are associated with the expression level of gene transcripts [2]. Assuming different regulatory mechanisms, eQTLs proximal to and distant from the transcription start site (TSS) of genes are called *cis*-eQTLs (< 1Mb) and *trans*-eQTLs (> 5Mb), respectively [3]. By combining high-throughput gene expression data and genetic phenotype information, eQTLs can be identified systematically using a GWAS (genome-wide association study) approach [4]. It has been shown that genetic variants (single nucleotide polymorphisms) associated with complex traits, including human diseases, are more likely to be eQTLs [5]. The genetic variants located in *cis*-regulatory elements (CREs), in particular, can influence the expression of targeted genes. In fact, eQTLs are associated with many classes of CREs that are enriched in promoters, enhancers, insulators, transcription factor (TF) binding sites, and DNase hypersensitive sites (DHSs) [6–10].

Gene expression level is regulated at both the transcriptional and post-transcriptional levels. At the transcriptional level, TFs regulate the transcription rate of genes by interacting with their promoters and enhancers [11,12]. TF binding and histone modification signals in the TSS proximal regions account for over 50% of variation of gene expression [13–15]. Genetic variants with functional impacts on TF binding motifs or promoter/enhancer accessibility are also expected to have effects on the transcription rate of related genes [16,17]. On the other hand, at the post-transcriptional level, the stability of mRNAs is under intensive regulation by microRNAs and RNA-binding proteins (RBPs) [18,19]. Genetic variants can also affect mRNA stability by interacting with microRNAs or RBPs. For example, the variant T of rs907091 located in the 3'UTR of *IZKF3* confers a miR-326 binding site, which leads to decreased mRNA stability and down-regulation of the gene; however, this is not seen with the variant C [20]. Additionally, some intronic genetic variants might also affect gene expression by interacting with splicing factors or other types of RBPs [21]. Therefore, it is often difficult to precisely interpret the eQTLs identified from high-throughput analysis. Namely, for many eQTLs, it is difficult to determine whether they influence gene expression through affecting transcriptional rate or mRNA stability. This problem is further complicated by linkage disequilibrium (LD) between neighboring genetic variants. Although high-throughput technologies that measure mRNA decay rates have been developed [22–24], there are no QTL studies that identify genetic variants associated with mRNA stability due to the lack of matched stability and genotype data.

In many eQTL studies, gene expression was determined by RNA sequencing (RNA-Seq) experiments [25–27]. Despite the protocol being designed to generate cDNA fragments from mature mRNAs, there was also a significant proportion of reads captured from intronic sequences in RNA-seq data [25]. Several studies proposed that the intronic reads of RNA-Seq were related to nascent transcription and transcriptional activity [27–30]. Based on this concept, computational methods have been developed to calculate mRNA stability based on RNA-seq data [27,31,32]. Gaidatzis

et al [27] proposed a method called exon-intron split analysis (EISA) to discriminate transcriptional and post-transcriptional regulation of gene expression. Given the RNA-seq data in two experiment conditions, EISA calculates changes in reads mapped to exons (Δexon) and introns (Δintron) for each gene. It was shown that Δexon-Δintron was significantly correlated with experimentally measured mRNA stability changes between ESCs and terminal neurons. The EISA method was then further improved and then implemented in a software package, REMBRANDTS, to measure the stability of mRNAs more correctly [32].

Motivated by these methods, we developed a framework to simultaneously identify genetic variants associated with gene expression (eQTL) or mRNA stability (stQTL). We applied this framework to the lung tissue RNA-Seq data produced by the Genotype-Tissue Expression (GTEx) project [33]. For this data, we estimated the mRNA stability using REMBRANDTS and gene expression, and then performed association analysis to 15,122,700 genetic variants for 13,476 genes. We then identified a total of 186,132 eQTLs for 4,751 genes and 142,801 stQTLs for 3,942 genes. From our analysis, we found that both the stQTLs and eQTLs are enriched in the 3'UTR and CDS regions, while eQTLs are also enriched in the 5'UTR and upstream region of TSS. Compared to eQTLs, stQTLs more frequently overlapped with the binding sites of RBPs and miRNAs. To explore the role of stQTLs in mRNA stability, we took a few examples to investigate the effect of genetic variants on the binding of RBPs or TFs. Together, this study suggested that the simultaneous identification of stQTLs and eQTLs can provide a useful method to better understand the molecular mechanisms underlying genetic variants.

Result

Overview of this study

Fig 1 shows the rationale underlying this study. During gene expression, a gene is transcribed into a pre-mRNA, after which the introns are removed while the exons are connected into the mature mRNA. The mature mRNA is under post-transcriptional regulation by miRNAs and other mechanisms. As shown, genetic variants can not only regulate mRNA splicing but also regulate gene expression-related traits by affecting transcription rate or mRNA stability (stability QTL, denoted as stQTL hereafter). From RNA-seq data, we are able to determine the reads mapped to exonic regions to obtain gene expression levels. The mRNA stability can also be calculated by combining the reads aligned to exonic and intronic regions using the REMBRANDTS [32] algorithm. Through the eQTL analysis, genetic variants associated with gene expression are identified to obtain eQTLs. In fact, eQTLs are a mixture of QTLs that affect transcription and stQTLs, as gene expression is controlled by both transcription rate and mRNA stability. Performing an association analysis of gene expression or stability on genetic variation can identify eQTLs and stQTLs, respectively. Simultaneous identification of eQTLs and stQTLs can provide a higher resolution to understand how genetic variants affect gene expression, as well as the information to infer that a genetic variant regulates gene expression by affecting transcription activity or RNA stability. As a proof-of-concept, in this study we applied this framework to GTEx data to simultaneously investigate the eQTLs and stQTLs in lung tissue.

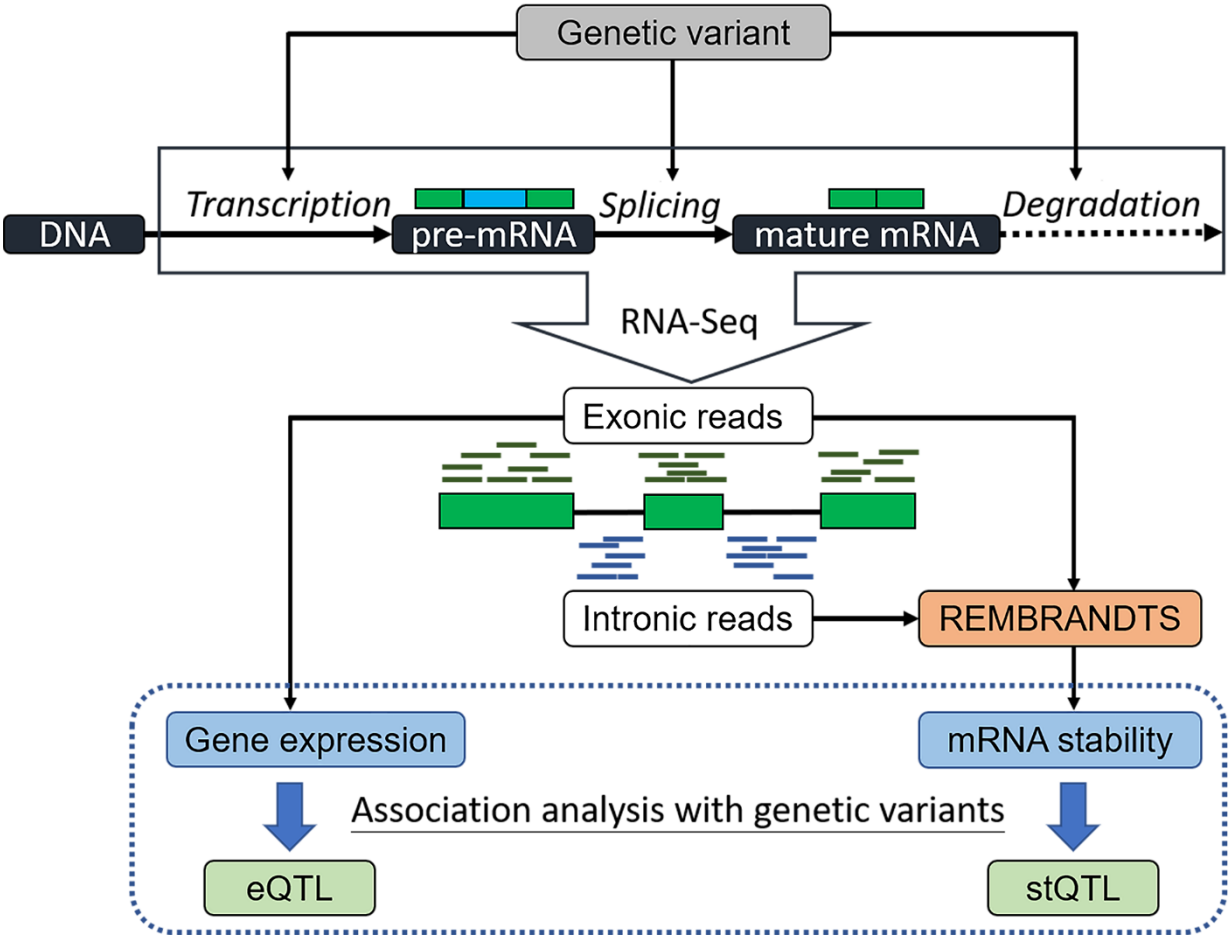


Fig 1. The workflow for identification of stQTL and eQTL using RNA-Seq. A genetic variant may regulate gene expression by affecting transcription, splicing, or stability at different stages of the life cycle of an mRNA. Both gene expression and mRNA stability can be estimated from RNA-Seq. Therefore, both expression quantitative trait loci (eQTLs) and stability quantitative trait loci (stQTLs) can be identified with genetic variations using the association analysis. By comparing the stQTL and eQTL, it is possible to distinguish the regulatory mechanisms underlying an eQTL.

Expression QTLs and Stability QTLs of human lung tissue

To identify and explore stQTLs and eQTLs, we processed the raw RNA-seq data for lung tissues generated by the GTEx project. After performing quality trimming, alignment, and replicate merging from the same donors, we obtained

the expression profiles of genes for a total of 289 subjects with matched genetic variation data. With REMBRANDTS, for each subject, we calculated the relative mRNA stability for 13,476 genes with intronic regions and constitutive exons. For QTL identification, we performed the association analysis on 15,122,700 genetic variants located within 100Kb upstream of TSS and 100 Kb downstream of TTS for 13,476 genes using gene expression or mRNA stability as the traits. We identified a total of 142,801 stQTLs and 186,132 eQTLs at the significance level of FDR < 5%. The numbers of QTLs were summarized in Table 1 according to the location of genetic variants on each QTL's corresponding genes.

Ideally, we would expect that all stQTLs are also eQTLs since a genetic variant that regulates RNA stability should also affect gene expression. However, in practice, identification of different QTL types is complicated by multiple factors, including differential statistical power and LD between genetic variants. Nevertheless, we still observed that there is a very high proportion (70,105) of overlap between stQTLs and eQTLs (Fig 2A). We also found that 49% of stQTLs were also eQTLs (Fig 2B), suggesting that nearly half of stQTLs do also significantly affect gene expression. On the contrary, only 37% of eQTLs were also stQTLs. This indicated that although a considerable part of eQTLs were derived from genetic variants that significantly affect stability, more of them were regulated by genetic variants that affect other factors related to gene expression.

Table 1. The summary of the stQTLs and eQTLs identification in GTEx lung tissue samples. The location indicates the genomic position of the genetic variant in its corresponding gene in a QTL. The percentage was calculated from the number of QTLs at each location divided by the total number of QTLs.

Location	Number of stQTL (%)	Number of eQTL (%)
Upstream	50,109 (35.09%)	73,647 (39.57%)
5'UTR	3,877 (2.71%)	5,802 (3.12%)
CDS	3,949 (2.77%)	4,404 (2.37%)
Intron	32,675 (22.88%)	37,232 (20.00%)
3'UTR	2,506 (1.75%)	2,411 (1.30%)
Downstream	49,685 (34.79%)	62,636 (33.65%)
Total	142,801 (100.00%)	186,132 (100.00%)

By investigating stQTL and eQTL together, it is possible to determine the regulatory mechanisms underlying an eQTL. For example, genetic variant rs3167757 is significantly associated with the *HMGN1* expression level (eQTL, FDR = 5e-18) with CC>CT/TT (Fig 2C). As shown, this genetic variant is also associated with *HMGN1*'s mRNA stability (stQTL, FDR=3.7e-30). This result indicated that rs3167757 might regulate the expression level of *HMGN1* by affecting its mRNA stability. Indeed, *HMGN1*-rs3167757 has also been reported as an eQTL in lymphoblastoid cell lines (LCLs) [34,35]. The rs3167757 is located at the 3'UTR of the *HMGN1* gene and overlaps with binding sites of 20 different RBPs [36]. According to the analysis using RBPmap [37] (S1 Table), while the variant C of rs3167757 confers a motif for eight RBPs (CUG-BP, HNRNPF, MBNL1, SFPQ, TRA2B, HNRNPL, SRSF3, and YBX2), the variant T disrupts the binding motifs of five of the RBPs (HNRNPF, MBNL1, SFPQ, TRA2B, and YBX2). Notably, among them, HNRNPF [38,39], MBNL1 [40,41], and YBX2 [42] are known to contribute to mRNA stabilization. This is consistent with the observation that genotype CC is associated with higher stability of *HMGN1* mRNA than CT and TT. As another example, genetic variant rs34873612 is significantly associated with *DDX11* expression level (eQTL, FDR = 2e-60) but not with *DDX11* mRNA stability (FDR > 0.1) with GG>GA/AA (Fig 2D). This result suggested that rs34873612 might regulate the expression level of *DDX11* by affecting the transcription rate rather than its mRNA stability. According to the PROMO prediction [43], the rs34873612 is located at the 5'UTR of the *DDX11* gene and overlaps with the binding site of three TFs: GR-alpha, GATA2, and GATA3. While the variant G contributes to the binding motifs of these TFs,

the variant A disrupts the binding motif of GATA3, which potentially contributes to the decreased *DDX11* expression seen in the GA and AA genotypes (Fig 2D). mRNA stability only contributes partially to gene expression level; consistently, many genetic variants are found to be stQTLs but not eQTLs. For example, rs1062976 is significantly associated with the mRNA stability of *SCYL3* (stQTL, FDR = 5e-07) but not its expression level (not an eQTL, FDR > 0.1) with CC>CT/TT (Fig 2E). Overall, our results indicated that simultaneous identification of stQTLs and eQTLs can provide us with more detailed biological insights on the regulatory effects of genetic variants on a large scale.

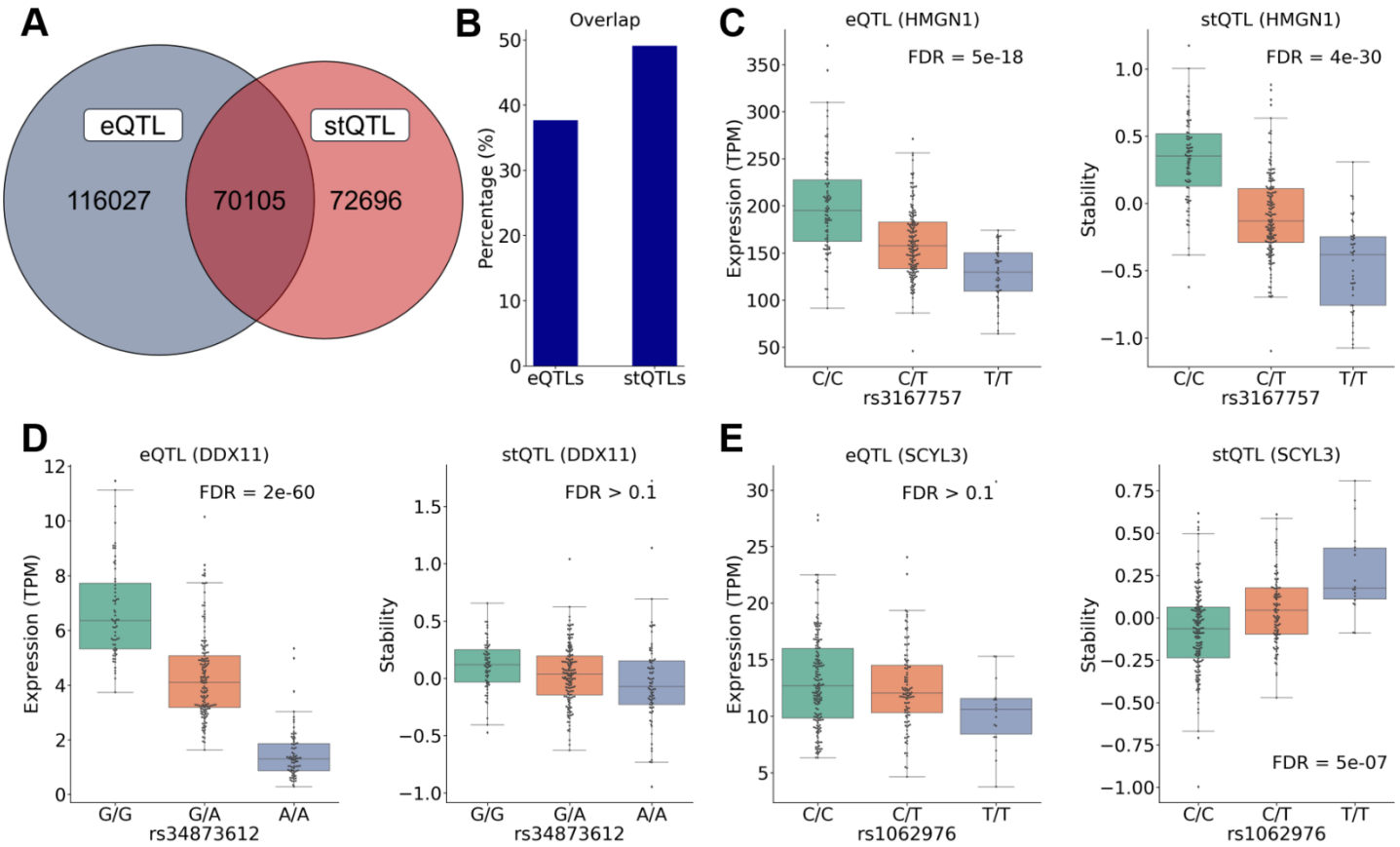


Fig 2. The simultaneous identification of stQTLs and eQTLs using GTEx lung tissue samples shows highly overlapped QTLs and provides additional information for investigating regulatory effects of genetic variants. (A) The Venn diagram between eQTLs and stQTLs shows that 70,105 genetic variants are both eQTLs and stQTLs. (B) The bar plot shows the percentage of the overlapped QTLs in stQTLs and eQTLs, respectively. (C) *HMGN1*-rs3167757

is an eQTL and a stQTL. The expression level and RNA stability of *HMGN1* will decrease as rs3167757 changes with CC>CT/TT. The rs3167757 is located on the binding sites of several RBPs in the 3'UTR region of *HMGN1*. (D) *DDX11*-rs34873612 is an eQTL but not a stQTL. The expression level, but not RNA stability, of the *DDX11* will decrease as rs34873612 changes with GG>GA/AA. The rs34873612 overlaps the binding sites of several TFs in 5'UTR of *DDX11*. (E) *SCYL3*-rs1062976 is an stQTL but not an eQTL. The RNA stability of the *SCYL3* will be affected by rs1062976, which is located in the 3'UTR region of *SCYL3*. The variant T of rs1062976 disrupts the binding motif of PTBP1 (destabilizer) but confers the binding motif of YBX1 (stabilizer).

Distributions of eQTLs and stQTLs across genic regions

stQTLs are associated with mRNA stability while eQTLs are associated with gene expression by affecting either mRNA stability or gene transcription. Therefore, we expect that their distributions in genes would be different. To examine this, we looked at the distribution of eQTLs and stQTLs in the DNA regions surrounding TSS and TTS of genes. We found that eQTLs are more likely to be located upstream of TSS of their corresponding genes while stQTLs tend to be located downstream of TSSs (Fig 3A). On the other hand, stQTLs are more likely to be located in the region from TTS to its 10Kb upstream than eQTLs. Both stQTLs and eQTLs are more likely to be located in the upstream region of TTS rather than the genes' downstream regions (Fig 3B).

Subsequently, we divided genomic regions associated with genes into upstream, 5'UTR, CDS, Intron, 3'UTR, and downstream regions and then examined the distributions of eQTLs and stQTLs in these regions. Using the distributions of all genetic variants as the background, we calculated the enrichment ratio of stQTLs and eQTLs by using a

194 hypergeometric test [44]. As shown in Fig 3C, stQTLs are enriched by 2.89-fold in the CDS (P < 2e-308) and by 2.25-
195 fold in 3'UTR (P = 2e-152). This result is consistent with the fact that genetic variants located in these regions might
196 have functional impacts on mRNA stability by affecting RNA secondary/tertiary structure or RBP/microRNA binding.
197 stQTLs are also slightly enriched in intron regions (ER = 1.19 and P = 2e-150). In contrast, eQTLs are enriched in the
198 CDS (ER = 2.22, P = 4e-274), upstream (ER = 1.10, P = 3e-65), 5'UTR (ER = 1.37, P = 5e-76), and 3'UTR (ER = 1.30,
199 P = 6e-20, Fig 3D), respectively. The enriched eQTLs in these regions may be due to the fact that gene expression can
200 be determined not only by transcriptional activity (genetic variants in upstream, 5'UTR, or CDS regions) but also by
201 RNA stability (genetic variants in CDS or 3'UTR regions). We compared the enrichment ratios of stQTLs and eQTLs
202 and found that stQTLs are more likely to be located in the CDS, intron, and 3'UTR regions, while eQTLs are enriched
203 in the upstream and 5'UTR regions (Fig 3E).

204

205 It should be noted that the resolution of QTL analysis is affected by linkage disequilibrium (LD) between
206 neighboring genetic variants. Based on the genotype data for lung samples used in this study, we performed LD analysis
207 and observed that many eQTL/stQTL loci were in high LD ($r^2 > 0.9$) with each other (S1 Fig). To best exclude the
208 influence of LD, we determined all LD blocks ($r^2 > 0.9$) and within each block selected the most significant genetic
209 variant as the representative stQTL/eQTL. Following that, we re-evaluated the distribution of stQTLs and eQTLs. After
210 LD filtering, we found that the stQTLs are more enriched in CDS (ER = 4.93 and P < 2e-308), 3'UTR (ER = 2.64 and
211 P = 1e-164), and slightly enriched at 5'UTR (ER = 1.14 and P = 9e-07). Of note, we no longer observed the enrichment
212 of stQTLs in intron regions (Fig 3F). In contrast, the distribution of eQTLs in each gene region is similar to that before
213 LD filtering (Fig 3G). When the distributions of stQTLs and eQTLs were directly compared, eQTLs were more likely

than stQTLs to locate in the 5'UTR and upstream of genes (Fig 3H), while stQTLs are more likely to locate in the CDS and 3'UTR. No obvious difference was observed between stQTLs and eQTLs in the intron and downstream regions.

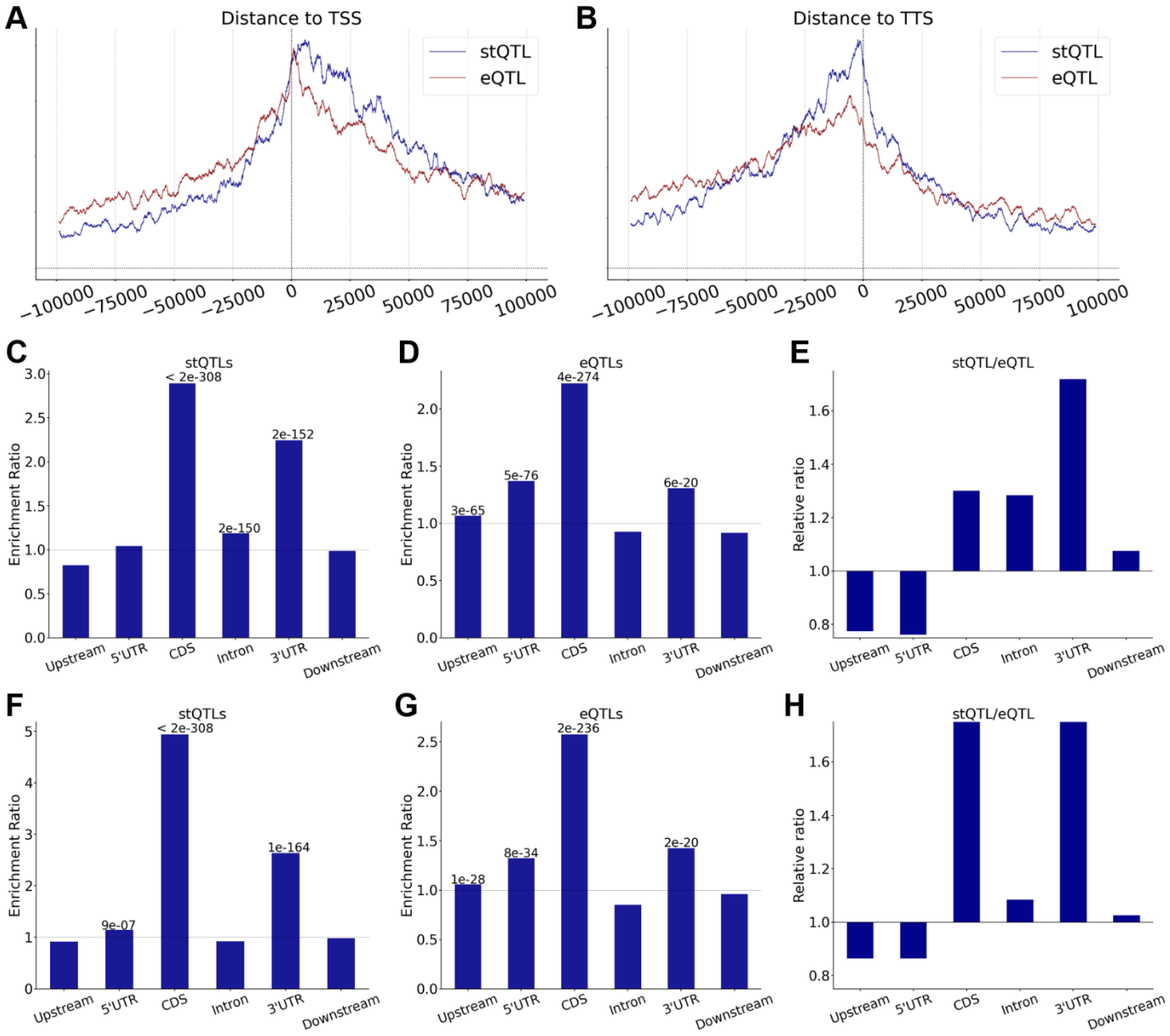


Fig 3. There are biased distributions in different genic regions of eQTLs and stQTLs. (A) The distribution from the enrichment ratio of stQTLs and eQTLs to TSS. Plot with the bin size of 2000 bp and the sliding window of 50 bp. (B) The distribution from the enrichment ratio of stQTLs and eQTLs to TTS. Plot with the bin size of 2000 bp and the sliding window of 50 bp. (C) The enrichment ratio in different genomic locations of stQTLs before LD filtering. The upstream indicates the region of 100 Kb upstream from TSS, and the downstream indicates the region of 100 Kb downstream from TTS. (The following figs are the same) (D) The enrichment ratio in different genomic locations of

223 eQTLs before LD filtering. (E) The relative proportion of enrichment ratio in different genomic locations between
224 stQTLs and eQTLs before LD filtering. (F) The enrichment ratio in different genomic locations of stQTLs after LD
225 filtering. (G) The enrichment ratio in different genomic locations of eQTLs after LD filtering. (H) The relative
226 proportion of enrichment ratio in different genomic locations between stQTLs and eQTLs after LD filtering.

227

228 **stQTLs are significantly enriched in RBP binding sites**

229 Having shown the enrichment of stQTLs in the 3'UTR and CDS regions, we then examined whether stQTLs tend
230 to locate in the binding sites of RBPs or miRNAs, many of which are known to be involved in post-transcriptional
231 regulation of mRNAs. To this end, we investigated the binding sites of RBPs and miRNAs provided by Postar2 [36]
232 and TargetScan [45], respectively, to annotate the stQTLs identified in our analysis. Our results indicated that stQTLs
233 ($P = 3e\text{-}18$, Fisher's exact test) but not eQTLs ($P > 0.1$, Fisher's exact test) are enriched in RBP binding sites. In fact,
234 we found that 26.81% (2,770/10,332) of stQTLs overlap with the binding sites of at least one RBP, which is significantly
235 higher ($P = 7e\text{-}17$, Fisher's exact test) than 22.10% (2,788/12,617) for eQTLs (Fig 4A). In addition, we have also
236 examined the overlap with miRNA binding sites and observed a higher proportion of stQTLs (0.19%, 20/10,332) than
237 eQTLs (0.15%, 19/12,617) in the miRNA binding sites, although no statistical significance was detected due to very
238 small genomic regions covered by miRNA binding sites (Fig 4B).

239

240 Then we performed Fisher's exact test to identify RBPs whose binding sites were enriched for stQTLs (S2 Table)
241 or eQTLs (S3 Table). We identified a total of six significant RBPs ($P < 1e\text{-}04$) including SND1, YTHDC1, DDX3X,
242 ATXN2, RPS3, and UPF1 (as shown in Table 2). Interestingly, SND1 [46–48], DDX3X [49,50], ATXN2 [51,52], and

RPS3 [53] were known to stabilize their bound mRNAs, while UPF1 is the key factor of nonsense-mediated mRNA decay pathway [54–56]. Moreover, YTHDC1 is a well-known m^{6}A (N^6-Methyladenosine) reader [57], which has been found to regulate mRNA splicing [58,59], alternative polyadenylation [59], and stability [60,61] through recognizing m^{6}A. Similarly, we identified four RBPs whose binding sites were significantly enriched for eQTLs (P < 1e-04, Fig 4D and Table 3), among which the two most significant RBPs, DDX3X and SND1, were also enriched for stQTLs. NCBP3 can regulate gene expression by forming a cap binding complex that binds to the 5'cap of pre-mRNA to promote splicing, 3'-end processing, and mRNA exporting [62–64], and AGGF1 was found to repress the expression of pro-inflammatory molecules [65]. These results indicate that stQTLs or eQTLs located in the binding sites of RBPs in lung tissue are indeed likely to have significant regulation on gene stability or expression by affecting the binding of the RBPs.

Table 2. The RBPs whose binding sites were enriched for stQTLs. Six RBPs significantly overlap (Log2 Enrichment-ratio > 0.3 and p-value < 1e-04, Fisher's exact test) with stQTLs in mature mRNAs in lung. ER: Enrichment ratio.

RBPs	stQTL	non-stQTL	ER	p-value	FDR
SND1	64	1,504	2.17	4E-08	5E-06
YTHDC1	100	2,850	1.80	1E-07	7E-06
DDX3X	226	8,348	1.39	3E-06	1E-04
ATXN2	436	17,911	1.25	7E-06	3E-04
RPS3	91	2,923	1.59	3E-05	0.001
UPF1	198	7,533	1.35	5E-05	0.002

Table 3. The RBPs whose binding sites were enriched for eQTLs. Four RBPs significantly overlap (Log2 Enrichment ratio > 0.3 and p-value < 1e-04, Fisher's exact test) with eQTLs in mature mRNAs in lung. ER: Enrichment ratio.

RBPs	eQTL	non-eQTL	ER	p-value	FDR
DDX3X	250	8,331	1.69	5E-14	1E-11
SND1	58	1,509	2.15	2E-07	2E-05

NCBP2	60	1,819	1.84	1E-05	6E-04
AGGF1	30	689	2.43	2E-05	6E-04

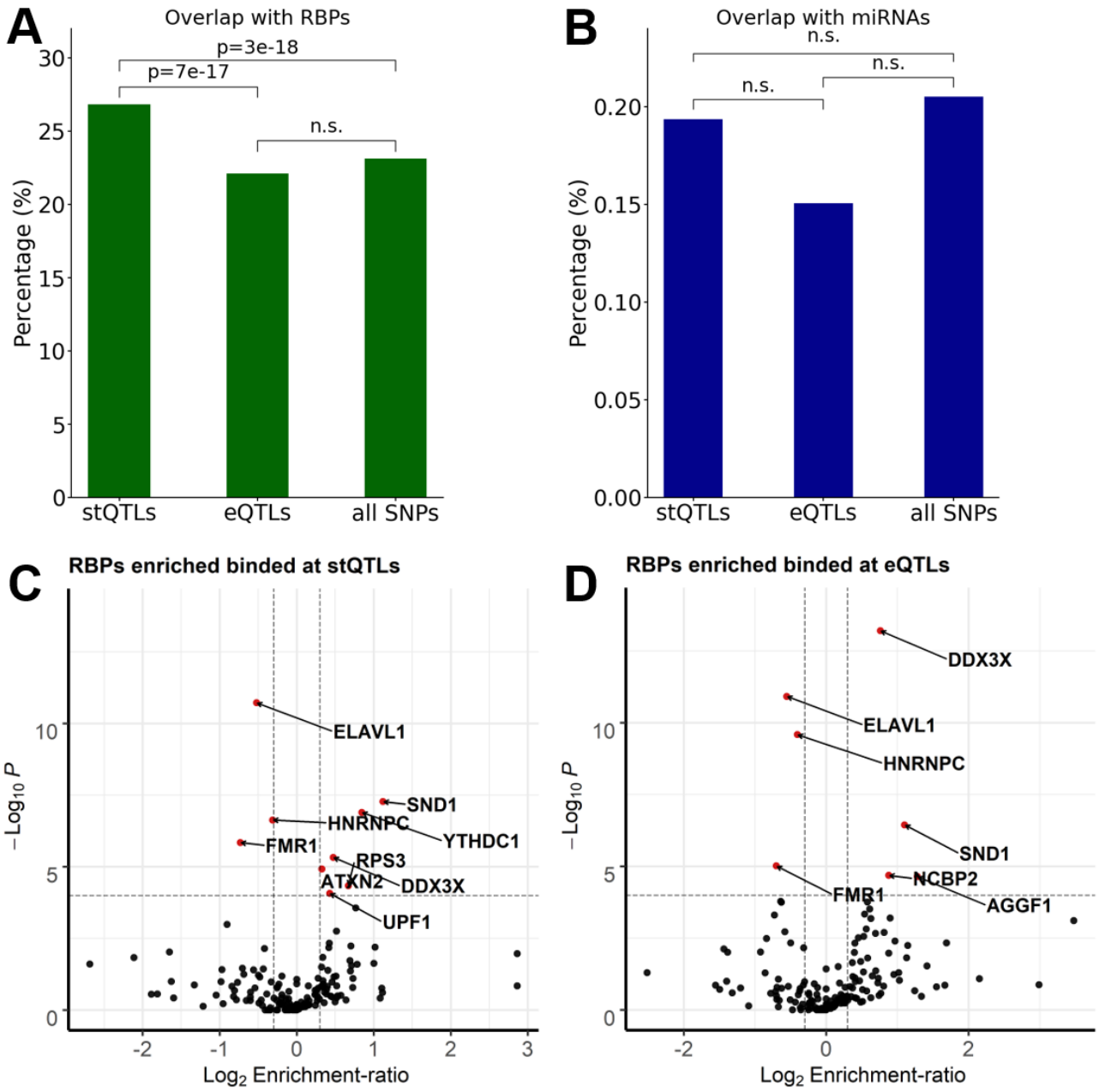


Fig 4. Enrichment of stQTLs and eQTLs in the binding sites of RBPs and miRNAs. (A) Proportion of overlap between stQTLs, eQTLs, and all genetic variants and RBP binding sites in mature mRNA. The statistical significance was calculated using Fisher's exact test. The n.s. indicates not significant. (B) Proportion of overlap between stQTLs, eQTLs, and all genetic variants and miRNA binding sites in mature mRNA. The statistical significance was calculated using Fisher's exact test. The n.s. indicates not significant. (C) The volcano plot shows that nine RBPs (red points)

whose binding sites were significantly ($-Log_{10}$ p-value > 4, 2-sides Fisher's exact test) enriched (six RBPs, Log_2 Enrichment-ratio > 0.3) or depleted (three RBPs, Log_2 Enrichment-ratio < -0.3) in stQTLs. (D) The volcano plot shows that seven RBPs (red points) whose binding sites were significantly ($-Log_{10}$ p-value > 4, 2-sides Fisher's exact test) enriched (four RBPs, Log_2 Enrichment-ratio > 0.3) or depleted (three RBPs, Log_2 Enrichment-ratio < -0.3) in eQTLs.

Gender-specific stQTLs

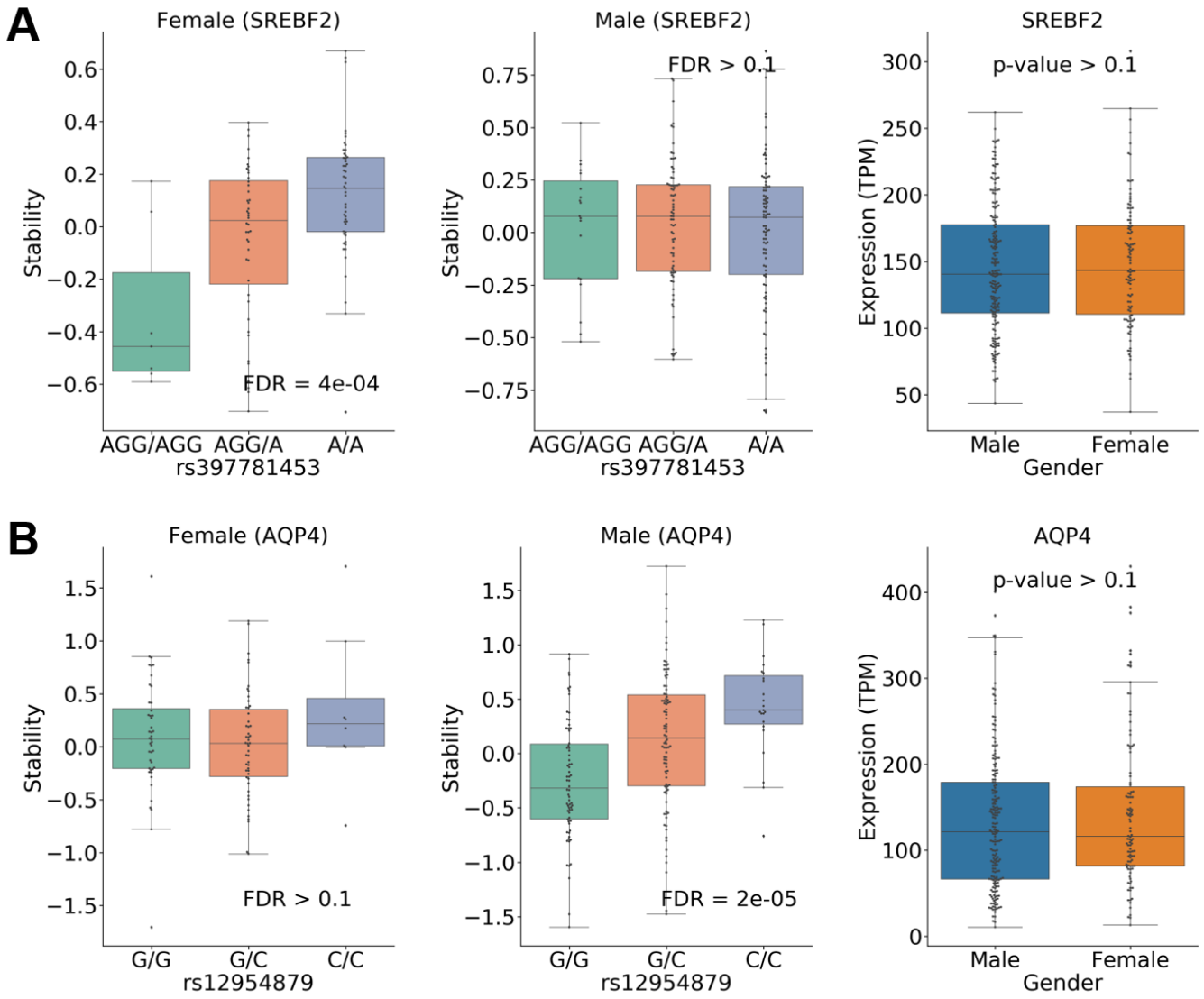
We then examined whether some genetic variants were associated with mRNA stability in a gender-specific manner and denoted them as gender-specific stQTLs. We divided 289 samples into 187 males and 102 females, and then performed association analysis with covariates to implement the gender-specific stQTL classification. If a gene is specifically expressed in males or females, then an stQTL/eQTL association can only be performed in the corresponding gender. Therefore, we focused our analysis on 13,116 genes that are not differentially expressed (FDR > 0.05, t-test) between both genders and then investigated a total of 14,987,511 genetic variants located from 100Kb upstream to 100Kb downstream of a gene. Out of these gene/genetic variants, we identified 71,694 stQTLs in males and 22,841 stQTLs in females (FDR < 0.05), as well as 117,065 eQTLs in males and 48,516 eQTLs in females (FDR < 0.05), respectively. Then we defined male-specific QTLs as those that are significant in males (FDR < 0.05) but not significant in females (P > 0.1), and similarly for female-specific QTLs. In total, we identified 18,893 male-specific and 2,879 female-specific stQTLs, and 32,716 male-specific and 7,484 female-specific eQTLs. After excluding intersection with gender-specific eQTLs, we finally identified 14,683 male-specific and 2,279 female-specific stQTLs. As an example, the association between genetic variant rs397781453 and the RNA stability of *SREBP2* is female-specific (Fig 5A). As shown, we detected a significant association in females with FDR = 4e-04 but not in males (FDR ≥ 0.1). On the other hand, the association between *AQP4* and genetic variant rs12954879 is male-specific (Fig 5B). The RNA stability of

286 *AQP4* is significantly associated (FDR = 2e-05) with genetic variant rs12954879 in males but not in females (FDR >

287 0.1). Of note, both *SREBP2* or *AQP4* have similar expression levels between males and females (the right panel of Fig

288 5A and 5B).

289



290

291 **Fig 5. Gender-specific stQTLs identification.** (A) The association between genetic variant rs397781453 and the RNA

292 stability of *SREBP2* is female-specific (stQTL, FDR = 4e-04), but this pattern does not occur in males (FDR ≥ 0.1). In

293 the right panel, the expression of *SREBP2* is not significantly different (P ≥ 0.1) between male and female samples. (B)

294 The association between genetic variant rs12954879 and the RNA stability of *AQP4* is male-specific (stQTL, FDR =

2e-05), but this pattern does not occur in females (FDR ≥ 0.1). In the right panel, the expression of *AQP4* is not significantly different (P ≥ 0.1) between male and female samples.

Discussion

In this study, we systematically identified stQTLs that are associated with mRNA stability in lung tissues and compared them with eQTLs using GTEx RNA-Seq data. Out of the 151,227,000 genetic variants within 100 Kb upstream from TSS to 100 Kb downstream from TTS of 13,476 corresponding genes, we identified a total of 186,132 eQTLs and 142,801 stQTLs. We found that stQTLs are mainly enriched in the 3'UTR and CDS regions, while eQTLs are enriched in the CDS, 5'UTR, 3'UTR, and upstream regions (Fig 3F and 3G). We also found that stQTLs are significantly located in the binding sites of RBPs (Fig 4A). Moreover, the different stQTL/eQTL variants will indeed change the motif to affect the bound RBPs, which then regulate RNA stability or gene expression (Fig 2C-2E). Our results suggest that stQTLs may significantly affect RNA stability, mostly because they are located in the 3'UTR [66,67] and CDS [68,69] regions that most often interact with other molecules. These results are consistent with previous studies, which have found that the codon usage and changes on CDS could affect its stability [68,70–72], and the sequence in 3'UTR affected mRNA stability since it includes binding sites of RBPs [73,74]. On the contrary, eQTLs are a group of complex mechanisms and may regulate expression levels by affecting stability [75,76], transcriptional activity [77–79], and even the addition of 5'cap or polyA tail [62]. Therefore, while eQTLs largely resemble stQTLs but are less enriched at the 3' UTR and CDS regions, eQTLs are also enriched in the 5'UTR and upstream regions where the enhancers and promoters that regulate transcriptional activity are located [80–82].

Although we have shown that identifying stQTLs provides additional insights, it is worthwhile to note that determining regulatory mechanisms is largely limited by the LD between proximal genetic variants. Due to LD, it is sometimes difficult to identify the exact genetic variants that regulate gene expression. For the same reason, it is also hard to clearly distinguish genetic variants controlling gene transcription from those controlling mRNA stabilities solely based on association analysis. This analysis is expected to improve with further consideration of the location and function impact of genetic variants. In addition, the power of stQTL analysis is also limited by the computational methods used for mRNA stability inference. Although previous studies have demonstrated that the EISA algorithm [27] and its improved REMBRANDTS package [32,83,84] used in this study achieve fairly high accuracy for mRNA stability evaluation, the accuracy of inferred mRNA stability may vary significantly between different genes. First, the differential expressed long noncoding RNAs (lncRNAs) [85,86] or perturbated factors involved in intron degradation [27,87] could cause the changes of difference intronic read counts (Δintron) to affect the stability estimate. Adding the annotation of non-coding RNAs in the alignment of RNA-Seq may improve the accuracy of the mRNA stability inference. Secondly, it is difficult to accurately calculate stability for the genes with low aligned read counts because the stability inference is based on the relative change of exonic and intronic reads (Δexon–Δintron) [32]. Of note, the REMBRANDTS provides a stringency parameter to filter genes with low read counts. In our study, we set the stringency to 0.01 to include 13,429 genes for comprehensive stQTLs identification since the stability of only 2,593 genes can be calculated when the stringency is 0.9. Interestingly, we found that 41.88% (634/1,514) of stQTLs with stringency ≥ 0.9 overlap with the RBP binding sites, which is significantly higher (P = 6e-44, Fisher's exact test) than 23.69% (1,546/6,527) of stQTLs with stringency ≤ 0.5. This result demonstrates that the stability of genes with low read counts is less accurate and can lead to potential false positives stQTLs. Finally, it should be of note that the mRNA stability calculated from RNA-Seq using REMBRANDTS is not an actual absolute value, but a differential mRNA stability

relative to the average of all samples for a given gene [32,87,88]. Therefore, we suggest that it is necessary to keep these limitations in mind before evaluating mRNA stability using RNA-Seq data, and to record the stringency of the gene as a reference for the reliability of stQTLs identification.

The identification of stQTLs provides a higher resolution to better understand the molecular mechanism of genetic variants regulating gene expression, and accurate estimation of mRNA stability is very important for the identification of stQTLs. Although some high-throughput technologies, such as BRIC-Seq [24,89], have been developed to determine the decay rate of mRNA, these methods are often limited to only being used in cell culture conditions [32] and there are not enough samples available for QTLs research. Therefore, despite the limitations of computational approaches, such as SnapShot-Seq [29], EISA, and REMBRANDTS, our analysis for mRNA stability inference using RNA-seq by REMBRANDTS shows that the stQTL genic distribution and overlap with RBP binding sites is indeed consistent with biological theories. Furthermore, computer algorithms based on RNA-Seq are still under continuous development. For example, INSPEcT [87] was recently designed to calculate RNA kinetic rates based on time course RNA-seq data, or to estimate stability by calculating the difference between premature and mature RNA expression [90]. Going forward, the stQTLs which are identified with more accurate mRNA stability profiles estimation may further our understanding of how genetic variants regulate gene expression.

In conclusion, we present a large-scale identification for eQTLs and stQTLs using RNA-Seq data in lung tissues. Our results demonstrate that there are differential genic distributions as well as interactions with RBPs or TFs between eQTLs and stQTLs. We show in this study that simultaneous identification of eQTLs and stQTLs provides more

biological insights for better understanding the regulatory mechanisms underlying genetic variants associated with gene expression.

Methods

Collection of datasets

The genotype data and RNA-Seq data of lung tissues produced by the Genotype-Tissue Expression project [33] (release 7) were used in this study. RNA-Seq SRA files and genetic variants data were downloaded from NCBI dbGaP [91] (Study Accession: phs000424.v7.p2). Subject phenotypes were collected from the GTExPortal (https://www.gtexportal.org/home/datasets). The data contains a total of 318 RNA-Seq runs and 404 genetic variant samples from 289 different subjects. We calculated the average dosages for the genetic variants data from the same subjects to represent the genotype data of subjects. For RNA-Seq analysis, the human reference genome and annotation were collected from Ensembl [92], version GRCh37.87. For RNA stability analysis, the annotation GTF files recording the coordinates of intronic and constitutive exonic segments of genes was generated using the shell script modified from the first step of the https://github.com/csglab/CRIES [32].

Processing of RNA-Seq data

The 318 RNA-Seq SRAs were dumped into FASTQ files using SRA Toolkit (http://ncbi.github.io/sra-tools). The read quality and retained adapters were checked with FastQC [93]; then, the adapters and low-quality reads were trimmed using Trimmomatic v0.39 [94]. The alignment was performed using HISAT2 v2.1.0 [95] with default parameters. Alignment files from the same subjects were merged. Read counts of introns or exons were extracted

376 separately using the HTSeq-count script of the HTSeq v0.12.4 [96] with the parameter --stranded=no. The RNA stability
377 profiles for 289 subjects were estimated using the REMBRANDTS [32] with the parameter of linear method and stability
378 stringency of 0.01. The TPM (transcripts per million) [97] was used as the expression unit to measure the expression
379 level of 13,476 genes which have stability profiles.

380

381 **Identifying QTLs by associating genetic variants with traits derived from RNA-seq data**

382 For covariates construction, the plink [98] (version 1.90 beta, https://www.cog-genomics.org/plink/1.9/) was
383 performed with the parameter: --indep-pairwise 200 100 0.2 to prune a subset of genetic variants. The PCA analysis
384 was performed after removing strand ambiguous variants (AT/CG) and genetic variants located in the MHC region. The
385 first three PCs were selected as covariates with gender and age. For *cis*-QTL identification, genetic variants that were
386 located within 100Kb upstream from the TSS (transcription start site) to 100Kb downstream from the TTS (transcription
387 termination site) of Ensembl annotated genes (GRCh37.87) were selected. The expression profile was then converted
388 with log10(TPM*100 + 1), and the linear association analysis was performed between the dosage of each genetic variant
389 and the value of expression or stability of each gene. The Benjamini-Hochberg Procedure [99] was implemented to
390 calculate the false discovery rate (FDR), and the genetic variant with the association of FDR less than 0.05 was regarded
391 as a QTL.

392

393 **Enrichment analysis of QTLs in different genic regions**

394 To determine whether eQTLs and stQTLs were evenly distributed in different genic regions, we performed the
395 following analyses. Here we use the stQTL as an example. First, we counted the number of all genetic variants in the

TSS-upstream (from TSS to 100Kb upstream), 5'UTR, CDS, 3'UTR, intronic and TSS-downstream (from TTS to 100Kb downstream) regions. Let us use N^k to denote the number of all genetic variants in the kth region (k=1, …, 6). We then counted the number of stQTLs in each of these regions and used Q^k to denote the number in the kth region. Third, to determine whether stQTLs are enriched in region k, we consider the following numbers: Q^k, $Q^{(-k)}$, N^k-Q^k, and $N^{(-k)}$-$Q^{(-k)}$, which (-k) indicate all regions other than k. Fisher's exact test was then used to calculate the significance of enrichment. The enrichment was performed separately for stQTLs and eQTLs.

Estimation of linkage disequilibrium effect

We performed the plink [98] to all genetic variants of 289 subjects with the parameter: (--r2 --ld-window 50 --ld-window-kb 100000 --ld-window-r2 0.9) to estimate the linkage disequilibrium (LD) between each genetic variant. We then constructed LD blocks, in which r^2 of LD between each genetic variant must be greater than 0.9. To reduce the influence of LD on the gene distribution of QTLs, we selected the QTLs with the lowest FDR of the association analysis in each LD block and then performed the enrichment analysis in different genic regions as the previous section.

Identification of QTLs located at binding sites of miRNAs or RBPs

stQTLs and eQTLs were mapped to the binding sites of RNA binding proteins (RBPs) and microRNAs (miRNAs). RBP binding site data were retrieved from Postar2 [36] (http://lulab.life.tsinghua.edu.cn/postar/). miRNA binding site data were downloaded from targetScanHuman [45] (http://www.targetscan.org/vert_72/). Both databases are based on the human genome reference version GRCh38. To match with our analysis, we performed LiftOver [100] (https://genome-store.ucsc.edu/) to convert genome coordinates into GRCh37. To evaluate QTLs that locate on the

binding sites of RBPs or miRNAs, we selected stQTLs or eQTLs on mature mRNA to align the binding sites data, and then used the Fisher's exact test [101] to identify RBPs whose binding sites were enriched located.

Acknowledgment

All RNA-Seq data and genetic variants data used for the analyses described in this manuscript were obtained from dbGaP (accession number: phs000424.v7.p2). The phenotype data was downloaded from GTEx Portal (https://www.gtexportal.org/home/). This study is supported by the Cancer Prevention Research Institute of Texas (CPRIT) (RR180061 to CC) and the National Cancer Institute of the National Institutes of Health (1R21CA227996 to CC, U19CA203654 to CA). CC is a CPRIT Scholar in Cancer Research.

Reference

1. Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. Philos Trans R Soc B Biol Sci [Internet]. 2013 Jun 19 [cited 2020 Dec 10];368(1620). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3682727/

2. Rockman MV, Kruglyak L. Genetics of global gene expression. Nat Rev Genet. 2006 Nov;7(11):862–72.

3. Lowe WL, Reddy TE. Genomic approaches for understanding the genetics of complex disease. Genome Res. 2015 Oct;25(10):1432–41.

4. Porcu E, Rüeger S, Lepik K, Santoni FA, Reymond A, Kutalik Z. Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. Nat Commun. 2019 Jul 24;10(1):3300.

5. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be

eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. 2010 Apr 1;6(4):e1000888.

6. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell. 2007 Mar 23;128(6):1231–45.

7. Gaffney DJ, Veyrieras J-B, Degner JF, Pique-Regi R, Pai AA, Crawford GE, et al. Dissecting the regulatory architecture of gene expression QTLs. Genome Biol. 2012 Jan 31;13(1):R7.

8. Degner JF, Pai AA, Pique-Regi R, Veyrieras J-B, Gaffney DJ, Pickrell JK, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. Nature. 2012 Feb 5;482(7385):390–4.

9. Brown CD, Mangravite LM, Engelhardt BE. Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. PLoS Genet. 2013;9(8):e1003649.

10. Mitchelmore J, Grinberg NF, Wallace C, Spivakov M. Functional effects of variation in transcription factor binding highlight long-range gene regulation by epromoters. Nucleic Acids Res. 2020 Apr 6;48(6):2866–79.

11. Maniatis T, Goodbourn S, Fischer JA. Regulation of inducible and tissue-specific gene expression. Science. 1987 Jun 5;236(4806):1237–45.

12. Akbari OS, Bae E, Johnsen H, Villaluz A, Wong D, Drewell RA. A novel promoter-tethering element regulates enhancer-driven gene expression at the bithorax complex in the Drosophila embryo. Dev Camb Engl. 2008 Jan;135(1):123–31.

13. Cheng C, Gerstein M. Modeling the relative relationship of transcription factor binding and histone modifications to gene expression levels in mouse embryonic stem cells. Nucleic Acids Res. 2012 Jan 1;40(2):553–68.

14. Dong X, Greven MC, Kundaje A, Djebali S, Brown JB, Cheng C, et al. Modeling gene expression using chromatin features in various cellular contexts. Genome Biol. 2012 Jun 13;13(9):R53.

15. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K-K, Cheng C, et al. Architecture of the human regulatory

network derived from ENCODE data. Nature. 2012 Sep;489(7414):91–100.

16. Johnston AD, Simões-Pires CA, Thompson TV, Suzuki M, Greally JM. Functional genetic variants can mediate their regulatory effects through alteration of transcription factor binding. Nat Commun. 2019 Aug 2;10(1):3472.

17. Nishizaki SS, Ng N, Dong S, Porter RS, Morterud C, Williams C, et al. Predicting the effects of SNPs on transcription factor binding affinity. Bioinformatics. 2020 Jan 15;36(2):364–72.

18. Palanisamy V, Jakymiw A, Van Tubergen EA, D'Silva NJ, Kirkwood KL. Control of Cytokine mRNA Expression by RNA-binding Proteins and microRNAs. J Dent Res. 2012 Jul;91(7):651–8.

19. Michlewski G, Cáceres JF. Post-transcriptional control of miRNA biogenesis. RNA. 2019 Jan;25(1):1–16.

20. Richardson K, Lai C-Q, Parnell LD, Lee Y-C, Ordovas JM. A genome-wide survey for SNPs altering microRNA seed sites identifies functional candidates in GWAS. BMC Genomics. 2011 Oct 13;12:504.

21. Cooper DN. Functional intronic polymorphisms: Buried treasure awaiting discovery within our genes. Hum Genomics. 2010 Jun 1;4(5):284–8.

22. Wang Y, Liu CL, Storey JD, Tibshirani RJ, Herschlag D, Brown PO. Precision and functional specificity in mRNA decay. Proc Natl Acad Sci. 2002 Apr 30;99(9):5860–5.

23. Munchel SE, Shultzaberger RK, Takizawa N, Weis K. Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. Mol Biol Cell. 2011 Jun 16;22(15):2787–95.

24. Tani H, Mizutani R, Salam KA, Tano K, Ijiri K, Wakamatsu A, et al. Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. Genome Res. 2012 May;22(5):947–56.

25. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, et al. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science. 2008 Aug 15;321(5891):956–60.

478 26. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian
479 transcriptomes by RNA-Seq. Nat Methods. 2008 Jul;5(7):621–8.
480 27. Gaidatzis D, Burger L, Florescu M, Stadler MB. Analysis of intronic and exonic reads in RNA-seq data
481 characterizes transcriptional and post-transcriptional regulation. Nat Biotechnol. 2015 Jul;33(7):722–9.
482 28. Ameur A, Zaghlool A, Halvardson J, Wetterbom A, Gyllensten U, Cavelier L, et al. Total RNA sequencing
483 reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nat Struct Mol Biol.
484 2011 Dec;18(12):1435–40.
485 29. Gray JM, Harmin DA, Boswell SA, Cloonan N, Mullen TE, Ling JJ, et al. SnapShot-Seq: A Method for
486 Extracting Genome-Wide, In Vivo mRNA Dynamics from a Single Total RNA Sample. PLOS ONE. 2014 Feb
487 26;9(2):e89673.
488 30. Hendriks G-J, Gaidatzis D, Aeschimann F, Großhans H. Extensive Oscillatory Gene Expression during C.
489 elegans Larval Development. Mol Cell. 2014 Feb 6;53(3):380–92.
490 31. Gosline SJC, Gurtan AM, JnBaptiste CK, Bosson A, Milani P, Dalin S, et al. Elucidating microRNA regulatory
491 networks using transcriptional, post-transcriptional and histone modification measurements. Cell Rep. 2016 Jan
492 12;14(2):310–9.
493 32. Alkallas R, Fish L, Goodarzi H, Najafabadi HS. Inference of RNA decay rate from transcriptional profiling
494 highlights the regulatory programs of Alzheimer's disease. Nat Commun. 2017 Oct 13;8(1):909.
495 33. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, et al. A Novel Approach to High-Quality
496 Postmortem Tissue Procurement: The GTEx Project. Biopreservation Biobanking. 2015 Oct 1;13(5):311–9.
497 34. Yu C-H, Pal LR, Moult J. Consensus Genome-Wide Expression Quantitative Trait Loci and Their Relationship
498 with Human Complex Trait Disease. OMICS J Integr Biol. 2016 Jul 1;20(7):400–14.

35. Liang L, Morar N, Dixon AL, Lathrop GM, Abecasis GR, Moffatt MF, et al. A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. Genome Res. 2013 Apr;23(4):716–26.

36. Zhu Y, Xu G, Yang YT, Xu Z, Chen X, Shi B, et al. POSTAR2: deciphering the post-transcriptional regulatory logics. Nucleic Acids Res. 2019 08;47(D1):D203–11.

37. Paz I, Kosti I, Ares M, Cline M, Mandel-Gutfreund Y. RBPmap: a web server for mapping binding sites of RNA-binding proteins. Nucleic Acids Res. 2014 Jul 1;42(Web Server issue):W361–7.

38. Chaudhury A, Chander P, Howe PH. Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: Focus on hnRNP E1's multifunctional regulatory roles. RNA. 2010 Aug;16(8):1449–62.

39. Geuens T, Bouhy D, Timmerman V. The hnRNP family: insights into their role in health and disease. Hum Genet. 2016;135:851–67.

40. Davis J, Salomonis N, Ghearing N, Lin S-CJ, Kwong JQ, Mohan A, et al. MBNL1-mediated regulation of differentiation RNAs promotes myofibroblast transformation and the fibrotic response. Nat Commun. 2015 Dec 16;6(1):10084.

41. Fish L, Pencheva N, Goodarzi H, Tran H, Yoshida M, Tavazoie SF. Muscleblind-like 1 suppresses breast cancer metastatic colonization and stabilizes metastasis suppressor transcripts. Genes Dev. 2016 Feb 15;30(4):386–98.

42. Xu D, Xu S, Kyaw AMM, Lim YC, Chia SY, Siang DTC, et al. RNA Binding Protein Ybx2 Regulates RNA Stability During Cold-Induced Brown Fat Activation. Diabetes. 2017 Dec 1;66(12):2987–3000.

43. Farré D, Roset R, Huerta M, Adsuara JE, Roselló L, Albà MM, et al. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. Nucleic Acids Res. 2003 Jul 1;31(13):3651–3.

44. Rivals I, Personnaz L, Taing L, Potier M-C. Enrichment or depletion of a GO category within a class of genes: which test? Bioinformatics. 2007 Feb 15;23(4):401–7.

45. Agarwal V, Bell GW, Nam J-W, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Izaurralde E, editor. eLife. 2015 Aug 12;4:e05005.

46. JARIWALA N, RAJASEKARAN D, SRIVASTAVA J, GREDLER R, AKIEL MA, ROBERTSON CL, et al. Role of the staphylococcal nuclease and tudor domain containing 1 in oncogenesis (Review). Int J Oncol. 2014 Nov 18;46(2):465–73.

47. Jariwala N, Mendoza RG, Garcia D, Lai Z, Subler MA, Windle JJ, et al. Posttranscriptional Inhibition of Protein Tyrosine Phosphatase Nonreceptor Type 23 by Staphylococcal Nuclease and Tudor Domain Containing 1: Implications for Hepatocellular Carcinoma. Hepatol Commun. 2019 Jul 15;3(9):1258–70.

48. Santhekadur PK, Akiel M, Emdad L, Gredler R, Srivastava J, Rajasekaran D, et al. Staphylococcal nuclease domain containing-1 (SND1) promotes migration and invasion via angiotensin II type 1 receptor (AT1R) and TGFβ signaling. FEBS Open Bio. 2014 Jan 1;4:353–61.

49. Somma D, Mastrovito P, Grieco M, Lavorgna A, Pignalosa A, Formisano L, et al. CIKS/DDX3X interaction controls the stability of the Zc3h12a mRNA induced by IL-17. J Immunol Baltim Md 1950. 2015 Apr 1;194(7):3286–94.

50. Song H, Ji X. The mechanism of RNA duplex recognition and unwinding by DEAD-box helicase DDX3X. Nat Commun. 2019 Jul 12;10(1):3085.

51. Yokoshi M, Li Q, Yamamoto M, Okada H, Suzuki Y, Kawahara Y. Direct binding of Ataxin-2 to distinct elements in 3' UTRs promotes mRNA stability and protein expression. Mol Cell. 2014 Jul 17;55(2):186–98.

52. Ostrowski LA, Hall AC, Mekhail K. Ataxin-2: From RNA Control to Human Health and Disease. Genes [Internet]. 2017 Jun 5 [cited 2021 Feb 3];8(6). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5485521/

53. Zhao L, Cao J, Hu K, Wang P, Li G, He X, et al. RNA-binding protein RPS3 contributes to hepatocarcinogenesis by post-transcriptionally up-regulating SIRT1. Nucleic Acids Res. 2019 Feb 28;47(4):2011–28.

54. Plank T, Wilkinson MF. A hidden talent for a RNA decay factor: UPF1 directs protein decay. BioEssays News Rev Mol Cell Dev Biol [Internet]. 2018 Jan [cited 2021 Feb 3];40(1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5843485/

55. Kim YK, Maquat LE. UPFront and center in RNA decay: UPF1 in nonsense-mediated mRNA decay and beyond. RNA. 2019 Apr;25(4):407–22.

56. Fiorini F, Bagchi D, Le Hir H, Croquette V. Human Upf1 is a highly processive RNA helicase and translocase with RNP remodelling activities. Nat Commun. 2015 Jul 3;6(1):7581.

57. Roundtree IA, Luo G-Z, Zhang Z, Wang X, Zhou T, Cui Y, et al. YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs. Proudfoot NJ, editor. eLife. 2017 Oct 6;6:e31311.

58. Xiao W, Adhikari S, Dahal U, Chen Y-S, Hao Y-J, Sun B-F, et al. Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. Mol Cell. 2016 Feb 18;61(4):507–19.

59. Kasowitz SD, Ma J, Anderson SJ, Leu NA, Xu Y, Gregory BD, et al. Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. PLOS Genet. 2018 May 25;14(5):e1007412.

60. Shima H, Matsumoto M, Ishigami Y, Ebina M, Muto A, Sato Y, et al. S-Adenosylmethionine Synthesis Is Regulated by Selective N6-Adenosine Methylation and mRNA Degradation Involving METTL16 and YTHDC1. Cell Rep. 2017 Dec 19;21(12):3354–63.

61. Lee Y, Choe J, Park OH, Kim YK. Molecular Mechanisms Driving mRNA Degradation by m6A Modification. Trends Genet. 2020 Mar 1;36(3):177–88.

562 62. Galloway A, Cowling VH. mRNA cap regulation in mammalian cell function and fate. Biochim Biophys Acta
563 BBA - Gene Regul Mech. 2019 Mar 1;1862(3):270–9.

564 63. Gebhardt A, Habjan M, Benda C, Meiler A, Haas DA, Hein MY, et al. mRNA export through an additional cap-
565 binding complex consisting of NCBP1 and NCBP3. Nat Commun. 2015 Sep 18;6(1):8192.

566 64. Moore MJ, Proudfoot NJ. Pre-mRNA Processing Reaches Back toTranscription and Ahead to Translation. Cell.
567 2009 Feb 20;136(4):688–700.

568 65. Hu F-Y, Wu C, Li Y, Xu K, Wang W-J, Cao H, et al. AGGF1 is a novel anti-inflammatory factor associated with
569 TNF-α-induced endothelial activation. Cell Signal. 2013 Aug 1;25(8):1645–53.

570 66. Koh WS, Porter JR, Batchelor E. Tuning of mRNA stability through altering 3′-UTR sequences generates distinct
571 output expression in a synthetic circuit driven by p53 oscillations. Sci Rep. 2019 Apr 12;9(1):5976.

572 67. Misquitta CM, Iyer VR, Werstiuk ES, Grover AK. The role of 3'-untranslated region (3'-UTR) mediated mRNA
573 stability in cardiovascular pathophysiology. Mol Cell Biochem. 2001 Aug;224(1–2):53–67.

574 68. Bazzini AA, Del Viso F, Moreno-Mateos MA, Johnstone TG, Vejnar CE, Qin Y, et al. Codon identity regulates
575 mRNA stability and translation efficiency during the maternal-to-zygotic transition. EMBO J. 2016 Oct
576 4;35(19):2087–103.

577 69. Narula A, Ellis J, Taliaferro JM, Rissland OS. Coding regions affect mRNA stability in human cells. RNA. 2019
578 Dec;25(12):1751–64.

579 70. Mishima Y, Tomari Y. Codon Usage and 3' UTR Length Determine Maternal mRNA Stability in Zebrafish. Mol
580 Cell. 2016 Mar 17;61(6):874–85.

581 71. Cheng J, Maier KC, Avsec Ž, Rus P, Gagneur J. Cis-regulatory elements explain most of the mRNA stability
582 variation across genes in yeast. RNA. 2017 Nov;23(11):1648–59.

583 72. Presnyak V, Alhusaini N, Chen Y-H, Martin S, Morris N, Kline N, et al. Codon optimality is a major determinant
584 of mRNA stability. Cell. 2015 Mar 12;160(6):1111–24.
585 73. Hasan A, Cotobal C, Duncan CDS, Mata J. Systematic analysis of the role of RNA-binding proteins in the
586 regulation of RNA stability. PLoS Genet. 2014 Nov;10(11):e1004684.
587 74. Shalgi R, Lapidot M, Shamir R, Pilpel Y. A catalog of stability-associated sequence elements in 3' UTRs of
588 yeast mRNAs. Genome Biol. 2005;6(10):R86.
589 75. Pai AA, Cain CE, Mizrahi-Man O, Leon SD, Lewellen N, Veyrieras J-B, et al. The Contribution of RNA Decay
590 Quantitative Trait Loci to Inter-Individual Variation in Steady-State Gene Expression Levels. PLOS Genet. 2012
591 Oct 11;8(10):e1003000.
592 76. Nickless A, Bailis JM, You Z. Control of gene expression through the nonsense-mediated RNA decay pathway.
593 Cell Biosci. 2017 May 19;7(1):26.
594 77. Garieri M, Delaneau O, Santoni F, Fish RJ, Mull D, Carninci P, et al. The effect of genetic variation on promoter
595 usage and enhancer activity. Nat Commun. 2017 Nov 7;8(1):1358.
596 78. Sun W, Yu T, Li K-C. Detection of eQTL modules mediated by activity levels of transcription factors.
597 Bioinformatics. 2007 Sep 1;23(17):2290–7.
598 79. Ranjan A, Budke JM, Rowland SD, Chitwood DH, Kumar R, Carriedo L, et al. eQTL Regulating Transcript
599 Levels Associated with Diverse Biological Processes in Tomato1[OPEN]. Plant Physiol. 2016 Sep;172(1):328–
600 40.
601 80. Chen H, Levo M, Barinov L, Fujioka M, Jaynes JB, Gregor T. Dynamic interplay between enhancer-promoter
602 topology and gene activity. Nat Genet. 2018 Sep;50(9):1296–303.
603 81. Lenardo M, Pierce JW, Baltimore D. Protein-binding sites in Ig gene enhancers determine transcriptional activity

and inducibility. Science. 1987 Jun 19;236(4808):1573–7.

82. Andersson R, Sandelin A. Determinants of enhancer and promoter activities of regulatory elements. Nat Rev Genet. 2020 Feb;21(2):71–87.

83. Yu J, Navickas A, Asgharian H, Culbertson B, Fish L, Garcia K, et al. RBMS1 suppresses colon cancer metastasis through targeted stabilization of its mRNA regulon. Cancer Discov. 2020 Sep;10(9):1410–23.

84. Baird TD, Cheng KC-C, Chen Y-C, Buehler E, Martin SE, Inglese J, et al. ICE1 promotes the link between splicing and nonsense-mediated mRNA decay. Green R, editor. eLife. 2018 Mar 12;7:e33178.

85. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. Genome Res. 2012 Jan 9;22(9):1775–89.

86. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. Nature. 2012 Sep;489(7414):101–8.

87. Furlan M, Galeota E, Gaudio ND, Dassi E, Caselle M, Pretis S de, et al. Genome-wide dynamics of RNA synthesis, processing, and degradation without RNA metabolic labeling. Genome Res. 2020 Jan 10;30(10):1492–507.

88. Furlan M, de Pretis S, Pelizzola M. Dynamics of transcriptional and post-transcriptional regulation. Brief Bioinform [Internet]. 2020 Dec 22 [cited 2021 Mar 15];(bbaa389). Available from: https://doi.org/10.1093/bib/bbaa389

89. Imamachi N, Tani H, Mizutani R, Imamura K, Irie T, Suzuki Y, et al. BRIC-seq: A genome-wide approach for determining RNA stability in mammalian cells. Methods. 2014 May 1;67(1):55–63.

90. de Pretis S, Furlan M, Pelizzola M. INSPEcT-GUI Reveals the Impact of the Kinetic Rates of RNA Synthesis,

Processing, and Degradation, on Premature and Mature RNA Species. Front Genet [Internet]. 2020 Jul 17 [cited 2021 Mar 15];11. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7379887/

91. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, et al. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet. 2007 Oct;39(10):1181–6.

92. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. Nucleic Acids Res. 2020 Jan 8;48(D1):D682–8.

93. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Internet]. 2010. Available from: https://qubeshub.org/resources/fastqc

94. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.

95. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019 Aug;37(8):907–15.

96. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinforma Oxf Engl. 2015 Jan 15;31(2):166–9.

97. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011 Aug 4;12(1):323.

98. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience [Internet]. 2015 Dec 1 [cited 2020 Dec 7];4(1). Available from: https://academic.oup.com/gigascience/article/4/1/s13742-015-0047-8/2707533

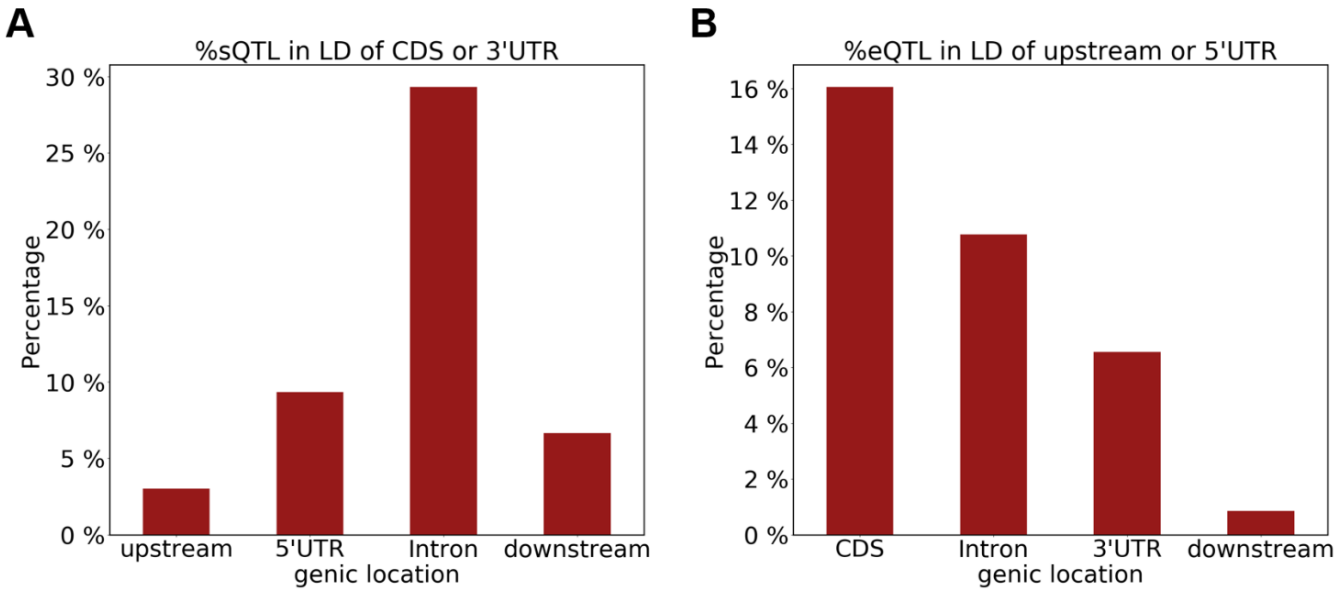
99. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple

Testing. J R Stat Soc Ser B Methodol. 1995;57(1):289–300.

100. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, et al. The UCSC Genome Browser Database. Nucleic Acids Res. 2003 Jan 1;31(1):51–4.

101. Fisher RA. On the Interpretation of χ2 from Contingency Tables, and the Calculation of P. J R Stat Soc. 1922;85(1):87–94.

Supporting information



S1 Fig. There is a considerable part of stQTLs/eQTLs have strong LD ($r^2 > 0.9$) with stQTLs/eQTLs in other regions. (A) The proportion of stQTLs in intron, 5'UTR, downstream, and upstream regions that have a strong LD with stQTLs in CDS or 3'UTR. (B) The proportion of eQTLs in CDS, intron, 3'UTR, and downstream regions that have a strong LD with eQTLs in 5'UTR or upstream regions.

S1 Table. RBP motifs of rs3167757 variant C.

S2 Table. Fisher's exact test between stQTL and not stQTL on mature mRNA (two-sides).

S3 Table. Fisher's exact test between eQTL and not eQTL on mature mRNA (two-sides).

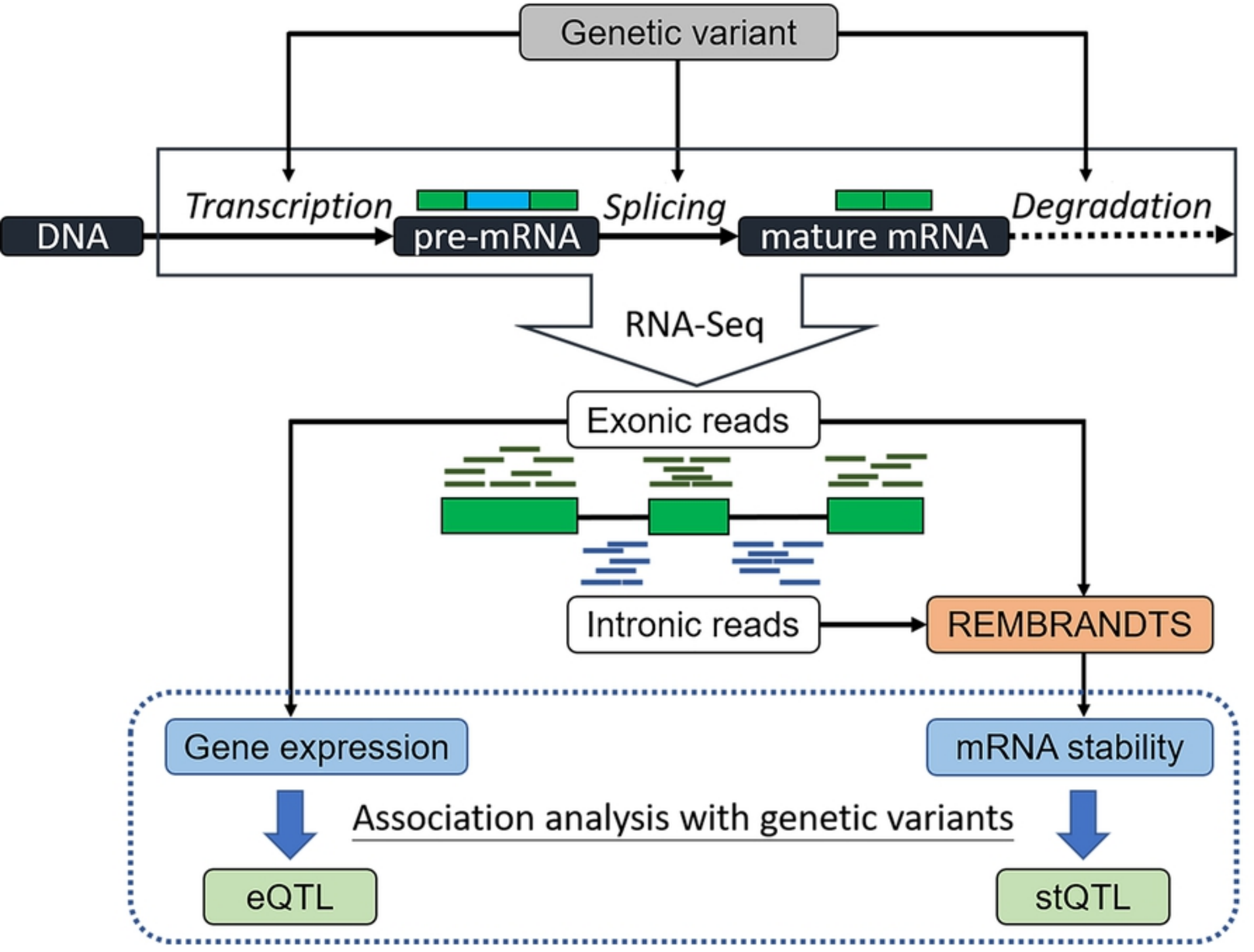

Genetic variant
Transcription
Splicing
Degradation
DNA
pre-mRNA
mature mRNA
RNA-Seq
Exonic reads
Intronic reads
REMBRANDTS
Gene expression
mRNA stability
Association analysis with genetic variants
eQTL
stQTL

Figure 1

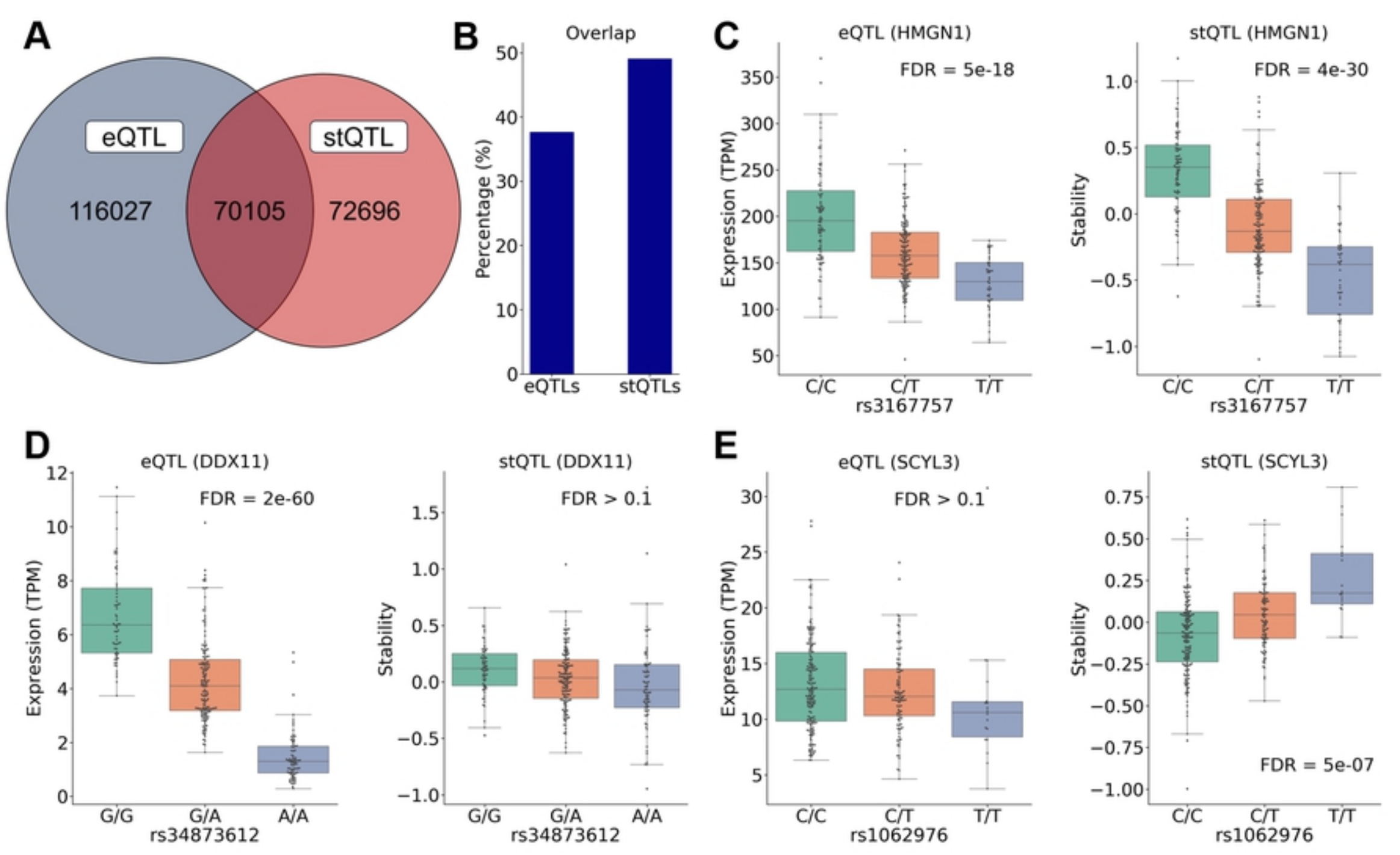

A
eQTL
stQTL
116027
70105
72696
B
Overlap
Percentage (%)
eQTLs
stQTLs
C
eQTL (HMGN1)
FDR = 5e-18
Expression (TPM)
C/C
C/T
T/T
rs3167757
stQTL (HMGN1)
FDR = 4e-30
Stability
D
eQTL (DDX11)
FDR = 2e-60
G/G
G/A
A/A
rs34873612
stQTL (DDX11)
FDR > 0.1
E
eQTL (SCYL3)
FDR > 0.1
rs1062976
stQTL (SCYL3)
FDR = 5e-07

Figure 2

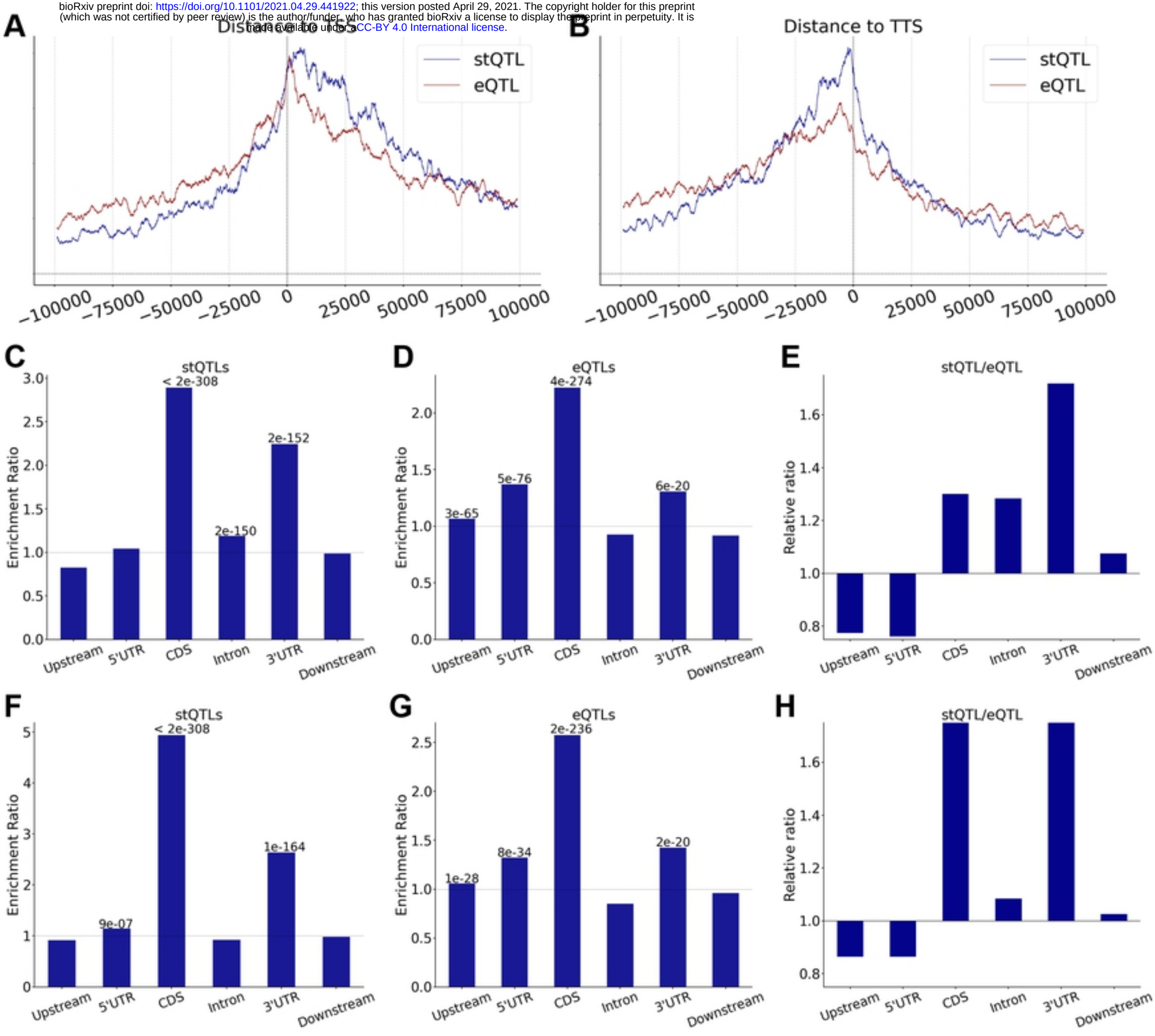


Figure 3

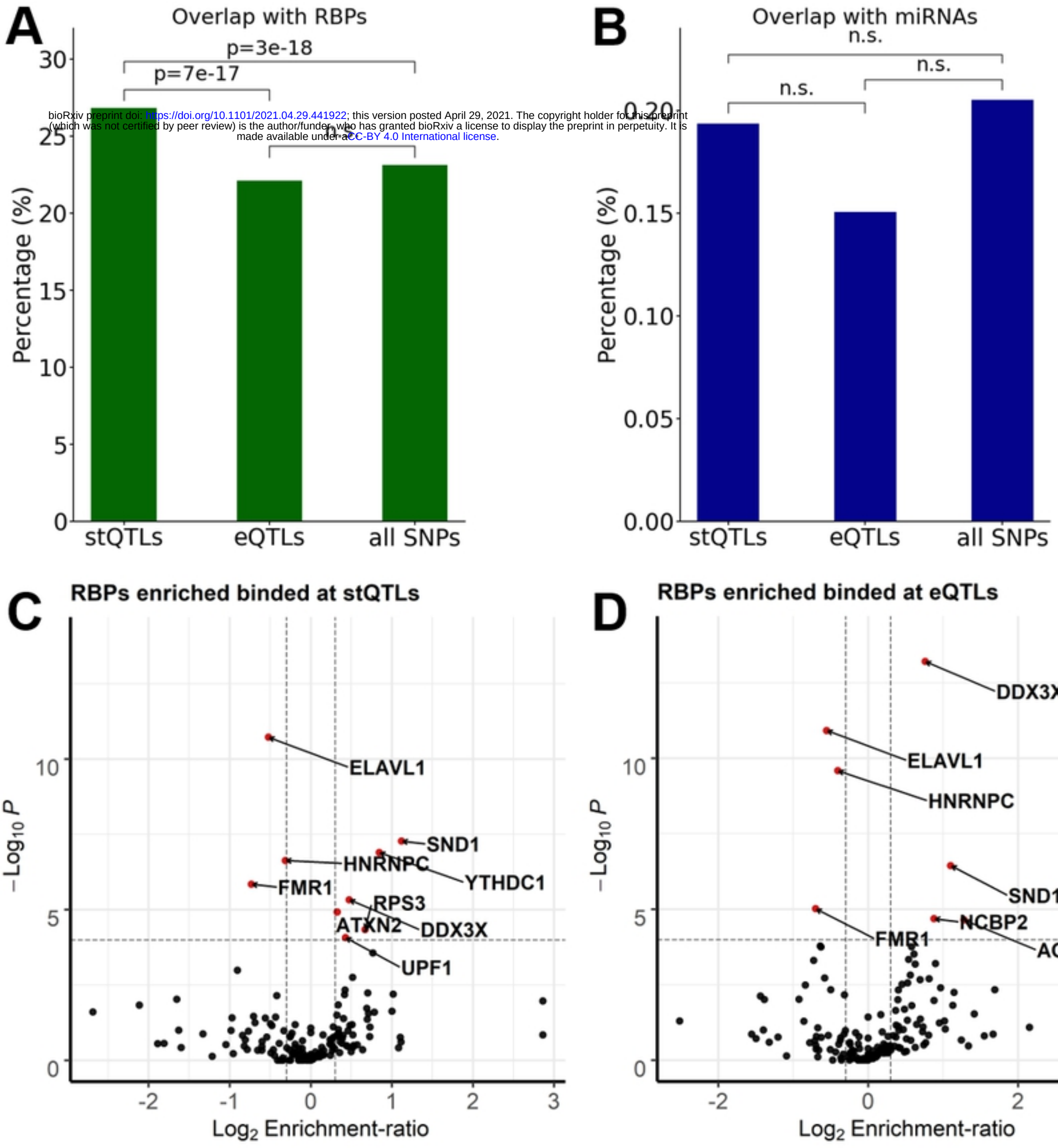


Figure 4

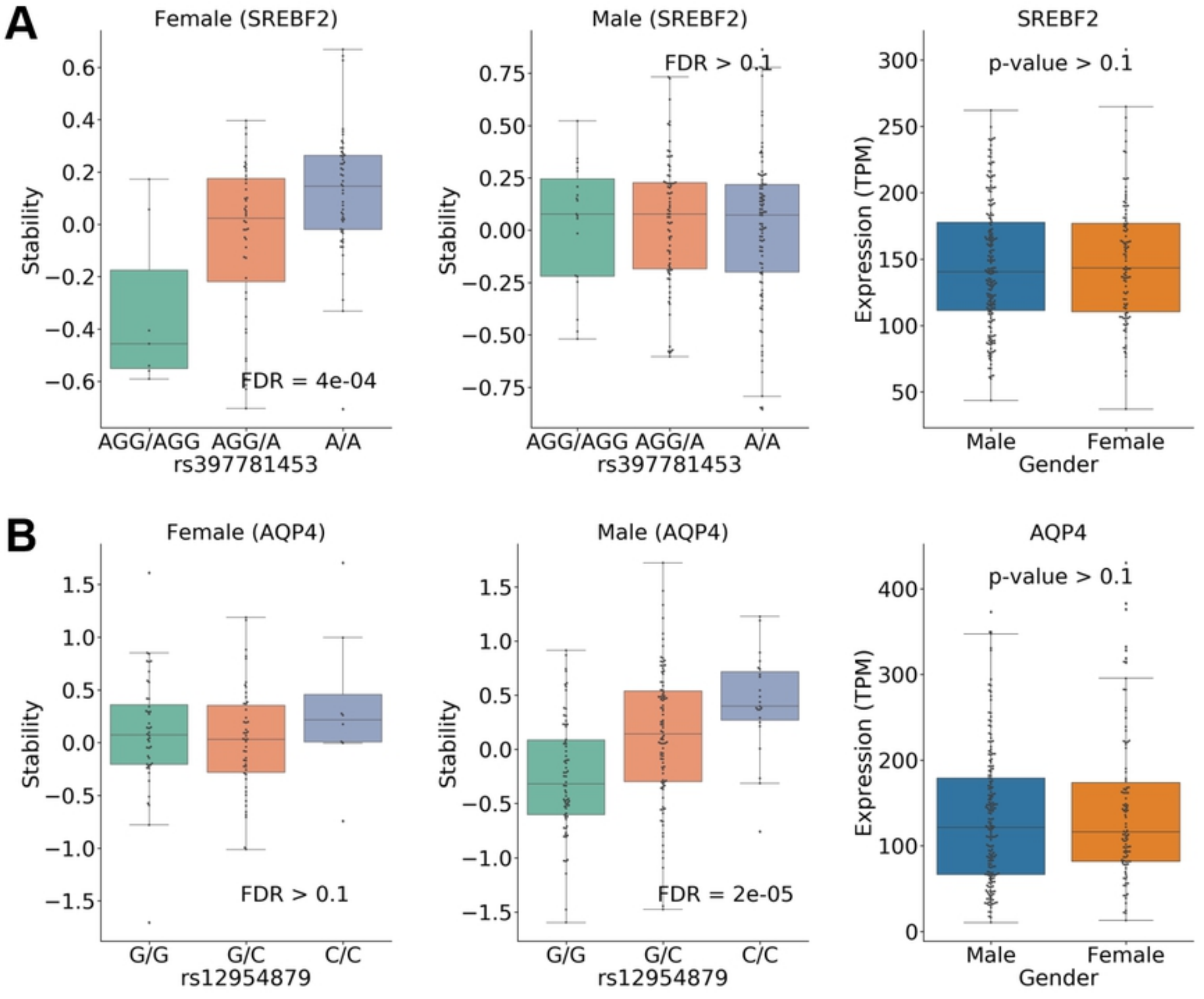
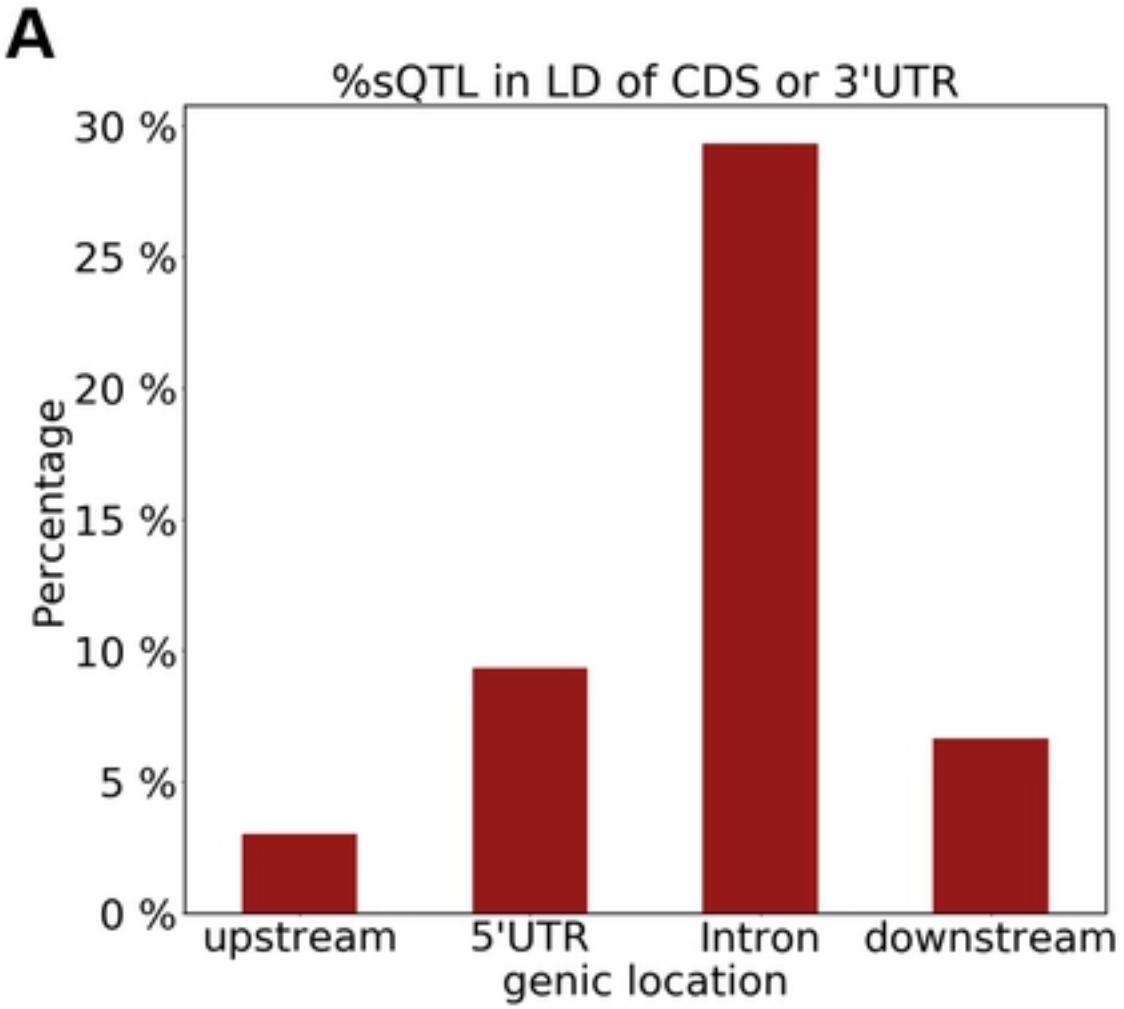
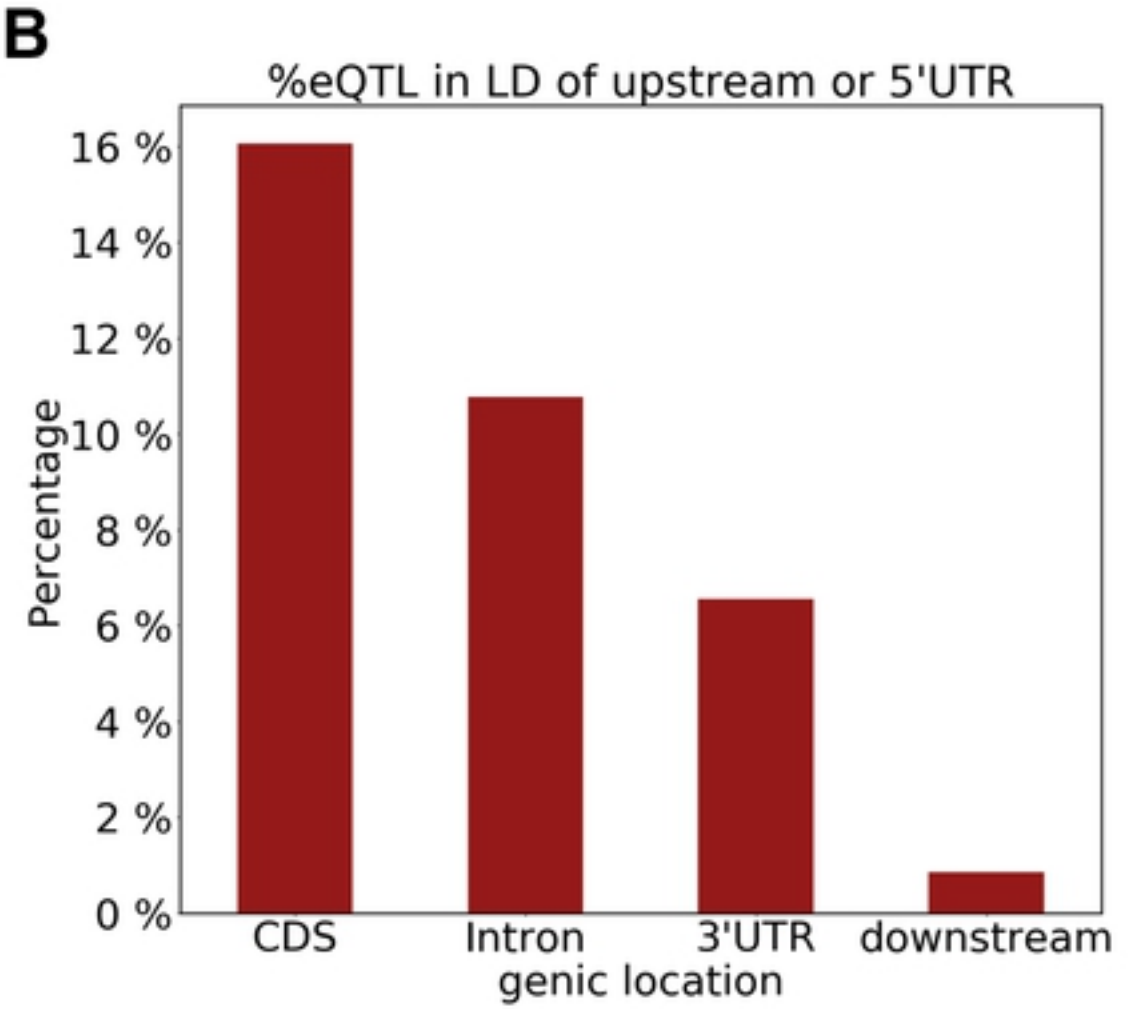

A
Female (SREBF2)
Stability
FDR = 4e-04
AGG/AGG
AGG/A
A/A
rs397781453
Male (SREBF2)
FDR > 0.1
SREBF2
p-value > 0.1
Expression (TPM)
Male
Female
Gender
B
Female (AQP4)
FDR > 0.1
G/G
G/C
C/C
rs12954879
Male (AQP4)
FDR = 2e-05
AQP4
p-value > 0.1

Figure 5


A
%sQTL in LD of CDS or 3'UTR
30 %
25 %
20 %
15 %
10 %
5 %
0 %
Percentage
upstream
5'UTR
Intron
downstream
genic location


B
%eQTL in LD of upstream or 5'UTR
16 %
14 %
12 %
10 %
8 %
6 %
4 %
2 %
0 %
Percentage
CDS
Intron
3'UTR
downstream
genic location

S1 Fig