

Introduction to DGE

[View on GitHub](#)

Approximate time: 20 minutes

Learning Objectives

- Summarizing significant differentially expressed genes for each comparison
- Extracting lists of significant genes

Summarizing results

To summarize the results table, a handy function in DESeq2 is `summary()`. Confusingly it has the same name as the function used to inspect data frames. This function when called with a DESeq results table as input, will summarize the results using a default threshold of $\text{padj} < 0.1$. However, since we had set the `alpha` argument to 0.05 when creating our results table threshold: $\text{FDR} < 0.05$ (padj/FDR is used even though the output says $\text{p-value} < 0.05$). Let's start with the OE vs control results:

```
## Summarize results
summary(res_tableOE, alpha = 0.05)
```

In addition to the number of genes up- and down-regulated at the default threshold, **the function also reports the number of genes that were tested (genes with non-zero total read count), and the number of genes not included in multiple test correction due to a low mean count.**

Extracting significant differentially expressed genes

Let's first create variables that contain our threshold criteria. We will only be using the adjusted p-values in our criteria:

```
### Set thresholds
padj.cutoff <- 0.05
```

We can easily subset the results table to only include those that are significant using the `filter()` function, but first we will convert the results table into a tibble:

```
# Create a tibble of results
res_tableOE_tb <- res_tableOE %>%
  data.frame() %>%
  rownames_to_column(var="gene") %>%
  as_tibble()
```

Now we can subset that table to only keep the significant genes using our pre-defined thresholds:

```
# Subset the tibble to keep only significant genes
sigOE <- res_tableOE_tb %>%
  filter(padj < padj.cutoff)
```

```
# Take a quick look at this tibble
sigOE
```

Exercise

MOV10 Differential Expression Analysis: Control versus Knockdown

1. Using the same p-adjusted threshold as above (`padj.cutoff < 0.05`), subset `res_tableKD` to report the number of genes that are up- and down-regulated in Mov10_knockdown compared to control.
2. How many genes are differentially expressed in the Knockdown compared to Control? How does this compare to the overexpression significant gene list (in terms of numbers)?

Now that we have extracted the significant results, we are ready for visualization!

This lesson has been developed by members of the teaching team at the [Harvard Chan Bioinformatics Core \(HBC\)](#). These are open access materials distributed under the terms of the [Creative Commons Attribution license \(CC BY 4.0\)](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Some materials and hands-on activities were adapted from [RNA-seq workflow](#) on the Bioconductor website

DGE_workshop_salmon_online is maintained by **hbctraining**.

This page was generated by [GitHub Pages](#).