

Assessing Proxy Gene Expression Following Expression CROP-Sequencing

Garcia Martinez, Katherine (School: Benjamin E. Mays High School)

Disease pathways are in great need of study. The understanding of disease pathways can result in the development of more individualized methods for diagnosis and treatment of various diseases. One way to pinpoint disease associated genes is through Genome Wide Association Studies (GWAS), which find regions of the genome associated with a specific disease. The majority of GWAS hits actually map to noncoding regions of the genome, indicating SNPs are likely regulatory (Young et al. 2019). Expression quantitative loci (eQTL) can be helpful in identifying likely causal variants. However, linkage disequilibrium causes it to be unclear which variants are truly causal (Huang et al. 2017). Therefore, experimental methods are needed to determine causality. One such method is eCROP-seq, which has shown to effectively identify causal SNPs (Pan et al. 2020). Using eCROP-sequencing, specific variants for a target gene in eQTL are knocked out, and the gene expression of cells with and without the perturbation are compared. Because these target genes are nearby other genes and can likely be coregulated by the variants tested, we further investigated the effect these variants have on nearby proxy genes of the 91 target genes originally tested. The variants are tested for their impact on 1,239 (on average 13) nearby genes surrounding the gene body of each target gene to investigate if eCROP-seq can be used to uncover even more causal relationships. After testing, it was found that there are additional causal relationships for variants in eQTL for nearby proxy genes of the 91 target genes.