Broad\_acs\_mini

Created by David de la Cerda

Files here were provided from Dr. Grace Tiao and were used to generate a PCA plot and estimate the ancestry of unknown individuals in the dataset provided.

The primary files of interest are:

**mini\_final.sh**

This executable was used to filter, generate PCs, and estimate ancestry

**mini\_acs\_PCA.R**

This script was used to visualize stats from the raw and filtered VCF files as well as visualize the PCA of individuals from unknown and predicted ancestry. It was also used to export tables related to ancestry classification and PC scores.

The directories are:

**vcf\_stats**

This contains the descriptive statistics used to visualize the data provided.

**vcf\_stats\_filtered**

This contains the descriptive statistics used to visualize the filtered data once the raw data were visualized.

**deliverables**

This contains necessary files asked to be provided. The descriptions are:

acs\_mini\_project\_predicted\_labels.txt

This contains the information where the columns are original individual ids, original ancestry assignment, and predicted ancestry assignment.

pca\_compare.pdf

This contains visualization of the distribution of PC values with the unknown and then the predicted classifications side-by-side. PC1 versus PC2, PC3, and PC4 are visualized to get a sense of separation.

pruned\_pca\_scores.txt

This file contains PC scores of each individual after pruning.