Unbiased comparison of alignment tools for splice junction detection from RNA-Seq data

Alberto Gatto[1], Fátima Sánchez-Cabo[1], Carlos Torroja[1] and Enrique Lara[1]

[1] Centro Nacional de Investigaciones Cardiovasculares, Spain
{agatto, fscabo@cnic.es}

Abstract. RNA-Seq makes it possible not only to measure gene expression but also to identify and quantify transcript isoforms in different experimental conditions. Although a large number of tools allow to infer isoform expression levels from RNA-Seq data, the quantification of alternative splicing variants remains challenging. Aligning reads that span (possibly un-annotated) exon-exon junctions in an effective manner is the crucial and most demanding step in a typical analysis workflow. Yet it is unclear which of the many available methods, or combination of methods, is the most suited to accurately detect junction-spanning reads based on the experimental setup. To address this question we systematically compared the performance of 5 alignment algorithms (TopHat, GSNAP, SOAPsplice, OLego, STAR) using simulated data for different sequencing protocols. Based on our data we propose clear guidelines for RNA-Seq data processing that will ease standard analysis aiming at the detection and quantification of alternative splicing events.