**Gene differential expression analysis by RNA-seq**

The RNA-seq raw counts data had gene-level read counts for 13 healthy participants and 19 HIV-infected participants. Each gene was labeled by an Ensembl ID, together with a gene symbol and its gene length. The genes which have at least five counts per sample on average were kept for differential expression (DE) analysis. The trimmed mean of M values (TMM) normalization method from edgeR (version 3.24.3) was chosen from several other methods such as: Transcripts Per Kilobase Million (TPM), DESeq2 (version 1.22.2) etc. DE analysis was called using normalized counts with edgeR according to the package vignettes and with an FDR of 5%.

**Analysis and graphic display**

PCA was performed by R (version 3.5.1) in Rstudio (Version 1.1.456).

**References:**

For edgeR

Robinson MD, McCarthy DJ, Smyth GK (2010). “edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.” Bioinformatics, 26(1), 139-140.

McCarthy, J. D, Chen, Yunshun, Smyth, K. G (2012). “Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation.” Nucleic Acids Research, 40(10), 4288-4297.