Alternative splicing analysis of RNA-seq data using SAJR

Thomas Doktor (thomaskd@bmb.sdu.dk)
Andresen's lab, University of Southern Denmark
Caroline Heintz
Harvard T. H. Chan School of Public Health
Brage Andresen
Andresen's lab, University of Southern Denmark
William Mair
Harvard T. H. Chan School of Public Health

Method Article

Keywords: RNA-seq, alternative splicing

Posted Date: March 12th, 2018

DOI: https://doi.org/10.1038/protex.2018.029

Abstract

Here we describe a fast an efficient protocol for analyzing alternative splicing using SAJR.

Introduction

Here we describe a fast an efficient protocol for analyzing alternative splicing using ‘cutadapt’ \[1] to trim reads before alignment with ‘STAR’ \[2], subsequent merging of samples using ‘samtools’ \[3] and finally analysis of splicing with ‘SAJR’ \[4]. We also added annotation of novel splicing events and conversion of SAJR specific ids to standard gene ids using ‘bedtools’ \[5] and custom perl scripts.

Procedure

First part of the protocol is preparing and mapping reads. 1. Trim reads to remove adapter sequences. Example using ‘cutadapt’ and Nextera adapters: cutadapt –trim-n -m 15 -o trimmed.S1_1.fastq.gz -p trimmed.S1_2.fastq.gz -a CTGTCTCTTATACACATCTCCGAGCCCACGAGA -A CTGTCTCTTATACACATCTGACGCTGCCGACGA S1_1.fastq.gz S1_2.fastq.gz 2. Align the samples to the genome using ‘STAR.’ 3. Merge all BAM files into a single BAM file using ‘samtools merge’. Second part of the protocol is preparing a reference as well as identifying novel splicing patterns and annotating these. 4. Convert a GTF reference to an SAJR specific GFF reference using SAJR’s annotation conversion mode. 5. Run SAJR in de novo annotation mode to find novel splice-forms using the merged BAM file and the known annotation to produce a novel annotation, novel.gff 6. Run SAJR in annotation comparison mode to compare the novel annotation with the known annotation and use get_genename_from_junction_comparison.pl to filter the results: get_genename_from_junction_comparison.pl sajr.comp > sajr.novel2known.tsv 7. Use bedtools and get_genename_from_segment_overlap.pl to associate SAJR ids with known gene ids from the reference: bedtools intersect -s -f 1.0 -loj -a novel.gff -b known.gff > novel_overlap_known.gff get_genename_from_segment_overlap.pl novel_overlap_known.gff > novel2known_from_overlap.tsv 8. Use bedtools and annotate_novel_segments.pl to annotate novel spliced regions: bedtools intersect -s -f 1.0 -r -loj -a novel.gff -b known.gff > novel_overlap_known_stringent.gff annotate_novel_segments.pl novel_overlap_known_stringent.gff > novel_overlap_known_stringent_novel.tsv The final part of the protocol is estimating inclusion levels in each sample, and testing for differences between groups of samples. 9. Run SAJR in count mode for each sample using the novel.gff reference. 10. Use the R package part of SAJR to identify alternative splicing, see sajr_analysis.R for an example workflow incorporating annotation of novel spliced regions.

Timing

The whole analysis can be completed within 24 hours for 36 samples with a total of app. 450 mio reads running on 16 cores.

Anticipated Results

The expected outcome is a list of significant alternative splicing events with with optional indication of novel splicing patterns.

References

1 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011 17, doi:10.14806/ej.17.1.200 pp. 10-12 \(2011). 2 Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15-21, doi:10.1093/bioinformatics/bts635 \(2013). 3 Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079, doi:10.1093/bioinformatics/btp352 \(2009). 4 Mazin, P. et al. Widespread splicing changes in human brain development and aging. Molecular systems biology 9, 633, doi:10.1038/msb.2012.67 \(2013). 5 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842, doi:10.1093/bioinformatics/btq033 \(2010).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement0.zip