





# SnakeSplice: A seamless workflow for splice variant discovery and visualization in RNA-Seq data



Oliver Küchler<sup>1,2</sup>, Felix Boschann<sup>1,3</sup>, Johannes Kopp<sup>1,4,5</sup>, Björn Fischer-Zirnsak<sup>1,4</sup>, Dominik Seelow<sup>1,2</sup>

1. Institute of Medical Genetics and Human Genetics, Charité - Universitätsmedizin Berlin, Freie Universität zu Berlin, Germany; 2. Exploratory Diagnostic Sciences, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Germany; Berlin, Germany; 3. BIH Biomedical Innovation Academy, BIH Charité Clinician Scientist Program, Berlin Institute of Health (BIH) at Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany; 4. Max Planck Institute for Molecular Genetics, Berlin, Germany; <sup>5.</sup> Institute of Chemistry and Biochemistry, Department of Biology, Chemistry and Pharmacy, Freie Universität Berlin, Berlin, Germany

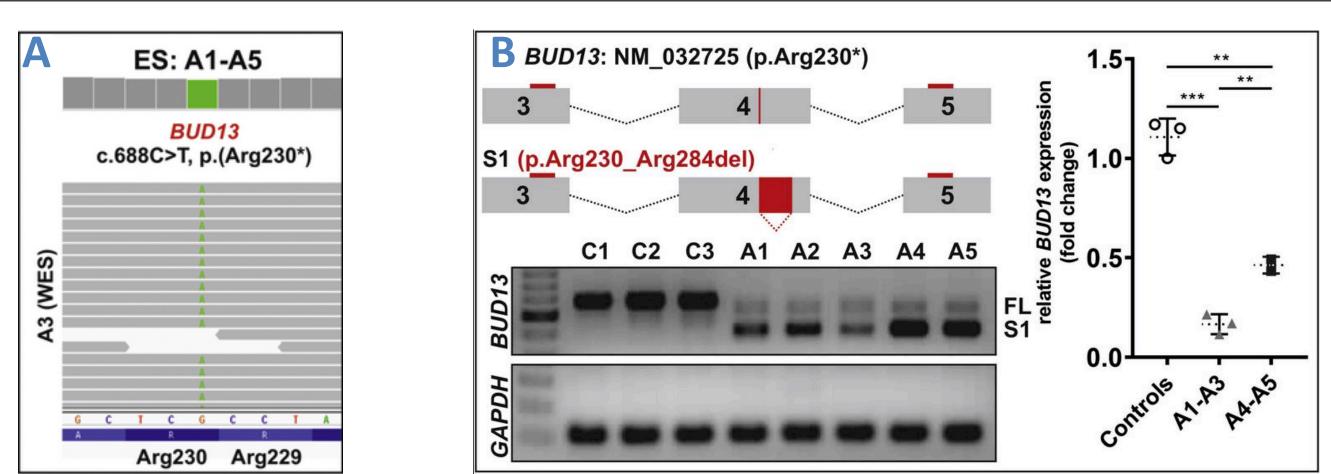
### **Abstract**

Splicing defects are a major cause of human genetic disease, yet detecting them from RNA-Seq data remains a complex and fragmented process. SnakeSplice is a scalable, Snakemake-based workflow that streamlines RNA-Seq analysis from quality control to splice variant detection. It combines robust preprocessing, diverse splicing analysis tools, and novel junction discovery in a single, scalable workflow optimized for high-performance computing. It enables efficient analysis of large RNA-Seq datasets and generates dynamic, interactive HTML reports that make results easy to interpret across research and clinical settings. By bridging deep analysis with intuitive visualization, SnakeSplice empowers transcriptome-driven disease gene discovery and variant prioritization.

#### SnakeSplice pipeline Input **Module 2: Detect Gene Fusions** Configuration file Reference files (Genome & Transcriptome) Alignmen<sup>.</sup> for fusion events output Gene annotation file **ARRIBA Fusion Table** Sample annotation file Module 1: Quality Control, Preprocessing and Alignment 5. Overview 4. Assess 2. Quality Control 1. Adaptor & Low-quality 3. Alignment alignment report sequence trimming ssential steps QualiMap <u>Multi⊕</u>C STAR / OLego **Trimmomatic** MultiQC FastQC QualiMap E. Alignment A. Analyse B. Check for potential C. Explore feature D. Alignmentsample contaminations distribution based PCA statistics strandedness ALFA check\_strandedness.py Kraken2 DeepTools - PCA Samtools: BamStats **Module 3: Transcript Quantification & Expression Analysis Transcript Discovery B:** Leveraging A: Expectation Reference **Module 5: Variant Calling** pseudo-alignments maximisation Reference variants RSEM Kallisto / Salmon **Gene Expression Analysis** Catch expression **DEG** analysis outliers **OUTRIDER** DROP - Outrider **Module 4: Splice Pattern Analysis Differential exon Identification of** Identification of **Detection of Private** Summary Differential splicing analysis Reference Alignment intron retentions splicing outliers **Splice Junctions** output **DEXSeq** FRASER Leafcutter Whippet rMATS **DEXSeq DROP - FRASER** PJD-Tool **IRFinder** HTML Uncovering both known and novel step: Generation of an

## **Application case:** Alternative splicing of *BUD13*

interactive, shareable HTML report



splicing events with confidence

A: A homozygous nonsense variant in BUD13 (c.688C>T; p.Arg230\*) was identified in five individuals presenting with syndromic lipodystrophy and variable disease severity\*. B: The variant leads to aberrant splicing and production of a truncated transcript (S1), whose abundance correlates with patient survival. C: HTML report displaying SnakeSplice's analysis of RNA-Seq data from individual A1-A5 highlights this splicing defect and its molecular signature.

\* KORNAK, Uwe, et al. Alternative splicing of BUD13 determines the severity of a developmental disorder with lipodystrophy and progeroid

features. Genetics in Medicine, 2022, 24. Jg., Nr. 9, S. 1927-1940.

## SnakeSplice report: Alternative splicing of BUD13

