Day 8: Advanced DESeq2 Experimental Design (Worksheet)

By: Samuel Hunter

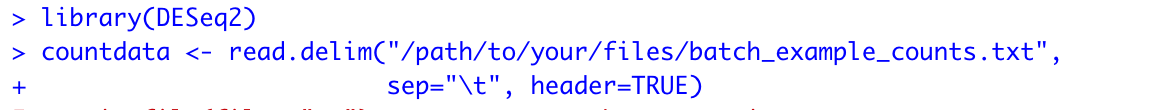
We will be working through some “advanced” DESeq2 designs. On your first day with DESeq2, you learned how to run a pairwise comparison and interpret/visualize your results. But our experimental designs are often more complex than just a simple pairwise comparison. How can we incorporate more design elements into DESeq2?

Download all files in /scratch/Shares/public/sread2022/data\_files/day8 from AWS to your local computer. You’ll need to replace all of the paths to these files in the instructions with the local path you stored them to. So, for the first step, I would replace “/path/to/your/files/batch\_example\_counts.txt “ with

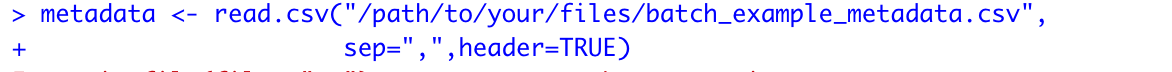
“/Users/samuelhunter/sread2023/day8/data\_files/batch\_example\_counts.txt”

**Part 1: Batch Effect Correction**

1. Load in the counts file for batch correction, along with our DESeq2 library:



1. Load in the metadata file for batch correction:



1. Organize the counts file into a DESeq2 compatible matrix, and check the metadata file:

Table

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

A picture containing graphical user interface

Description automatically generated

1. For now we aren’t using batch information, only sample treatment information.
2. Run DESeq2:

A picture containing text

Description automatically generated

1. View “NM\_016817” in results file. Take note of the Log2FoldChange and LfcSE values (This is a strong IFN response gene)

Text, letter

Description automatically generated

Visualize the results with an MA plot:



And get a summary printout of the results



1. Now fetch the normalized counts for each sample, and make a plot of this gene. You’ll notice some substantial batch effects



Chart, box and whisker chart

Description automatically generated

1. You can also use a PCA plot to view this on a global scale:



1. Run DESeq2 again, adding batch as a term to our design formula. View “NM\_016817” again.

Text, letter

Description automatically generated

You’ll notice that the same gene now has a smaller lfcSE but a similar Log2FoldChange. This is because we’ve explained some of that standard error by the differences in basal levels in each batch. But, each batch responded similarly, so the log2FC doesn’t shift much. We can now more confidently call this gene as significant (padj is much smaller now)

Visualize the results with an MA plot and summarize results as before:



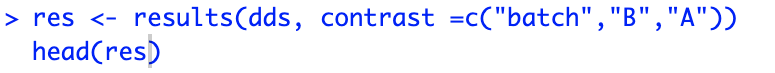


**Part 1.1: ComBat-Seq**

**MARY ENTERS COMBAT-SEQ PRACTICE**

**Part 1.2: Contrasts**

1. You can also take a look at the magnitude of the batch effects by using a contrast:



1. Contrasts have the following format: c(“name of metadata column”, “name of entry you want in the numerator”, “name of entry you want in the denominator”). Reverse the entries and look at the differences in the fold change

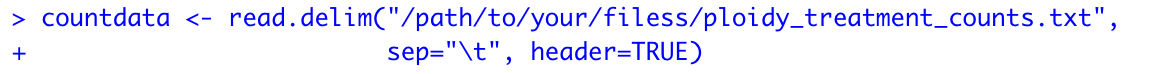
A screenshot of a computer

Description automatically generated with medium confidence

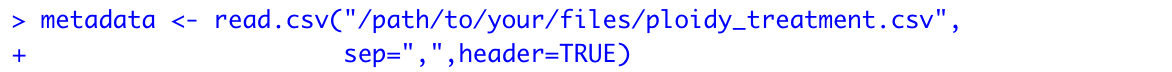
1. Notice that the signs of the fold changes are reversed, but everything else remains the same. All we’ve done is flip the numerator and denominator

**Part 2: Interaction Coefficients**

1. Load in the counts file containing multiple treatments



1. Load in the metadata. Notice we have multiple columns of interest. How can you load in all of the metadata information into the design matrix?



1. Run DESeq2 with an interaction term in the design matrix:

Graphical user interface, application

Description automatically generated

1. Find the results for the interaction term for the interferon-response gene “NM\_016817”

Text

Description automatically generated with medium confidence

How do we interpret these results? Put short, the interaction term here is the treatment effect of IFN in T21 minus the treatment effect of IFN in D21. Since our log2FoldChange here is close to 0, it means that this gene responds similarly to IFN for both T21 and D21.