# Interpreting uncertainty in differential expression with DESeq2

Modrák Martin1

1 Bioinformatics Core Facility, Institute of Microbiology of the Czech Academy of Sciences

While differential expression analysis is often reported primarily as a list of genes where the hypothesis of no/low change can be rejected, further insights may be gained by examining the reported uncertainty of the log fold change. In a simulation study we investigate the coverage properties of confidence intervals derived from the reported standard errors in DESeq2. We note that with the most straightforward normal approximation, the tail (> 90%) confidence intervals are overly narrow and have coverage around 1 – 5 percentage points lower than nominal across a range of settings. Using a student-t approximation can overcorrect and result in too wide confidence intervals. Moreover, if only differentially expressed genes are selected for reporting, an adjustment of the confidence intervals for multiple comparisons is necessary to avoid overconfident claims. We note that this mild discrepancy is completely compatible with p-values from DESeq2 providing tight control of false positives.

The same simulation setup also allows us to judge how closely the empirical Bayes shrinkage methods implemented in DESeq2 approximate an exact Bayesian posterior. We note that while the approximation is imperfect, Bayesian interpretation of uncertainty intervals from DESeq2 as “likely” containing a “true” value of the fold change is not completely unwarranted, especially if higher number of replicates and/or a student-t approximation is used.