

This folder contains the necessary files and R scripts needed to carry out the various example whole blood differential gene expression analyses presented in O'Connell et al.

ethnicity_comparisons_script.r contains the annotated R script used to compare whole blood gene expression between Hispanic and non-Hispanic donors under each of the following three conditions: without correcting for circulating leukocyte counts, correcting for circulating leukocyte counts using actual cell counts measured with flow cytometry, and correcting for circulating leukocyte counts using inferred cell counts generated via PCA-based deconvolution method described in the *project-007/pca_deconvolution* folder of this repository.

whole_blood_raw_counts.csv is an abbreviated example input file illustrating the format whole blood RNA sequencing data should be provided in. Data should be in the form of gene-level summarized read counts. *whole_blood_sample_meta_data.csv* is an abbreviated example input file illustrating format that whole blood sample metadata should be provided in, including donor age, race, ethnicity, as well as actual and inferred leukocyte counts. Actual cell count data should be in the form of relative cell counts (proportion of total WBC count). The exact RNA sequencing data and metadata used for the analyses presented in the manuscript can be retrieved from the National Center for Biotechnology Information via BioProject accession number PRJNA949611.