SPLICE-q: a Python tool for Genome-Wide Quantification of Splicing Efficiency

Verônica Melo Costa [1,2], Julianus Pfeuffer [3], Evgenia Ntini [2], Ulf Andersson Ørom[4], Rosario M. Piro [2,5]

1-Max Planck Institute for Molecular Genetics, Berlin, Germany; 2-Freie Universität Berlin, Berlin, Germany
3-Universität Tübingen, Tübingen, Germany; 4-Aarhus Universitet, Aarhus C, Denmark
5-Charité - Universitätsmedizin Berlin, Berlin, Germany

Contact: melo@molgen.mpg.de veronica_rmc

1. INTRODUCTION

Eukaryotic genes are mainly composed of a number of exons intercalated by introns that are generally removed from pre-mRNAs to form mature RNA molecules **(Figure 1).** This post-transcriptional process is called splicing which consists basically of a series of hydrolysis and ligation reactions led by the spliceosome.

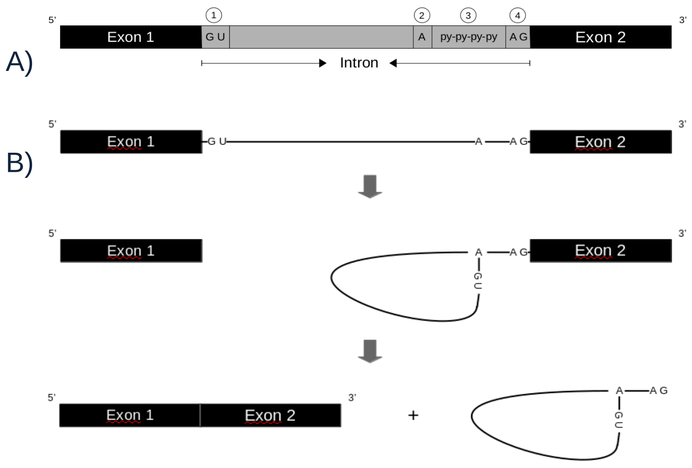

Figure 1: Exons and introns in pre-mRNA. Exons are represented in black and introns grey. (1) Donor-site with consensus sequence GU at the 5' region; (2) Branchpoint, which includes an adenine nucleotide involved in lariat formation in the splicing reaction. This region consists in a sequence of 20-50 nucleotides; (3) Polypyrimidine tract; (4) Acceptor-site with consensus sequence AG at the 3' region.

2. METHODS

An efficient pre-mRNA splicing is essential and its mis-regulation is related to numerous human diseases. To better understand the dynamics of this process and the perturbations that might be caused by aberrant transcript processing it is important to quantify its efficiency (q). RNA-seq allows these analyses from a genome-wide point of view (**Figure 2**).

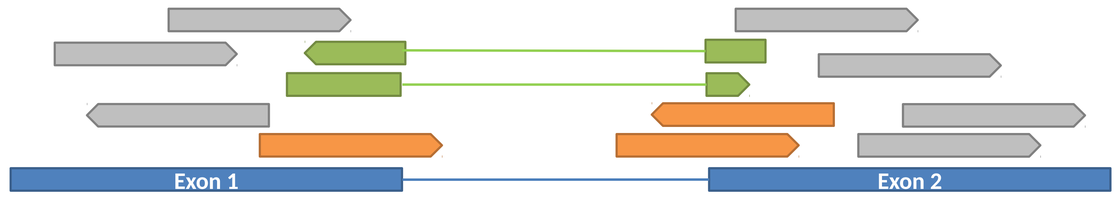

Figure 2: Read alignment visualization scheme.

$$q = \frac{\text{5'SJ split reads + 3'SJ split reads}}{\text{[(5'SJ split reads + 3'SJ split reads)+(5'SJ nonsplit reads + 3'SJ nonsplit reads)]}}$$

3. SPLICE-q

There are different frameworks for calculating splicing efficiency (SE) from RNA-seq data, but the bioinformatics steps involved might be challenging, especially for experimental biologists. Thus, we present **SPLICE-q** - a complete, up-to-date and easy-to-use python tool for genome-wide quantification of SEs from total RNA-seq data **(Figure 3).**

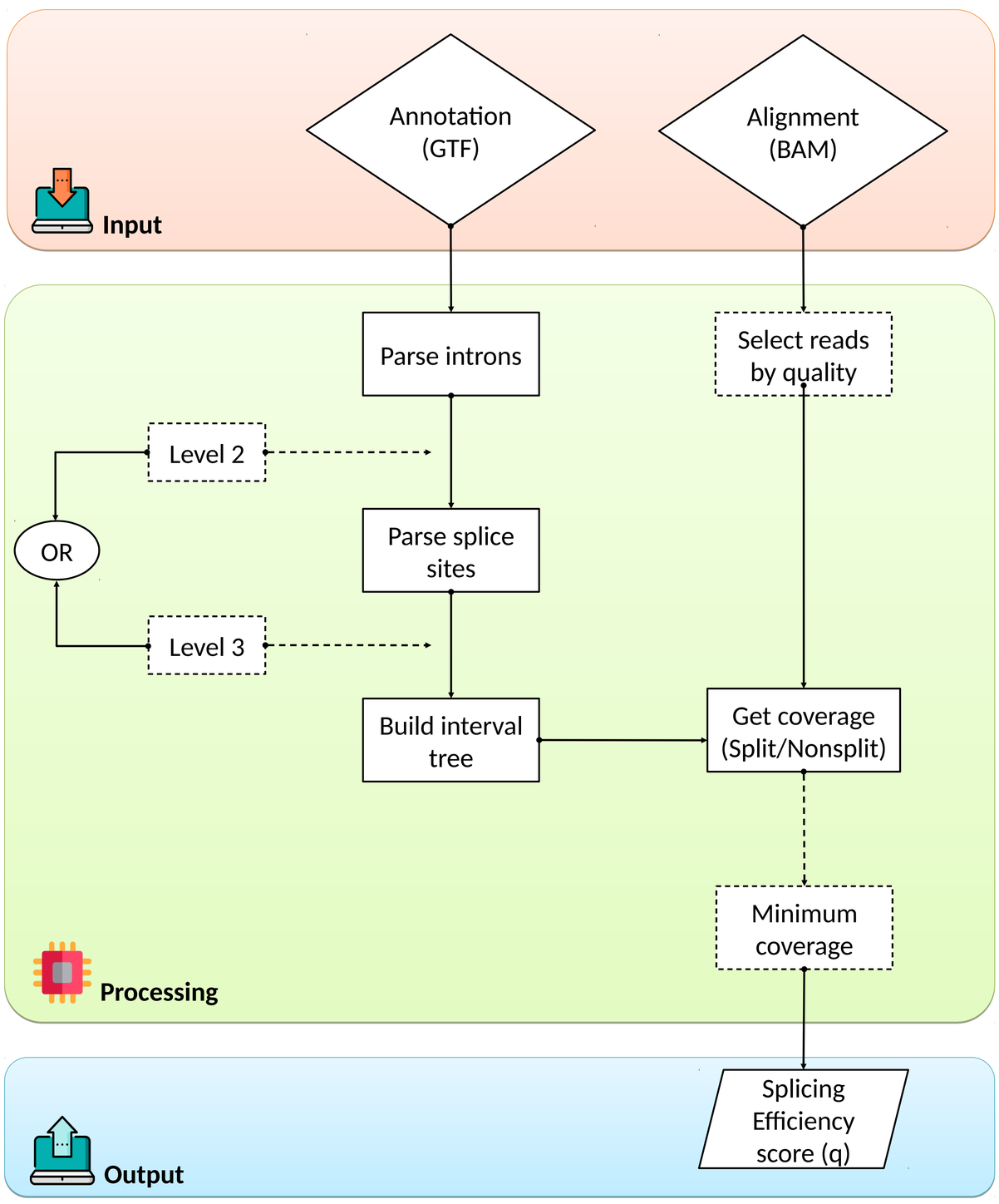

Figure 3: SLICE-q's Workflow.

SPLICE-q allows the user to select different levels of restrictiveness for filtering, including:

Level 1	Keep all introns in the genome regardless of overlaps with other genomic elements.
Level 2	Select only introns whose splice junctions do not overlap any exon in different genes.
Level 3	Select only introns that do not overlap with any exon of the same or different gene.

Other filtering can also be set up according to minimum coverage, read quality, and minimum intron length.

4. RESULTS: Splicing Dynamics in Gliomas

A glioma is a type of brain or spinal cord tumor that comprises about 30% of all brain and central nervous system tumors, and 80% of all malignant brain tumors.

- Low-grade: well-differentiated (not anaplastic); benign tendencies.
- High-grade: undifferentiated or anaplastic; these are malignant and carry a worse prognosis.

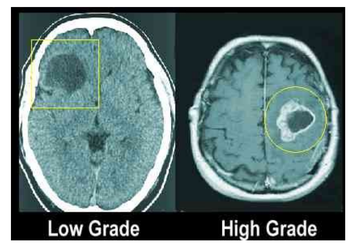

Figure 4: Appearance of a Glioblastoma on an MRI Brain Scan (About Cancer 2019).

A 47-year-old Asian male patient was diagnosed with oligodendroglioma and 72 months later anaplastic oligodendroglioma.

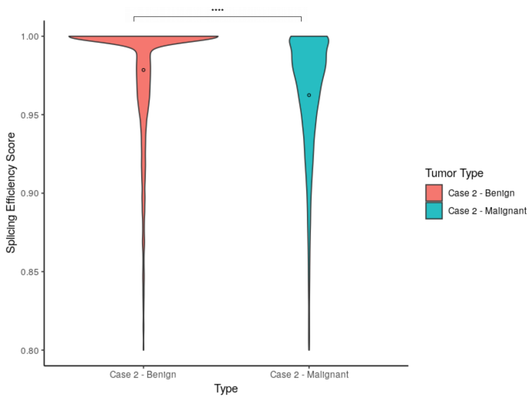

Figure 5: **Splicing Efficiency in Gliomas**. Benign Vs. Malignant.

Future work: What biological features might be associated with different patterns of splicing dynamics?

Peptidylprolyl isomerase B (PPIB) participates in many biological processes as well as in cancer. Currently, PPIB is employed as a biomarker for various types of tumors.

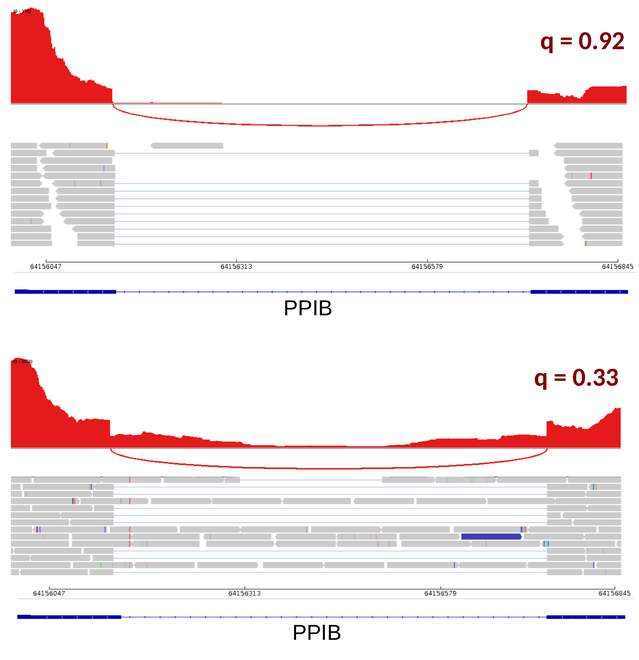

Figure 6: Splicing Efficiency - PPIB Intron 4.

REFERENCES

- Park, Chul-Kee, et al. "Genomic dynamics associated with malignant transformation in IDH1 mutated gliomas." Oncotarget 6.41 (2015): 43653.
- Mukherjee, Neelanjan, et al. "Integrative classification of human coding and noncoding genes through RNA metabolism profiles." Nature structural & molecular biology 24.1 (2017): 86.
- Louloupi, Annita, et al. "Transient N-6-methyladenosine transcriptome sequencing reveals a regulatory role of m6A in splicing efficiency." Cell reports 23.12 (2018): 3429-3437.

FUNDING

CAPES Max-Planck-Institut für molekulare Genetik

12TH ANNUAL INTERNATIONAL CONFERENCE ON SYSTEMS BIOLOGY OF HUMAN DISEASE
May 2019