The American Journal of Human Genetics, Volume *111*

Supplemental information

Regulatory features aid interpretation of 3′UTR variants

Lindsay Romo, Scott D. Findlay, and Christopher B. Burge

SUPPLEMENTAL FIGURES

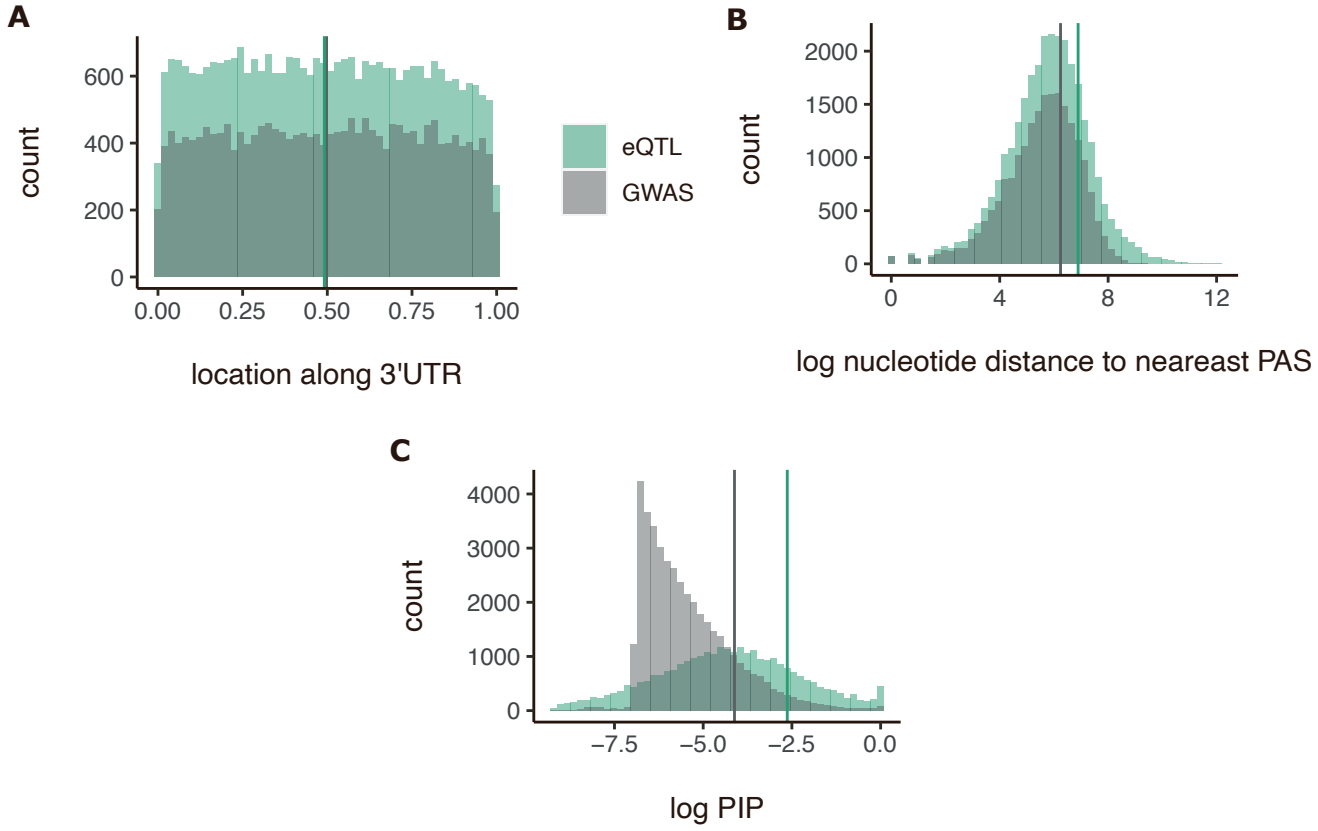

Figure S1: Comparison of eQTL variant and GWAS hit distribution along 3'UTR (A), proximity to nearest polyA signal (B), and PIPs (C). Vertical lines represent means for each distribution.

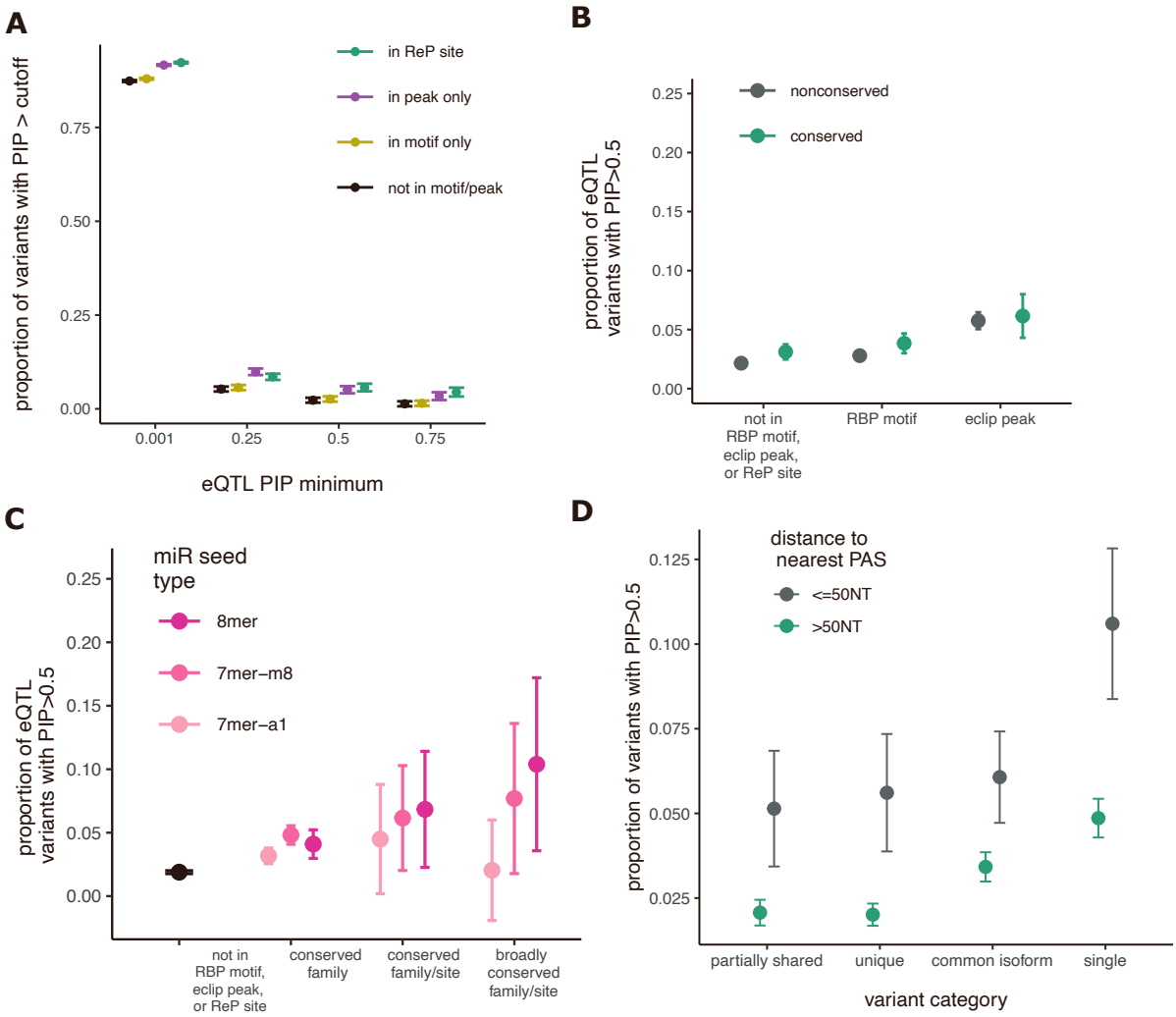

Figure S2: eQTL findings are robust even with a more stringent summary statistic PIP threshold. A Proportion of eQTLs with PIP greater than a minimum cutoff for variants not in RBP motifs or eCLIP peaks compared to variants in RBP motifs, eCLIP peaks, and ReP sites, with 95% confidence intervals. **B** Fraction causal (proportion of eQTLs with PIP>0.5) for variants not in RBP motifs or eCLIP peaks compared to variants in RBP motifs or eCLIP peaks. **C** Fraction causal for variants not in miRNA sites compared to variants in miRNA sites with increasing predicted seed strength. **D** Fraction causal for eQTL variants in genes with various numbers of canonical alternatively polyadenylated (APA) isoforms.

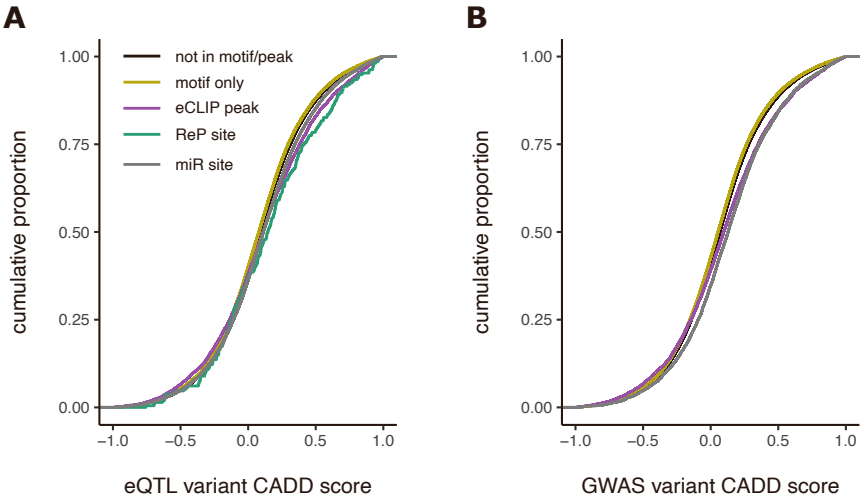

Figure S3: Variants in putative regulatory elements have higher CADD scores. Comparison of raw combined annotation dependent depletion (CADD) score distributions for eQTLs (A) or GWAS hits (B) in various putative regulatory elements versus controls.

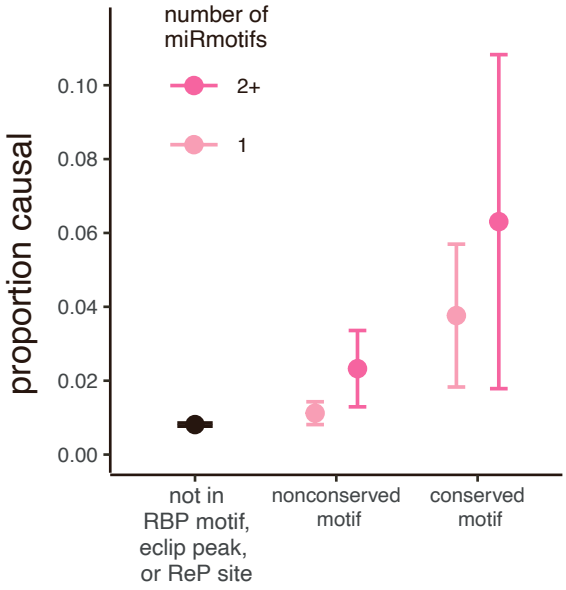

Figure S4: Trend towards higher PIP for variants predicted to disrupt more than one miRNA site. Fraction causal with 95% confidence intervals for GWAS variants not in miRNA sites compared to variants in increasing number of sites.

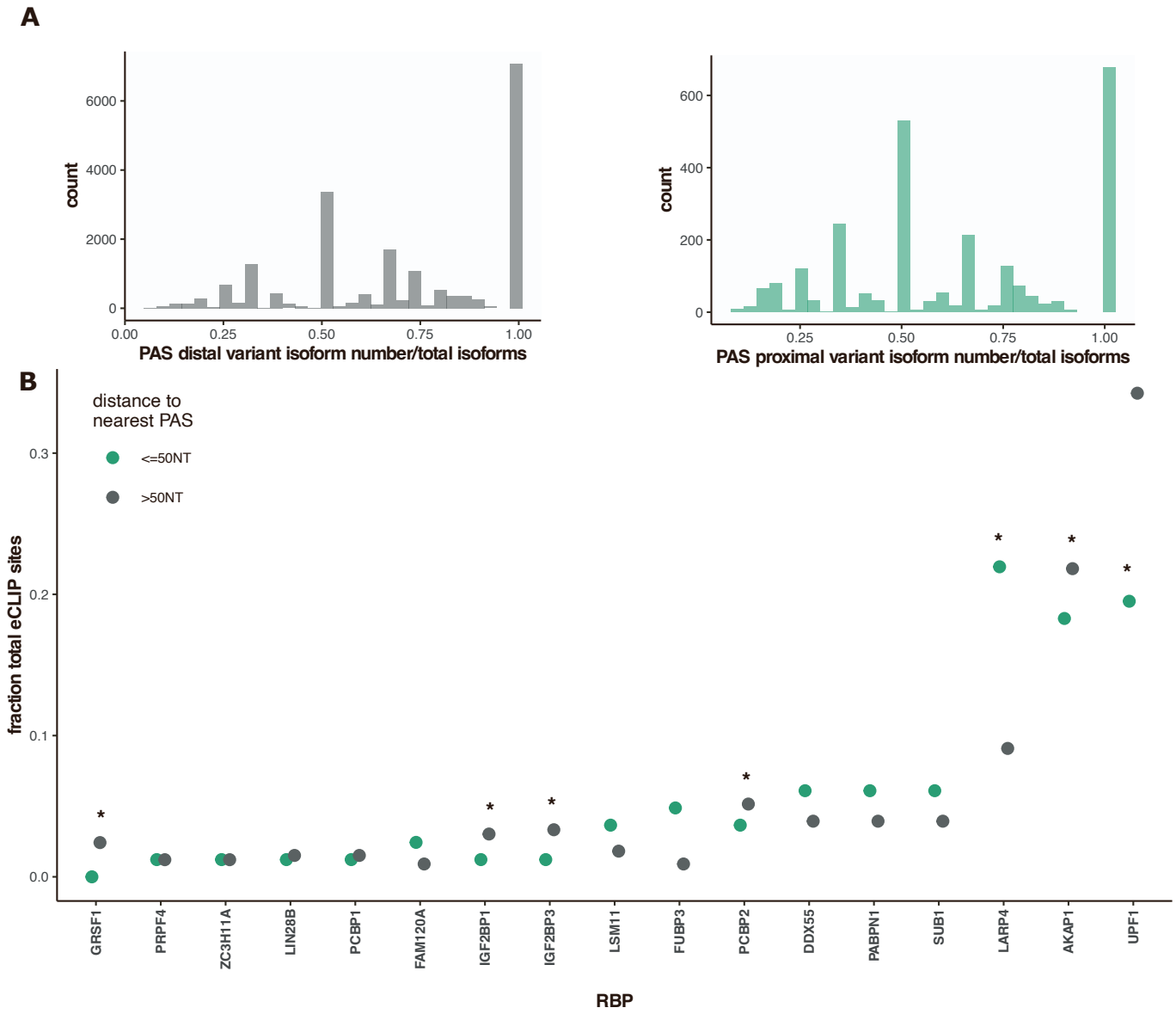

Figure S5: Comparison of eQTLs proximal versus distal to PAS demonstrates enrichment for known polyA binding proteins. **A** distribution of eQTLs distal (>50 nt, left panel) and proximal (<50nt, right panel) to PAS (x axis is isoform ordinal number divided by total isoforms per gene). Differences are nonsignificant. **B** Comparison of fraction of PAS-distal versus PAS-proximal eQTLs falling in RBP eCLIP sites, with 95% confidence intervals. * indicates p<0.01.

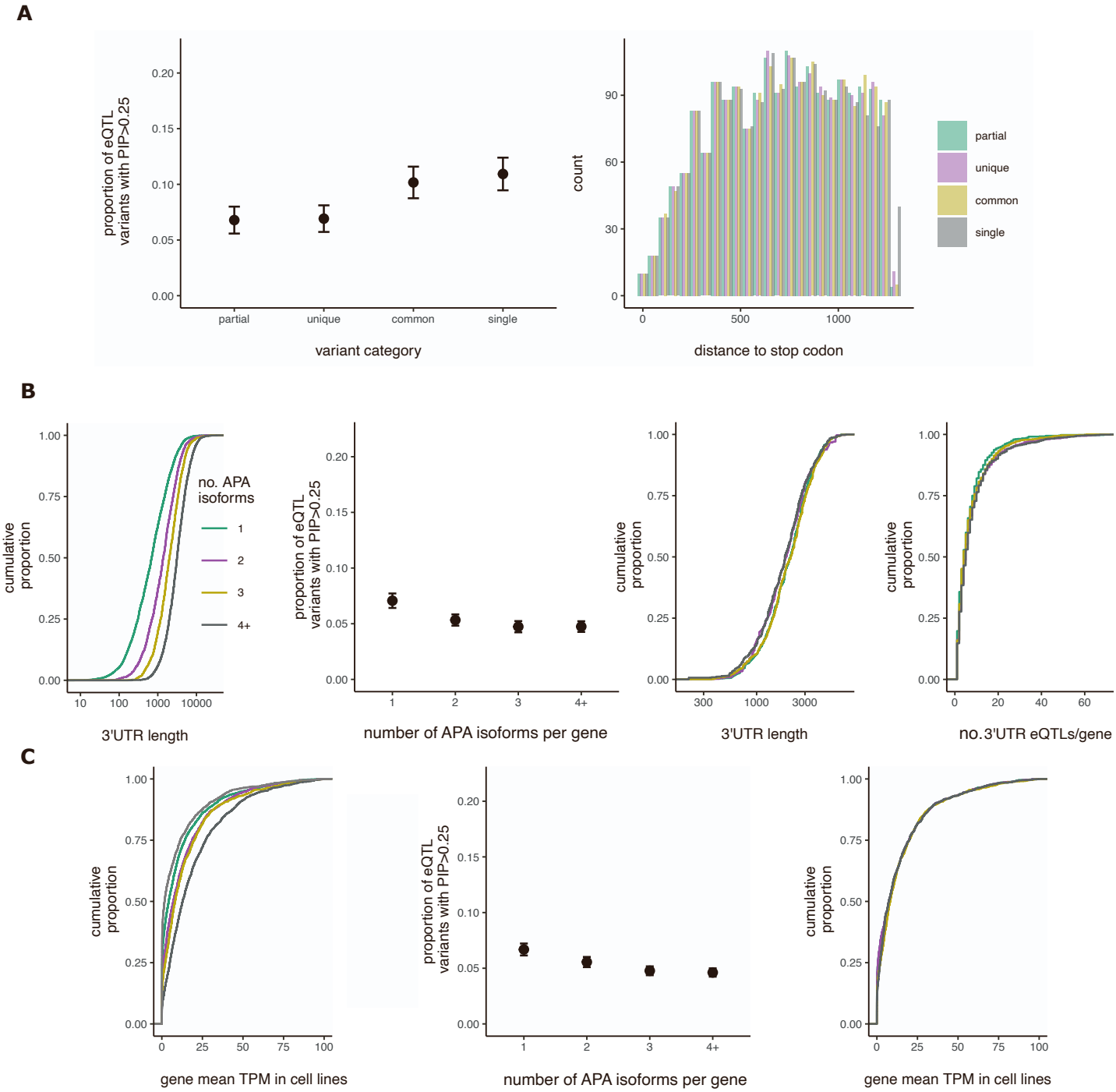

Figure S6: eQTL findings are not due to stop proximity, 3'UTR length, number of eQTLs per gene, or gene expression. A Fraction causal (proportion of eQTL variants with PIP greater than 0.25) with 95% confidence intervals for variants in various 3'UTR regions (left), after matching distance to canonical stop codon (right). **B** Distribution of 3'UTR length (left) with fraction causal and 95% confidence intervals for eQTL variants in genes with various numbers of canonical alternatively polyadenylated (APA) isoforms (middle left) after matching gene 3'UTR length (middle right). On right is the distribution of number of eQTLs per gene for genes with varying isoform numbers. **C** Distribution of gene expression (left) with fraction causal and 95% confidence intervals for eQTL variants in genes with various numbers of canonical alternatively polyadenylated (APA) isoforms (middle) after matching gene expression (right).

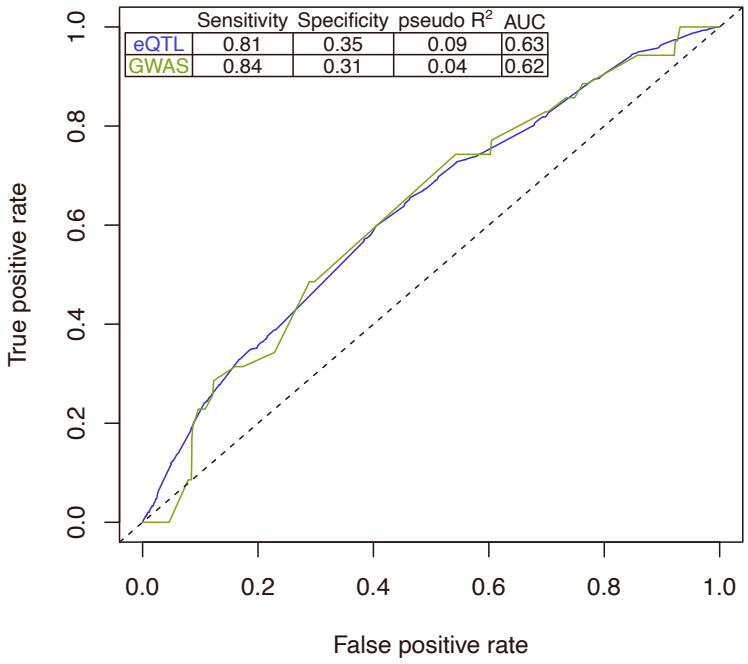

Figure S7: Performance of generalized linear models. Logistic regression analysis was performed to predict GWAS and eQTL variants (PIP>0.5). A variant was predicted to be an eQTL or GWAS hit if its log-odds was greater than 0.01 (eQTL) or 0.0075 (GWAS). These thresholds maximized sensitivity and specificity. Goodness of fit was assessed via Hosmer-Lemeshow Test with a chi squared of 1.0204 and p-value of 0.9981 for the eQTL model and a chi squared of 13.262 and p-value of 0.1032 for the GWAS model.

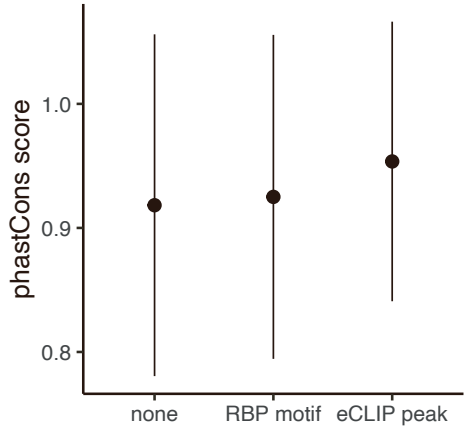

Figure S8: enrichment for pathogenic variants in regulatory elements is not solely due to conservation. Shown is mean phastCons score with standard deviation for variants in each category.

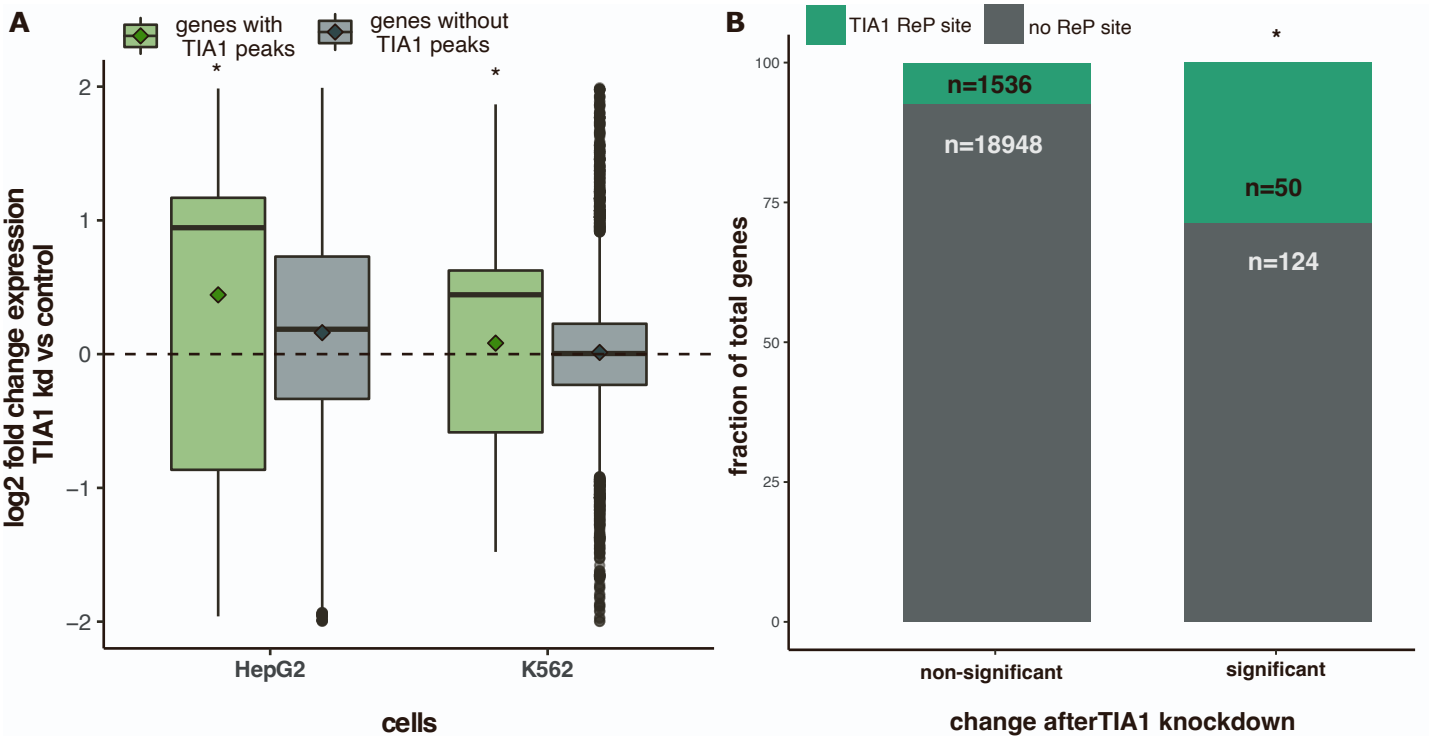

Figure S9: TIA1 binding sites regulate gene expression via TIA1 binding. A Genes with TIA eCLIP peaks exhibit increased expression after TIA1 knockdown in cell lines. **B** Genes that are differentially expressed after TIA1 knockdown are significantly enriched for TIA1 motifs.