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

Kristoffer Vitting-Seerup[1,2], Bo Torben Porse[2,3,4], Albin Sandelin[1,2*] and Johannes Waage[1,2,3,4,*]

[1]The Bioinformatics Centre, Department of Biology, Ole Maaloes Vej 5, University of Copenhagen, DK2200, Copenhagen, Denmark

[2]Biotech Research and Innovation Centre (BRIC), Ole Maaloes Vej 5, University of Copenhagen, DK-2200 Copenhagen, Denmark

[3]The Finsen Laboratory, Rigshospitalet, Faculty of Health Sciences, Ole Maaloes Vej 5, University of Copenhagen, DK2200 Copenhagen, Denmark

[4] The Danish Stem Cell Centre (DanStem) Faculty of Health Sciences, Ole Maaloes Vej 5, University of Copenhagen, DK2200 Copenhagen Denmark

[*]To whom correspondence should be addressed.

Abstract

Background

With the increasing depth and decreasing costs of RNA-sequencing researchers are now able to profile the transcriptome with unprecedented detail. These advances not only allow for precise approximation of gene expression levels, but also for the characterization of alternative transcript usage/switching between conditions. 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 transcripts output from RNA-seq assemblers, to detect single- and 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. Furthermore, isoform fraction values are calculated for effective post-filtering, i.e. identification of 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 a easy-to-use tool that allows detection of alternative splicing, transcript switching and NMD sensitivity from RNA-seq data, extending the usability of RNA-seq and assembly technologies. 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, Alternative Splicing, RNA-Seq, Nonsense Mediated Decay (NMD).

Background

Alternative splicing is considered to be one of the most important mechanism of pre-RNA modification, leading to protein diversification, and contributing to the complexity of higher organisms [1]. Recent advances in sequencing technology of RNA (RNA-seq), combined with modern RNA-seq assembly software such as Cufflinks [2], now allows for high-resolution unbiased profiling of the RNA landscape. In addition, an exhaustive catalog of transcripts originating from alternative splicing within the same gene unit can be inferred. Based on RNA-seq data, it is estimated that at least 95% of all human genes undergo alternative splicing [3] and recently, the ENCODE Consortium demonstrated that in humans, alternative splicing generates an average of 6.3 different transcripts per gene [4], underlining the importance of the focus on alternative splicing, when analyzing RNA-seq data. A classical example of functional changes, as a consequence of alternative splicing, is the cancer-related isoform switching of the pyruvate kinase protein,

[*]To whom correspondence should be addressed.

where a single exon skipping event in the corresponding pre-mRNA results in a change from the adult isoform, which promotes oxidative phosphorylation, to the embryonic isoform, which promotes aerobic glycolysis [5].

To gain a deeper understanding of alternative splicing it is necessary to both classify the type of splicing event and to annotate the differentially spliced regions. The identification and annotation of differentially spliced regions is imperative if we want to assess the functional consequences, such as loss/gain of domains, miRNA response elements, localization signals etc. Furthermore sequence analysis of the differentially spliced regions, and its surroundings could be used to investigate both the underlying mechanism and regulation of the corresponding splicing event.

Alternative splicing and nonsense mediated decay (NMD) are two tightly linked mechanisms, and the mechanism by which even small changes in alternative splicing can result in degradation via NMD is well described [6]. Since NMD sensitive transcripts will be degraded, and therefore most likely will not have any functional consequences for the cell, it is necessary to assesses the NMD status of all identified RNA transcripts in order to better predict functional consequences based on RNA-seq data.

There are, however, at present no tools that can perform these analyses adequately. Existing tools, such as MISO [7], Astalavista [8] and DiffSplice[9] etc., either cannot output the genomic coordinates of differentially spliced regions [7, 10–14], have insufficient classification of alternative splicing (i.e., only a subset of alternative splicing classes are supported) [7–14], or cannot assess novel features [7, 10, 11, 15]. Furthermore, none of the existing tools for analyzing alternative splicing can predict the NMD status of transcripts [7–15].

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, MSI), alternative donor and acceptor sites (A5, A3), intron retention (IR), alternative first or last exon usage (ATSS, ATTS), and mutually exclusive exon events (MEE). For each of these events spliceR annotates the genomic coordinates of the differentially spliced elements facilitating downstream sequence analysis. Lastly, spliceR predicts the coding potential of transcripts, calculates UTR and ORF lengths and predicts whether transcripts are NMD sensitive based on compatible open reading frames and stop codon positions.

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 fully based on standard Bioconductor [16] 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 can use the output from any full-length RNA-seq assembler, but is especially well integrated with the popular assembler Cufflinks, since spliceR has a dedicated function that retrieves all relevant information from the SQL database generated by Cufflinks axillary R package cummeRbund. A standard workflow based on Cufflinks data is illustrated in Figure 1, while a workflow using output from other full-length RNA-seq assemblers can be found in the spliceR documentation.

Classification of alternative splicing

For each gene, spliceR initially constructs the hypothetical pre-RNA based on the exon information from all transcripts originating from that gene. Subsequently, all transcripts are, in a pairwise manner, compared to

this hypothetical pre-RNA and alternative splicing is 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, which can be an advantage in scenarios where splicing patterns between samples are expected to change substantially. 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.

IF values

For each transcript and condition, spliceR calculates an isoform fraction (IF) value. This IF value is calculated as (transcript expression) / (gene expression) *100% and represents the contribution of a transcript to the expression of the parent gene. Furthermore, a delta-IF (dIF) value, giving the absolute change in IF values between conditions, is also calculated. Since IF values are normalized to changes in gene expression, the size of an IF value represents the relative abundance of the corresponding transcript, in the context of the parent gene. Hence, the size (and sign) of a dIF value represents the change in the relative abundance of a transcript between conditions.

Analysis of coding potential

To assess the coding potential of individual transcripts, spliceR initially retrieves the genomic exon sequences of each transcript from one of the Bioconductor annotation files. Next, ORF annotation is 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 positional data about the most upstream stop codon, including distance to final exon-exon junction, is recorded and returned to the user. Based on the literature consensus [17], 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), although this setting is user-configurable. Positional data about the start codon is also annotated which, in combination with the stop codon information and the annotated transcript lengths, enables users to calculate 5'UTR lengths, ORF lengths and 3'UTR lengths. In future versions, we plan to implement non-CDS dependent alternative methods of coding potential, such as CPAT [18].

Post-analysis

To help users visualize transcripts and to allow integration of the RNA-seq analysis with the vast information found in online repositories, the spliceR package can generate a GTF file that can be uploaded to any genome browser. This SpliceR GTF file has two advantages over the corresponding GTF file generated by Cufflinks' Cuffmerge tool. First, spliceR have the option to filter out transcripts based on a wide range of criteria, e.g. requiring that the transcript should be expressed, or that Cufflinks marks the transcript deconvolution as successful. Such an approach removes as much as 80% of transcripts belonging to the same gene compared to the Cufflinks GTF (data not shown), making the GTF file generated by spliceR more suitable for visual analysis. Second, spliceR has the option of color-coding 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.

Finally, the tabulated output of spliceR facilitates various downstream analyses, including filtering on transcripts that exhibits major changes between samples, filtering for specific splicing classes, or sequence analysis on elements that are spliced in or out between samples. The latter could for example be detection of enriched motifs in, or surrounding, such elements, or identification of protein domains that are spliced in/out. spliceR facilitates this type of analysis by outputting the genomic coordinates of each alternatively spliced element.

Results and discussion

To show the power and versatility of spliceR we analyzed previously published RNA-seq data from Zhang et al [19]. This study identified Usp49 as a histone H2B deubiquitinase, and to assess the importance of this protein RNA-seq following Usp49 knock-down (KD) was performed. To make the original analysis and the one performed here comparable, we used the Cufflinks output from their study, downloaded (GEO: GSE38100) and transcripts marked as successfully deconvoluted by cufflinks was used as input to spliceR. The resulting dataset contained 5,675 single-transcript genes, and 2,497 multi-transcript genes. Combined, the multi-transcripts genes expressed 6,048 unique transcripts, corresponding to 2.42 transcripts per gene. All analysis presented here are based on the output obtained by simply using the five lines of code, modified to hg18, shown in Figure 1 B-D. For reference, this analysis took ~30 minutes on a typical laptop(MacBook pro with a 2.5Ghz i5 processor and 8 GB ram).

Analysis of splicing pattern

In an effort to validate the transcripts generated by Cufflinks, the two first and two last nucleotides of all introns, corresponding to the splice site consensus sequences, were extracted from both the transcripts generated by Cufflinks as well as transcripts originating from RefSeq and Gencode. The extracted dinucleotides were then compared to the canonical splicing motifs and the faction of dinucleotides agreeing with the classical splice site motif was analyzed (Figure 3). The results of this analysis clearly shows that the transcripts generated by Cufflinks spliced in accordance with the hyper-conserved splicing motifs just as often as transcripts originating form both RefSeq and Gencode. This indicates that that the splicing pattern observed in Cufflinks derived transcripts are genuine splicing patterns.

Alternative splicing and NMD

From the Usp49 KD dataset spliceR identified a total of 7,095 alternative splicing events, distributed across the different splicing classes as seen in Figure 2. spliceR furthermore found 8,742 (74.6%) transcripts without a PTC (PTC-), 830 (7.1%) transcripts with a PTC (PTC+) while 2,151 (18.3%) 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. Combined, these results indicates that a non-neglectable faction of transcripts could be NMD sensitive, which illustrates the importance of assessing the NMD sensitivity of transcripts before making conclusions solely based on RNA transcripts. By using the annotated positional information about 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

By analyzing the Usp49 KD data, Zhang and colleagues found that transcripts were enriched for intron retention following Usp49 depletion, leading to the hypothesis that Usp49 KD reduced the splicing efficiency of pre-RNA molecules [19]. 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 Zhang et al.'s hypothesis by comparing the distributions of IF values from the subset of transcripts with IR (Figure 4). Furthermore, we used the same approach to analyze changes in the subset of transcripts that are predicted to be NMD sensitive and the subset of transcript that both have an IR and are predicted to be NMD sensitive. As seen from figure 4, the subset of transcripts with IR shows a general tendency towards lower values in the Usp49 KD, indicating that on a global level the relative abundance of transcripts with IR are reduced. This result suggests that the splicing efficiency actually improves when Usp49 is knocked down. It should however be noted that the relative abundance of a small proportion of the transcripts with IR have increased in the Usp49 KD, and it is possible that this is this subset of transcripts that Zhang et al. identified. The results in figure 4 also indicate that on a global level the relative abundance of the subset of transcripts that are predicted as NMD sensitive is reduced in the Usp49 KD. Interestingly this tendency towards lower numbers is even more

pronounced in the subset of transcripts that both have IR and are PTC+ (Figure 4). This indicates a functional connection between IR and NMD sensitivity, which could be explained by the high probability of either including a stop codons or breaking the reading frame when an introns are retained [20].

Transcript switching

Since dIF values measures changes in the relative abundance of transcripts, they can be used for easy identifying transcript switching by filtering for genes that both have large positive and large negative dIF values (corresponding to a binary transcript-switch). Applying this approach to the Usp49 KD data, 183 high confidence transcript switches was found, 18 of which upregulated a PTC+ transcript while downregulating a PTC- transcript in the KD.

To investigate these transcript switches further, the GTF file generated by spliceR was uploaded to the UCSC genome browser whereby other data, such as known protein domains, could be incorporated. One of the findings was that in the isoform switch of the SQSTM1 gene, where both transcripts are PTC- (data not shown), the KD of Usp49 caused a switch form the long transcript, where the protein domain PB1 is truncated, to the short transcript, where the entire domain is included, indicating a possible gain of function (Figure 5).

The results presented here illustrates how spliceR extends the usability and power of RNA-seq and assembly technologies by extraction information about alternative splicing, coding potential as well as the relative abundance of isoforms, in just a few lines of R code (Figure 1 B-D). Furthermore, we have shown that spliceR can generate new experimentally testable hypothesis, such as gain or loss of function due to changed protein content.

Conclusion

Here, we have introduced the R package spliceR, which harnesses the power of current RNA-seq and assembly technologies, and provides a full overview of alternative splicing events and protein coding potential of transcripts. spliceR is flexible and easily integrated in existing workflows, supporting input and output of standard Bioconductor data types, and enables researches 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

Competing interests

The authors declare that they have no competing interests.

Authors Contribution

KVS and JW developed the R package. BP, AS, KVS and JW planned the development and wrote the article.

Acknowledgements

KVS, JW and AS were supported by grants from the Lundbeck and Novo Nordisk Foundations 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)

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**:1587–93.

2. Trapnell C, Williams B a, 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. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ: **Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing.** *Nat. Genet.* 2008, **40**:1413–5.

4. The ENCODE Project Consortium: **An integrated encyclopedia of DNA elements in the human genome.** *Nature* 2012, **489**:57–74.

5. Christofk HR, Vander Heiden MG, Harris MH, Ramanathan A, Gerszten RE, Wei R, Fleming MD, Schreiber SL, Cantley LC: **The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth.** *Nature* 2008, **452**:230–3.

6. 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.

7. Katz Y, Wang ET, Airoldi EM, Burge CB: **Analysis and design of RNA sequencing experiments for identifying isoform regulation.** *Nat. Methods* 2010, **7**:1009–15.

8. Foissac S, Sammeth M: **ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets.** *Nucleic Acids Res.* 2007, **35**:W297–9.

9. 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.* 2012:1–18.

10. 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.

11. 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**:2385–7.

12. Sacomoto G a T, 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 Bioinformatics* 2012, **13 Suppl 6**:S5.

13. Shen S, Park JW, Huang J, Dittmar K a, 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.

14. 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 Suppl 8**:S10.

15. Liu Q, Chen C, Shen E, Zhao F, Sun Z, Wu J: **Detection, annotation and visualization of alternative splicing from RNA-Seq data with SplicingViewer.** *Genomics* 2012, **99**:178–82.

16. 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.

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**:1581–95.

20. Wong JJ-L, Ritchie W, Ebner OA, Selbach M, Wong JWH, Huang Y, Gao D, Pinello N, Gonzalez M, Baidya K, Thoeng A, Khoo T-L, Bailey CG, Holst J, Rasko JEJ: **Orchestrated Intron Retention Regulates Normal Granulocyte Differentiation**. *Cell* 2013, **154**:583–595.

21. Nicholson P, Yepiskoposyan H, Metze S, Zamudio Orozco R, Kleinschmidt N, Mühlemann O: **Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors.** *Cell. Mol. Life Sci.* 2010, **67**:677–700.

22. Punta M, Coggill 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**:D290–301.

Figure 1: Code example. The R code necessary to run a standard spliceR analysis based on Cuffdiff output standard spliceR analysis based on Cuffdiff output.

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.

Figure 3: Frequency of splice site consensus sequences: 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.

Figure 4: Changes in relative abundance of PTC+ transcripts containing IR. Transcripts were divided into three subsets; IR, PTC+ and IR + PTC+, and the distribution of IF values was visualized for WT and Usp49 KD.

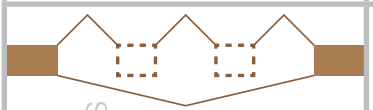
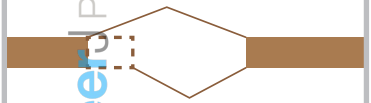
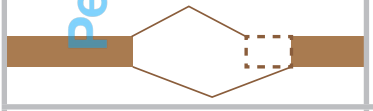
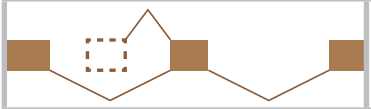
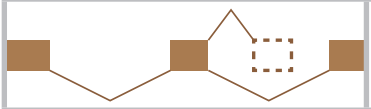
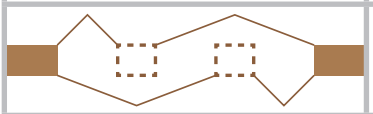
Figure 5: Example of transcript switching. Screen shot from the UCSC genome browser showing the transcript switch found in the SQSTM1 gene. The two top tracks are the transcripts generated by the generateGTF() function for WT (top) and Usp49KD (bottom). Darker transcripts are expressed at higher levels. The two bottom tracks indicates RefSeq genes (top) and protein domains identified via pfam [22] respectively (bottom).

PeerJ PrePrints | https://peerj.com/preprints/80v2/ | v2 received: 17 Dec 2013, published: 20 Nov 2013, doi: 10.7287/peerj.preprints.80v2

```
# 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

PeerJ PrePrints | https://peerj.com/preprints/80v2/ | v2 received: 17 Dec 2013, published: 20 Nov 2013, doi: 10.7287/peerj.preprints.80v2

	Event type	No. of events
	Exon skipping/inclusion (ESI)	1,805
	Mult. exon skipping/inclusion (MESI)	216
	Intron skipping/inclusion ISI	377
	Alternative 5’ splice site A5	825
	Alternative 3’ splice site A3	878
	Alternative transcription start site ATSS	1,585
	Alternative transcription termination site ATTS	1,385
	Mutually exclusive exons MEE	24
	All events	7,095

PeerJ PrePrints | https://peerj.com/preprints/80v2/ | v2 received: 17 Dec 2013, published: 20 Nov 2013, doi: 10.7287/peerj.preprints.80v2

Figure 2

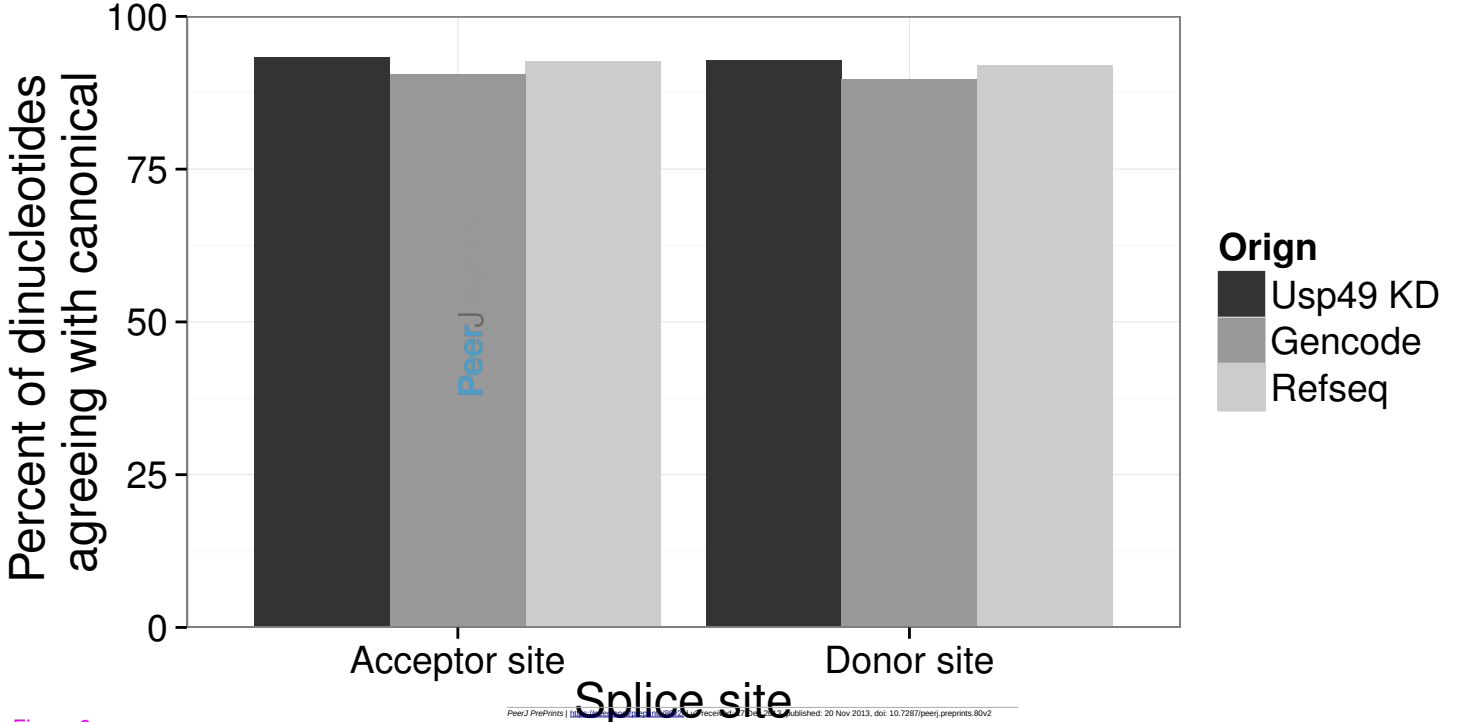
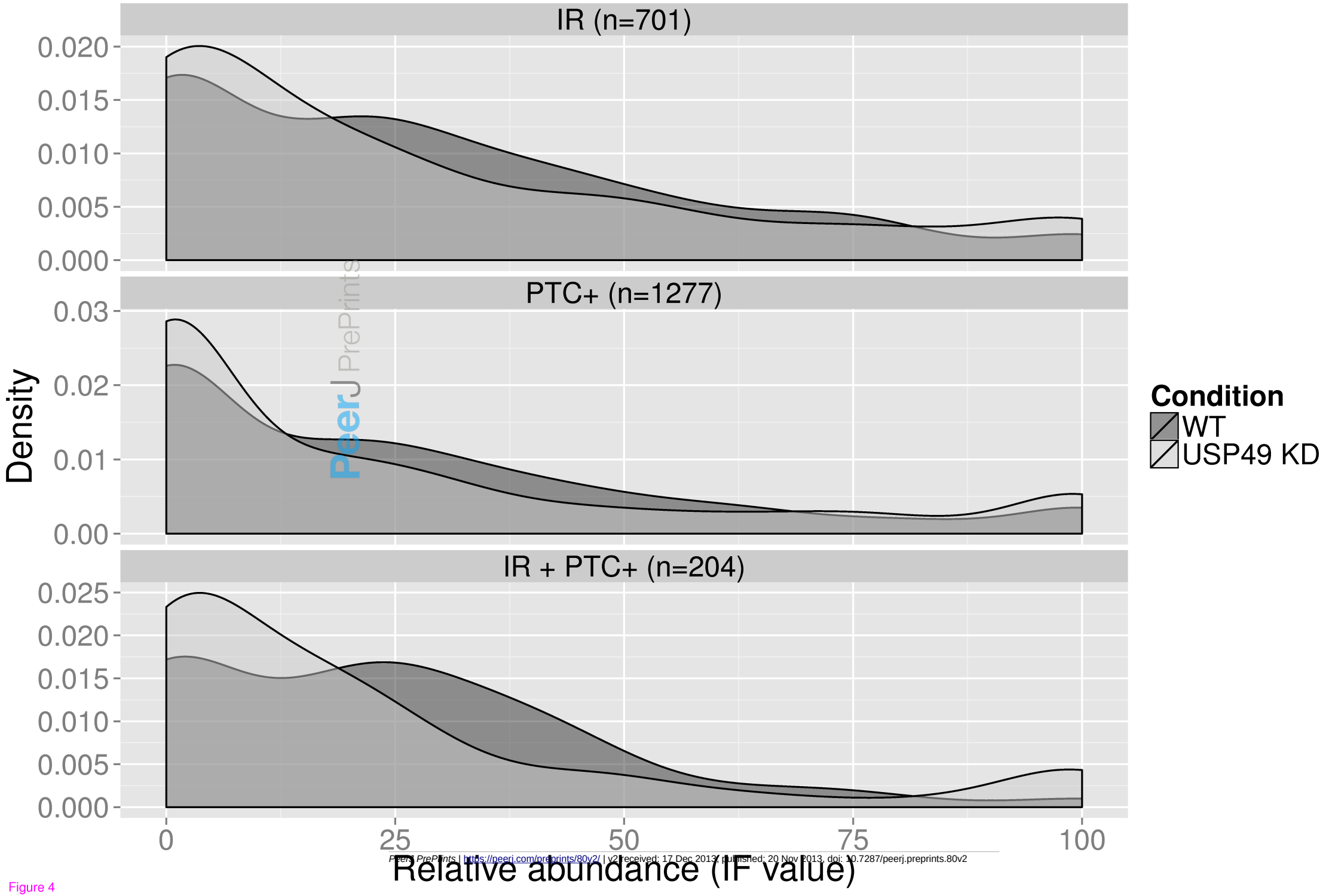
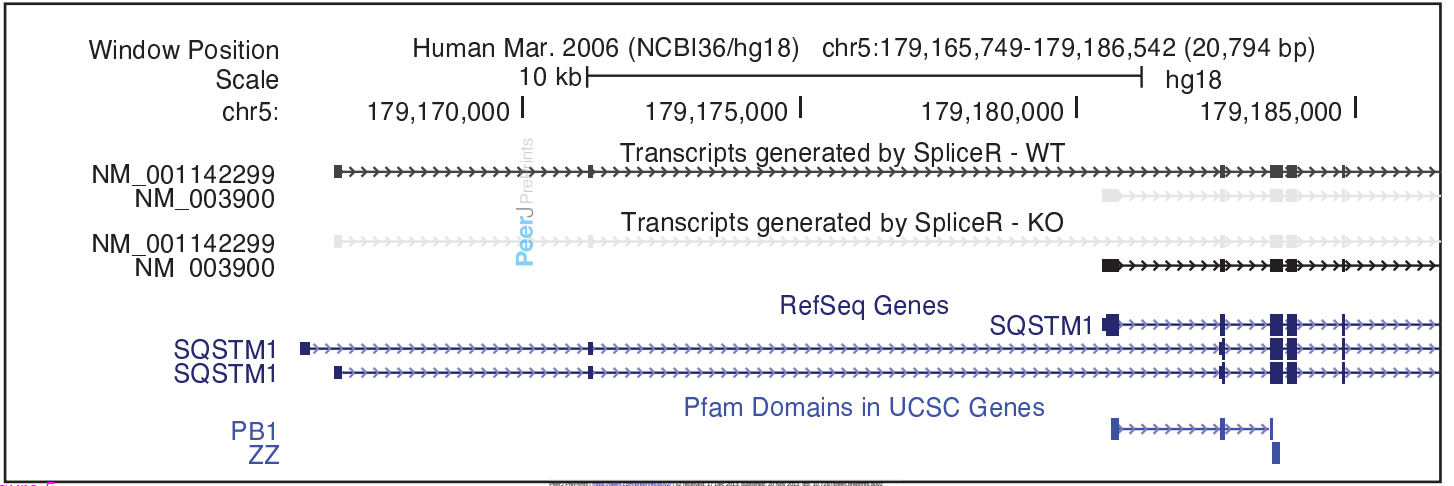


Figure 3



PeerJ PrePrints | https://peerj.com/preprints/80v2/ | v2 received: 17 Dec 2013, published: 20 Nov 2013, doi: 10.7287/peerj.preprints.80v2

Figure 4


Window Position
Human Mar. 2006 (NCBI36/hg18) chr5:179,165,749-179,186,542 (20,794 bp)
Scale
10 kb
hg18
chr5:
179,170,000
179,175,000
179,180,000
179,185,000
Transcripts generated by SpliceR - WT
NM_001142299
NM_003900
Transcripts generated by SpliceR - KO
NM_001142299
NM_003900
RefSeq Genes
SQSTM1
SQSTM1
SQSTM1
Pfam Domains in UCSC Genes
PB1
ZZ

igure 5