DEgenes Hunter - Differential expression analysis report

December 1, 2016

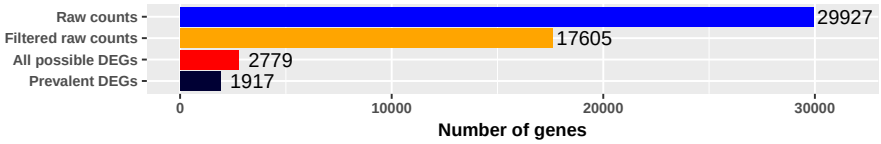
1 General description of the analysis workflow

1.1 Main concepts

In DEgenes Hunter, genes are labeled according to the following considerations:

- **Prevalent DEGs**: Differentially expressed genes considered by all packages employed but one.
- **All possible DEGs**: Differentially expressed genes considered by at least one of the R-packages but not enough of them to be considered as "prevalent DEGs".
- **Raw counts**: A matrix containing the raw counts (without any filtering).
- **Filtered raw counts**: Raw count matrix after filtering.
- **Filtered out**: Genes discarded during the filtering process.
- **Not DEGs**: Genes not considered DEGs in any package (by a minimum of $n-1$ packages employed if we used more than three algorithms).

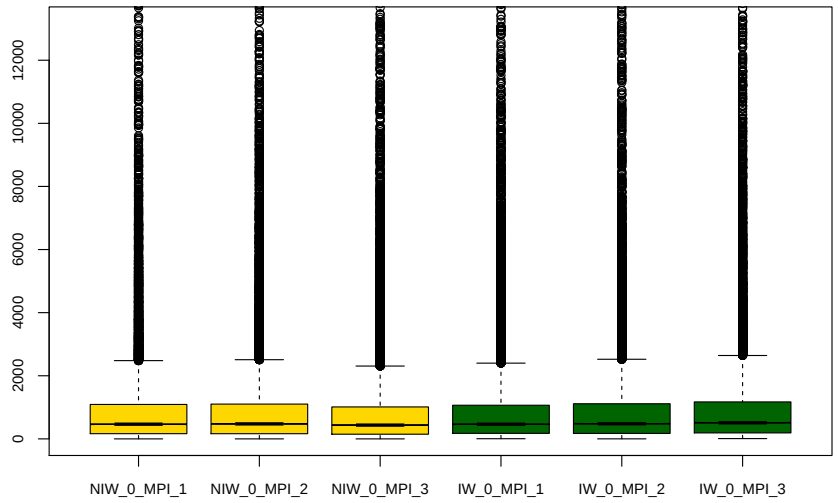
2 Data quality control (QC)



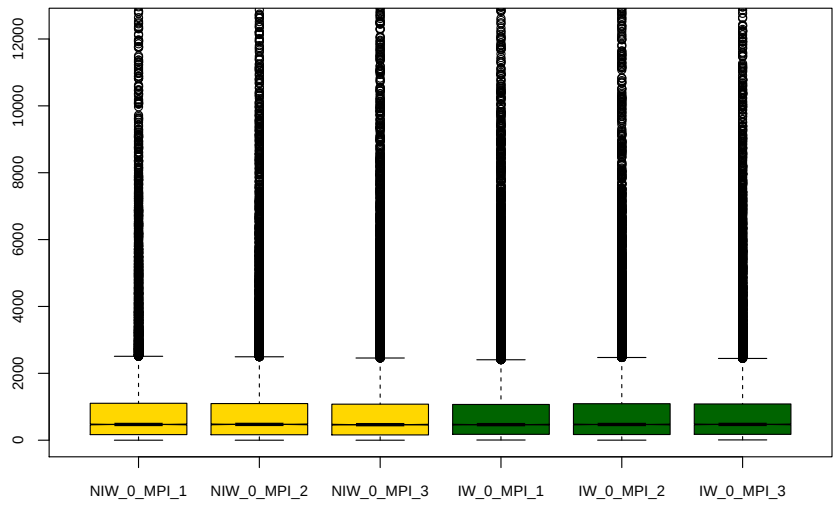
DEgenes Hunter creates box plots, principal component (PCA) plots, and multidimensional scaling (MDS) plots before and after raw count normalization to assess the consistence of comparison groups.

2.1 Box plots

It is expected that samples belonging to the same treatment appear together at least after normalization, guarantying that the treatment groups to compare are sufficiently different.
This is a boxplot before normalization of the count data (Image extracted from **"boxplot_before_normalization.pdf"** file):



This is a boxplot after normalization of the count data (Image extracted from **"boxplot_normalized_data.pdf"** file):

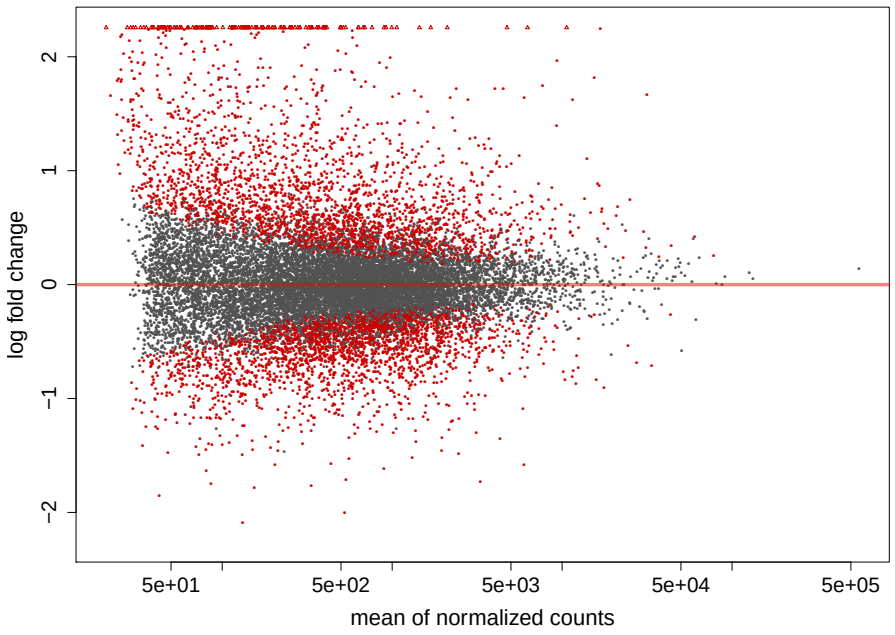


3 Graph outputs of the employed analysis methods

Different plots concerning every differential gene expression algorithm are shown.

3.1 DESeq2 MA plot

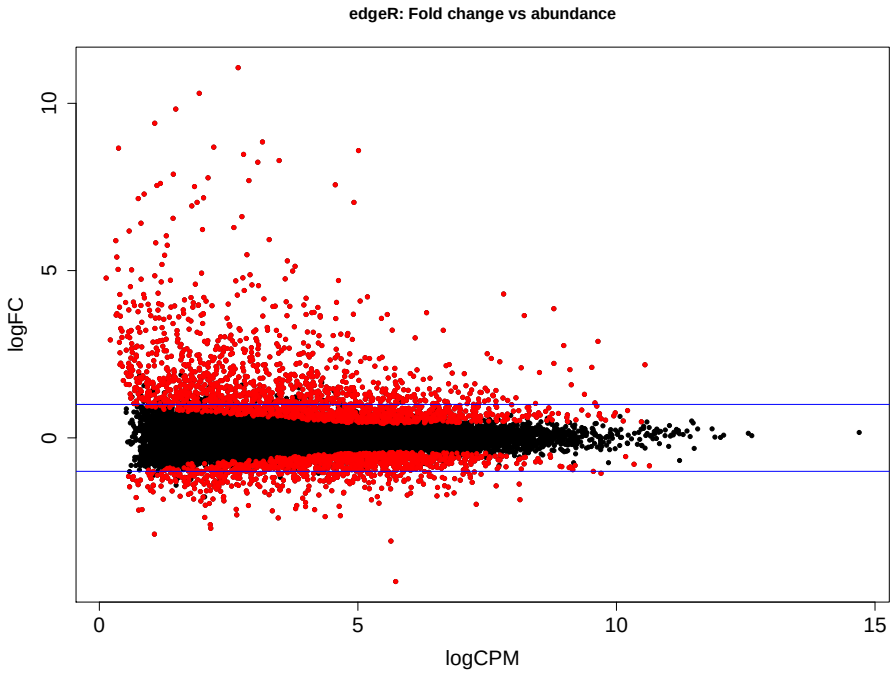
This is the MA plot from DESeq2 package (Image extracted from **"MA_plot_DESeq2.pdf"** file):



In DESeq2, the MA-plot (log ratio versus abundance) shows the log2 fold changes are attributable to a given variable over the mean of normalized counts. Points will be colored red if the adjusted p value is less than 0.1. Points which fall out of the window are plotted as open triangles pointing either up or down. A table containing the DESeq2 DEGs is provided in **"DEgenes_DESEq2.txt"**. A table containing the DESeq2 normalized counts is provided in **"Normalized_counts_DESEq2.txt"**.

3.2 edgeR MA plot

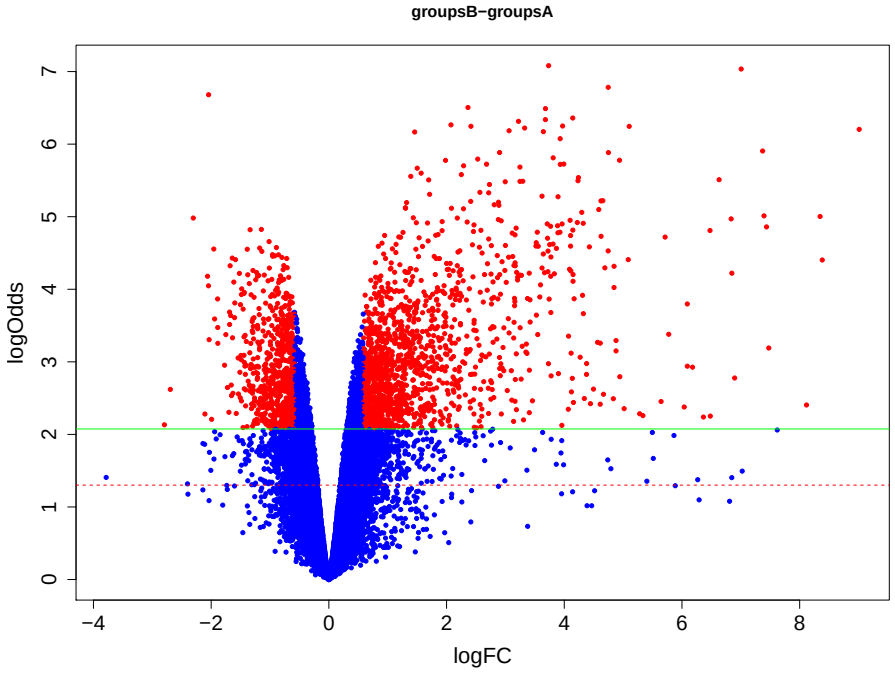
This is the MA plot from DESeq2 package (Image extracted from **"MA_plot_edgeR.pdf"** file):



Differential gene expression data can be visualized as MA-plots (log ratio versus abundance) where each dot represents a gene. The differentially expressed genes are colored red and the non-differentially expressed ones are colored black. A table containing the edgeR DEGs is provided in **"DEgenes_edgeR.txt"**. A table containing the edgeR normalized counts is provided in **"Normalized_counts_edgeR.txt"**.

3.3 limma Volcanoplot

This is the volcano plot of log-fold changes versus log-odds (log10 of adjusted p values) of differential expression from limma package (Image extracted from **"Volcanoplot_limma.pdf"** file):

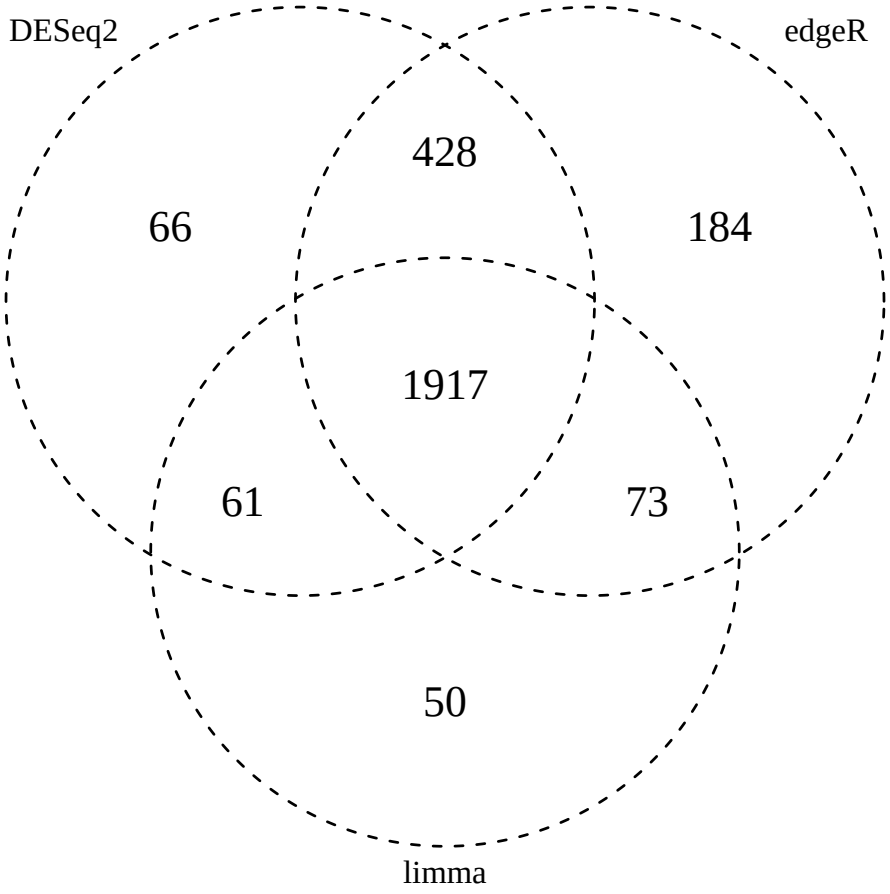


A table containing the limma DEGs is provided in **"DEgenes_limma.txt"**. A table containing the limma normalized counts is provided in **"Normalized_counts_limma.txt"**.

4 Graph concerning common results

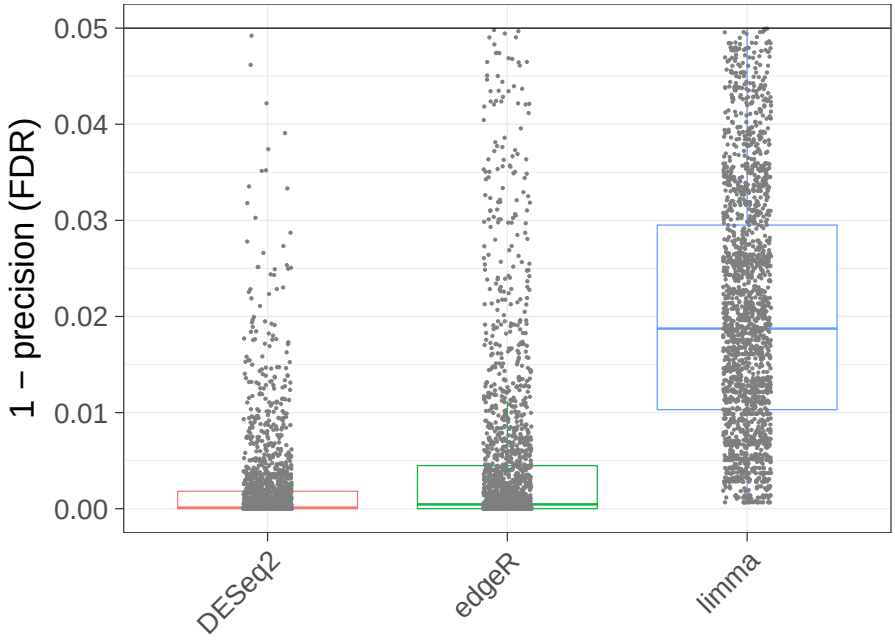
4.1 Venn Diagram

This is the Venn Diagram of the all possible DEGs encountered in the experiment (Image extracted from **"VennDiagram.pdf"** file):



4.2 FDR gene-wise benchmarking

Benchmark of false positive calling (Image extracted from **"padj_prevalent_DEGs.pdf"** file):



Estimates of P(p-value < 0.05) under the null hypothesis among prevalent DEGs are shown.

The complete results of the DEgenes Hunter differential expression analysis can be found in the "hunter_results_table.txt"