Rough Analysis Plan 03/09/2022

1. Determine biological sex from RNA-Seq data
   1. Using PCA analysis of gene expression and verify with expression of SRY gene expression.
   2. Cheyret will look into methylation to see if predicted sex has already been calculated
2. Check if center and vape status correlated? If yes, by how much?
3. RUV normalize the data
   1. Variables of interest to retain: age, sex, vape status, and center
   2. Do we want to include lung function as variables of interest?
   3. Include plots for determining number of k’s to include.
4. Gene expression and vape status
   1. Gene expression ~ vape status + age + sex + center + RUV factors
      1. Histogram of p-values with vape status and center
         1. Check for model fit issues
      2. Plots of top genes (expression ~ vape status) color code by center might also be helpful
      3. Id gene that are only significant in vape status and not center (are there any?)
   2. Table of cilia pathway gene list (provided by Sunita)
5. Pseudo Sensitivity analysis
   1. Rerun model with just subjects from Pueblo (removing center variable)
6. Enrichment analysis (GSEA)
   1. We will need to decide on a ranking parameter at some point, I’ve used sign of beta \* -log2(p-value) in the past.
7. Gene expression and lung function (without vape status)
   1. Gene expression ~ lung function + age + sex + center + RUV factors
      1. Histogram of p-values lung function and center
      2. Plots of top genes (expression ~ lung function) color code by center might also be helpful
8. Targeted Methylation
   1. Use list of significant genes associated with vape status to narrow which probes to test