**Custom Code**

**File 1: clustering\_clues.R**

Contains code to perform unsupervised clustering of CLUES study data (ACR phenotypes) and compare clinical variables between clusters

Lines 116- 148: Train random forest model to predict cluster label from ACR variables

**File 2: differential\_methylation\_analysis.R**

Contains code to perform differential methylation analysis for CLUES data with covariates, make figures, and perform race-enrichment analysis

Lines 92 – 148: Annotate cluster-associated CpGs using EPIC annotation file

Lines 151- 165: Create volcano plots for pairwise cluster comparisons

Lines 167 – 199: Pathway analysis of cluster-associated CpGs mapped to genes using EPIC annotation file

Lines 203 – 232: Generate heatmap of cluster-associated CpGs

Lines 235 – 263: QQ plot

Lines 265 – 308: Calculate race-association enrichment statistic by permuting self-reported race

**File 3: clustering\_validation\_data.R**

Contains code to apply random forest model to ACR phenotypic data from validation cohort and find demographic and clinical differences between clusters

**File 4: methylation\_mediation\_analysis.R**

Contains code to perform methylation mediation analysis using a causal inference test (CIT R package)

**File 5: helper\_diff\_meth.R**

Internal helper functions (do not modify)

**File 6: load\_data\_meqtl.R**

Internal helper function to load all data required for meQTL and mediation analysis (do not modify)