Gene expression

Jutils: a visualization toolkit for differential alternative splicing events

Guangyu Yang¹, Leslie Cope², Zitong He¹ and Liliana Florea ¹, 3, *¹

¹Department of Computer Science, Johns Hopkins University, Baltimore, MD 21205, USA, ²Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA and ³Department of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA

*To whom correspondence should be addressed.

Associate Editor: Can Alkan

Received on January 3, 2021; revised on May 17, 2021; editorial decision on May 20, 2021; accepted on May 21, 2021

Abstract

Motivation: Gene alternative splicing plays an important role in development, tissue specialization and disease and differences in splicing patterns can reveal important factors for phenotypic differentiation. While multiple computational methods exist to determine splicing differences, there is a need for user-friendly visualizations that present an intuitive view of the data and work across methods.

Results: We developed a toolkit, Jutils, for visualizing differential splicing events at the intron (splice junction) level. Jutils is method-agnostic, converting individual tools’ output into a unified representation and using it to create visualizations. Jutils creates three types of visualizations, namely heatmaps of absolute and Z-score normalized splice ratios, sashimi plots and Venn diagrams of results from multiple comparisons. Jutils is lightweight, relying solely on the unified data file for visualizations.

Availability and implementation: Jutils is implemented in Python and is available from https://github.com/Splicebox/Jutils.

Contact: florea@jhu.edu

1 Introduction

Alternative splicing is a fundamental gene regulatory mechanism in plants and animals. On the evolutionary scale, alternative splicing has contributed to species diversification, and within organisms it has been broadly implicated in physiology and disease. Alternative splicing has high prevalence in eukaryotic species, with more than 95% of human genes having multiple splice isoforms (Pan et al., 2008) and high levels reported in other species. Characterizing splicing variation between cellular conditions is therefore important to identify molecular markers of phenotypic differentiation.

While multiple methods have been developed to determine differential splicing patterns from RNA-seq data (LeafCutter, MAJIQ, rMATS, MtJULiP; Li et al., 2018; Shen et al., 2014; Vaquero-Garcia et al., 2016; Yang et al., 2020), there is a scarcity of tools to present the results to the user in a way that is intuitive and easy to explore. Moreover, most visualization tools are designed for a particular differential splicing method, such as rmat2sashimiplot (https://github.com/Xinglab/rmat2sashimiplot) and LeafViz (https://leafcutter.shinyapps.io/leafviz/), and are not adapted for general use. To fill this gap, we developed Jutils, a toolkit for visualizing alternative splicing differences that can be used across methods.

2 System design

Jutils works with the output of a differential splicing tool, converting it into a unified data file that contains the information necessary for the visualizations. (Additional information, such as the BAM files, can be optionally provided.) Metadata about experiment design, such as the condition associated with each sample, can be provided in a specification file. Jutils then extracts events to include in the visualizations based on user specified criteria. Lastly, it generates one of three types of visualizations: heatmap, sashimi plot and Venn diagram. Details of each component are provided below.

2.1 The unified file format

Jutils uses an intermediate Tab Separated Values (TSV) file format to collect event information generated by a differential splicing program, which it then uses to create visualizations. Jutils has built-in output conversion modules for several analysis programs, including LeafCutter, MAJIQ, MtJULiP and rMATS, and users can develop their own conversion scripts for other programs of interest.

The TSV file has 14 columns containing, in order: gene name, group id, feature id, feature type (e.g. intron, exon skipping), feature label (derived from the chromosomal location), strand, p-value, q-value, dPSI (difference in Percent Splice In values), read count 1,
read count 2 and PSI values, per sample and averaged by condition. The default feature for Jutils is introns, but the program can represent more complex events such as those reported by rMATS, for instance exon inclusion and exon exclusion constituted as a single exon skipping event. Programs may further aggregate features into groups, for instance LeafCutter and MntJULiP group introns that share a splice junction. Jutils supports identifiers and operations on individual features as well as groups. Read count 1 and read count 2 represent vectors of per sample values and correspond to the paired splice forms in a complex feature (e.g. exon skipping), whereas for simple features (e.g. introns) read count 2 is marked with ‘.’.

2.2 Heatmaps

Jutils generates heatmaps of differential splicing events represented in the TSV file. A metadata file contains the classification of each sample. Jutils generates heatmaps of PSI values, either Z-score normalized or absolute values (Fig. 1A). The software allows clustering by rows (events) and columns (samples), using different distance metrics and clustering methods. By default, the ‘cityblock’ (Manhattan distance) metric with the ‘weighted’ (weighted pair group with arithmetic mean) method is used. Events can be filtered at run time based on quality and confidence measures such as p-value, q-value and dPSI and the user may choose to visualize all relevant features or select a representative feature per group or per gene. Lastly, while Jutils is intended to work primarily with the output of differential splicing tools, it can also be used to display the features with the highest variance (option ‘unsupervised’).

2.3 Sashimi plots

Sashimi plots have been previously introduced to visually represent differences in splicing. A traditional sashimi plot shows raw RNA-seq densities along with exons and junctions for multiple samples. The Jutils sashimi visualization utilizes a modified version of the ggsashimi package (Garrido-Martín et al., 2018) to display graphical representations of intron read counts within a specified genomic region, intron or intron group (Fig. 1B). By default, Jutils provides a lightweight representation based solely on the intron read counts provided in the TSV file, without the flanking exon read depth information. When alignment files are also provided, Jutils extracts alignments from the BAM files and provides full sashimi representations reflecting the accurate exonic coverage.

2.4 Venn diagrams

Methods for differential splicing detection employ a variety of models for features and objective functions. Therefore, it becomes desirable to compare the outputs of different programs, and for different parameter settings, to obtain a complete view of the predicted differential splicing and to gauge support from multiple methods. Jutils provides a Venn diagram visualization of the predicted gene sets (Fig. 1C), along with a text file containing the list of genes in each category.

In summary, we developed a lightweight toolkit, Jutils, for visualizing differential alternative splicing between cellular conditions. Jutils can be used automatically with the popular differential splicing analysis tools LeafCutter, MAJiQ, MntJULiP and rMATS, and can be easily adapted to any other program, and thus represents a useful and practical tool to explore the landscape of alternative splicing.

Funding

This work was supported in part by grants [R01GM129085, R01GM124531] from the National Institutes of Health.

Conflict of Interest: none declared.
References

Garrido-Martín, D. et al. (2018) Ggsashimi: sashimi plot revised for browser and annotation-independent splicing visualization. PLoS Comput. Biol., 14, e1006360.

Li, Y. I. et al. (2018) Annotation-free quantification of RNA splicing using LeafCutter. Nat. Genet., 50, 151–158.

Pan, Q. et al. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat. Genet., 40, 1413–1415.

Shen, S. et al. (2014) rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-seq data. Proc. Natl. Acad. Sci. USA, 111, E593–E5601.

Vaquero-García, J. et al. (2016) A new view of transcriptome complexity and regulation through the lens of local splicing variations. Elife, 5, e11752.

Yang, G. et al. (2020) Comprehensive and scalable quantification of splicing differences with MntJULiP. bioRxiv, 2020.10.26.355941.