Supplemental Information

Supplemental Figures

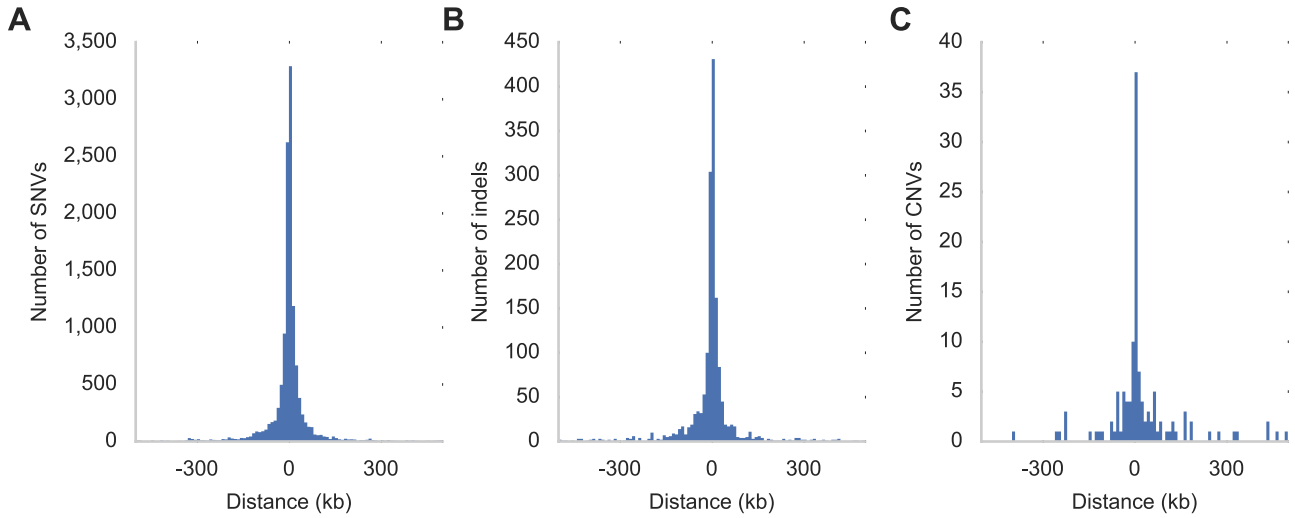

Figure S1. Related to Figure 1. Distance from lead variants to transcription start sites. (A-C) Distribution of distance from (A) SNVs, (B) indels, and (C) CNVs to the nearest transcription start site of the associated eGene for lead variants. If a gene had multiple lead variants due to *p*-value ties, all lead variants were included.

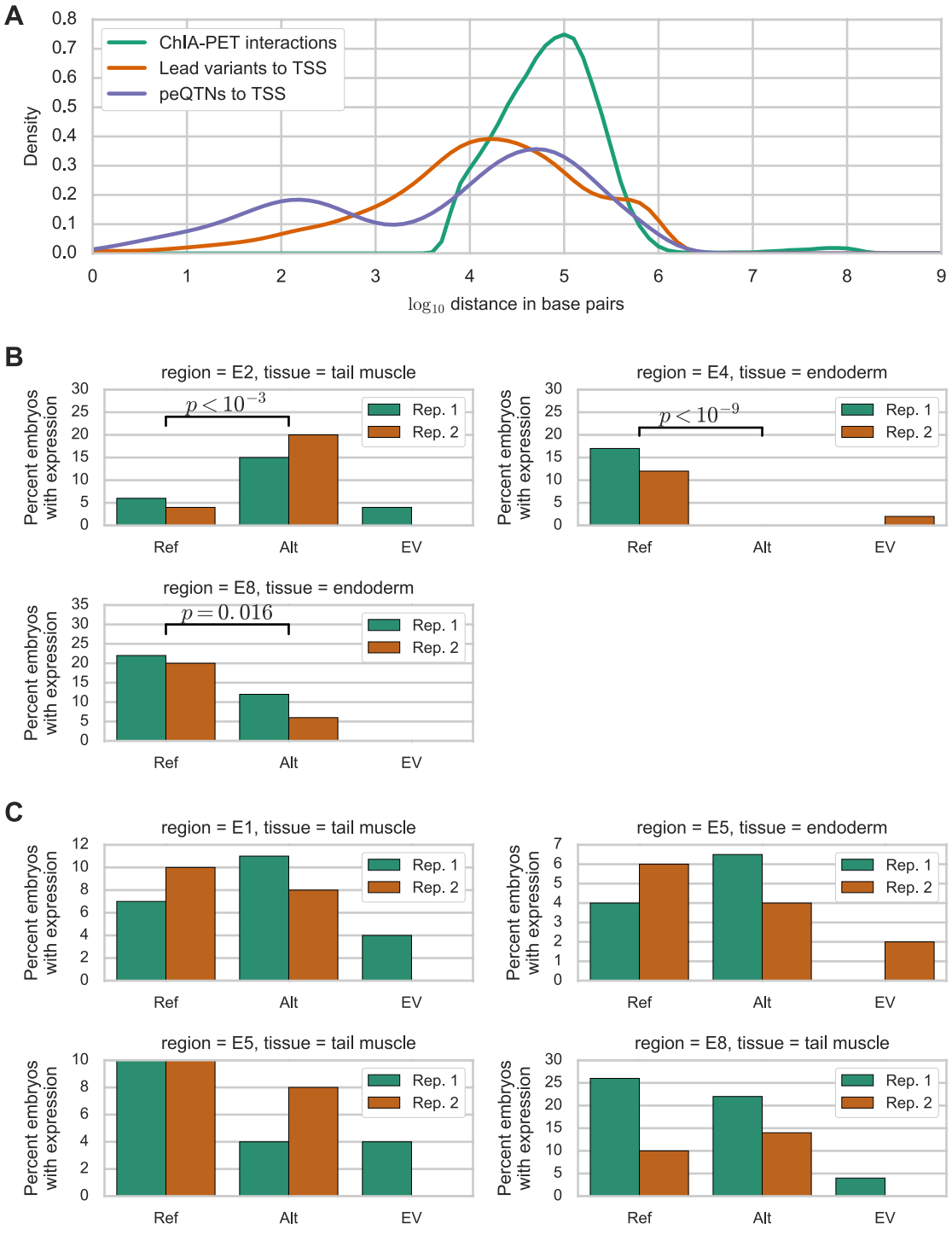

Figure S2. Related to Figure 3. Distance histograms and counting data for electroporation of regulatory regions into *Ciona intestinalis* embryos. (A) Density histograms of the $\log_{10}$ distances in base pairs from lead variants (purple) and peQTNs (red) to the transcription start site (TSS) of the associated gene. The green histogram is the size of ChIA-PET interactions from (Ji et al., 2016). (B and C) Percent of *Ciona* embryos with expression in the specified tissue for regions containing the reference allele (Ref), alternate allele (Alt), or empty vector (EV). (B) shows regions/tissues with a significant difference in expression between reference and alternate or alleles and (C) shows regions/tissues that drive expression but without a significant difference between the reference and alternate alleles. Rep. 1 and Rep. 2 refer to completely independent repeats for each experiment. See Table S5 for more information.

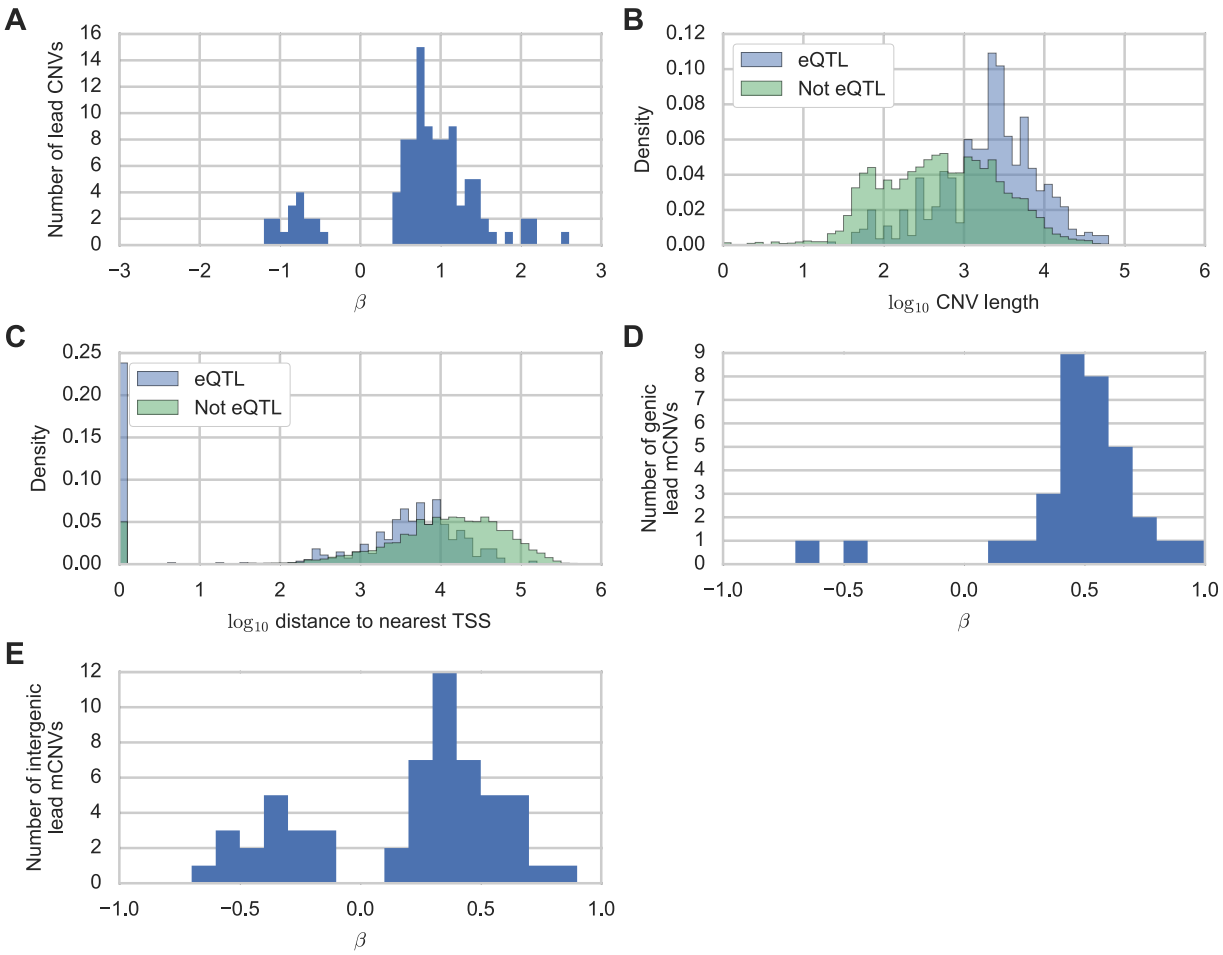

Figure S3. Related to Figure 5. CNV eQTL characteristics. (A) Effect sizes for lead CNVs eQTLs. Distribution of (B) CNV length and (C) distance to nearest transcription start site (TSS) in $\log_{10}$ base pairs for 550 CNVs that had at least one significant gene expression association and 14,731 CNVs that did not have any significant associations. Distribution of effect sizes for (D) 33 lead mCNV eQTLs where a significant mCNV overlapped the eGene and (E) 57 lead mCNV eQTLs where no significant mCNV overlapped the eGene.

Supplemental Tables

Table S1. Related to Figure 1. Classification of 5,746 eGenes and GWAS enrichment results. Number of genes tested and significant in primary eQTL analysis using 215 subjects stratified by gene type according to Gencode v19. Percent eGenes is the percentage of the eGenes that the indicated gene type comprise. Lead eQTL variant GWAS enrichment results for 33 phenotypes from the GRASP v2 database.

Table S2. Related to Figure 1. eQTL mapping results. Lead variants and all significant variants for 5,746 eGenes as well as 709 eGenes with second eQTLs and 175 eGenes with third eQTLs. "leads01" contains the lead variants for the primary eQTL analysis and "all01" contains all significant associations for the primary eQTL analysis. "leads02" and "all02" contain the lead variants and all significant associations for the second eQTL analysis conditioning on the lead variant from the first analysis. "leads03" and "all03" contain the lead variants and all significant associations for the third eQTL analysis conditioning on the lead variants from the first and second analyses.

Table S3. Related to Figure 2. Noncoding lead variant enrichment results. Results for enrichment of 4,616 noncoding SNVs and indels (Fisher exact test) in Roadmap and ENCODE DNase hypersensitivity sites and ENCODE transcription factor ChIP-seq peaks.

Table S4. Related to Figure 3. Putative expression quantitative trait nucleotides. Putative expression quantitative trait nucleotides (peQTNs) that overlap transcription factor ChIP-seq peaks and disrupt an associated motif. Associated motifs were defined using {Kheradpour, 2014 #604} which provides both annotated and novel motifs for different TFs. There are 3,140 distinct peQTNs although some are associated with the expression of more than one gene so there are more than 3,140 rows in this table.

Table S5. Related to Figure 3. *Ciona* peQTN experiments. Primers and regions used for *Ciona* experiments. Summary of expression observed in *Ciona* embryos for all regions and raw data for each region. Constructs containing the putative regulatory region with either the reference or alternate peQTN were electroporated into *Ciona* embryos. The percent of embryos with expression was calculated for each tissue where the region drove expression (enhancer regions E5 and E8 drove expression in more than one tissue). The total number of embryos measured (n) is noted for each experiment. The experiment was repeated twice for each allele. For region P1, the region was also tested without the Snail enhancer upstream of the promoter element.