**Supplemental Material**

**Structural variants are a major source of gene expression differences in humans and often affect multiple nearby genes**

Alexandra J. Scott, Colby Chiang, Ira M. Hall

**Supplemental Fig. S1:** eQTL effect size distributions 2

**Supplemental Fig. S2:** Enrichment of SV-eQTLs in annotated genomic features 3

**Supplemental Fig. S3:** Enrichment of SV-eQTLs in Roadmap epigenomic segmentation states 4

**Supplemental Fig. S4:** Distribution of eQTL tissue specificity across tissues 5

**Supplemental Fig. S5:** Distribution of eQTLs with unknown status 6

**Supplemental Fig. S6:** Distribution of eQTL effect sizes and effect size standard errors 7

**Supplemental Fig. S7:** Enrichment of singleton SVs near multi-tissue expression outliers 8

**Supplemental Fig. S8:** Distribution of gene expression outlier effect sizes 9

**Supplemental Fig. S9:** Distribution of outlier-associated SV impact scores 10

*Supplemental Tables are separate from this file and can be downloaded individually.*

****

**Supplemental Fig. S1.** eQTL effect size distributions for coding and noncoding variants of each type with the number of eQTLs shown above the distribution. Deletions and duplications are separated by evidence used for variant discovery, either breakpoint (BP) or read-depth (RD).



**Supplemental Fig. S2.** Feature enrichment of SV-eQTLs. Fold enrichment and 95% confidence intervals (based on 100 random shuffled sets of the positions of SVs in each bin) for the overlap between the most significant SV for each eGene and various annotated genomic features. (**A**) Enrichment of SVs in each causality bin for intersections with exons of the affected eGene. (**B-T**) For the remaining plots, SVs that overlapped with an exon of the affected eGenes were excluded. Enrichment was only observed in the 10-kb regions upstream (**N**) and downstream (**O**) of TSSs and in segmental duplications (**S**), which is consistent with the known concentration of SVs in architecturally complex genomic regions.

****

**Supplemental Fig. S3.** Enrichment of SV-eQTLs in Roadmap Epigenomics segmentation states. Fold enrichment and 95% confidence intervals (based on 100 random shuffled sets of the positions of SVs in each bin) for the overlap between the most significant SV for each eGene and various annotated genomic states. (**A**) Enrichment of SVs in each causality bin for intersections with exons of the affected eGene. (**B-P**) For the remaining plots, SVs that overlapped with an exon of the affected eGenes were excluded. We identified genomic intervals where each of the 15 Roadmap Epigenomics segementation states is found in at least 10 of the 127 available epigenomes and identified SVs in each causality bin that overlapped with these collapsed genomic intervals.

****

**Supplemental Fig. S4.** Distribution of the tissue specificity of eQTLs across tissues, as evaluated by METASOFT, for eQTLs in which the activity status is known in at least 43 of 48 evaluated tissues. Red lines indicate the distribution of SV-eQTLs that are active in the fraction of evaluated tissues indicated on the x-axis. Blue lines indicate the same for SNV- and indel-eQTLs. Solid lines denote coding eQTLs where the eVariant intersects the coding region of the associated eGene and dashed lines show the distributions for noncoding eQTLs.

****

**Supplemental Fig. S5.** Distribution of eQTLs with unknown status (0.1≤*m*≤0.9) across the indicated number of tissues as evaluated by METASOFT. (**A**) Distribution for coding SV-eQTLs where the eSV intersects with the coding region of the associated eGene.(**B**) Distribution for coding SNV/indel-eQTLs. (**C**) Distribution for noncoding SV-eQTLs where the eSV does not intersect the coding region of the associated eGene. A large number of noncoding SV-eQTLs (1,094/9,274) have unknown status in all 48 tissues evaluated. (**D**) Distribution for noncoding SNV/indel-eQTLs**.**



**Supplemental Fig. S6.** Distribution of eQTL effect sizes and the standard errors of effect sizes for eQTLs evaluated by METASOFT. (**A-D**) Distribution of eQTL effect sizes (beta) for all eQTLs evaluated by METASOFT across all 48 tissues. Coding eSVs (**A**) have much larger effect sizes compared to coding eSNVs and eIndels (**B**), noncoding eSVs (**C**) and noncoding eSNVs/eIndels (**D**). (**E-H**) Distribution of eQTL effect size standard errors for all eQTLs evaluated by METASOFT across all 48 tissues. Both coding eSVs (**E**) and noncoding eSVs (**G**) have larger standard errors compared to coding eSNVs/eIndels (**F**) and noncoding eSNVs/eIndels (**G**).



**Supplemental Fig. S7.** Fold enrichment of European singleton SVs within the indicated distance of multi-tissue expression outliers (**A**) and fold enrichment of multi-tissue outliers within indicated distance of European singleton SVs (**B**). Enrichments calculated between the observed set of 26,289 autosomal multi-tissue outliers and 1,000 random permutations of outlier sample names. All included SVs are smaller than 1 Mb in size. Only European ancestry samples were included in this analysis. 95% confidence intevals are indicated by the error bars and p<0.001 for the enrichment at all distances.

****

**Supplemental Fig. S8.** Distribution of gene expression outlier effect sizes. The effect sizes for all outliers are shown in green. Outliers associated with each SV type are also shown, separated by whether the SV causes a complete dosage change (red), partial dosage change (blue) or is noncoding for the associated outlier gene (yellow). Effect sizes shown are the most extreme effect size for each outlier gene across all tissues with available expression data. Counts below the x-axis indicate the number of unique SV/outlier pairs with the indicated SV type. Percentages above distributions indicate the fraction of SV/outlier pairs with the relevant variant type found in each dosage category.

****

**Supplemental Fig. S9.** Distribution of SV impact scores calculated with SVScore. (**A-B**) Distribution of mean phastCONS scores for outlier-associated (**A**) and control-associated (**B**) SVs. (**C-D**) Distribution of mean LINSIGHT scores for outlier-associated (**C**) and control-associated (**D**) SVs.