**Statistical Analysis**

**Differential expression analysis**

The RNA-seq raw counts data had gene-level read counts for 11 healthy participants and 15 ART-infected and treated participants. Each gene was labeled by an Ensembl ID, together with a gene symbol and its gene length. 19454 out of 42995 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) and DESeq2 (version 1.22.2). TMM method calculates the scaling normalization factor for each sample by accounting for library size and observed counts, while under the assumption that the majority of genes are not DE. In our case, the TMM method provides the best normalization results in terms of removing the unwanted variation, according to the relative log expression (RLE) plots. Two-group comparison DE analysis was called using normalized counts with edgeR according to the package vignettes and with an FDR (false discovery rate) of 5%. This analysis is to test whether a gene is significantly altered between healthy controls and ART-infected and treated donors.