Differential gene expression data output

DESEQ2 output file column information

Col1 : many times the raw output will contain just numbers in this column. This will be cleaned most of the times when data is delivered to the collaborator

Col2: geneid/row.names: Names of genes

Col3: baseMean: average of the normalized count values over all samples.

Col4: log2FoldChange: It is the effect size estimate. It tells us how much the gene’s expression has changes due to the treatment (relative to the control or reference).

Col5: lfcSE : This column represents the standard error estimate for the log2FoldChange estimate

Col6: stat: Wald test statistics

Col7: pval : This column represents the pvalue for the pairwise comparison

Col8: padj: This column represents the Benjamini Hochberg adjustment to deal with false positives in your dataset. You have to apply a false discovery rate of 0.05 (5%FDR) or 0.1 (10%FDR) to obtain a list of significantly expressed genes.

In addition to the above column, which is the standard DeSeq2 output, you will see multiple columns (one for each sample that went into the analysis). These additional columns contain values of normalized read counts for the individual samples and these values will be identical for all the files.

The following table indicated what condition is up/down expressed relative to the other condition in the pairwise condition.

|  |  |
| --- | --- |
| **File** | **what does the data represent?** |
| **refA\_B.norm.SIGS** | **Significantly expressed genes with padj 0.05 or 5% FDR (irrespective of direction)** |
| **refA\_B.norm.SIGS.NEGL2FC** | **Up expressed in A relative to B** |
| **refA\_B.norm.SIGS.POSL2FC** | **Down expressed in A relative to B** |