




Check for updates

REVIEW

Selecting differential splicing methods: Practical considerations

[version 1; peer review: awaiting peer review]

Ben J. Draper ¹, Mark J. Dunning², David C. James¹¹Department of Chemical and Biological Engineering, Mappin St., The University of Sheffield, Sheffield, S1 3JD, UK²Bioinformatics Core Bioinformatics Core, The Faculty of Health,, The University of Sheffield, Sheffield, S10 2HQ, UK**V1** First published: 08 Jan 2025, 14:47
<https://doi.org/10.12688/f1000research.155223.1>Latest published: 08 Jan 2025, 14:47
<https://doi.org/10.12688/f1000research.155223.1>

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

Abstract

Alternative splicing is crucial in gene regulation, with significant implications in clinical settings and biotechnology. This review article compiles bioinformatics RNA-seq tools for investigating differential splicing; offering a detailed examination of their statistical methods, case applications, and benefits. A total of 22 tools are categorised by their statistical family (parametric, non-parametric, and probabilistic) and level of analysis (transcript, exon, and event). The central challenges in quantifying alternative splicing include correct splice site identification and accurate isoform deconvolution of transcripts. Benchmarking studies show no consensus on tool performance, revealing considerable variability across different scenarios. Tools with high citation frequency and continued developer maintenance, such as DEXSeq and rMATS, are recommended for prospective researchers. To aid in tool selection, a guide schematic is proposed based on variations in data input and the required level of analysis. Additionally, advancements in long-read RNA sequencing are expected to drive the evolution of differential splicing tools, reducing the need for isoform deconvolution and prompting further innovation.

Keywords

Bioinformatics, Alternative Splicing, RNASeq, Transcriptomics, Differential Expression

This article is included in the **Bioinformatics** gateway.

Corresponding author: Ben J. Draper (bjdraper1@sheffield.ac.uk)

Author roles: **Draper BJ:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Dunning MJ:** Investigation, Resources, Supervision, Validation, Writing – Review & Editing; **James DC:** Funding Acquisition, Project Administration, Supervision

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by funding from Lonza Biologics Inc & The University of Sheffield. <https://www.lonza.com/> I extend my gratitude for their financial assistance in facilitating this research project. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Draper BJ *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Draper BJ, Dunning MJ and James DC. **Selecting differential splicing methods: Practical considerations [version 1; peer review: awaiting peer review]** F1000Research 2025, 14:47 <https://doi.org/10.12688/f1000research.155223.1>

First published: 08 Jan 2025, 14:47 <https://doi.org/10.12688/f1000research.155223.1>

Introduction

Alternative splicing (AS) can be best described as fine-tuning gene expression by rearranging exons and introns in pre-mRNA. With 90-95% of human multi-exon genes estimated to possess some form of alternative splicing, it is a widespread regulatory process in cellular biology.¹ The cell utilises a large ribonucleoprotein (RBP) complex known as the spliceosome which is guided to target sites through the interaction of sequence elements (splice sites, enhancers & silencers and the polypyrimidine tract) and/or splicing factors. Pre-mRNA splicing can also occur without the spliceosome as in the case of self-splicing group I & II introns, tRNA splicing and trans-splicing.² This ultimately results in genome-wide transcript diversity and subsequently, measurable changes to protein functionality.

Previous research has uncovered the phenotypic consequences of alternative splicing in disease. In humans, clinical research has shown alternative splicing (AS) as a key instigator in several forms of cancer and neurodegenerative disorders.³⁻⁵ One notable discovery in Microtubule-associated protein tau's (MAPT) possession of mis-spliced isoforms causing abnormal TAU accumulation progressing to Alzheimer's disease.⁶ In cancer, numerous mis-spliced variants of tumour suppressors, apoptotic and angiogenic proteins have been discovered to contribute to tumour progression.^{7,8} Beyond clinical research, the utility of alternative transcripts for bioengineering purposes has been explored. For example, an alternatively spliced version of the transcription factor X-box binding protein 1 (XBP1) coexpressed in production cell lines has been shown to increase productivity in the biomanufacturing of recombinant proteins.⁹⁻¹¹ In bio-agriculture, the CRISPR-mediated directed evolution of SF3B1 mutants (a spliceosomal component) in rice has improved crop traits through better resistance to splicing inhibitors.¹² Increasingly, the value of AS in both clinical and biotechnology applications has been recognised; highlighting the need for robust bioinformatics pipelines to identify variants.

For prospective researchers to investigate AS, the transcriptomic data is usually generated using next-generation sequencing. Short-read RNAseq is the most commonly used experimental technique to interrogate a transcriptome owing to its versatility and cost-effectiveness.^{13,14} It involves sequencing short fragments of RNA molecules, providing insights into the respective expression levels of genomic features assembled from reference genomes. These features may be coding sequences, genes, transcripts, exons, introns, codons or even untranslated regions. A typical RNAseq pre-processing pipeline will consist of quality control (QC), read alignment & quantification before statistical analysis begins. QC assesses the quality of the raw fragmented reads using a standardised tool such as FastQC and trims low-quality reads or adaptor sequences.¹⁵ Then for alignment, a reference genome/transcriptome arranges the subsequent sequences into feature bins such as genes, transcripts, exons and coding sequences using software such as STAR or HISAT.^{16,17} Alignment files (usually in the form of Sequence Alignment Maps: SAMs) can then be quantified to these features using a quantification tool such as HTSeq, Salmon or featureCounts usually normalising for library size and sequencing depth.¹⁸⁻²⁰ Depending on the purpose of analysis, normalisation may be scaled by total number of reads (CPM: Counts per Million), per length of transcript (TPM: Transcripts per Million), by paired-end fragments (RPKM: Fragments Per Kilobase of Transcript) or by using a median of ratios (DESeq2's method).²¹ Commonly, a differential expression analysis will be performed at the gene or transcript level between groups of samples to identify statistically significant changes in expression. The pre-processing steps for RNA-seq have been extensively researched over many years, and there is a consensus within the community regarding the gold-standard set of tools. Projects like nf-core enable the execution of RNA-seq pre-processing pipelines with minimal intervention and limited bioinformatics expertise.²² However, these tend to be focused on the use-case of conventional differential expression rather than the more bespoke AS pipelines as discussed here.

A growing repertoire of tools now annotate and quantify changes to splicing events. Quantification of features such as splice sites, and exon/intron junctions found in alignment files are commonly used to annotate splicing events. Although the true repertoire of splicing events is difficult to capture, conventional processes can be categorised into distinct groups. The most common events are exon skipping, retained introns, mutually exclusive exons, alternative 5' and 3' splice sites. More complex regulatory events involve genomic features beyond exons and introns, such as alternative promoter and polyadenylation sites, which result in varying mRNA 5' and 3' UTR ends. However, these events are seldom included in most bioinformatics analyses, tools such as CAGER (Cap Analysis of Gene Expression) and DaPars (Dynamic Analysis of Alternative PolyAdenylation from RNA-Seq) are available for niche research.^{23,24} Visualisation of AS is predicated upon the level of detail required in the analysis. If a highly detailed analysis of individual gene structure is needed, splice graphs, sashimi plots and junction maps are commonly used.^{25,26} To visualize changes to groups of transcripts, typically MA and Volcano plots are used much the same way as in differential expression level analysis.²¹

Current statistical methods for differential splicing

Commonly, researchers are interested in comparisons of two or more groups of samples known as differential analyses. Differential gene/transcript expression (DGE/DTE) of genes or transcripts involves taking raw read count data, normalizing or scaling it, and calculating whether the changes in expression levels between different biological groups

are statistically significant. Differential transcript/exon usage (DTU/DEU), however, uses gene-level group modelling to assess whether the proportional use of the feature (exon or transcript) is statistically significant. Differential splicing events (DSE) on the other hand use a diverse array of statistical methods to quantify and infer splicing events. A comprehensive summary of differential splicing tools is described in the supplementary table (**Supplementary Table 1**) and in the following sections.

Parametric & mixed methods

Differential expression analysis tools began in the early 2000s coinciding with the development of high throughput technologies such as microarrays. An early example was LIMMA (Linear Models for Microarray Data), developed by Gordon Smyth and colleagues in 2003, which utilises a linear regression framework and empirical Bayes techniques to identify differentially expressed genes.²⁷ Whilst initially only utilised for microarrays, the functionality thus extended to RNASeq data and has been one of the most cited RNASeq methods. As the field shifted from microarray technology to RNASeq, methods were developed such as DESeq (Differential Expression Analysis for Sequence Count Data) and edgeR to capture the nature of count data better and improve modelling.^{21,27,28} A major change incorporated in DESeq2 was empirical Bayes-based shrinkage to improve gene-wise variance estimation enhancing accuracy (**Figure 1**). Secondly, GLMs (Generalized Linear Models) replaced the simple linear models as these were shown to adapt well to non-normally distributed count-based data.²¹ The flexibility of GLMs allowed algorithms to effectively deal with issues such as overdispersion, shrinkage, heteroscedasticity and covariates. To date, GLMs are usually fitted to the NB (Negative Binomial) distribution which confers some strong advantages. The NB distribution effectively captures overdispersion (the empirical variability of counts) and can handle a large excess of zero values commonly seen in transcript or exon-level count data. However, limma, DESeq2 and edgeR were not developed to specifically address the challenges of identifying AS.

In 2014, DEXSeq was introduced by Michael Love and colleagues, a framework based on DESeq2's GLM NB model becoming the de-facto tool for parametric splicing-based analysis. Instead of analysing gene-level differential expression, DEXSeq identifies exons within genes that exhibit significant changes in their usage across conditions. This is particularly useful for studying the exonic composition of alternatively spliced transcripts. The development of tools such as DSGseq, rDiff-parametric, JunctionSeq and SeqGSEA has expanded the functionality of the GLM NB family of differential splicing tools.^{29–32} DSGseq utilises a holistic approach considering splicing events not as individual elements but as comprehensive gene-wise splice graphs that more accurately reflect complex splicing dependencies.²⁹ The tool rDiff-parametric on the other hand utilises isoform-specific loci such as restricted exonic regions to identify significant differences in isoform composition.³⁰ The proposed advantage of this approach is in the smaller exonic regions rather than full isoform deconvolution. Assigning reads to isoforms is challenging because these transcripts are practically identical, making it difficult to definitively attribute a read from an overlapping region to a particular region without supplementary data. Therefore, full isoform deconvolution is significantly biased against genes with many isoform variants.³³

A few newer methods such as DRIMSeq and DTUrtle have progressed onto non-parametric or mixed Dirichlet Multinomial Models (DMM) which have been argued to capture better the complex variability of count data and better estimate isoform abundance^{34,35} (**Figure 1 & Supplementary Table 1**). Other methods such as IsoformSwitchAnalyzeR and some custom DEXSeq workflows now incorporate modularity allowing users a selection of bioinformatics tools for filtering, hypothesis testing and posterior calculations.^{36,37} An example of the usage of parametric analysis was in the discovery of a chimeric fusion transcript of PRKACA and DNAJB1 in a rare liver tumour FL-HCC (fibrolamellar hepatocellular carcinoma) using DEXSeq's differential exon usage framework.³⁸ The discovery of differential exon usage of PRKACA's exons 2-10 and subsequent decreased usage of DNAJB1's exons 2-3 led the researchers to identify a chimeric transcript in FL-HCC patients. This demonstrated the utility of smaller exon-based analysis in identifying differences in transcript structure which would not be detected in larger gene or transcript-based analysis alone.

Probabilistic & non-parametric methods

Non-parametric or probabilistic techniques such as MAJIQ, SUPPA, WHIPPET and rMATS frequently utilize Bayesian inference and/or probabilistic methodologies.^{26,39–41} By avoiding assumptions about the data's underlying distribution, these methods enable more sophisticated modelling. Consequently, in contrast to the predominantly standardized parametric exon/transcript-based techniques, event-based methods often showcase a broader array of statistical approaches (**Figure 1**). A few common features can be identified, however. Often the targets for event annotations are not labelled in gene-transfer format such as splice sites, exon/intron junctions and splicing quantitative trait loci (QTLs) which must be calculated. This then allows the "Percent spliced in" (PSI) to be calculated per exon, representing the ratio of the number of transcripts containing an alternative exon versus the total number of transcripts per any given splice site. By comparing PSI values, different splicing events can then be identified and explored through

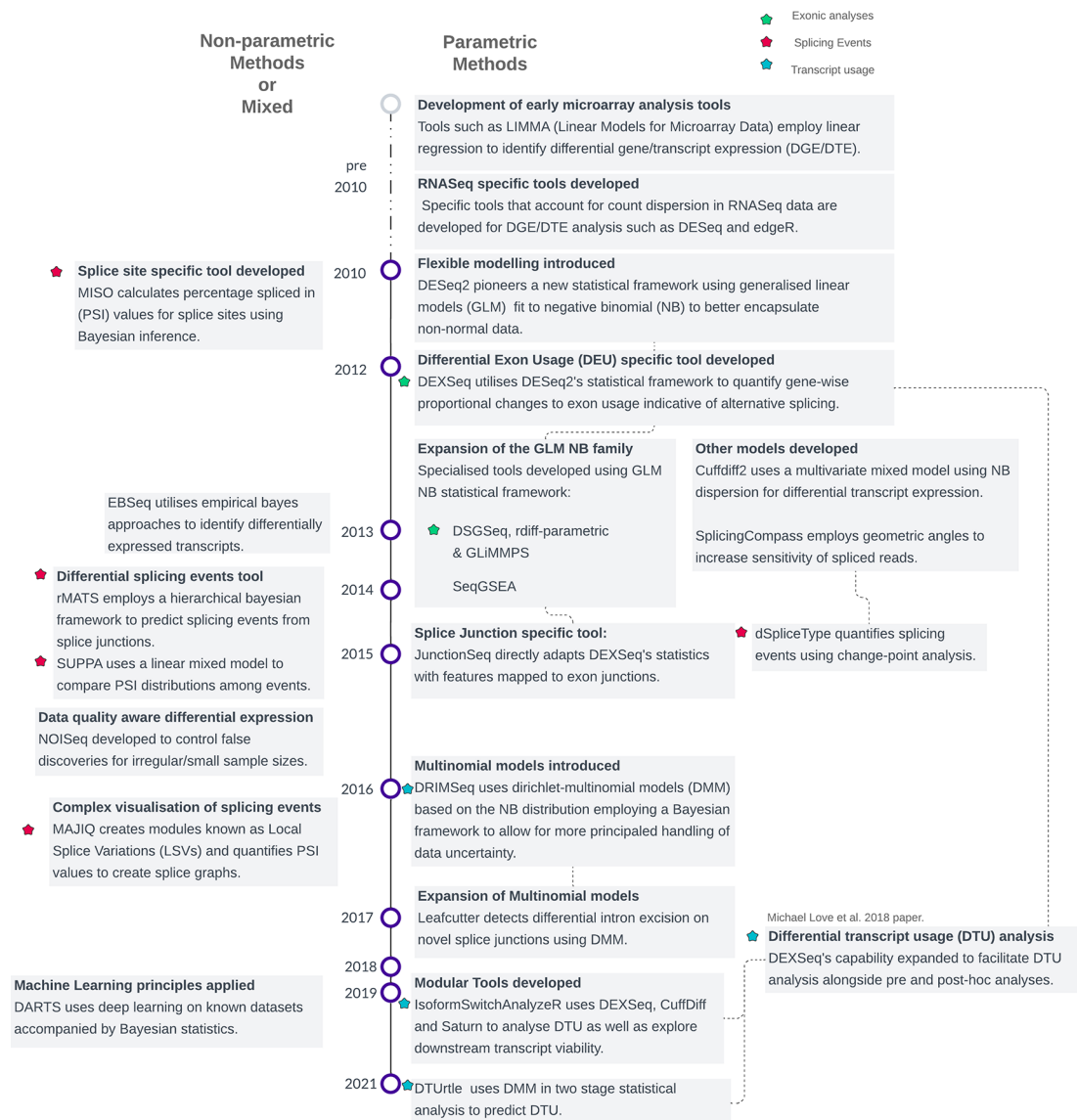


Figure 1. Timeline of statistical methods in differential splicing tool development. Methods are categorized into parametric and non-parametric approaches, grouped by methodological families. The classification is based on the underlying statistical procedures used for modelling or hypothesis testing, as detailed in Supplementary Table 1. Note that some methods incorporate elements of both parametric and non-parametric frameworks, resulting in overlapping features.

splice graphs and sashimi plots. An example of non-parametric tool usage was in the mapping of splicing events in the rice (*Oryza sativa*) transcriptome, revealing prevalent AS under deprived nutrient conditions.⁴² Importantly, this study utilised rMATS to reveal the underlying exon-intron structure of key nutrient transporter genes.

Some tools possess features for specific utility in certain scenarios. NOISeq is a non-parametric differential expression tool that is specifically designed to handle smaller numbers of biological replicates through its noise model.⁴³ For more complex modelling, tools such as GLIMMPS (Generalized Linear Mixed Model for Pedigree Data with Population Substructure) employ mixed-effects models to account for both fixed and random effects such as genetic family substructure.⁴⁴ Beyond splicing, the modular tool IsoformSwitchAnalyzeR facilitates analysis on spliced transcript quality such as Nonsense Mediated Decay (NMD) sensitivity, Intrinsically Disordered Regions (IDR) and protein domains.³⁶ Increasingly, deep learning-based approaches are being utilised to improve the accuracy of differential splicing predictions leveraging publicly available RNASeq data such as with DARTs and Bisbee.^{45,46}

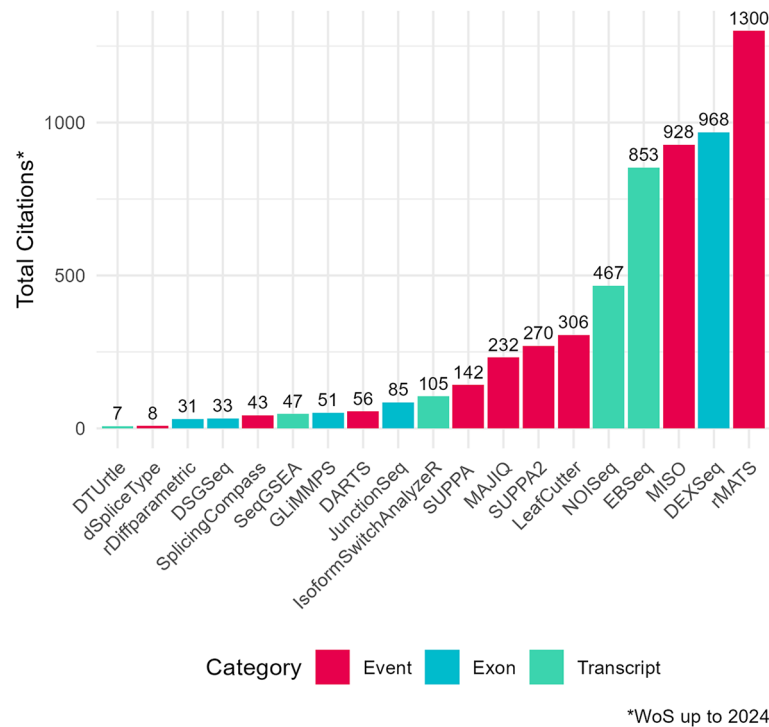


Figure 2. Citation counts of differential splicing tools (2010–2024) from Web of Science (WoS) Data. Total citation counts for surveyed differential splicing tools (2010–2024) from the Web of Science Data Portal (WoS). Tools are categorized by analysis level: event, exon, or transcript. DRIMSeq's original paper was excluded from the citation frequency analysis as it was not indexed in WoS. Certain data included herein are derived from Clarivate Web of Science. © Copyright Clarivate 2023. All rights reserved. Total citation counts for surveyed differential splicing tools (2010–2024) from the Web of Science Data Portal (WoS). Tools are categorized by analysis level: event, exon, or transcript. DRIMSeq's original paper was excluded from the citation frequency analysis as it was not indexed in WoS. Certain data included herein are derived from Clarivate Web of Science. © Copyright Clarivate 2023. All rights reserved.

Popularity & developer maintenance of methods

To assess the academic popularity of tools, a citation and developer engagement analysis of original research articles within the Web of Science (WoS) domain and the respective GitHub website domains (if applicable). The assessment spanned from 2010 to 2024 and encompassed 19 original papers on various differential splicing tools. Notably, the citation counts for these splicing tools were considerably lower compared to conventional RNA-Seq differential expression analysis tools. For instance, while the general purpose DGE/DTE tool DESeq2 amassed a total of 35,887 citations during the same period, citations for differential splicing tools ranged from 7 to 1300 (Figure 2). This discrepancy may pose challenges for researchers seeking resources and workflows specific to differential splicing analysis. Additionally, the importance of developer support cannot be understated, as it directly influences the usability and longevity of software tools. Notably, differential splicing tools such as DEXSeq, EBSeq, rMATS, SUPPA2, and MAJIQ^{26,39,40,47,48} have shown increasing usage and ongoing developer engagement, as evidenced by their growing citation counts and sustained support (Figure 3; Figure 4). One possible explanation for the lower citation rates observed in exon/transcript-based methodologies could be the broader adoption of general-purpose differential expression workflows, like DESeq2 that can employ DTE.²¹ Researchers may prefer more explicit splicing event-based tools for targeted splicing analyses and defer to DTE for transcript-based analyses. While the nuances between DTU and DTE may not be a primary focus for many researchers, it is a distinction worth noting in the context of differential splicing analysis.

The decision between exon/transcript-level (typically parametric) and event-level (typically non-parametric) analyses hinges on several factors, including the particular scientific inquiry, data accessibility, and the level of granularity required to address the research goal. In certain scenarios, integrating both methodologies could offer a more holistic understanding of splicing control mechanisms and their biological significance.



Figure 3. Citation trends of differential splicing tools (2010–2024) from Web of Science (WoS) Data. Annual citation frequency for current differential splicing tools (2010–2024) from Web of Science (WoS). Tools are categorized by analysis level: event, exon, or transcript. DRIMSeq's original paper is excluded as it is not indexed in WoS. Certain data included herein are derived from Clarivate Web of Science. © Copyright Clarivate 2023. All rights reserved.

Benchmarking of methods is difficult

To evaluate the quality of differential splicing bioinformatics tools, several benchmarks have been conducted to date. Benchmarking either the scientific accuracy or the computational power of methods can be challenging due to several factors. The main issue is the lack of ground truth to set as a reference to compare measurements to. Commonly, a small subset of experimentally validated splicing events is used as a gold standard to compare against. This was demonstrated in a recent systematic evaluation of 10 differential splicing tools in 2019, where a total of 62 qPCR-validated differentially spliced genes were tested.⁴⁹ The results from this benchmark revealed weak consensus over tool quality as the performance was markedly different across the 4 human and mouse cancer datasets. This demonstrates another issue with these evaluations: inherent heterogeneity in RNASeq data. Often, the performance of methods will depend on the upstream RNASeq pre-processing steps such as in library size, sequence depth, positional bias and annotation quality. To mitigate these issues, some papers use simulated data to i) increase the number of differentially spliced genes to reference and ii) achieve finer control over ground truth and variability within the data.^{50–52} One such benchmark used RSEM-based simulated data based on a human prostate cancer dataset (GSE22260⁵³).⁵¹ Another comparison utilised a combination of experimental and simulated Arabidopsis heat shock RNASeq datasets using the Flux Simulator tool.⁵⁴ However, it is important to note that simulated data lacks the complexity of typical biological data. Confounding factors such as outliers, and technical/procedural biases cannot be modelled in current simulations.

The consensus drawn from these three benchmarks is that the performance of differential splicing tools exhibits considerable variability depending on the outlined factors. The ongoing evolution and upkeep of tools by developers introduce a time-dependent aspect to benchmarking. Community-led maintenance efforts consistently enhance the

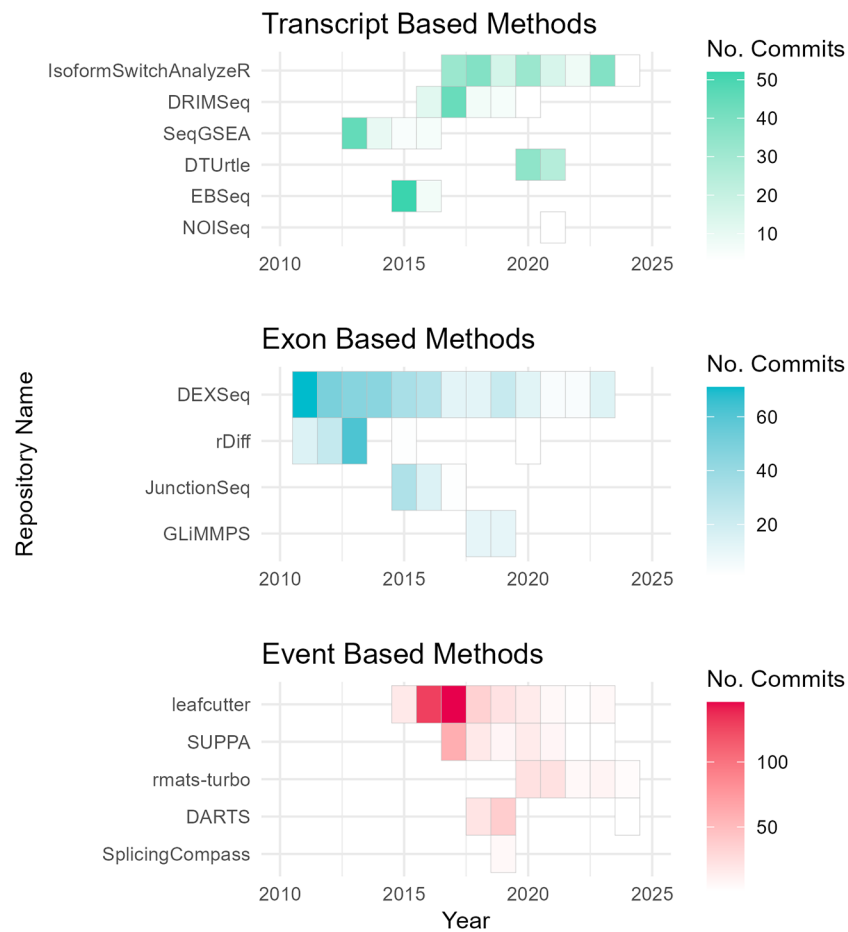


Figure 4. Developer maintenance of differential splicing tools. Annual GitHub repository commits (2010–2024) by category, highlighting community-led maintenance of differential splicing tools. Tools without GitHub pages (MAJIQ, MISO, DSGseq, and dSpliceType) were excluded from the analysis.

functionality and reliability of tools over time. Rather than aiming for a singular optimal tool for differential splicing analysis, researchers should contemplate employing a suite of tools tailored to address specific inquiries.

Method recommendations

A diagram outlining optimal tool selection is provided to guide prospective alternative splicing (AS) researchers (Figure 5). Initially, researchers should evaluate the scope and objectives of their analysis. For instance, if the aim is to identify known transcripts, it is advisable to opt for a parametric transcript-based tool like DEXSeq or DRIMSeq and execute a DTU study following Michael Love’s protocol.³⁷ Nonetheless, variations in experimental parameters such as sample size or covariate inclusion may necessitate alternative approaches.

If the objective is to uncover novel transcripts, an exon-based parametric approach might be better suited. This choice circumvents the challenges associated with isoform deconvolution and the breadth of transcript annotation, given the smaller exonic regions. For general-purpose differential exon usage (DEU) analysis, DEXSeq remains the preferred protocol due to its robust and flexible statistical methods, as well as its actively maintained software.²¹ However, again intricacies within the data may prompt the usage of more specialised alternatives. Transcript and exon-based methods offer top-down visualizations such as MA/Volcano plots, heatmaps and proportional transcript/exon graphs. If the analysis aims to visualise the movement of exons/introns and splice sites, then an event-based protocol would be more appropriate. Generally, tools such as rMATs, SUPPA2 and MISO offer comprehensive and detailed splicing event analysis.^{39,40,55}

Commonly, sashimi plots are the best method to visualise splice junctions from aligned data with events annotated, although this can also be plotted separately in IGV.⁵⁶ For user-friendly visualization, MAJIQ offers a summative

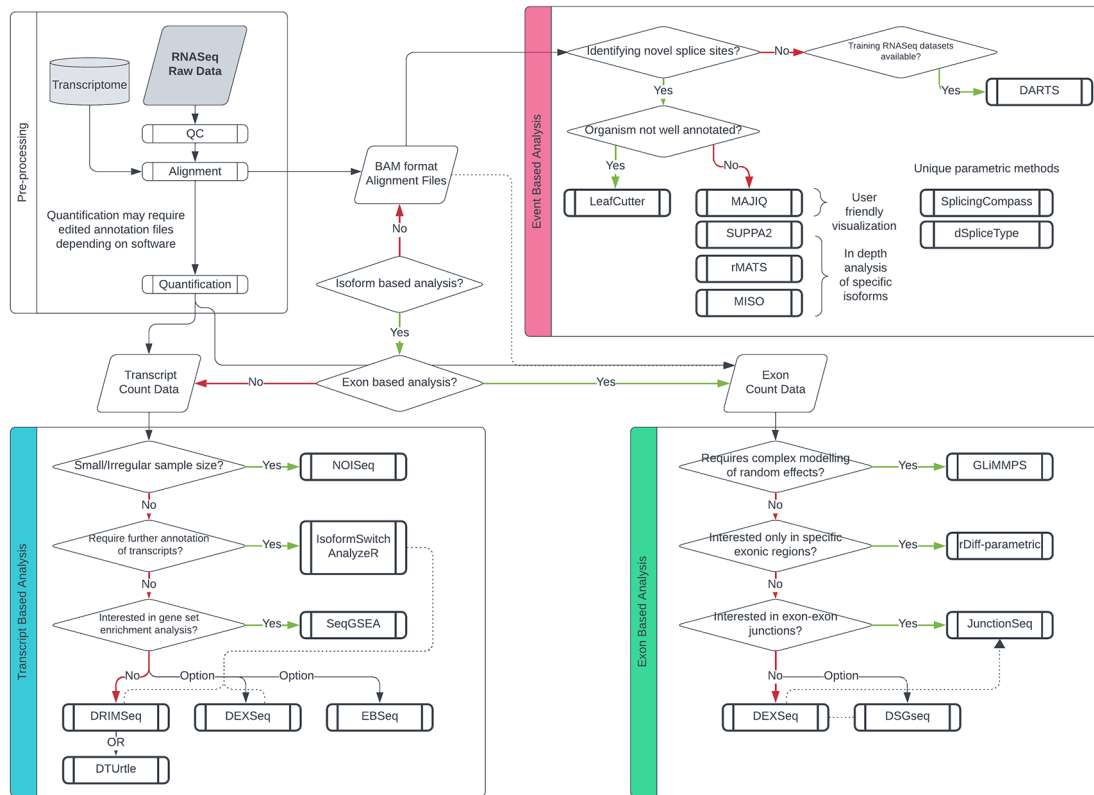


Figure 5. Guideline for differential splicing tool selection based on experimental parameters. Decision tree for differential splicing analysis, categorized by three branches based on the level of analysis. Transcript-based methods are represented in blue, exon-based methods in pink, and event-based methods in yellow.

HTML-based visualizer for complex events such as exitrons or orphan junctions.²⁶ Another factor to consider is the annotation quality of the organism/tissue being studied. If researchers are not confident in the quality of annotations and would like annotation-free analysis, methods such as LeafCutter are a good alternative to conventional methods.^{57,58} Overall, event-based methods are more suited to advanced programmers owing to their use of command-line tools over interpreters that use IDEs (Integrated development environments). For most analyses, however, a DEU or DTU-based analysis is recommended for simple interpretability and robustness. Optional steps for AS-specific analyses can also be performed to enhance the data quality. For example, Portcullis enables the accurate filtering of false splice junctions that are often incorrectly characterized by common aligners.⁵⁹

Discussion

While the repertoire of tools to accommodate differential splicing analysis has grown in the past two decades, they are ultimately limited by the capabilities of the RNASeq technology available to date. Since 2010 however, the development of nanopore sequencing technology such as Oxford Nanopore Technologies (ONT) and PacBio's single-molecule real-time (SMRT) has facilitated the development of long-read RNASeq.^{60–62} Long read lengths typically fall within the range of 10kb to 100kb, with ultra-long read lengths now up to 1-2 Mb.⁶³ The main benefit this technology confers is the ability to bypass the aforementioned deconvolution issue stemming from multiple mapping and reconstruct full-length transcript isoforms in a single read. This can not only more accurately identify known transcripts but also novel or splice variants as well as fusion genes. Most current parametric DS tools can therefore be utilised in long-read-based analyses. A recent study utilized IsoformSwitchAnlyzeR's DEXSeq-based DTU workflow on ONT long reads, demonstrating the capability of current long-standing methods on long-read data.^{21,36,64} This was facilitated through long-read custom annotation of the transcriptome using TALON to identify novel transcripts.⁶⁵ Additionally, specific novel technologies such as LIQA have been developed to analyse long reads.⁶⁶

Long-read RNASeq still possess notable disadvantages, however. Early on, long-read RNASeq possessed error rates of 10-20%.^{67,68} The development of HiFi sequencing by PacBio using circular consensus sequencing has since reduced the error rate to a reported 0.5%.⁶⁹ While the development of deep-learning algorithms such as DeepConsensus has sought to

push HiFi accuracy further bringing it on par to short read.⁷⁰ However, this is still highly dependent on the depth of sequencing. The most efficient error correction method involves hybridising the analysis with short-read RNAseq methods.⁷¹ This ultimately means that while accuracy can now be brought to close to 99.5%, error correction drives the cost of long-read RNAseq methods up significantly. The field is progressing towards optimal error correction and is now focusing on lowering costs which is currently the largest hurdle for practical use for common research.

As interest in alternative splicing grows, researchers have access to an expanding array of tools. Advances in statistical methods and longer RNA sequencing read lengths are overcoming technical limitations. This leads to more precise transcript alignment and reduces the need for complex computational steps. With workflows becoming streamlined and modular, platforms like Nextflow enable researchers to create tailored pipelines for their specific goals and data types.⁷² These developments promise a brighter future for alternative splicing analysis, facilitating a deeper exploration of transcriptomic regulation and its functional significance.

Ethical approval and consent statement

Ethical approval and consent were not required.

Data availability statement

Underlying data

No data associated with this article.

Extended data

Zenodo: Selecting differential splicing methods: Practical considerations <https://doi.org/10.5281/zenodo.14293573>.⁷³

The repository contains the following underlying data:

- Supplementary Table 1.docx: Statistical details on differential splicing tools.
- citations_2023.csv: WoS citation count for differential splicing tools.
- citations_year_plot_new.R: R script to visualise citation trends.
- github_repos_txt: Github repository locations cloned on 20.02.2024.
- github_repos.R: Github maintenance analysis and visualisation.
- citations_2023.xlsx

Software availability statement

- <https://github.com/bjdraper/Selecting-differential-splicing-methods-Practical-Considerations---R-Scripts-and-Data>
- Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 1.0).

Archived software available from: [10.5281/zenodo.14293573](https://doi.org/10.5281/zenodo.14293573)

[github_repos.R](#)

[citations_year_plot_new.R](#)

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

References

1. Baralle FE, Giudice J: **Alternative splicing as a regulator of development and tissue identity.** *Nat. Rev. Mol. Cell Biol.* 2017 Jul;

18(7): 437–451.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

2. Matera AG, Wang Z: **A day in the life of the spliceosome.** *Nat. Rev. Mol. Cell Biol.* 2014 Feb; **15**(2): 108–121.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
3. Singh B, Eyraes E: **The role of alternative splicing in cancer.** *Transcription.* 2017 Mar 15; **8**(2): 91–98.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Bonnal SC, López-Oreja I, Valcárcel J: **Roles and mechanisms of alternative splicing in cancer — implications for care.** *Nat. Rev. Clin. Oncol.* 2020 Aug; **17**(8): 457–474.
[PubMed Abstract](#) | [Publisher Full Text](#)
5. Zhang Y, Qian J, Gu C, *et al.*: **Alternative splicing and cancer: a systematic review.** *Signal Transduct. Target. Ther.* 2021 Feb 24; **6**(1): 1–14.
[Publisher Full Text](#)
6. Kar A, Kuo D, He R, *et al.*: **Tau Alternative Splicing and Frontotemporal Dementia.** *Alzheimer Dis. Assoc. Disord.* 2005; **19**(Suppl 1): S29–S36.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Yanagisawa M, Huvelde D, Kreinest P, *et al.*: **A p120 Catenin Isoform Switch Affects Rho Activity, Induces Tumor Cell Invasion, and Predicts Metastatic Disease.** *J. Biol. Chem.* 2008 Jun 27; **283**(26): 18344–18354.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. McEvoy J, Ulyanov A, Brennan R, *et al.*: **Analysis of MDM2 and MDM4 Single Nucleotide Polymorphisms, mRNA Splicing and Protein Expression in Retinoblastoma.** *PLoS One.* 2012 Aug 20; **7**(8): e42739.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Cain K, Peters S, Hailu H, *et al.*: **A CHO cell line engineered to express XBP1 and ERO1- α has increased levels of transient protein expression.** *Biotechnol. Prog.* 2013 Jun; **29**(3): 697–706.
[PubMed Abstract](#) | [Publisher Full Text](#)
10. Johari YB, Estes SD, Alves CS, *et al.*: **Integrated cell and process engineering for improved transient production of a “difficult-to-express” fusion protein by CHO cells.** *Biotechnol. Bioeng.* 2015; **112**(12): 2527–2542.
[PubMed Abstract](#) | [Publisher Full Text](#)
11. Torres M, Dickson AJ: **Reprogramming of Chinese hamster ovary cells towards enhanced protein secretion.** *Metab. Eng.* 2022 Jan; **69**(69): 249–261.
[Publisher Full Text](#)
12. Butt H, Eid A, Momin AA, *et al.*: **CRISPR directed evolution of the spliceosome for resistance to splicing inhibitors.** *Genome Biol.* 2019 Apr 30; **20**(1): 73.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nat. Rev. Genet.* 2009 Jan; **10**(1): 57–63.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. Stark R, Grzelak M, Hadfield J: **RNA sequencing: the teenage years.** *Nat. Rev. Genet.* 2019 Nov; **20**(11): 631–656.
[Publisher Full Text](#)
15. **Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data.** [cited 2022 Aug 7].
[Reference Source](#)
16. Dobin A, Davis CA, Schlesinger F, *et al.*: **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics.* 2013 Jan 1; **29**(1): 15–21.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
17. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory requirements.** *Nat. Methods.* 2015 Apr; **12**(4): 357–360.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
18. Patro R, Duggal G, Love MI, *et al.*: **Salmon provides fast and bias-aware quantification of transcript expression.** *Nat. Methods.* 2017 Apr; **14**(4): 417–419.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Putri GH, Anders S, Pyl PT, *et al.*: **Analysing high-throughput sequencing data in Python with HTSeq 2.0.** *Bioinformatics.* 2022 May 13; **38**(10): 2943–2945.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Liao Y, Smyth GK, Shi W: **featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.** *Bioinformatics.* 2014 Apr 1; **30**(7): 923–930.
[PubMed Abstract](#) | [Publisher Full Text](#)
21. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.** *Genome Biol.* 2014 Dec 5; **15**(12): 550.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
22. Ewels PA, Peltzer A, Fillinger S, *et al.*: **nf-core: Community curated bioinformatics pipelines.** *bioRxiv.* 2019 [cited 2024 Mar 15]; p. 610741.
[Publisher Full Text](#)
23. Kawaji H, Lizio M, Itoh M, *et al.*: **Comparison of CAGE and RNA-seq transcriptome profiling using clonally amplified and single-molecule next-generation sequencing.** *Genome Res.* 2014 Apr; **24**(4): 708–717.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
24. Xia Z, Donehower LA, Cooper TA, *et al.*: **Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types.** *Nat. Commun.* 2014 Nov 20; **5**(1): 5274.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Katz Y, Wang ET, Silterra J, *et al.*: **Quantitative visualization of alternative exon expression from RNA-seq data.** *Bioinformatics.* 2015 Jul 1; **31**(14): 2400–2402.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. Vaquero-García J, Barrera A, Gazzara MR, *et al.*: **A new view of transcriptome complexity and regulation through the lens of local splicing variations.** Valcárcel J, editor. *elife.* 2016 Feb 1; **5**(5): e11752.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. Ritchie ME, Phipson B, Wu D, *et al.*: **Limma powers differential expression analyses for RNA-sequencing and microarray studies.** *Nucleic Acids Res.* 2015 Apr 20; **43**(7): e47.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.** *Bioinformatics.* 2010 Jan 1; **26**(1): 139–140.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
29. Wang W, Qin Z, Feng Z, *et al.*: **Identifying differentially spliced genes from two groups of RNA-seq samples.** *Gene.* 2013 Apr 10; **518**(1): 164–170.
[PubMed Abstract](#) | [Publisher Full Text](#)
30. Drewe P, Stegle O, Hartmann L, *et al.*: **Accurate detection of differential RNA processing.** *Nucleic Acids Res.* 2013 May 1; **41**(10): 5189–5198.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
31. Hartley SW, Mullikin JC: **Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq.** *Nucleic Acids Res.* 2016 Sep 6; **44**(15): e127.
[PubMed Abstract](#) | [Publisher Full Text](#)
32. Wang X, Cairns MJ: **SeqGSEA: a Bioconductor package for gene set enrichment analysis of RNA-Seq data integrating differential expression and splicing.** *Bioinforma Oxf. Engl.* 2014 Jun 15; **30**(12): 1777–1779.
[PubMed Abstract](#) | [Publisher Full Text](#)
33. Hiller D, Jiang H, Xu W, *et al.*: **Identifiability of isoform deconvolution from junction arrays and RNA-Seq.** *Bioinformatics.* 2009 Dec 1; **25**(23): 3056–3059.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Nowicka M, Robinson MD: **DRIMSeq: a Dirichlet-multinomial framework for multivariate count outcomes in genomics.** *F1000Research.* 2016 Dec 6; **5**: 1356.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. Tekath T, Dugas M: **Differential transcript usage analysis of bulk and single-cell RNA-seq data with DTUrtle.** *Bioinformatics.* 2021 Nov 1; **37**(21): 3781–3787.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
36. Vitting-Seerup K, Sandelin A: **IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences.** *Bioinformatics.* 2019 Nov 1; **35**(21): 4469–4471.
[PubMed Abstract](#) | [Publisher Full Text](#)
37. Love MI, Soneson C, Patro R: **Swimming downstream: statistical analysis of differential transcript usage following Salmon quantification.** *F1000Res.* 2018 [cited 2023 Feb 16].
[Publisher Full Text](#) | [Reference Source](#)
38. Honeyman JN, Simon EP, Robine N, *et al.*: **Detection of a Recurrent DNAJB1-PRKACA Chimeric Transcript in Fibrolamellar Hepatocellular Carcinoma.** *Science.* 2014 Feb 28; **343**(6174): 1010–1014.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
39. Trincado JL, Entizne JC, Hysenaj G, *et al.*: **SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions.** *Genome Biol.* 2018 Mar 23; **19**(1): 40.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
40. Shen S, Park JW, Lu Z, *et al.*: **rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data.** *Proc. Natl. Acad. Sci.* 2014 Dec 23; **111**(51): E5593–E5601.
[Publisher Full Text](#)
41. Sterne-Weiler T, Weatheritt RJ, Best AJ, *et al.*: **Efficient and Accurate Quantitative Profiling of Alternative Splicing Patterns of Any Complexity on a Laptop.** *Mol. Cell.* 2018 Oct 4; **72**(1): 187–200.e6.
[PubMed Abstract](#) | [Publisher Full Text](#)

42. Dong C, He F, Berkowitz O, *et al.*: **Alternative Splicing Plays a Critical Role in Maintaining Mineral Nutrient Homeostasis in Rice (*Oryza sativa*)**. *Plant Cell*. 2018 Oct 1; **30**(10): 2267–2285.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
43. Tarazona S, Furió-Tarí P, Turrà D, *et al.*: **Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package**. *Nucleic Acids Res*. 2015 Dec 2; **43**(21): e140.
[PubMed Abstract](#) | [Publisher Full Text](#)
44. Zhao K, Lu Z, Xiang, *et al.*: **GLIMMPS: robust statistical model for regulatory variation of alternative splicing using RNA-seq data**. *Genome Biol*. 2013 Jul 22; **14**(7): R74.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
45. Halperin RF, Hegde A, Lang JD, *et al.*: **Improved methods for RNAseq-based alternative splicing analysis**. *Sci. Rep.* 2021 May 24; **11**(1): 10740.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
46. Zhang Z, Pan Z, Ying Y, *et al.*: **Deep-learning augmented RNA-seq analysis of transcript splicing**. *Nat. Methods*. 2019 Apr; **16**(4): 307–310.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
47. Anders S, Reyes A, Huber W: **Detecting differential usage of exons from RNA-seq data**. *Genome Res*. 2012 Oct 1; **22**(10): 2008–2017.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
48. Leng N, Dawson JA, Thomson JA, *et al.*: **EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments**. *Bioinformatics*. 2013 Apr 15; **29**(8): 1035–1043.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
49. Mehmood A, Laiho A, Venäläinen MS, *et al.*: **Systematic evaluation of differential splicing tools for RNA-seq studies**. *Brief. Bioinform.* 2020 Dec 1; **21**(6): 2052–2065.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
50. Liu R, Loraine AE, Dickerson JA: **Comparisons of computational methods for differential alternative splicing detection using RNA-seq in plant systems**. *BMC Bioinformatics*. 2014 Dec 16; **15**(1): 364.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
51. Merino GA, Conesa A, Fernández EA, *et al.*: **A benchmarking of workflows for detecting differential splicing and differential expression at isoform level in human RNA-seq studies**. *Brief. Bioinform.* 2019 Mar 25; **20**(2): 471–481.
[PubMed Abstract](#) | [Publisher Full Text](#)
52. Jiang M, Zhang S, Yin H, *et al.*: **A comprehensive benchmarking of differential splicing tools for RNA-seq analysis at the event level**. *Brief. Bioinform.* 2023 May 1; **24**(3): bbad121.
[PubMed Abstract](#) | [Publisher Full Text](#)
53. Kannan K, Wang L, Wang J, *et al.*: **Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing**. *Proc. Natl. Acad. Sci.* 2011 May 31; **108**(22): 9172–9177.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
54. Griebel T, Zacher B, Ribeca P, *et al.*: **Modelling and simulating generic RNA-Seq experiments with the flux simulator**. *Nucleic Acids Res*. 2012 Nov 1; **40**(20): 10073–10083.
[PubMed Abstract](#) | [Publisher Full Text](#)
55. Katz Y, Wang ET, Airolidi EM, *et al.*: **Analysis and design of RNA sequencing experiments for identifying isoform regulation**. *Nat. Methods*. 2010 Dec; **7**(12): 1009–1015.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
56. Robinson JT, Thorvaldsdóttir H, Winckler W, *et al.*: **Integrative genomics viewer**. *Nat. Biotechnol.* 2011 Jan; **29**(1): 24–26.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
57. Li YI, Knowles DA, Humphrey J, *et al.*: **Annotation-free quantification of RNA splicing using LeafCutter**. *Nat. Genet.* 2018 Jan; **50**(1): 151–158.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
58. Benegas G, Fischer J, Song YS: **Robust and annotation-free analysis of alternative splicing across diverse cell types in mice**. *Eyras E, Manley JL, editors. eLife*. 2022 Mar 1; **11**: e73520.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
59. Mapleson D, Venturini L, Kaithakottil G, *et al.*: **Efficient and accurate detection of splice junctions from RNA-seq with Portcullis**. *GigaScience*. 2018 Dec 1; **7**(12): giy131.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
60. Eid J, Fehr A, Gray J, *et al.*: **Real-Time DNA Sequencing from Single Polymerase Molecules**. *Science*. 2009 Jan 2; **323**(5910): 133–138.
[Publisher Full Text](#)
61. Derrington IM, Butler TZ, Collins MD, *et al.*: **Nanopore DNA sequencing with MspA**. *Proc. Natl. Acad. Sci.* 2010 Sep 14; **107**(37): 16060–16065.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
62. Feng Y, Zhang Y, Ying C, *et al.*: **Nanopore-based Fourth-generation DNA Sequencing Technology**. *Genomics Proteomics Bioinformatics*. 2015 Feb 1; **13**(1): 4–16.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
63. Jain M, Koren S, Miga KH, *et al.*: **Nanopore sequencing and assembly of a human genome with ultra-long reads**. *Nat. Biotechnol.* 2018 Apr; **36**(4): 338–345.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
64. Wright DJ, Hall NAL, Irish N, *et al.*: **Long read sequencing reveals novel isoforms and insights into splicing regulation during cell state changes**. *BMC Genomics*. 2022 Jan 10; **23**(1): 42.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
65. Wyman D, Balderrama-Gutierrez G, Reese F, *et al.*: **A technology-agnostic long-read analysis pipeline for transcriptome discovery and quantification**. *bioRxiv*. 2020 [cited 2023 Sep 7]; p. 672931.
[Publisher Full Text](#)
66. Hu Y, Fang L, Chen X, *et al.*: **LIQA: long-read isoform quantification and analysis**. *Genome Biol*. 2021 Jun 17; **22**(1): 182.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
67. Carneiro MO, Russ C, Ross MG, *et al.*: **Pacific biosciences sequencing technology for genotyping and variation discovery in human data**. *BMC Genomics*. 2012 Aug 5; **13**(1): 375.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
68. Jain M, Fiddes IT, Miga KH, *et al.*: **Improved data analysis for the MinION nanopore sequencer**. *Nat. Methods*. 2015 Apr; **12**(4): 351–356.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
69. Hon T, Mars K, Young G, *et al.*: **Highly accurate long-read HiFi sequencing data for five complex genomes**. *Sci. Data*. 2020 Nov 17; **7**(1): 399.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
70. Baid G, Cook DE, Shafin K, *et al.*: **DeepConsensus improves the accuracy of sequences with a gap-aware sequence transformer**. *Nat. Biotechnol.* 2023 Feb; **41**(2): 232–238.
[PubMed Abstract](#) | [Publisher Full Text](#)
71. Amarasinghe SL, Su S, Dong X, *et al.*: **Opportunities and challenges in long-read sequencing data analysis**. *Genome Biol*. 2020 Feb 7; **21**(1): 30.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
72. Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows**. *Nat. Biotechnol.* 2017 Apr; **35**(4): 316–319.
[PubMed Abstract](#) | [Publisher Full Text](#)
73. Draper BJ: **Selecting differential splicing methods: Practical Considerations - R Scripts and Data**. *Zenodo*. 2024.
[Publisher Full Text](#)

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research