Sequencing: RNA-seq data intro

Koen Van den Berge

8/6/2021

Contents

Start with challenges below, and make a section for each. This will simultaneously follow a full RNA-seq analysis pipeline.

Challenges:

- Choice of modeling assumptions (distribution)
- Normalization
- Parameter estimation under limited information setting
- High dimensionality (many genes), multiple testing

Following code should be part of a next file where we start working with RNA-seq data, show the mean-variance trend, discuss normalization, and introduce a DE analysis.

1 Properties of (RNA-seq) count data

Goals:

- Work with real RNA-seq count data, gene by sample matrix
- Explore the data for a single gene: univariate (large range, zeroes, discrete, skewed) as well as bivariate wrt covariate.
-
- Demonstrate the mean-variance relationship by plotting mean and variance across genes. Show that the Poisson doesn't hold across biological replicates and use this as introduction for the negative binomial distribution.

In this lecture, we will introduce working with count data, using a real bulk RNA-seq dataset from Haglund *et al.* (2012). We will be importing this dataset using the data package parathyroidSE from Bioconductor.

```
if (!requireNamespace("BiocManager", quietly = TRUE)){
  install.packages("BiocManager")
}
if(!"SummarizedExperiment" %in% installed.packages()){
  BiocManager::install("SummarizedExperiment")
}
# install package if not installed.
if(!"parathyroidSE" %in% installed.packages()) BiocManager::install("parathyroidSE")

library(parathyroidSE)
library(SummarizedExperiment)

# import data
data("parathyroidGenesSE", package="parathyroidSE")
# rename for convenience
se <- parathyroidGenesSE
rm(parathyroidGenesSE)
```

- Count data are inherently discrete.

```
y <- assays(se)$counts[1,]
hist(y, breaks = ncol(se),
     xlab = "Gene expression")
```

1.1 Mean-variance relationship

1.2 Relative uncertainty, offsets and count scaling

Defer to normalization.

2 Variance-stabilizing transformations

Defer to when talking about dim red.