An introduction to RNA-seq differential expression analysis

Antti Honkela

Helsinki Institute for Information Technology HIIT,
Department of Computer Science, University of Helsinki, Finland

8 January 2014

Differential expression analysis using RNA-seq

Objective: analyse differences in expression between two biological conditions.

Assay: sequence cDNA derived from mRNA from multiple biological replicates from both conditions.

Analysis: identify entities (genes, exons, transcripts, . . .) that are differentially expressed

What is differential expression?

What is differential expression?

- Difference in expression between two conditions
- But
  - Expression values in each condition form a distribution, how to compare these?
  - Expression is a continuous quantity, hence it will always be different.
- NB: without replicates it is impossible to quantify the difference!

Simple RNA-seq data analysis workflow

1. Align reads from each sequencing experiment to a reference genome/transcriptome
2. Quantify the expression of entities of interest
   - Often by counting the number of overlapping reads
3. Perform statistical testing to find statistically significantly differentially expressed entities

Differential expression analysis checklist

- Sufficient replication needed!
- Expression quantification needs to be done with care
  - Different ways to count reads for different purposes
  - Accurate analysis needs to take alternatively spliced transcripts into account
- Normalisation is very important
  - Sequencer gives relative proportions, they may not be comparable if total mRNA volume has changed
- Multiple testing issues in statistical analysis
  - Drawing the line between differentially expressed and non-differentially expressed entities can be tricky