

Modelling Biological Data – RiboMed Summer School

RNA-Seq Practical

Objective

On this practical, you will run a differential expression analysis workflow using the DESeq2 package. It uses the negative binomial model that we discussed earlier. However, it fits one model per gene that is available on the experiment (i.e., if you have data for 10,000 genes, then DESeq2 fits 10,000 negative binomial models).

By the end of this section, you're expected to understand:

- The nature of the negative binomial model for RNA-Seq counts
- The format of Bioconductor containers
- The principles of pre-filtering
- The basic results of a DESeq2 run
- The problems of multiple comparisons
- Explore the results of DESeq2
- The interpretation of multifactorial models

Activities

1. Access the “**Analyzing RNA-Seq Data with DESeq2**” vignette :
<https://bioconductor.org/packages/release/bioc/vignettes/DESeq2/inst/doc/DESeq2.htm>
2. Start at the “**Count matrix input**” section
 - a. What does the `system.file()` command do?
 - b. What does the `factor()` command do?
 - c. What is the result of the `DESeqDataSetFromMatrix()` command?
 - d. What are the components of a `DESeqDataSet` object?
 - i. How do you subset it?
 - ii. What’s in the **countData** slot?
 - iii. What’s in the **colData** slot?
 - iv. What does **design** mean?
3. Skip to the “**Pre-filtering**” section
 - a. Why would we remove genes prior to the analysis?
4. Go to the “**Differential expression analysis**” section
 - a. What are the steps performed by the `DESeq()` command?
 - b. What is size factor (s_j) and how is it used in the negative binomial model?
 - c. Interpret the output returned by `results()`.
 - d. What is the multiple comparison problem?
 - e. How do you compare conditions of interest using `DESeq2`?
5. At the “Exploring and exporting results” section
 - a. How do you interpret the MA-plot?
 - b. What is the relationship between the log-foldchange and the parameters estimated by the negative binomial model?
 - c. How can you extract (sample) data for a specific gene to create a plot of the counts?
6. Skip to the “Multi-factor designs” section
 - a. How do you add more predictors to the negative binomial model?
 - b. How do you compare gene expressions for different predictors?