

Todd Gavin  
Dr. Peter Calabrese  
QBIO401  
16 November 2022

### Assignment #10 – Final Project Proposal

For my final project, I will be analyzing how the number of certain cancer-risk SNPs correlate with specific ethnic groups.

My specific research question is as followed: ***Are the genotypic frequencies of cancer-risk SNPs different in certain ethnic groups compared to others?***

Below I have a list of 3 identified cancer-risk SNPs, with the associating cancer that they create a predisposition risk for, and the chromosome and read location of where the SNP is isolated.

Cancer Risk SNPs:

- rs6983267
  - Associated with colorectal cancer.
  - 8:127401060 (GRCh38) --- 8:128413305 (GRCh37)
  - <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2763485/>
- rs4713266
  - Associated with prostate cancer.
  - Location: 6:11218797 (GRCh38) --- 6:11219030 (GRCh37)
  - <https://aacrjournals.org/cancerres/article/81/14/3764/670224/How-Cancer-Risk-SNPs-May-Contribute-to-Prostate>
- rs72699833
  - Associated with breast, cervical, colorectal and non-small-cell lung cancer
  - Location: 1:121539689 (GRCh38) --- 1:121281487 (GRCh37)
  - <https://www.nature.com/articles/s41416-019-0614-3>

For my analysis, I will be collecting population SNP data from the 1000 genomes project from the locations of the cancer-risk SNPs.

My analysis is follows:

- For each ethnic group, I will calculate the frequency of homozygous recessive (0|0), homozygous dominate (1|1), and heterozygous (0|1, 1|0) for the cancer-risk SNPs.
- Once the frequency and percentages are calculated for each group, I will perform a statistical test to determine if the differences in SNPs frequency between ethnic groups statistically significant.
- If there is time, based on the genotypic frequencies, I will calculate if certain ethnic groups are at a greater risk to developing certain cancers compared to other groups.