

SOFTWARE

Open Access

# spliceR: an R package for classification of alternative splicing and prediction of coding potential from RNA-seq data

Kristoffer Vitting-Seerup<sup>1,2</sup>, Bo Torben Porse<sup>2,3,4</sup>, Albin Sandelin<sup>1,2\*</sup> and Johannes Waage<sup>1,2,3,4\*</sup>

## Abstract

**Background:** RNA-seq data is currently underutilized, in part because it is difficult to predict the functional impact of alternate transcription events. Recent software improvements in full-length transcript deconvolution prompted us to develop spliceR, an R package for classification of alternative splicing and prediction of coding potential.

**Results:** spliceR uses the full-length transcript output from RNA-seq assemblers to detect single or multiple exon skipping, alternative donor and acceptor sites, intron retention, alternative first or last exon usage, and mutually exclusive exon events. For each of these events spliceR also annotates the genomic coordinates of the differentially spliced elements, facilitating downstream sequence analysis. For each transcript isoform fraction values are calculated to identify transcript switching between conditions. Lastly, spliceR predicts the coding potential, as well as the potential nonsense mediated decay (NMD) sensitivity of each transcript.

**Conclusions:** spliceR is an easy-to-use tool that extends the usability of RNA-seq and assembly technologies by allowing greater depth of annotation of RNA-seq data. spliceR is implemented as an R package and is freely available from the Bioconductor repository (<http://www.bioconductor.org/packages/2.13/bioc/html/spliceR.html>).

**Keywords:** spliceR, RNA-Seq, Alternative splicing, Nonsense mediated decay (NMD), Isoform switch

## Background

Alternative splicing is one of the most important RNA modifications, leading to protein diversification and contributing to the complexity of higher organisms [1]. Recent advances in RNA sequencing (RNA-seq), combined with modern RNA-seq assembly software such as Cufflinks [2], allows for high-resolution profiling of the RNA landscape. The technological and computational advances enables identification of a catalog of distinctly spliced transcripts originating from the same transcription unit/gene. These full-length transcripts are however underutilized, in part because it is difficult to predict the functional impact of alternate transcription events leading to the different transcripts. In addition to the potential for impact on protein domains, alternate splicing

may also alter RNA processing, stability and localization. Nonsense mediated decay (NMD) is tightly linked to alternate splicing, and the mechanism by which even small changes in alternative splicing can result in degradation via NMD is well described [3]. To predict splicing events that may lead to these functional changes, it is necessary to classify the type of event as well as annotate the genomic position of the regions that are differentially spliced. Such classifications also enable systematic follow-up analyses, such as sequence analysis of the differentially spliced regions to infer the underlying mechanisms. However, at present there are no available tools that adequately perform these analyses. Existing methods, including MISO [4], Astalavista [5] and DiffSplice [6], do not output the genomic coordinates of differentially spliced regions [4,7-11], have insufficient classification of alternative splicing (i.e., only a subset of alternative splicing classes are supported) [4-11], or cannot assess novel features [4,7,8,12]. Furthermore, none of the existing tools for analyzing alternative splicing support NMD predictions [4-12].

\* Correspondence: [albin@binf.ku.dk](mailto:albin@binf.ku.dk); [johannes.waage@gmail.com](mailto:johannes.waage@gmail.com)

<sup>1</sup>Department of Biology, The Bioinformatics Centre, University of Copenhagen, Ole Maaloes Vej 5, Copenhagen DK2200, Denmark

<sup>2</sup>Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Ole Maaloes Vej 5, Copenhagen DK-2200, Denmark

Full list of author information is available at the end of the article

This prompted us to develop the R package spliceR. spliceR uses the full-length transcripts created by RNA-seq assemblers to detect single- and multiple exon skipping/inclusion (ESI, MESI), alternative donor and acceptor sites (A5, A3), intron retention (IR), alternative first or last exon usage (ATSS, ATTS), and mutually exclusive exon events (MEE). spliceR annotates the genomic coordinates of the differentially spliced elements, facilitating downstream sequence analysis. Finally, spliceR predicts the coding potential of transcripts, calculates untranslated region (UTR) and open reading frame (ORF) lengths and predicts whether transcripts are NMD-sensitive based on compatible annotated start codon positions and their downstream ORF.

## Implementation

### Retrieval of data

spliceR is implemented as an R package and is freely available from the Bioconductor repository (<http://www.bioconductor.org/packages/2.13/bioc/html/spliceR.html>). It is based on standard Bioconductor [13] classes such as GRanges, allowing for full flexibility and modularity, and support for all species and versions supported in the Bioconductor annotation packages. An example dataset is included to allow easy exploration of the package.

spliceR is compatible with the output from any full-length RNA-seq assembler, but was designed to integrate with Cufflinks and includes a dedicated function that retrieves all relevant information from the SQL database generated by Cufflinks axillary R package cummeRbund. In future versions new functions dedicated to import data from other RNA-seq assemblers will be introduced. The R code for a standard workflow based on Cufflinks data is illustrated in Figure 1. A workflow using output from other full-length RNA-seq assemblers is provided in the spliceR Bioconductor documentation.

### Classification of alternative splicing

For each gene, spliceR constructs the hypothetical pre-RNA based on the exon information from all transcripts originating from that gene. Subsequently, all transcripts are compared to this hypothetical pre-RNA in a pairwise manner, and alternative splicing events are classified and annotated (see Figure 2 for a schematic overview). Alternatively, spliceR can be configured to use the most expressed transcript as the reference transcript instead of the theoretical pre-RNA. This may be useful in perturbation scenarios where investigators are interested in deviation from normal conditions. For statistical assessment of differential splicing, users can access the transcript fidelity status and *P*-values of Cufflinks, or can easily apply other R-packages that are tailored for this purpose, including edgeR [14], deseq [15], and baySeq [16].

### Isoform fraction values

For each transcript and condition, spliceR calculates an isoform fraction (IF) value, which is calculated as (transcript expression) / (gene expression) \*100 to represent the contribution of a transcript to the expression of the parent gene. Furthermore delta-IF (dIF) values, which measure the absolute change in IF values between conditions, are also calculated. Since these dIF values measure changes in relative abundance of isoforms within a gene, they are well suited for identifying and analyzing changes in the isoform usage. Such analysis does however require that the expression values of transcripts and genes have been normalized to length, sequencing depths, and possibly other biases, a standard feature of the Cufflinks FPKM (fragments per kilobase of transcript per million fragments sequenced) metric. For data from other assemblers, the user may need to accommodate this requirement manually. These IF values are highly dependent on the underlying data quality and relative expression

```
# A) Retrieve data from Cufflinks
cuffDB <- readCufflinks(dir='./cuffdiff_output/', rebuild=TRUE, gtfFile='./cuff-
merge/merged.gtf') # create SQL database via cummeRbund

mySpliceRList <- prepareCuff(cuffDB) # Extract data from SQL database

# B) Identify ORFs and annotate PTCs in transcripts
require("BSgenome.Hsapiens.UCSC.hg19", character.only = TRUE) # load genome sequence

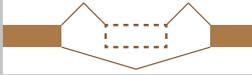
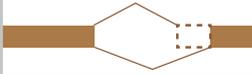
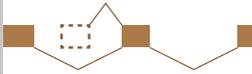
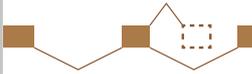
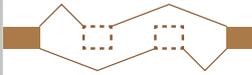
ucscCDS <- getCDS(selectedGenome="hg19", repoName="UCSC") # Get annotated ORFs

mySpliceRList <- annotatePTC(mySpliceRList, ucscCDS, Hsapiens) # Analyze ORFs

# C) Analyze alternative splicing in transcripts
mySpliceRList <- spliceR(mySpliceRList, compareTo='preTranscript', filters='isoOK')

# D) Create GTF file
generateGTF(mySpliceRList, filters="isoOK", filePrefix='./outputPaht/outputName')
```

**Figure 1** The R code for a standard spliceR analysis. The R code necessary to run a standard spliceR analysis based on Cuffdiff output. **A)** The R code to generate a spliceRList from Cuffdiff output. **B-D)** The R code for making a standard spliceR analysis.

	Event type	No. of events
	Exon skipping/inclusion (ESI)	1,619
	Mult. exon skipping/inclusion (MESI)	190
	Intron retention (IR)	256
	Alternative 5' splice site (A5)	755
	Alternative 3' splice site (A3)	733
	Alternative transcription start site (ATSS)	1,381
	Alternative transcription termination site (ATTS)	1,125
	Mutually exclusive exons (MEE)	22
	All events	6,121

**Figure 2 Number of individual alternative splicing events identified.** A schematic structure of each alternative splicing type, along with the associated name, abbreviation and the number of classified events in Usp49 KD RNA seq data.

strengths, and should be viewed as a helpful ranking statistic for identifying changes in alternative splicing between conditions.

#### Analysis of coding potential

spliceR retrieves the genomic exon sequences of each transcript from one of the Bioconductor annotation files. ORF annotation is then retrieved from the UCSC Genome Browser repository either from RefSeq, UCSC or Ensembl, as specified by the user. Alternatively, a custom ORF-table can be supplied. For each transcript, the most upstream compatible start codon is identified, the downstream sequence is translated, and the relative position of the first in-frame stop codon and the distance to the final exon-exon junction is recorded and returned to the user. Transcripts are marked NMD-sensitive if the stop codon falls more than 50 nt upstream of the final exon-exon junction, indicating a pre-mature stop codon (PTC). This is based on the literature consensus [17], however this setting is user-configurable. The position of the start codon is also annotated, which, in combination with the found stop codon and the annotated transcript lengths, enables users to calculate 5'UTR lengths, ORF lengths and 3'UTR lengths. In future versions, we plan

to include alternative methods of coding region prediction, such as the logistic regression model implemented in the program CPAT [18].

#### Visualization and data integration tools

The spliceR package generates a GTF file that can be uploaded to the UCSC genome browser, to help users visualize transcripts and to allow for integration of the RNA-seq analysis with external annotation sources. This spliceR GTF file has two main advantages over the corresponding GTF file generated by Cufflinks' Cuffmerge tool: Firstly, spliceR's optional filters uses stringent criteria, e.g. requiring that the transcript should be expressed, or that Cufflinks marks the transcript deconvolution as successful. In our experience, this removes up to 80% of transcripts originally predicted to belong to the same gene in the Cufflinks GTF file (data not shown), making the GTF file generated by spliceR more suitable for visual analysis.

Secondly, spliceR can color-code transcripts according to their expression level within the parent gene. This feature facilitates easy visualization of changes in gene and transcript expression both within and between conditions.

The tabulated output of spliceR facilitates a number of downstream analyses, including identification of transcripts that exhibit major changes between samples, the ability to filter the output for specific splicing classes, as well as sequence analysis in or around regions that are spliced in or out between samples. Examples include detection of enriched motifs, or identification of protein domains, miRNA response elements or localization signals that are spliced in or out. spliceR facilitates these types of analyses by outputting the genomic coordinates of each alternatively spliced element.

## Results and discussion

To show the power and versatility of spliceR we reanalyzed the publicly available RNA-seq dataset from Zhang *et al.* [19], which compared two experimental conditions: the human colorectal cell line HT116 with and without a siRNA directed towards Usp49, a histone H2B deubiquitinase. To compare our approach with the original analysis, we used the original Cufflinks data (GEO: GSE38100). Transcripts successfully deconvoluted by cufflinks were used as the input into spliceR. The resulting dataset contained 5,496 single-transcript genes, and 1,867 multi-transcript genes. Combined, the multi-transcripts genes were predicted to express 4,612 unique transcripts 2.47 transcripts per gene). The analysis and results presented here are based on only five lines of R code, (modified to hg18), shown in Figure 1B-D, which illustrates the power and ease-of-use of spliceR. For reference, the spliceR analysis took ~30 minutes on a typical laptop (MacBook Pro 2.5Ghz i5, 8 GB RAM).

### Splicing pattern and transcript structure

To validate the transcripts generated by Cufflinks, the two first and two last nucleotides of all introns, corresponding to the locations of splice site consensus sequences, were extracted from both the transcripts generated by Cufflinks as well as reference transcripts from RefSeq and Gencode. The extracted dinucleotides were then compared to the canonical splicing motifs and the percentage of dinucleotides agreeing with the classical splice site motif was analyzed (Table 1). The transcripts predicted by Cufflinks were spliced in accordance

**Table 1 Frequency of splice site consensus sequences**

	Usp49 RNA-seq	Gencode	Refseq
5' end (GT)	93.23	89.72	92.08
3' end (AG)	93.70	90.64	92.71

The two first and last nucleotides, corresponding to the splice site consensus sequences, were extracted from all exons originating from the RNA-seq data, Gencode, and RefSeq. The percentage of dinucleotides identical to the canonical motif was calculated.

with the hyper-conserved splicing motifs as frequently as transcripts originating from RefSeq or Gencode. To further validate the transcripts obtained through Cufflinks we used Cufflinks' cuffcompare tool (v 2.1.1) against all RefSeq transcripts. This showed that the sensitivity and specificity for both introns and exons detection are high (always over 0.91 and 0.81, respectively, and typically much higher) (Table 2). These two analyses indicate that the splicing pattern and transcript structure observed in the Cufflinks derived transcripts are of high quality and suitable for further analyses.

### Alternative splicing and NMD

From the Usp49 KD dataset spliceR identified a total of 6,121 alternative splicing events (Additional file 1: Table S1), distributed across the different splicing classes shown in Figure 2. spliceR found 8,179 (80.9%) transcripts without a PTC (PTC-), 642 (6.4%) transcripts with a PTC (PTC+) while 1,287 (12.7%) transcripts did not have any annotated compatible start codons. Similar fractions of transcripts were predicted to be NMD sensitive when all transcripts from RefSeq (8.2%) and Gencode (9.7%) were analyzed with spliceR, indicating that a non-neglectable fraction of transcripts could be NMD sensitive. By using spliceR's annotation of start and stop codons, the length of both 5'UTRs, 3'UTRs and ORFs were analyzed, but no changes between conditions were found (data not shown).

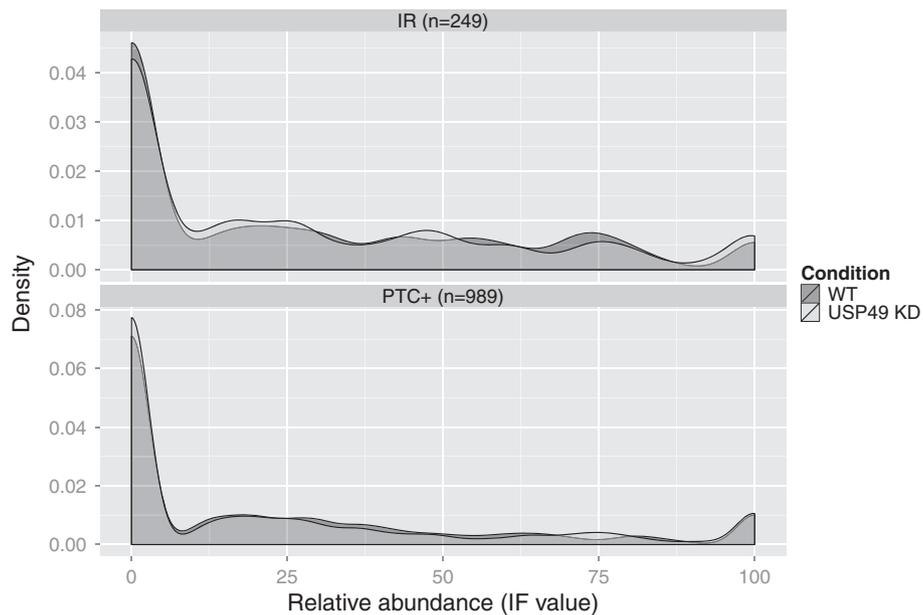
### Splicing efficiency

Zhang and colleagues reported that a subset of transcripts were enriched for intron retention following Usp49 depletion [19], leading to the hypothesis that Usp49 KD reduced the splicing efficiency of pre-RNA molecules. If Usp49 KD impaired splicing efficiency, an increase in the relative abundance of transcripts with IR when comparing WT and KD would be expected. Since the relative abundance of transcripts is measured by IF values, we tested this hypothesis by comparing the distributions of IF values from the subset of transcripts with IR (Figure 3). This analysis shows that transcripts with IR had no global changes in their relative abundance ( $P = 0.71$ , Mann-Whitney U test) indicating that the global splicing efficiency was unchanged. This type

**Table 2 Comparison of the analyzed transcripts to RefSeq**

	WT GTF		Usp49KD GTF	
	Sensitivity	Specificity	Sensitivity	Specificity
Exon level	97.90	89.20	98.00	88.70
Intron level	96.40	97.00	96.40	96.90
Intron chain level	91.70	81.90	91.80	81.90

Cufflinks' cuffcompare tool (v. 2.1.1.) was used to compare RefSeq to the obtained transcripts and the sensitivity and specificity (as defined [20]) for exons, introns, and intron chains are shown.



**Figure 3 Relative abundance of transcripts.** All NMD + transcripts (bottom) and all transcripts with IR (top) was extracted and the density distributions of the IF values from WT and Usp49 KD were plotted.

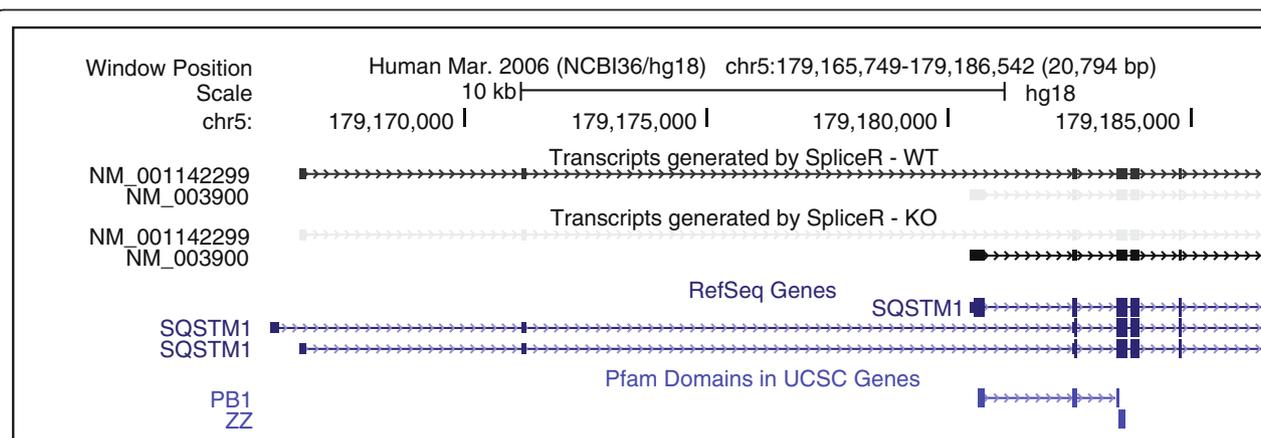
of analysis could however be used to analyze changes in isoform usage in any subset of transcripts that the user could find interesting, for example all NMD sensitive transcripts (Figure 3).

### Transcript switching

We next assessed transcripts whose relative abundance was altered by the Usp49KD, by filtering for genes that had both a large positive and large negative dIF value (corresponding to a binary transcript-switch). 183 high confidence transcript switches were found: in 18 instances (~9.8%), an NMD-negative transcript was down-regulated while a NMD-sensitive transcript was up-regulated. This

illustrates that failing to assess the NMD sensitivity can lead to overestimation of the number of functionally relevant transcript switches.

The transcript switch in the SQSTM1 gene (Figure 4) illustrates the utility of integrating the spliceR data with information in the UCSC genome browser to identify functional changes conferred by alternate splicing. Visual inspection of the isoform switch was possible by uploading the GTF file generated by spliceR. As seen in Figure 4, KD of Usp49 caused a switch from the long transcript predicted to contain a truncated PB1 domain, to the short transcript predicted to encode an intact PB1 domain.



**Figure 4 Example of transcript switching.** Screen shot from the UCSC genome browser showing the transcript switch found in the SQSTM1 gene. The two top tracks show transcripts generated by the generateGTF() function for WT (top) and Usp49KD (bottom). Darker transcripts are expressed at higher levels. The two bottom tracks indicate RefSeq genes (top) and protein domains identified via Pfam [21] respectively (bottom).

## Conclusion

Here, we have introduced the R package spliceR, which increases the usability and power of RNA-seq and assembly technologies by providing a full overview of alternative splicing events and protein coding potential of transcripts. spliceR is flexible and easily integrated in existing workflows, supports input and output of standard Bioconductor data types, and enables investigators to perform many different downstream analyses of both transcript abundance and differentially spliced elements. We demonstrate the power and versatility of spliceR by showing how new conclusions can be made from existing RNA-seq data.

## Availability and requirements

SpliceR is implemented as an R package, is freely available from the Bioconductor repository and can be installed simply by copy/pasting two lines into an R console.

- Project name: spliceR
- Project home page: <http://www.bioconductor.org/packages/2.13/bioc/html/spliceR.html>
- Operating system(s): Platform independent
- Programming language: R and C
- Other requirements: R v 3.0.2 or higher
- License: GPL
- Any restrictions to use by non-academics: No limitations

## Additional file

**Additional file 1: Table S1.** Tabulated output of the spliceR analysis.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

KVS and JW developed the R package. BP, AS, KVS and JW planned the development and wrote the article. All authors read and approved the final manuscript.

## Acknowledgements

KVS, JW and AS were supported by grants from the Lundbeck Foundation, the Novo Nordisk Foundation, and the RiMod-FTD Joint EU program for Neurodegenerative research to AS. Work in the BTP lab was supported through a center grant from the Novo Nordisk Foundation (The Novo Nordisk Foundation Section for Stem Cell Biology in Human Disease). We thank Dr Christine Wells, Glasgow University, for comments on the manuscript.

## Author details

<sup>1</sup>Department of Biology, The Bioinformatics Centre, University of Copenhagen, Ole Maaloes Vej 5, Copenhagen DK2200, Denmark. <sup>2</sup>Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Ole Maaloes Vej 5, Copenhagen DK-2200, Denmark. <sup>3</sup>The Finsen Laboratory, Rigshospitalet, Faculty of Health Sciences, University of Copenhagen, Ole Maaloes Vej 5, Copenhagen DK2200, Denmark. <sup>4</sup>The Danish Stem Cell Centre (DanStem) Faculty of Health Sciences, University of Copenhagen, Ole Maaloes Vej 5, Copenhagen DK2200, Denmark.

Received: 15 November 2013 Accepted: 17 March 2014

Published: 23 March 2014

## References

1. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R, Kim T, Misquitta-Ali CM, Wilson MD, Kim PM, Odom DT, Frey BJ, Blencowe BJ: **The evolutionary landscape of alternative splicing in vertebrate species.** *Science* 2012, **338**:1587–1593.
2. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: **Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.** *Nat Biotechnol* 2010, **28**:511–515.
3. Guthrie C, Steitz J, Lejeune F, Maquat LE: **Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells.** *Curr Opin Cell Biol* 2005, **17**:309–315.
4. Katz Y, Wang ET, Airolidi EM, Burge CB: **Analysis and design of RNA sequencing experiments for identifying isoform regulation.** *Nat Methods* 2010, **7**:1009–1015.
5. Foissac S, Sammeth M: **ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets.** *Nucleic Acids Res* 2007, **35**:W297–W299.
6. Hu Y, Huang Y, Du Y, Orellana CF, Singh D, Johnson AR, Monroy A, Kuan P-F, Hammond SM, Makowski L, Randell SH, Chiang DY, Hayes DN, Jones C, Liu Y, Prins JF, Liu J: **DiffSplice: the genome-wide detection of differential splicing events with RNA-seq.** *Nucleic Acids Res* 2012, **41**:1–18.
7. Rogers MF, Thomas J, Reddy AS, Ben-Hur A: **SpliceGrapher: detecting patterns of alternative splicing from RNA-Seq data in the context of gene models and EST data.** *Genome Biol* 2012, **13**:R4.
8. 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**:2385–2387.
9. Sacomoto GAT, Kielbassa J, Chikhi R, Uricaru R, Antoniou P, Sagot M-F, Peterlongo P, Lacroix V: **KISSPLICE: de-novo calling alternative splicing events from RNA-seq data.** *BMC Bioinforma* 2012, **13**(6):S5.
10. Shen S, Park JW, Huang J, Dittmar KA, Lu Z, Zhou Q, Carstens RP, Xing Y: **MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data.** *Nucleic Acids Res* 2012, **40**:e61.
11. Zhou A, Breese MR, Hao Y, Edenberg HJ, Li L, Skaar TC, Liu Y: **Alt event finder: a tool for extracting alternative splicing events from RNA-seq data.** *BMC Genomics* 2012, **13**(8):S10.
12. Liu Q, Chen C, Shen E, Zhao F, Sun Z, Wu J: **Detection, annotation and visualization of alternative splicing from RNA-Seq data with Splicing-Viewer.** *Genomics* 2012, **99**:178–182.
13. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J: **Bioconductor: open software development for computational biology and bioinformatics.** *Genome Biol* 2004, **5**:R80.
14. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a bioconductor package for differential expression analysis of digital gene expression data.** *Bioinformatics* 2010, **26**:139–140.
15. Anders S, Huber W: **Differential expression analysis for sequence count data.** *Genome Biol* 2010, **11**:R106.
16. Hardcastle TJ, Kelly KA: **baySeq: empirical Bayesian methods for identifying differential expression in sequence count data.** *BMC Bioinformatics* 2010, **11**:422.
17. Weischenfeldt J, Waage J, Tian G, Zhao J, Damgaard I, Jakobsen JS, Kristiansen K, Krogh A, Wang J, Porse BT: **Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns.** *Genome Biol* 2012, **13**:R35.
18. Wang L, Park HJ, Dasari S, Wang S, Kocher J-P, Li W: **CPAT: coding-potential assessment tool using an alignment-free logistic regression model.** *Nucleic Acids Res* 2013, **41**:e74.
19. Zhang Z, Jones A, Joo H-Y, Zhou D, Cao Y, Chen S, Erdjument-Bromage H, Renfrow M, He H, Tempst P, Townes TM, Giles KE, Ma L, Wang H: **USP49 deubiquitinates histone H2B and regulates cotranscriptional pre-mRNA splicing.** *Genes Dev* 2013, **27**:1581–1595.

20. Buset M, Guigo R: Evaluation of gene structure prediction programs. 1996, **367**:353–367.
21. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD: The Pfam protein families database. *Nucleic Acids Res.* 2012, **40**:D290–D301.

doi:10.1186/1471-2105-15-81

**Cite this article as:** Vitting-Seerup *et al.*: spliceR: an R package for classification of alternative splicing and prediction of coding potential from RNA-seq data. *BMC Bioinformatics* 2014 **15**:81.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

