# Analysis Report

**Purpose**:

This is to provide a writeup about the RNA-Seq data analysis that has been done as part of the **two** homework for the course “*Setting Bioinformatics Pipelines*”. The report has two components: implementing limma-trend, voom, ORA, GSEA, and conducting some comparisons between the mentioned tools. The analysis was done using the dataset *GSE198256*.

**Pre-implementation thoughts:**

During the second class some of the students’ groups discussed some of the differences between limma-trend and voom and later we discussed when it would be appropriate to use either one. Perhaps, one elemental difference would the variabilities of the samples’ library size. It is important that the source of this variability should be technical not biological or related to the sample’s quality or conditions. In short, voom would suite the analysis if the level of variability is high, otherwise limma-trend would be faster algorithm to use.

We also discussed ORA (Over-Representation Analysis) and GSEA (Gene Set Enrichment Analysis) for the biological interpretation. One the key differences between the two is GSEA considers the expression levels of “all” genes and their rank ordering. This could allow for the identification of coordinated changes in genes expressions. In other words, GSEA would show a broader network of interactions than ORA.

**Implementation**:

1. Voom – ORA-KEGG.
2. Trend – ORA-KEGG.
3. Voom – GSEA-KEGG.
4. Trend – GSEA-KEGG.

All used the same parameters (like contrast group, Normalized Enrichment Score (NES), …etc.)

**Post-implementation analysis:**

First, the top 100 differentially expressed genes in voom and limma-trend were compared, and only 64% were found to be shared. Using only the top 20 genes, we found 61% shared genes.

ORA produced almost the same associated pathways when used with voom & limma-trend (Figure 1 & 2). While GSEA produced multiple pathways when used with voom, it only produced one pathway when implemented with limma-trend. GSEA-Voom produced multiple pathways with positive and negative enrichment scores. GSEA-limma-trend only found the ribosome pathway (Figure 3 & 4).

This is very limited way of doing comparison (based on a single study & dataset). But in this case, it could be better to start with “simpler” pipeline, i.e. limma-trend with ORA. And based on the results, it would be worth checking another pipeline voom with GSEA.

\_

Figure 1:limma-trend with ORA (enrichment plot).



\_

Figure 2:voom with ORA (enrichment plot)



\_

Figure 3:limma-trend with GSEA.



\_

Figure 4:voom with GSEA (NES > 0).