Comprehensive comparison of RNA-seq based methods for differential splicing analysis

G. Speshilov

IITP RAS, Bolshoy Karetny per. 19, build.1, Moscow, 127051 Russia, gspeshilov@kirby.unsw.edu.au

Post-transcription maturation of many eukaryotic mRNAs requires splicing: removal of some parts of pre-mRNA called introns and joining of remaining parts called exons. Alternative Splicing (AS), cell type- or condition-specific removal or retention of particular introns, was shown to play an important role in cell differentiation and organ development, whilst its disruption my serve as a possible reason of many pathological conditions.

A great number of computational methods have been developed for detecting and analysis of AS events from RNA-seq data. The assumption that mRNA concentration in the original biological sample is proportional to the number of RNA-seq reads mapped to corresponding genome region probably is the only common feature of all these methods. The aim of this work is to benchmark existing computational differential splicing detection methods so that users can choose the most suitable tools according to their tasks.

We have performed unbiased comparison of 23 different up-to-date AS evaluating tools: DEXSeq, Cuffdiff2, SAJR, MISO, MATS, SpliceTrap, SplicingCompass, GLiMMPS, AltAnalyze, FDM, PSGInfer, DiffSplice, rDiff, DSGSeq, GPSeq, SOLAS, ALEXA-Seq, JuncBASE, JETTA, SpliceSeq, RSEM-EBSeq, BitSeq, ARHseq, which allow usage of pre-existing exon annotation on datasets. 13 methods could not have been launched due to internal software errors or lack of proper manuals. Two types of datasets was used in this benchmark: experimental dataset obtained from NCBI SRA database (SRP002628) and several simulated sets with controlled properties (the presence or absence of differential expression along with AS and the magnitude of both these effects).

Huge differences between approaches put results in a strong dependence on method used and complicate comparison of outcomes obtained in different studies. Our results shows that while few methods such as Cuffdiff2, MISO, MATS, SpliceTrap, RSEM-EBSeq, BitSeq and

ARHseq exhibit low sensitivity and high false discovery rate (FDR), results of other methods such as DEXSeq, SAJR and DSGSeq have comparably high quality. Unfortunately, DEXSeq and, at lower degree Cuffdiff2 and DSGSeq, appear to produce a lot of false positive outcomes in the presence of differential expression. Strikingly, even in absence of differential expression, results of these best-performing methods exhibit very low agreement.

As a result no single method performs the best in all situations. Nevertheless, it was shown that FDR drops dramatically for the results found by several methods at a time. For instance, all but one AS events selected by at least 7 methods are true positives in generated dataset.