Ancestry Informative eQTLs

Creed JH, Gerke TA, Yamoah K, Berglund AE

**Key words:** eQTL, TCGA, racial disparities

**Objective**

Incidence rates of prostate cancer are highest in African American men, however most biomarker research is performed utilizing samples from Caucasian men. This problem is particularly evident in expression quantitative trait loci (eQTLs) analysis due to the underlying differences in allele distributions in African Americans versus Caucasians. Therefore this project serves as a preliminary analysis of eQTLS between African Americans and Caucasians, focusing on SNPs that are known to diverge in ancestral populations.

**Methods**

Gene expression values from prostate tumor tissue were downloaded from the Cancer Genome Atlas (TCGA) along with germ-line genotypes. If germ-line was unavailable, adjacent normal prostate tissue samples were used. Ancestry estimates for 3 ancestral populations were calculated from SNP genotypes using Admixture, with the dominant population (>50%) being used to determine race. Initially, SNPs were filtered based on call confidence and missingness, and then filtered by minor allele frequency < 0.01 and Hardy Weinberg p-value < 10-5 separately in African Americans and Caucasians. eQTLs were identified from linear regression models for tumor gene expression on the number of minor alleles for each genotype (minor allele in the combined cohort).

**Results**

Of the possible 499 individuals, 332 had both high quality genotype and gene expression data. From this subset, 282 were deemed Caucasian and 42 African American. Of the remaining 8 individuals, 6 were classified as having another ancestral background other than African American or European and 2 failed to have a single dominant ancestry.

Over 14 billion association tests were performed for each race, covering 709,002 SNPs and roughly 20,000 genes. Using the Bonferroni-adjusted p-value threshold of 3.13-12, 934,410 eQTLs were deemed significant in African Americans, while 108,914 were significant in Caucasians. From these, there were 42 overlapping eQTLS between both races. The eQTLS were then compared to a panel of 2,076 ancestry informative markers (AIMs). Of the resulting ancestral snp-gene pairs, the majority of the SNPs reside on chromosome X, while the genes were distributed throughout the genome. One marker, rs5943149 on chromosome X, was found to be associated with multiple genes in Europeans but none in African Americans. This difference could stem from the underlying variation in genome structure between ancestral populations, such as the G allele frequency for rs5943149, which is the major in allele in the CEU population but the minor allele in the YRI population.

**Conclusion**

The preliminary findings of this study demonstrate the different SNP-gene relationships between African Americans and Caucasians. A more in depth analysis will follow, such as examining eQTLs that share the same SNP for gene pathways. Additionally, these findings should be validated in additional cohorts that contain both genotype and gene expression data for prostate tissue, such as GTEx.

**Research Supported By**

Gerke Lab at Moffitt Cancer Center