

Identification of Tumor-Specific Alternative Splicing Events Using an Outlier-Based Computational Approach

Zsolt Kovács^{1*}, Katalin Szabo¹, Gabor Nagy², Eszter Toth³

¹ Department of Cancer Genomics and Alternative Splicing, Faculty of Medicine, Semmelweis University, Budapest, Hungary.

² Department of Outlier-Based Detection and Transcriptomics, Faculty of Medicine, University of Debrecen, Debrecen, Hungary.

³ Department of Tumor-Specific Splicing Events, Faculty of Medicine, University of Pécs, Pécs, Hungary.

*E-mail ✉ zsolt.kovacs@semmelweis.hu

Abstract

Alternative splicing generates protein diversity that plays a critical role in disease pathogenesis and tumor development. To address this, we created OutSplice, a straightforward, accessible computational tool that detects splicing outliers in cancerous samples by benchmarking them against patterns observed in normal tissues. Although multiple existing software packages have been introduced to identify splicing anomalies, they differ considerably in their underlying methods, technical demands, and overall usability, often complicating tool selection for researchers new to the field. We therefore evaluated several such algorithms through systematic benchmarking to clarify their respective strengths and to illustrate how OutSplice complements or diverges from them. Results show that, while OutSplice flags fewer genes with statistically significant splicing aberrations, those it does identify tend to carry strong biological significance. Moreover, we uncovered 17 genes with consistent and prominent splicing changes in tumor samples that were independently detected by at least five of the tested tools, highlighting them as high-priority targets for subsequent investigation. In conclusion, the most effective strategy is to integrate OutSplice with complementary splicing analysis programs to confirm irregular events and prioritize those of greatest biological importance.

Keywords: Algorithm, Benchmarking, Alternative splicing, Head and neck cancer, Software

Introduction

Alternative splicing represents a fundamental biological mechanism in eukaryotes, characterized by the selective inclusion, exclusion, or rearrangement of exons within pre-mRNA transcripts. This process enables the assembly of diverse coding sequences from a single gene, thereby substantially increasing protein diversity while also modifying protein function and expression levels [1]. The resulting protein variations arising from this

post-transcriptional regulation constitute a major driver of functional alterations and contribute to the development of various diseases, including cancer [2]. Extensive research has established that dysregulated splicing events are central to oncogenesis across multiple cancer types. For instance, the cell-surface receptor CD44 produces multiple splice variants, and their overexpression has been linked to tumor initiation and metastatic progression in breast and colorectal cancers [3-5]. Similarly, exon skipping within the proto-oncogene MDM2 generates the MDM2-ALT1 isoform, which promotes accelerated tumor formation in rhabdomyosarcoma [6].

Assessing the contribution of splicing abnormalities to cancer progression requires not only detecting individual variants but also evaluating their tumor specificity. In head and neck cancer, prior work has revealed that

Access this article online

<https://smerpub.com/>

Received: 20 March 2026; Accepted: 18 June 2026

Copyright CC BY-NC-SA 4.0

How to cite this article: Kovács Z, Szabo K, Nagy G, Toth E. Identification of Tumor-Specific Alternative Splicing Events Using an Outlier-Based Computational Approach. J Med Sci Interdiscip Res. 2026;6(1):250-63. <https://doi.org/10.51847/yAxde8CGmZ>

certain splicing alterations can activate oncogenic pathways independently of mutations [7]. Moreover, tumor-specific splicing changes can produce neoantigens, offering promising opportunities to improve immunotherapy efficacy [8-10]. Although somatic mutation analysis has successfully identified neoantigens in many cancers, such approaches may be less effective in tumors exhibiting low mutational loads [11, 12]. This limitation is especially relevant in head and neck cancer, where tumors often display heterogeneous mutational profiles alongside extensive splicing dysregulation [8, 13]. Consequently, a detailed understanding of tumor-specific alternative splicing is essential for elucidating disease mechanisms and for developing targeted immunotherapeutic strategies.

In earlier research, we introduced an algorithm to identify splicing outliers by detecting junction expression deviations in tumors relative to normal controls [7]. Over time, various computational methods and software tools have emerged for alternative splicing analysis. These encompass transcript isoform quantification [14-17], differential exon usage [18-21], percent spliced-in (PSI) calculations [22-26], and the construction of splicing graphs [27-31]. Although many of these approaches are robust, they often involve greater complexity than outlier-based junction analysis. They are primarily designed to detect shifts in average isoform usage, which may not adequately accommodate the inherent heterogeneity of tumor transcriptomes [32]. To overcome these limitations, we developed OutSplice, a Bioconductor package with a GenePattern interface that implements our outlier-based framework. This tool facilitates the straightforward detection of novel splicing events on a per-tumor basis [7, 33, 34]. Given the diversity of available differential splicing tools, it is important to evaluate which algorithms best align with individual research requirements. Prior comparative studies have mainly emphasized performance metrics such as precision, recall, and false discovery rate [35, 36]. While these measures are valuable, a comprehensive assessment should also consider the full analytical pipelines, including variations in usability, input requirements, and output formats. Such factors, particularly ease of implementation and reduced need for advanced scripting, are especially beneficial for researchers entering the field.

Accordingly, we benchmarked OutSplice against several established differential splicing detection algorithms—edgeR, LeafCutter, psichomics, rMATS, and Whippet—

using metrics that included the number of significant events detected, computational runtime, memory usage, input/output specifications, and overall usability. The evaluation employed an RNA-seq dataset comprising 47 oropharyngeal squamous cell carcinoma (OPSCC) tumors and 25 normal samples. In addition, a synthetic RNA-seq dataset was generated to assess each tool's accuracy against known ground-truth alternatively spliced transcripts. The selected algorithms represent major methodological categories in splicing analysis. Transcript isoform estimation methods were excluded from this comparison because leading tools in this category, such as cuffdiff and MISO, are hindered by prohibitively long runtimes or are restricted to pairwise sample analyses. Given the methodological overlap between psichomics and OutSplice in junction read quantification, psichomics was included alongside the widely used rMATS. We further compared OutSplice with other outlier-focused approaches, specifically FRASER and LeafCutterMD [37, 38]. This study delineates OutSplice's distinctive characteristics relative to these tools. It highlights 17 genes exhibiting aberrant splicing that were consistently identified across at least five algorithms, thereby conferring high confidence in their biological relevance.

Materials and Methods

Data preparation

Differential splicing analyses were conducted on 47 primary tumor specimens sourced from individuals with HPV-positive oropharyngeal squamous cell carcinoma (OPSCC) [39]. Matched normal controls consisted of 25 oropharyngeal mucosal specimens collected during uvulopalatopharyngoplasty (UPPP) procedures in non-cancer patients. RNA was isolated using the mirVana miRNA Isolation Kit (Ambion, Foster City, CA, USA), and concentration was measured with a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Sequencing libraries were generated via the Illumina TruSeq Stranded Total RNA with Ribo-Zero Gold kit (San Diego, CA, USA) and sequenced on the HiSeq 2500 system (Illumina, San Diego, CA, USA) utilizing the TruSeq Cluster Kit, achieving an average sequencing depth of 80 million 100×100 paired-end reads per sample. After sequencing, reads were trimmed to remove adapter sequences and low-quality bases. The processed compressed FASTQ files had file sizes ranging from 1.4 gigabytes (GB) to 6.6 GB.

Computational resources

All algorithms were executed on the National Resource for Network Biology (NRNB) cluster provided by the San Diego Supercomputer Center (SDSC). Job scheduling and resource monitoring were managed via the Simple Linux Utility for Resource Management (SLURM). Each step within every pipeline received a maximum allocation of 8 CPUs. We captured elapsed runtime and maximum resident set size (maxRSS) for each processing stage. When pipelines involved per-sample computations, array jobs facilitated parallel processing, and reported runtimes reflected the mean value computed across all samples.

Genome index building and alignment

For tools necessitating aligned reads in Binary Alignment Map (BAM) format, read alignment and genome indexing were performed using STAR version 2.7.1a [40]. Required reference files, including annotation and assembly data, were downloaded from GENCODE [41]. Trimmed reads extracted from FASTQ files were aligned to the hg38 human genome assembly. Alignment accuracy for splice junctions was enhanced through the application of STAR's two-pass mode [42].

Algorithm data formatting pipelines

Implementation of each algorithm followed the official user guides and documentation precisely, starting from raw FASTQ input files. This strategy was chosen to reflect the typical experience of researchers processing their own sequencing data. Default settings were used throughout, modified only where applicable to specify reverse-strand library type, utilize 8 processing threads, and enable paired-end read handling. Statistical significance was uniformly defined as an FDR-adjusted p-value of 0.05. All resulting splicing events received annotation based on the hg38 genome assembly.

OutSplice pipeline

OutSplice extracted junction read counts directly from the SJ.out.tab files produced by STAR alignments. Total raw counts were derived from the #Uniquely Mapped Reads statistic reported in the Log.Final.out file. Corresponding gene expression levels were prepared according to the TCGA RSEM v2 normalization procedure with RSEM v1.3.1 [15, 43], using the same GENCODE-derived reference files for transcript quantification.

A dedicated utility, the OutSplice Formatter, was created to help users prepare input data. This tool parses STAR and RSEM output files to generate the required junction counts, raw counts, and gene expression matrices. It further incorporates optional upper-quartile normalization through a Perl script obtained from the University of North Carolina Bioinformatics utilities (<https://github.com/mozack/ubu>, accessed on 16 March 2023). The formatter is freely available as open-source software on GitHub (<https://github.com/GuoLabUCSD/OutSpliceFormatter/tree/Version1.0.0>, accessed on 16 March 2023).

Upon execution of OutSplice version 1.0.1, STAR-derived raw junction counts are converted to reads per million (RPM) through division by total uniquely mapped reads. Normalization by gene expression follows by dividing these RPM values by the upper-quartile adjusted counts from RSEM. Outlier detection relies on a rank-sum statistic adapted from the Ghosh method, with a threshold requiring normalized RPM values above 0.00001 relative to gene expression levels [44, 45]. The OutSplice package is distributed as open-source software and is available through both Bioconductor and the GenePattern environment (<https://bioconductor.org/packages/release/bioc/html/OutSplice.html>, accessed on 26 April 2023).

edgeR pipeline

Differential splicing analysis was conducted with edgeR version 3.42.2. Read counts for edgeR were generated via the featureCounts function from the Rsubread package v2.14.2 [46, 47]. Default parameters were retained except for those required to accommodate paired-end reads, exclude meta-features, and allow multiple overlaps, in accordance with the edgeR user guide recommendations for alternative splicing studies (<https://bioconductor.org/packages/release/bioc/vignettes/edgeR/inst/doc/edgeRUsersGuide.pdf>, accessed on 16 March 2023). Gene filtering retained only genes with at least 10 read counts in the smallest sample and at least 15 counts across the entire dataset. Data normalization was carried out using the default trimmed mean of M-values (TMM) approach.

Leafcutter/LeafcutterMD pipeline

STAR-generated BAM files were indexed with SAMtools v1.15.1 and then converted to JUNC files using RegTools v0.5.2 [48, 49]. Parameters included an anchor length of 8 base pairs (bp) and intron lengths

ranging from 70 to 500,000 bp. Intron clustering was executed using a Python script from LeafCutter v0.2.9, requiring a minimum of 30 split reads to define a cluster for input into the differential splicing module. Annotation files derived from the hg38 GENCODE reference were supplied to LeafCutter's visualization script to support the LeafViz tool. The LeafCutterMD analysis repeated the same preparatory steps but was restricted to a subset of 25 tumor samples for direct comparison with FRASER.

Psichomics pipeline

Analysis of differential splicing utilized psichomics version 1.26.0 (https://nuno-agostinho.github.io/psichomics/articles/custom_data.html, https://nuno-agostinho.github.io/psichomics/articles/CLI_tutorial.html, accessed on 16 March 2023). Junction counts were sourced from the SJ.out.tab output produced by STAR. Consistent with the software documentation, events were retained only if they displayed a change in percent spliced in (PSI) exceeding 0.1, thereby reducing the inclusion of events lacking clear biological importance.

rMATS pipeline

Differential splicing detection was performed with rMATS version 4.1.0. Input consisted of BAM files aligned using STAR. A fixed read length of 100 was declared, and the variable read length parameter was activated to properly handle the trimmed sequencing reads.

Whippet pipeline

A Julia script included with Whippet v1.6.1 was first applied to quantify reads from the original FASTQ files. At the same time, a custom index was constructed utilizing the same GENCODE annotation and assembly files employed for STAR alignment. The generated PSI files were then submitted to Whippet's "delta" module for differential splicing evaluation. Detected events were filtered to retain only those showing a PSI change above 0.1 and a minimum splicing probability of 90%.

FRASER pipeline

Outlier detection was carried out using FRASER version 1.12.1 on a set of 25 tumor samples. Split and non-split reads were quantified from aligned BAM files through the countRNAData function. Filtering and normalization required at least 20 split reads per intron and an absolute

PSI change of 0.3. Model hyperparameters were optimized and fixed before fitting (<http://www.bioconductor.org/packages/release/bioc/vignettes/FRASER/inst/doc/FRASER.pdf>, accessed on 16 March 2023).

Simulated data creation

The initial 1000 genes located on chromosome 1 were extracted from the GENCODE hg38 GTF file. Genes lacking characterization, ribosomal protein S (RPS) genes, and entries without an assigned HGNC symbol were excluded. For each selected gene, we identified the Ensembl Canonical transcript from the GTF annotation. In addition, an alternative splice variant was chosen for 200 of these genes. These transcript definitions served as input for Polyester to simulate FASTA files containing reads for 20 "Normal" and 40 "Tumor" samples, with a baseline expression of 600 reads per transcript [50]. To mimic tumor heterogeneity and establish ground-truth overexpression, expression levels were varied relative to the baseline: canonical transcripts ranged from 1 to 2× baseline, while alternative transcripts were set at 1× in normal samples, 1–2× in 20 tumor samples, and 4–5× in the remaining 20 tumor samples. These 200 transcripts were designated as true positives for subsequent performance metric calculations. The Polyester-generated output was converted to FASTQ format using BBTools (<https://sourceforge.net/projects/bbmap/>, accessed on 16 May 2022). All algorithms were executed following the protocols described earlier, with the sole adjustment of using an unstranded library type. For OutSplice, the fold-change threshold was lowered to 2 (from the default of 10) to better match the reduced expression variation in the simulated dataset.

Results and Discussion

OutSplice

Tumor heterogeneity poses a substantial challenge in cancer splicing studies, as it complicates the reliable detection of both cohort-wide splicing patterns and sample-specific events. To address this issue, OutSplice was purposefully developed to accommodate the complexities of tumor transcriptomes when comparing splicing profiles between tumor and normal tissues. The algorithm incorporates a modified implementation of the Ghosh method within its software framework to identify outlier splicing events, applying a minimum expression threshold relative to normal controls [44, 45]. For each

tumor specimen, normalized junction expression values are evaluated against the distribution derived from normal samples, thereby flagging individual outlier events on a per-sample basis. Subsequently, Fisher's exact test with false discovery rate (FDR) correction is applied to each event to compare outlier frequencies between groups. This process enables the detection and enumeration of both overexpressed and underexpressed splicing alterations that occur in a statistically significant proportion of tumors. OutSplice also calculates the overall splice burden and categorizes event types within each sample, enabling rapid identification of specimens with the greatest degree of aberrant splicing and the predominant classes of events involved. This outlier-based strategy has previously proven effective in uncovering functionally consequential splice variants in AKT3, DOCK5, and LOXL2 that drive oncogenesis in head and neck cancer [7, 33, 34]. Integrating these outlier

analytics into a comprehensive, user-friendly software package with a GenePattern notebook interface significantly expands its utility for differential splicing investigations in oncology.

OutSplice ranks among the differential splicing tools that demand the fewest processing steps (**Figure 1**). Both the data formatter and core algorithm can be executed within a single function call, eliminating the need for users to perform manual normalization or intermediate data manipulation. Moreover, the package includes built-in compatibility with data directly downloaded from TCGA/Firebrowse. Firebrowse provides pre-processed TCGA datasets across numerous cancer types, including junction and gene expression information. This seamless integration offers considerable advantages for investigators exploring splicing alterations in diverse malignancies.

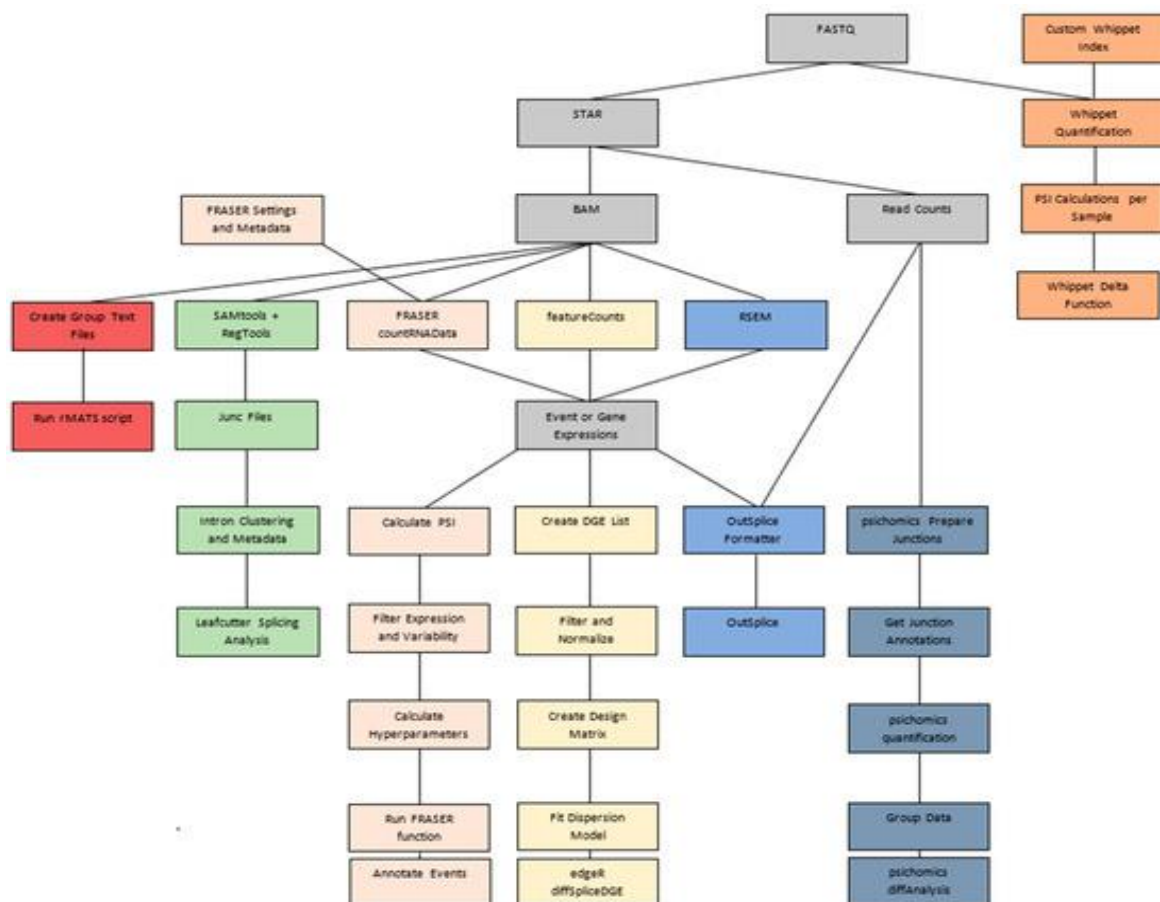


Figure 1. Algorithm pipeline comparisons starting with FASTQ file input. Gray = Input/Output/Tools required by more than one algorithm. Abbreviations: Orange = Whippet, Green = LeafCutter, Pink = FRASER, Yellow = edgeR, Blue = OutSplice, Aegean = psychomics, and Red = rMATS.

Benchmarking of computational performance showed that OutSplice requires relatively high memory usage and has longer total runtimes than several other differential splicing tools. This is primarily attributable to its reliance on external programs, STAR and RSEM, for generating the necessary expression counts. Specifically, OutSplice utilizes junction-spanning reads reported in STAR's SJ.out.tab files rather than all overlapping reads, as employed by tools such as edgeR. This junction-centric approach can more precisely delineate specific splicing events, rather than broader exon-level expression changes. Although using STAR's gene-level counts would accelerate processing, incorporating RSEM for gene expression normalization may offer greater accuracy because it better handles ambiguous and multimapping reads [15].

edgeR

edgeR identifies splicing changes by quantifying differences in exon-level expression between sample groups, calculated as the log fold change in exon counts relative to the log fold change observed at the gene level [18]. Exon expression is modeled using a negative binomial distribution fitted to a read count matrix generated by auxiliary tools such as STAR and featureCounts. These programs aggregate all reads mapping to each exon, thereby estimating the expression of individual coding segments.

When initiated from raw FASTQ files, the complete edgeR workflow exhibited one of the longest processing times and demanded the most extensive data preparation among the evaluated tools. The primary bottleneck was featureCounts, which took the longest to quantify reads across all 72 samples and assemble the count matrix. Excluding the alignment and counting phases, edgeR itself runs quickly; however, the preparatory stages require custom R scripting. Although comprehensive user guides are available, inexperienced researchers may encounter difficulties with data-specific issues. For instance, standard GENCODE GTF files often contain paralogous genes and ENSEMBL identifiers that produce duplicate or non-standard row names, triggering edgeR errors. Consequently, successful implementation depends more heavily on user expertise than several alternative tools.

Despite these considerations, edgeR provides substantial advantages. As a Bioconductor package, it offers straightforward installation and supports the broadest

range of data processing and statistical options. This flexibility is particularly valuable for investigators employing alternative normalization strategies, complex experimental designs, or large cohort sizes. Furthermore, edgeR generates clear, interpretable output as a single matrix listing all exons with automatically computed FDR-adjusted p-values, simplifying the identification of significant events.

Leafcutter

LeafCutter detects splicing differences by examining intron excision patterns derived from clusters of split reads spanning exon-exon junctions, rather than relying on exon expression levels. This strategy enables the discovery of splice junctions independently of potentially incomplete reference annotations [22]. Intron counts within each cluster are then modeled and compared using a Dirichlet-multinomial framework.

Excluding alignment and initial read quantification, LeafCutter required modestly longer computation times than most other tools. It involved additional manual pre-processing steps, including BAM file indexing with SAMtools and conversion to JUNC format using RegTools. These preparatory tasks are relatively quick but may require access to a computing cluster capable of parallel processing for large datasets, along with familiarity with additional bioinformatics utilities. Importantly, these steps perform read counting internally, removing the need for separate expression quantification software, and require modest memory. For users who already have aligned BAM files, LeafCutter is a practical option when computational resources are limited, though it may increase overall runtime.

The software produces results that include FDR-corrected cluster-level p-values for each gene, percent spliced in (PSI) values for individual events, and—when supplied with a GTF annotation—labels distinguishing canonical from cryptic junctions. Cluster-level FDR correction may complicate assessing significance for specific events. Mapping events to genes also requires users to generate custom annotation codes from a GTF file using the provided script. Nevertheless, event-level PSI values help evaluate the prevalence of individual splicing changes, and canonical junction annotations greatly facilitate identifying novel splicing events.

Psichomics

To detect splicing differences between groups, psichomics tools like LeafCutter rely on reads spanning exon-exon junctions to compute inclusion levels. However, rather than grouping overlapping introns through split-read clusters, psichomics determines the percent spliced in (PSI) value for every event in each sample by calculating the proportion of junction reads supporting the inclusion isoform [24].

Psichomics offers notable advantages in data handling. Although the workflow involves several sequential steps that require basic R knowledge when executed via the command line, the software provides a graphical user interface that substantially simplifies and accelerates the process. This feature makes it particularly suitable for researchers lacking programming experience. Furthermore, similar to OutSplice, psichomics supports immediate analysis of data obtained from TCGA/Firebrowse, enabling rapid evaluation of cancer datasets.

Performance benchmarking demonstrated that psichomics exhibited the lowest memory consumption and one of the shortest overall runtimes among the tested algorithms. The tool bypasses the need for separate programs to generate count matrices by directly utilizing junction information from STAR's SJ.out.tab files following alignment and performing sample-specific normalization during PSI computation. However, users without pre-existing junction count matrices must first run STAR, which increases memory requirements.

Additional strengths of psichomics lie in its comprehensive output features. These include FDR-corrected p-values for event detection, PSI values, and event classification. The package also incorporates built-in principal component analysis, allowing users to identify which splicing events contribute most strongly to group separation. A key limitation is its dependence on pre-existing alternative splicing annotations generated by external resources such as VAST-TOOLS [51]. Although default hg38 annotations are supplied, this reliance restricts the reliable discovery of previously uncharacterized splicing events [24].

rMATS

Like LeafCutter and psichomics, rMATS calculates inclusion levels using read counts that span splice junctions. It further incorporates read counts from the flanking exons to improve event detection. The underlying statistical model uses a hierarchical framework that combines binomial and logit-normal

distributions to account for estimation uncertainty and variability across replicates [26].

The primary advantage of the rMATS pipeline is its minimal number of required steps. After installation, users need only prepare simple text files defining the treatment and control groups before executing a single comprehensive function. Nevertheless, rMATS had one of the longest runtimes and the highest memory demands, requiring substantial computational resources to run effectively.

rMATS output includes several informative metrics. These encompass automatic gene symbol annotation for each event, FDR-adjusted P-values, the magnitude of difference in inclusion levels between groups, and classification of the splicing event type.

Whippet

Whippet performs differential splicing analysis by constructing splicing graph models that represent each sample's transcriptome [27]. It first generates an index from the user-supplied GTF annotation and genome FASTA sequence. It then aligns raw RNA-seq reads directly to this contiguous splicing graph, enabling the tool to process FASTQ files without requiring a separate aligner such as STAR.

One of Whippet's key strengths is that it operates without BAM files, unlike every other algorithm evaluated in this study. This characteristic makes it particularly advantageous for researchers with limited access to high-performance computing infrastructure or large memory resources. Whippet can also incorporate aligned BAM files when desired to improve index construction. Users experienced in merging BAM files may find this option helpful for uncovering unannotated splicing regions.

In terms of results, Whippet reports event categories and percent spliced in (PSI) values but does not generate statistical p-values for significance. Users are instead advised to apply filters based on a probability score of 90% or higher and a PSI difference exceeding 0.1, with customization possible depending on the experimental design. The software additionally computes entropy values that reflect uncertainty in the splicing calls. Although valuable, these metrics may be challenging for less experienced users to interpret and select appropriately. Another limitation is the lack of automatic conversion from Ensembl IDs to gene symbols, requiring supplementary annotation packages such as AnnotationDBI or biomaRt.

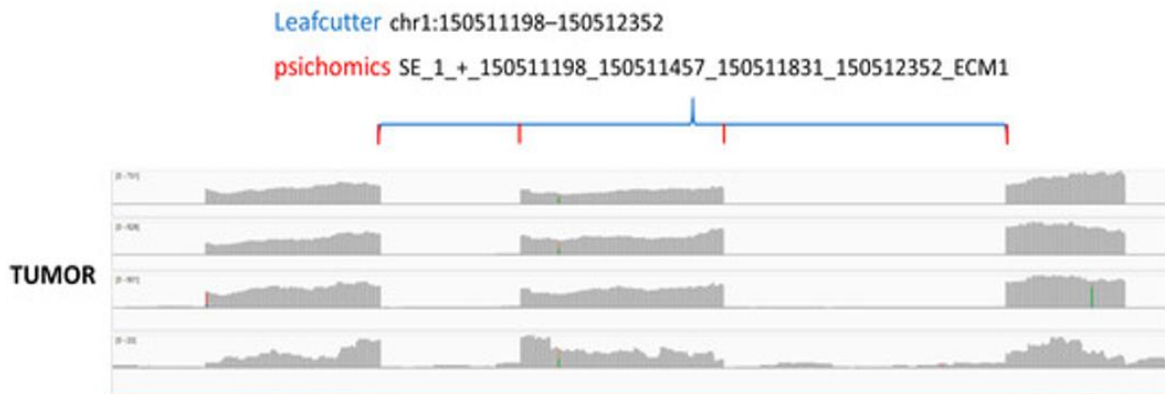
Gene overlap and algorithm comparisons

An initial comparison was performed between outcomes from standard differential splicing methods and the outlier detection strategy. After processing results from each algorithm, splicing events were mapped to their corresponding genes, and the counts of unique genes showing significant alterations were compared to evaluate concordance. LeafCutter and rMATS detected the largest sets of genes with significant splicing changes (3159 and 3691 genes, respectively), whereas OutSplice yielded the smallest number (261 genes). This disparity arises in part from OutSplice's default exclusion of junctions located on the X and Y chromosomes, since splicing differences in cancer are generally independent of biological sex. Consequently, OutSplice did not capture 86 genes significant under LeafCutter and 115 under rMATS. In addition, OutSplice imposes a normalized expression threshold relative to normal samples before classifying an event as an outlier, thereby filtering out many changes lacking strong biological relevance. Psychomics and Whippet likewise reported fewer genes (345 and 1021, respectively), as both apply minimum PSI change criteria that reduce the total count of significant calls. Although identifying more events can help discover novel or unannotated splicing, this quantity alone should not be the primary measure of analysis quality.

The highest gene-level overlaps were observed between LeafCutter and rMATS, and between LeafCutter and edgeR [52]. The shared methodological emphasis on split-read counts for inclusion-level estimation likely explains the strong agreement between LeafCutter and rMATS. The overlap with edgeR probably reflects the large total number of genes identified by both tools, increasing the chance of common detections. Most other

algorithm pairs, however, showed relatively modest overlap, highlighting how each method defines statistical significance. This variability underscores the value of applying several complementary tools, since another may identify a gene missed by one approach.

Seventeen genes were detected as significant by five or more of the evaluated algorithms (**Table 1**). Such multi-algorithm agreement increases confidence that the splicing changes are genuine and marks these genes as strong candidates for targeted validation experiments. Nevertheless, variations in how coordinates are reported and the possibility of multiple splice events per gene can make it challenging to verify that precisely the same alteration was detected across tools. Visualization of read alignments in the interactive genomics viewer helps resolve such ambiguities. For instance, the gene ECM1 showed a prominent exon skipping event identified by most algorithms, though labeled with differing coordinates (**Figure 2**) [53]. The exon located at chr1:150511457-150511831 showed markedly higher expression in tumor samples than in normal tissue, indicating skipping in normal cells and potential generation of a novel peptide in tumors. This case exemplifies the lack of unified event annotation standards across splicing software. Investigators using multiple tools should be mindful of this issue, as equivalent biological events may be reported in non-identical formats and easily overlooked. Although most algorithms flagged this as a major splicing change, edgeR did not assign statistical significance despite detecting the exon. This outcome may stem from the negative binomial model's sensitivity to dispersion and replicate variability, characteristics that are typically elevated in heterogeneous tumor populations.



(a)

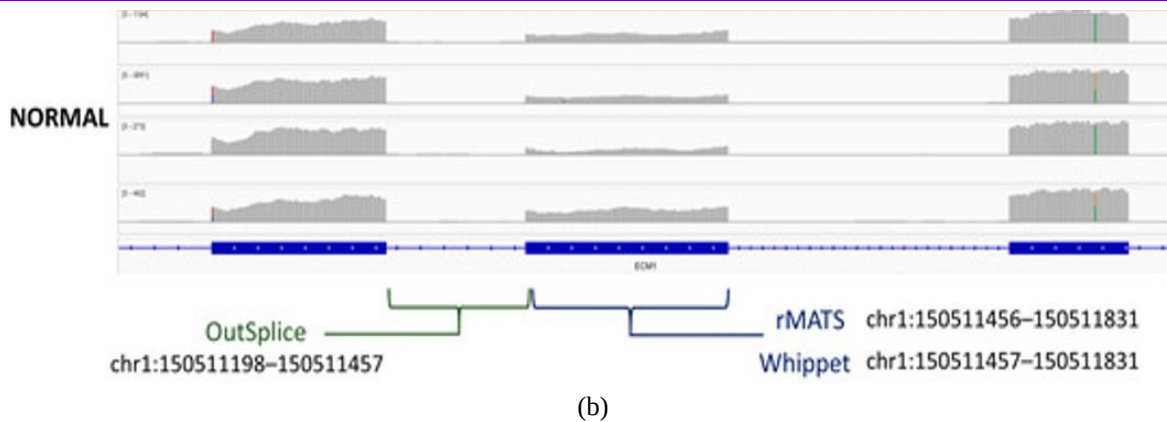


Figure 2. Five of the algorithms identified the ECM1 exon “skipping” event. Brackets indicate the unique coordinate annotation provided by each corresponding algorithm for the singular junction event.

Table 1. Genes with significant differential splicing found in ≥ 5 of the tested differential splicing algorithms, their descriptions, and the types of events found. “X” indicates which algorithms identified significant splicing in the given gene.

Gene showing significant splicing alterations	Gene function/Description	Detected alternative splicing event types	edgeR	LeafCutter	OutSplice	Psychomics	rMATS	Whippet
ECM1	Extracellular matrix protein 1	Exon skipping and alternative 5' splice site events		X	X	X	X	X
COL6A3	Collagen type VI alpha-3 chain	Exon skipping and mutually exclusive exon events	X	X		X	X	X
KIAA1217	Embryonic skeletal system development-related gene	Exon skipping, alternative first exon, retained intron, and mutually exclusive exon events	X	X		X	X	X
HDAC9	Histone deacetylase 9	Exon skipping, alternative first exon, tandem start site, and mutually exclusive exon events	X	X		X	X	X
MBNL1	Muscleblind-like splicing regulator 1	Exon skipping and mutually exclusive exon events	X	X		X	X	X
VPS39	VPS39 component of HOPS complex	Exon skipping and tandem start site events	X	X		X	X	X
PLEKHG1	Pleckstrin homology and RhoGEF domain-containing protein G1	Exon skipping, insertion, tandem start site, and mutually exclusive exon events	X	X	X	X	X	X
ITGB4	Integrin beta-4 subunit	Exon skipping	X	X	X	X	X	X
PTPN6	Protein tyrosine phosphatase non-receptor type 6	Exon skipping, alternative acceptor, alternative first exon, alternative 3' start site, mutually exclusive exons, and retained introns	X	X		X	X	X
MTMR1	Myotubularin-related protein 1	Exon skipping and mutually exclusive exon events	X	X		X	X	X
PARD3	Par-3 family cell polarity regulator	Exon skipping, alternative 5' splice site, alternative first exon, and mutually exclusive exon events	X	X		X	X	X
NUMA1	Nuclear mitotic apparatus protein 1	Alternative first exon, tandem transcription start site, retained intron, and mutually exclusive exon events	X	X		X	X	X
RABGAP1L	RAB GTPase-activating protein 1-like	Exon skipping and insertion events	X	X	X	X	X	

MDM2	Mouse double minute 2 homolog (proto-oncogene)	Exon skipping, alternative first exon, and mutually exclusive exon events	X	X	X	X	X
MCM7	Minichromosome maintenance complex component 7	Exon skipping, alternative first exon, and retained intron events	X	X	X	X	X
MEI1	Meiotic double-stranded break formation protein 1	Exon skipping, insertion, and deletion events	X	X	X	X	X
FCGR2B	Fc gamma receptor IIb	Exon skipping and retained intron events	X	X	X	X	X

This ECM1 splicing alteration can also be examined directly through the visualization capabilities of the individual algorithms. Except for Whippet and rMATS, each tool offers distinct built-in visualization options, leaving the final choice to user preference. LeafCutter supports the LeafViz application, which generates clear graphical representations of the relevant intron clusters and the constituent splicing events that collectively reach significance. For ECM1, this visualization highlights elevated PSI values for the flanking exons relative to the internally skipped exon (**Figure 3a**). OutSplice produces waterfall plots of normalized junction expression,

showing that multiple tumor samples have markedly higher expression of the ECM1 event than the distribution in normal tissues (**Figure 3b**). Psychomics, like LeafCutter, visualizes PSI differences for the ECM1 event, but it does so with distribution plots showing increased PSI values in tumors relative to normal controls (**Figure 3c**). Finally, edgeR displays log fold changes in exon expression. Although edgeR successfully detected the skipped ECM1 exon and indicated a positive log fold change consistent with upregulation in tumor samples, the event did not meet the threshold for statistical significance (**Figure 3d**).

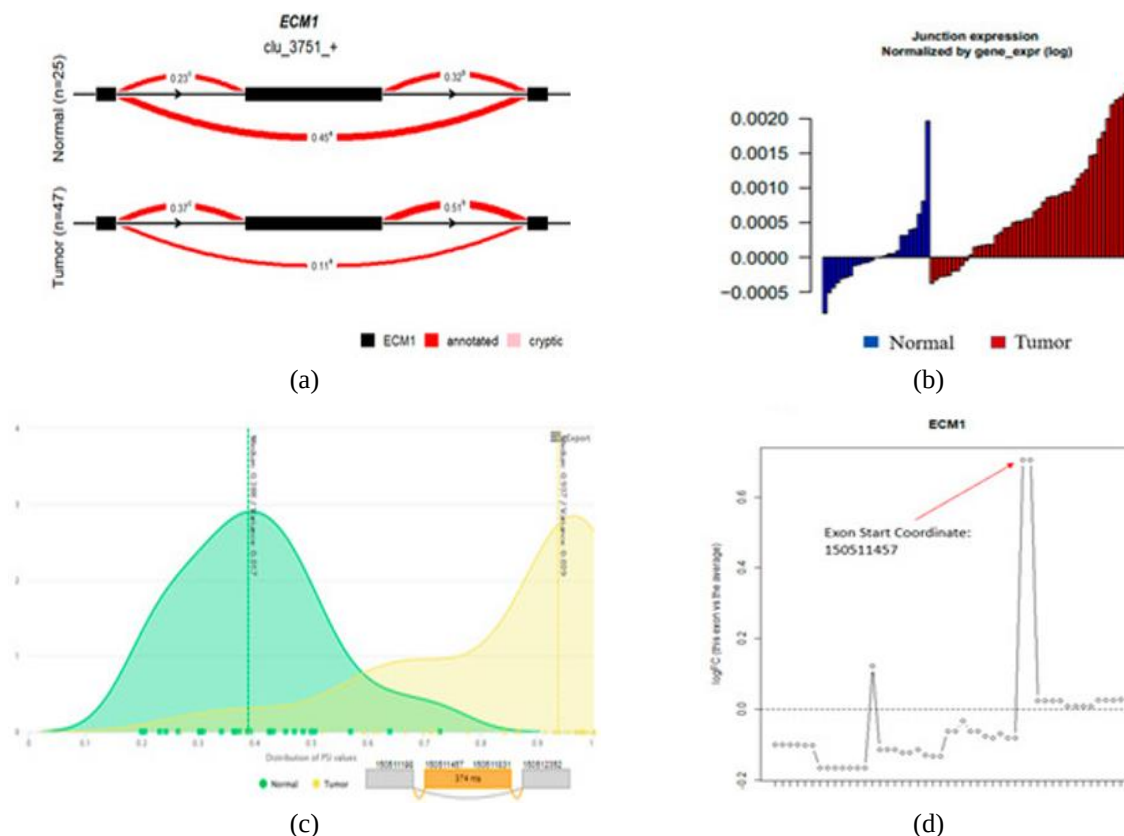


Figure 3. The ECM1 event visualizations were provided, where possible, by various algorithms. (a) LeafCutter splicing plot; (b) OutSplice normalized junction expression; (c) psychomics percent spliced-in distributions with

splicing plot; (d) edgeR log fold changes, where positive changes represent higher exon expression in tumor samples compared to normal.

To evaluate OutSplice against other differential splicing tools, we generated synthetic RNA-seq datasets. These simulated data were restricted to annotated transcripts and known splicing events, with elevated expression of the alternative transcripts introduced in half of the simulated “Tumor” samples. In this controlled setting, OutSplice performed moderately relative to the other algorithms, with reduced sensitivity and an elevated false discovery rate (Table 2). OutSplice was intentionally developed to detect pronounced outliers in heterogeneous

tumor datasets exhibiting greater than 10× differences in event expression between tumor and normal groups, and to identify novel, unannotated junctions. These specific capabilities could not be fully assessed using the Polyester-generated simulated data. Nevertheless, analyses of real biological datasets have shown OutSplice to be effective at detecting functionally relevant splicing alterations, which were subsequently confirmed through polymerase chain reaction and experimental functional studies [7, 33-55].

Table 2. Performance metrics for each algorithm evaluated on the simulated dataset. FDR denotes the false discovery rate. “Identified Genes” refers to genes containing splicing events detected by the algorithm irrespective of statistical significance. “Identified + Significant Genes” refers to genes containing splicing events that were both detected by the algorithm and reached statistical significance at an algorithm-defined p-value threshold of ≤0.05. Because Whippet does not output p-values, we defined significance as genes with a splicing probability of at least 90% and a difference in percent spliced-in value greater than 0.1.

Method	Identified genes: true positives	Identified genes: total detected	Identified genes: sensitivity	Identified genes: specificity	Identified genes: FDR	Significant identified genes: true positives	Significant identified genes: total detected	Significant identified genes: sensitivity	Significant identified genes: specificity	Significant identified genes: FDR
edgeR	200	1014	1	0.03	0.80	192	265	0.96	0.91	0.28
LeafCutter	185	213	0.93	0.97	0.13	181	194	0.91	0.98	0.07
OutSplice	91	366	0.46	0.67	0.75	78	336	0.39	0.69	0.77
Psychomics	108	123	0.54	0.98	0.12	102	115	0.51	0.98	0.11
rMATS	149	392	0.75	0.70	0.62	115	120	0.58	0.99	0.04
Whippet	200	972	1	0.07	0.79	77	79	0.39	1	0.03

Comparative outlier detection and method evaluation
Beyond OutSplice, the principal alternative strategies for detecting splicing outliers are LeafCutterMD and FRASER. In contrast to OutSplice, these two algorithms focus on pinpointing individual outliers, primarily in rare disease research. LeafCutterMD follows the same intron clustering and Dirichlet-Multinomial framework as the original LeafCutter to uncover splicing anomalies. Still, it also applies statistical evaluations to assess whether the observed intron counts align with the expected distribution [38]. FRASER, by contrast, quantifies alternative acceptor sites, donor sites, and splicing efficiencies by analyzing split and non-split reads at exon-exon boundaries. For each metric, the observations are modeled using a beta-binomial distribution, with

outliers defined as data points that substantially depart from this model [37].
A direct comparison between FRASER and LeafCutterMD was performed using the previously outlined benchmark criteria. Because FRASER’s split-read counting process imposes heavy memory demands, the analysis was limited to 25 tumor samples. Results showed that LeafCutterMD required far less memory and completed its tasks much faster. After completing STAR alignment and intron clustering, LeafCutterMD finished the outlier detection phase in only 8 minutes using 4 GB of memory. In contrast, FRASER’s pipeline, dominated by its RNA counting stage, required roughly 15 hours and consumed 197 GB of memory, posing significant challenges for researchers operating with modest computational capabilities. Moreover, LeafCutterMD

identified a substantially larger set of genes exhibiting significant splicing outliers (166) compared with the 33 genes detected by FRASER.

FRASER nevertheless provides distinct strengths in its results. LeafCutterMD users must separately obtain gene symbol annotations via packages such as GenomicRanges to interpret the often extensive gene coordinate lists and must also manually correct p-values. In addition, FRASER uniquely calculates splicing efficiency—a metric absent from both OutSplice and LeafCutterMD—which proves advantageous for studies examining intron retention events.

While LeafCutterMD and FRASER effectively support splicing outlier identification, they are not designed for comparative analyses involving tumor samples and matched controls, a core feature of OutSplice. These tools instead target rare events exclusively within the disease cohort. As a result, splicing changes common across many tumors (outliers relative to normal tissue but not among tumors themselves) are likely to be overlooked. Tumors exhibit pronounced heterogeneity and extensive splicing dysregulation compared with normal samples; therefore, reliable detection of both recurrent shared outliers and isolated individual outliers is vital. OutSplice is the only tool to integrate differential splicing analysis with outlier detection. This combined functionality allows investigators to distinguish between frequently altered splicing regions that may yield broadly targetable proteins for immunotherapy and highly patient-specific alterations that may benefit only select individuals.

Conclusion

Taken together, these findings indicate that OutSplice is especially effective for uncovering rare splicing events aligned with the molecular characteristics of tumors. When the goal is to survey the broadest possible range of potentially relevant changes across all documented junction events, LeafCutter offers a strong alternative. For settings with restricted computing power, Whippet remains a practical choice. Evaluation of all three algorithms revealed limited concordance in the splicing alterations they identified. Subsequent research would benefit from a unified workflow that systematically applies each tool, consolidates their outputs, and automatically aligns genes or precise junction coordinates, thereby increasing confidence in the reported events.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, et al. Function of alternative splicing. *Gene*. 2013;514(1):1-30.
2. Tazi J, Bakkour N, Stamm S. Alternative splicing and disease. *Biochim Biophys Acta*. 2009;1792(1):14-26.
3. Screaton GR, Bell MV, Jackson DG, Cornelis FB, Gerth U, Bell JL. Genomic structure of DNA encoding the lymphocyte homing receptor CD44 reveals at least 12 alternatively spliced exons. *Proc Natl Acad Sci USA*. 1992;89(24):12160-4.
4. Matsumura Y, Tarin D. Significance of CD44 gene products for cancer diagnosis and disease evaluation. *Lancet*. 1992;340(8827):1053-8.
5. Chen C, Zhao S, Karnad A, Freeman JW. The biology and role of CD44 in cancer progression: therapeutic implications. *J Hematol Oncol*. 2018;11(1):64.
6. Venkataramany AS, Schieffer KM, Lee K, Cottrell CE, Wang PY, Mardis ER, et al. Alternative RNA splicing defects in pediatric cancers: new insights in tumorigenesis and potential therapeutic vulnerabilities. *Ann Oncol*. 2022;33(6):578-92.
7. Guo T, Sakai A, Afsari B, Considine M, Danilova L, Favorov AV, et al. A novel functional splice variant of AKT3 defined by analysis of alternative splice expression in HPV-positive oropharyngeal cancers. *Cancer Res*. 2017;77(19):5248-58.
8. Kahles A, Lehmann KV, Toussaint NC, Huser M, Stark SG, Sachsenberg T, et al. Comprehensive analysis of alternative splicing across tumors from 8705 patients. *Cancer Cell*. 2018;34(2):211-24.e6.
9. Pan Y, Phillips JW, Zhang BD, Noguchi M, Kutschera E, McLaughlin J, et al. IRIS: discovery of cancer immunotherapy targets arising from pre-mRNA alternative splicing. *Proc Natl Acad Sci USA*. 2023;120(3):e2221116120.
10. Wood MA, Weeder BR, David JK, Nellore A, Thompson RF. Burden of tumor mutations,

- neopeptides, and other variants are weak predictors of cancer immunotherapy response and overall survival. *Genome Med.* 2020;12(1):33.
11. Ott PA, Hu ZT, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal neoantigen vaccine for patients with melanoma (vol 547, pg 217, 2017). *Nature.* 2018;555(7696):402.
 12. Martin SD, Brown SD, Wick DA, Nielsen JS, Kroeger DR, Twumasi-Boateng K, et al. Low mutation burden in ovarian cancer may limit the utility of neoantigen-targeted vaccines. *PLoS ONE.* 2016;11(5):e0155189.
 13. Yarchoan M, Hopkins A, Jaffee EM. Tumor mutational burden and response rate to PD-1 inhibition. *N Engl J Med.* 2017;377(25):2500-1.
 14. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol.* 2013;31(1):46-53.
 15. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323.
 16. Katz Y, Wang ET, Airoidi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods.* 2010;7(12):1009-15.
 17. Hu Y, Huang Y, Du Y, Orellana CF, Singh D, Johnson AR, et al. DiffSplice: the genome-wide detection of differential splicing events with RNA-seq. *Nucleic Acids Res.* 2013;41(4):e39.
 18. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139-40.
 19. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* 2012;22(10):2008-17.
 20. Hartley SW, Mullikin JC. Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq. *Nucleic Acids Res.* 2016;44(16):e127.
 21. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
 22. Li YI, Knowles DA, Humphrey J, Barbeira AN, Dickinson SP, Im HK, et al. Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet.* 2018;50(1):151-8.
 23. Wang Q, Rio DC. JUM is a computational method for comprehensive annotation-free analysis of alternative pre-mRNA splicing patterns. *Proc Natl Acad Sci USA.* 2018;115(35):E8181-90.
 24. Saraiva-Agostinho N, Barbosa-Morais NL. psichomics: graphical application for alternative splicing quantification and analysis. *Nucleic Acids Res.* 2019;47(1):e7.
 25. Trincado JL, Entizne JC, Hysenaj G, Singh B, Skalic M, Elliott DJ, et al. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol.* 2018;19(1):40.
 26. Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci USA.* 2014;111(51):E5593-601.
 27. Sterne-Weiler T, Weatheritt RJ, Best AJ, Ha KCH, Blencowe BJ. Efficient and accurate quantitative profiling of alternative splicing patterns of any complexity on a laptop. *Mol Cell.* 2018;72(1):187-200.e6.
 28. Vaquero-Garcia J, Barrera A, Gazzara MR, Gonzalez-Vallinas J, Lahens NF, Hogenesch JB, et al. A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife.* 2016;5:e11752.
 29. Denti L, Rizzi R, Beretta S, Vedova GD, Previtali M, Bonizzoni P. ASGAL: aligning RNA-Seq data to a splicing graph to detect novel alternative splicing events. *BMC Bioinformatics.* 2018;19(1):444.
 30. Kahles A, Ong CS, Zhong Y, Rätsch G. SplAdder: identification, quantification and testing of alternative splicing events from RNA-Seq data. *Bioinformatics.* 2016;32(12):1840-7.
 31. Ryan MC, Cleland J, Kim R, Wong WC, Weinstein JN. SpliceSeq: a resource for analysis and visualization of RNA-Seq data on alternative splicing and its functional impacts. *Bioinformatics.* 2012;28(18):2385-7.
 32. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol.* 2018;15(2):81-94.
 33. Liu C, Guo T, Xu G, Sakai A, Ren S, Fukusumi T, et al. Characterization of alternative splicing events in HPV-negative head and neck squamous cell

- carcinoma identifies an oncogenic DOCK5 variant. *Clin Cancer Res.* 2018;24(20):5123-32.
34. Liu C, Guo T, Sakai A, Ren S, Fukusumi T, Ando M, et al. A novel splice variant of LOXL2 promotes progression of human papillomavirus-negative head and neck squamous cell carcinoma. *Cancer.* 2020;126(4):737-48.
 35. Mehmood A, Laiho A, Venäläinen MS, McGlinchey AJ, Wang N, Elo LL. Systematic evaluation of differential splicing tools for RNA-seq studies. *Brief Bioinform.* 2020;21(6):2052-65.
 36. Jiang MH, Zhang SY, Yin HX, Zhuo ZY, Meng GY. A comprehensive benchmarking of differential splicing tools for RNA-seq analysis at the event level. *Brief Bioinform.* 2023;24(3):bbad121.
 37. Mertens C, Scheller IF, Yépez VA, Çelik MH, Liang Y, Kremer LS, et al. Detection of aberrant splicing events in RNA-seq data using FRASER. *Nat Commun.* 2021;12(1):529.
 38. Jenkinson G, Li YI, Basu S, Cousin MA, Oliver GR, Klee EW. LeafCutterMD: an algorithm for outlier splicing detection in rare diseases. *Bioinformatics.* 2020;36(17):4609-15.
 39. Guo T, Gaykalova DA, Considine M, Wheelan S, Pallavajjala A, Bishop JA, et al. Characterization of functionally active gene fusions in human papillomavirus related oropharyngeal squamous cell carcinoma. *Int J Cancer.* 2016;139(2):373-82.
 40. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21.
 41. Frankish A, Diekhans M, Jungreis I, Lagarde J, Loveland JE, Mudge JM, et al. GENCODE 2021. *Nucleic Acids Res.* 2021;49(D1):D916-23.
 42. Veeneman BA, Shukla S, Dhanasekaran SM, Chinnaiyan AM, Nesvizhskii AI. Two-pass alignment improves novel splice junction quantification. *Bioinformatics.* 2016;32(1):43-9.
 43. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature.* 2015;517(7536):576-82.
 44. Ghosh D. Discrete nonparametric algorithms for outlier detection with genomic data. *J Biopharm Stat.* 2010;20(2):193-208.
 45. Ochs MF, Farrar JE, Considine M, Wei Y, Meshinchi S, Arcenci RJ. Outlier analysis and top scoring pair for integrated data analysis and biomarker discovery. *IEEE/ACM Trans Comput Biol Bioinform.* 2014;11(3):520-32.
 46. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014;30(7):923-30.
 47. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* 2019;47(8):e47.
 48. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-9.
 49. Cotto KC, Feng YY, Ramu A, Richters M, Freshour SL, Skidmore ZL, et al. RegTools: integrated analysis of genomic and transcriptomic data for the discovery of splice-associated variants in cancer. *Nat Commun.* 2023;14(1):1589.
 50. Frazee AC, Jaffe AE, Langmead B, Leek JT. Polyester: simulating RNA-seq datasets with differential transcript expression. *Bioinformatics.* 2015;31(17):2778-84.
 51. Tapiá J, Ha KCH, Sterne-Weiler T, Gohr A, Braunschweig U, Hermoso-Pulido A, et al. An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Res.* 2017;27(10):1759-68.
 52. Wickham H. ggplot2: elegant graphics for data analysis. Springer: New York, NY, USA; 2009. p. 1–212.
 53. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 2013;14(2):178-92.
 54. Guo T, Zambo KDA, Zamuner FT, Ou T, Hopkins C, Kelley DZ, et al. Chromatin structure regulates cancer-specific alternative splicing events in primary HPV-related oropharyngeal squamous cell carcinoma. *Epigenetics.* 2020;15(9):959-71.
 55. Kelley DZ, Flam EL, Guo T, Danilova LV, Zamuner FT, Bohrsen C, et al. Functional characterization of alternatively spliced GSN in head and neck squamous cell carcinoma. *Transl Res.* 2018;202:109-19.