ORIGINAL ARTICLE

Differential analysis of alternative splicing events in gene regions using residual neural networks

Simone Ciccolella[1] · Luca Denti[1] · Jorge Avila Cartes[1] · Gianluca Della Vedova[1] · Yuri Pirola[1] · Raffaella Rizzi[1] · Paola Bonizzoni[1]

Abstract
Several computational methods for the differential analysis of alternative splicing (AS) events among RNA-Seq samples typically rely on estimating isoform-level gene expression. However, these approaches are often error-prone due to the interplay of individual AS events, which results in different isoforms with locally similar sequences. Moreover, methods based on isoform-level quantification usually need annotated transcripts. In this work, we leverage the ability of deep learning networks to learn features from images and propose `deepSpecas`, a novel method for event-based AS differential analysis between two RNA-Seq samples. Our method does not rely on isoform abundance estimation, neither on a specific annotation. `deepSpecas` employs an image embedding scheme to represent the alignments of the two samples on the same region and utilizes a residual neural network to predict the AS events possibly expressed within that region. To our knowledge, `deepSpecas` is the first deep learning approach for performing an event-based AS analysis of RNA-Seq samples. To validate `deepSpecas`, we also address the lack of high quality AS benchmark datasets. For this purpose, we manually curated a set of regions exhibiting AS events. These regions were used for training our model and for assessing the predictions of our method. Our results highlight that `deepSpecas` achieves higher precision at the expense of a small reduction in sensitivity. The tool and the manually curated regions are available at https://github.com/sciccolella/deepSpecas.

1 Introduction

A gene is a locus in the genome that contains the instructions for the synthesis of proteins. A gene exhibits a peculiar structure characterized by coding regions, known as *exons*, interleaved with non-coding regions, known as *introns*. During the *splicing*, which is a phase of protein synthesis, exons are retained and included in the messenger RNA (also known as *transcript* or *isoform*), whereas introns are spliced out. A transcript can therefore be described as a continuous sequence of exons which contains the information needed for the production of a protein. Alternative splicing is a regulatory process of gene expression in eukaryotic organisms which allows a single gene to produce multiple transcripts by combining its exons in different ways [1]. These ways are usually referred to as *alternative splicing events*. For instance (see Fig. 1), one or more exons can be excluded from the transcript (*exon skipping* event), only a portion (prefix or suffix) of a given exon is included in the transcript (*alternative 5' and 3' splicing sites* events), or an intron is retained (*intron retention* event). These events enable the production of a variety of isoforms from a single gene and contribute to expand the complexity of proteome, i.e., the set of proteins that an organism can synthesize [2]. Alternative splicing is known to play a crucial role in many biological processes, particularly in cell development and tissue differentiation [3]. Most importantly, alterations and dysfunctions of this mechanism are often associated with genetic diseases and various types of cancer [4–6].

✉ Simone Ciccolella
  simone.ciccolella@unimib.it

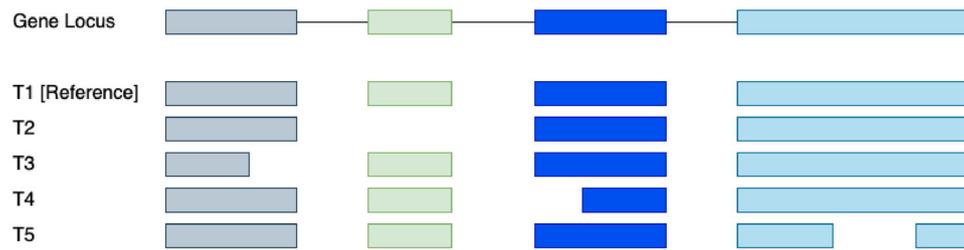1 DISCo, Universita degli Studi di Milano-Bicocca, Milan, Italy

Fig. 1 Example of gene locus composed of four constitutive exons (colored blocks) and three introns (thin lines). T1, T2, T3, T4, and T5 are its potentially expressed alternative isoforms. Each exon that makes up an isoform is represented aligned with the corresponding exon on the locus, even though an isoform is actually a concatenation of exons. Isoform T1, which includes all the exons, is considered the reference to illustrate the alternative splicing events taken into account. Specifically, T2 does not include the green exon and therefore exhibits an exon skipping event compared to T1. T3 includes only a prefix of the gray exon, thus manifesting a competing 5' event (the start of an intron is, in fact, a 5' splicing site). The dual competing 3' event (the end of an intron is, in fact, a 3' splicing site) is present in T4, which includes only a suffix of the blue exon. Finally, T5 presents an intron retention event, as only a prefix and a suffix of the light blue exon are included, and an intron within such exon is retained in T5 compared to T1

Due to the noteworthy effects of alternative splicing, many computational approaches have been developed to shed more light on this fundamental regulation mechanism [7]. Most of these approaches start from the analysis of sequencing data, i.e., fragments of DNA (or RNA in the case of RNA-Seq) which are obtained by "reading" a DNA (RNA) molecule [8]. The output of a sequencing experiments is known as *read sample* and can be computationally described as a set of strings built over the alphabet $\Sigma = \{A, C, G, T, U\}$, where each character corresponds to a nucleotide composing the molecule. The emergence of next-generation sequencing (NGS) technologies, with their high throughput, compared with pre-NGS technologies, brought forth the promise of revolutionizing gene expression analysis in two ways: (1) enabling the exploration of gene expression at the transcript level and (2) facilitating the discovery of novel genes and transcripts within the entire transcriptome, i.e., the full set of known transcripts of an organism [9, 10]. However, early NGS technologies were severely limited in read length, which negatively affected their ability to differentiate repeats. This limitation is relevant in RNA-Seq analyses since, due to alternative splicing, different isoforms usually share large portions of their sequence. As a consequence, *transcript quantification*, which is the task of determining the abundance of the transcripts expressed in a read sample, is very complex, because RNA-Seq reads cannot be unambiguously assigned to their originating transcript. Transcript quantification is typically followed by differential expression analysis that aims at identifying the genes and/or transcripts whose expression is significantly different between conditions. Differential expression is crucial since it enables to focus the analysis only on those genes or transcripts that, among tens of thousands, could potentially be involved in the expression of the phenotype under investigation. Many tools in the literature aim at quantifying transcripts. For example, StringTie [11], Cufflinks [12], Scripture [13], and IsoLasso [14] are tools that reconstruct (or assemble) transcripts from RNA-Seq reads and then perform the quantification of the reconstructed isoforms. Kallisto [15] and Salmon [16] perform transcript quantification taking as input a transcriptome and a set of RNA-Seq reads.

Due to the complexity of transcript reconstruction, the analysis of alternative splicing can be reduced to the analysis of the alternative splicing events that differentiate the transcripts expressed in the RNA-Seq samples. In this case, the focus is not on predicting and quantifying full-length transcripts but on detecting and quantifying AS events that differentiate two conditions. For this task, several alignment-based approaches have been proposed [17–22]. These approaches analyze how RNA-Seq reads are aligned to a reference genome to detect and quantify the events. On the other hand, SUPPA2 [23] analyzes AS events starting from transcript quantification. Approaches like JunctionSeq [24], MAJIQ [25], and leafcutter [26] perform quantification at the splice junction level, whereas DexSeq [27] identifies differentially expressed exons. Observe that all these methods start by aligning reads either to the genome or to the transcriptome and, therefore, they often need the support of an existing gene annotation. Such an annotation improves the quality of the predictions on annotated events, but it also makes it almost impossible to correctly manage unannotated events, which is one of the limitations we want to address in this paper.

We follow the research direction started by some state-of-art tools, exploiting deep learning (convolutional neural networks CNN, in particular, residual neural networks ResNet) to accomplish bioinformatics tasks such as variant calling, classifying or detecting splicing sites and gene expression analysis [28–31].

For example, DeepVariant [32] identifies variants in NGS reads, SpliceAI [33] exploits a deep learning network to annotate input variants with their predicted effect on
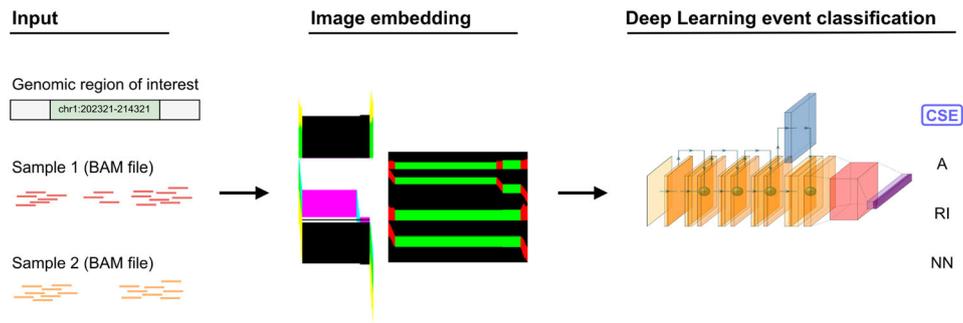
Fig. 2 Graphical representation of `deepSpecas` framework. The tool requires as input a genomic region of interest, represented in standard forward notation, and a pair of samples (conditions) of splice-aligned reads in BAM format; `deepSpecas` proceeds to

produce required image embedding of the two samples for the specified region and using the RNN model outputs the predicted differential AS events between the two samples, or *non-event* if none is expressed

splicing, SpliceRover [34] and DeepSplicer [35] make splice site prediction, and CI-SpliceAI [36] predicts disease causing splicing variants.

We propose a novel method for detecting AS events in two RNA-Seq samples, leveraging on an image embedding schema representing alignments of the samples to the genome. To our knowledge, this is the first method that uses a deep learning approach to detect AS events expressed by two RNA-Seq samples, and the first tool that does not need an annotation. A description of the method is depicted in Fig. 2. Our main goal is to avoid relying on a specific transcript annotation that often constitutes a limitation, as it might be incomplete or even unavailable for contexts such as detecting alternative splicing in tumor cells [37], where we do not have a reliable annotation. We implemented our deep learning method in the tool `deepSpecas` which takes as input the spliced alignments of two RNA-Seq samples to the reference genome and a list of genomic regions (*i.e.*, genomic locations) in order to detect, for each specified region, the AS event differentially expressed in the two conditions.

To validate the performance of `deepSpecas`, we curated a comprehensive benchmark dataset of differentially expressed AS events. Indeed, existing benchmark studies, which may help users in selecting the appropriate tools for their analyses, have several limitations. Usually, these studies use as ground truth only a small subset of well-studied genes or rely on simulated RNA-Seq data, where alternative splicing events are introduced randomly and encompass only few event types. We have evaluated `deepSpecas` with the classical ML accuracy metrics (precision, recall and F1-score), showing extremely high F1-scores (99%+) on (i) train-test-validation and (ii) with fivefold cross-validation with 70:30 splits and (iii) very high accuracy on real dataset extracted by consensus of the two state-of-the-art tools `rMATS` and `SUPPA2`.

The last contribution of this work is a detailed analysis of a real RNA-Seq datasets and a thorough examination of the supported AS events. This investigation consisted of both a visual inspection with the Integrative Genomics Viewer [38] and statistical examination of the (several) predicted AS events that we considered imprecise or doubtful. The resulting set of manually curated AS events can be used by future works as an accurate benchmarking dataset.

Alternative splicing may be particularly important in the study of specific splice isoforms in target genes that are mutated in cancer. In this framework indeed, typically only very few genes show aberrant forms of splicing that promote tumor growth. Since `deepSpecas` is specifically accurate when target regions are under investigation, `deepSpecas` could be useful to process those genes that in a cancer environment selects novel splice isoforms. As pointed out in a recent survey, [39] computing accurate isoform expression from high-throughput RNA-Seq data is still the challenging task, and `deepSpecas` aims to facilitate this task for specific target genes.

2 Methods

`deepSpecas` takes as input the alignments (given in BAM, binary alignment format[1]) of two RNA-Seq samples (denoted as *Sample1* and *Sample2* in the rest of the paper) to the reference genome and a list of genomic regions, referred to as *regions of interest*, where potentially the two samples exhibit an AS event among the following: Cassette Exon/Exon Skipping (CSE), Alternative Acceptor/Donor sites (A) and Intron Retention (IR). `deepSpecas` exploits a residual neural network (ResNet) to decide, for each input region of interest, whether the two samples exhibit a particular alternative splicing event (among the previous ones) relative to each other. For example, an exon included

in *Sample1* but not in *Sample2*, or vice versa (Cassette Exon/Exon Skipping, CSE), a donor/acceptor competing site between an exon in *Sample1* and an exon in *Sample2* (A), or an exon of *Sample1* within which is retained an intron in *Sample2*, or vice versa (RI). The tool will report a *non-event* NN if no alternative event has been detected between the two samples.

The method used by deepSpecas is the following. Alignments (from the two samples) falling within a given region are encoded through an image (bi- or four-dimensional *data tensor* depending on the encoding), and an image for each region of interest is produced. The trained ResNet is fed with each image produced, thereby yielding the response for each one of the considered regions of interest.

Section 2.1 describes the types of images we use for encoding the alignments inside a given region. Section 2.2 describes the generation of the training images, while Secions 2.3 and 2.4 present the model architecture and the model training.

2.1 Encoding the regions of interest

In this work, we explore the accuracy achieved by the ResNet using six different encodings (i.e., data tensors) of the read alignments on each region of interest. The encodings provide an image representation of the alignments in the region and mimic the commonly used views of spliced alignments on genomic viewers such as IGV. In particular, the six encodings we proposed are:

– **Cov:** The image encodes vertical bar plots representing the coverage (*y*-axis) across the genomic region (*x*-axis). Each pixel on the *x*-axis represents a single genomic position and the coverage at each position is the height (in pixels) of the bar. Two different coverages are computed and encoded as distinct channels of the image: the read coverage, *i.e.,*, the number of reads aligned on that position, and the *reference skip* coverage, *i.e.,*, the number of alignments that contain a N in their CIGAR string in that position (in other words, the number of alignments that "support" the presence of an intron in that position). The two coverages of each sample are represented by a single bar plot. The two bar plots of the two samples are vertically stacked. An example is shown in Fig. 3a, in which read coverage uses the red channel and reference skip coverage the green channel.
– **Squish:** The image displays the read alignments of the two samples in the region, one sample on top of the other. Each read alignment is represented as a

horizontal, 1 pixel high, line of the image. Alignments are ordered according to their starting position on the genome. Aligned positions and reference skips (*i.e.,*, Ns in the CIGAR string) are encoded on two separate channels. An example is shown in Fig. 3b, in which aligned positions use the red channel and skipped positions the green channel.
– **Comb:** vertically stacks the two previous encodings, grouped by sample. In other words, the resulting image is the *Cov* encoding and *Squish* encoding of the first sample on top of the *Cov* encoding and *Squish* encoding of the second sample. An example is shown in Fig. 3c.
– **Cov-4D:** Contains the same information as *Cov* but the two samples are superimposed on separate channels instead of being vertically stacked, resulting in a four-channel image (see Fig. 4a).
– **Squish-4D:** Contains the same information as *Squish* but the two samples are superimposed on separate channels instead of being vertically stacked, resulting in a four-channel image (see Fig. 4b).
– **Comb-4D:** contains the same information as *Comb* but the two samples are superimposed on separate channels instead of being vertically stacked, resulting in a four-channel image (see Fig. 4c).

Since maximum coverage and length of the region generally vary across the regions of interest, the images were resized to a width of 1000 pixels and height equal to 200, 400, and 800 depending of the encoding before being fed to the network that, by definition, requires same-sized inputs.

2.2 Generation of training images

Training images have been generated using synthetic RNA-Seq samples to ensure to obtain a comprehensive and curated set of labeled examples, necessary for supervised learning. For simulating the dataset, we constructed a pipeline that utilizes de facto standard methods for each step: (i) we constructed AS events directly from the gene annotation, making it impossible to have noise in the event construction step; (ii) we simulated the reads with the one of most accurate and used tool in the literature for RNA sequence simulation (Polyester [40]); (iii) we aligned the reads using the gold standard aligner in the field (STAR [41]). All these step should assure that the simulation follows the real data as much as possible.

We started from the GRCh38 v109 gene annotation keeping only protein-coding genes with multiple multi-exon transcripts annotated with the Havana/Havana TAGENE pipeline. Then, from the filtered annotation, we extracted the regions of interest by pairwise comparison of the transcripts of the same gene and keeping all the regions where an AS event, as defined in [23] between the two

---

[1] https://samtools.github.io/hts-specs/SAMv1.pdf.
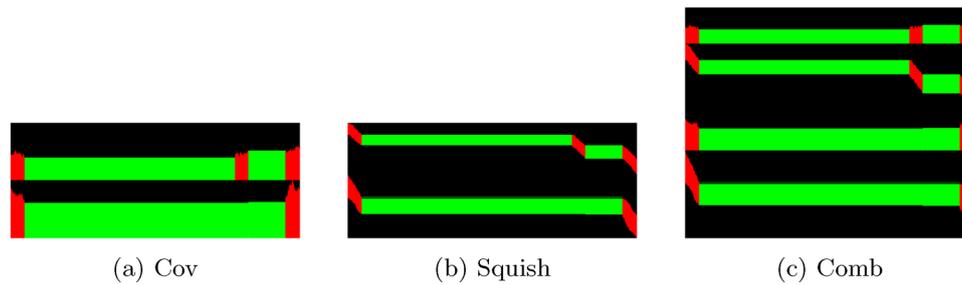
(a) Cov   (b) Squish   (c) Comb

Fig. 3 Three image representations of the read alignments of the same genomic region using two-channel images. Red pixels represent *aligned* portions, green pixel *reference skips* and black pixel no information. This example region exhibits an *Exon Skipping/Cassette Exon* event
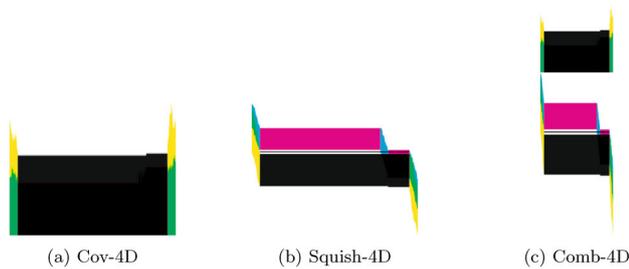

(a) Cov-4D   (b) Squish-4D   (c) Comb-4D

Fig. 4 Three image representations of the read alignments of the same genomic region using four-channel images. Each color channel (here depicted using the cyan, magenta, yellow, and black color model) encodes the following information, in order, (i) *aligned* portions for the first sample, (ii) *reference skip* portions for the first sample, (iii) *aligned* portions for the second sample, and (iv) *reference skips* portions for the second sample. This example region exhibits an *Exon Skipping/Cassette Exon* event

transcripts is present. For each pair of transcripts exhibiting a local AS event, we simulated RNA-Seq reads using Polyester [40] and we aligned them to the GRCh38 reference genome using STAR [41]. For each encoding defined in Sect. 2.1, we then generated four images for each pair t_i, t_j of transcripts and for each region of interest involving t_i and t_j by considering (1) the alignments of the reads simulated from t_i as sample 1 and the alignments of reads simulated from t_j as sample 2, (2) the opposite, *i.e.,*, reads of t_i as sample 2 and reads of t_j as sample 1, (3) reads of t_i as both sample 1 and sample 2, and (4) reads of t_j as both sample 1 and sample 2. Cases (1)–(2) correspond to the presence of the AS event between the two samples, and the training images generated from these cases are labeled with the corresponding local AS event they exhibit, using the following nomenclature: *A* (Alternative Acceptor/Donor sites), *CSE* (Exon Skipping/Cassette Exon), and *RI* (Intron Retention). On the other hand, cases (3)–(4) correspond to the *absence* of AS events between the two samples, and the images generated from these two cases are used to train the model to classify also the *non-event* (labeled *NN*).

We have added complexity to the training images, so that those images better resemble real data. More precisely, we generated some cross-fed samples, that is, we have randomly selected a portion of the reads from one sample and we have added those reads to the other sample. Such operation yields two interesting and useful results; (i) it adds noise to the images, making the network more robust to external noise and (ii) it simulates cases in which both the two different splicing events are concurrently present, albeit at different expression levels. In our experiments, we have a 20% cross-feed in both directions. An example of the images resulting from this process is shown in Fig. 5.

2.3 Deep learning model architecture

In order to perform region classification within our study, we employed a residual neural network architecture known as ResNet50 [42]. Since our datasets comprises two-color-channel images, and four-color-channel images, an adapted version of ResNet50 for each case was used. This adaptation of the ResNet50 architecture only considers the change in the input size to handle both input images, *i.e.,* intermediate layers remain unaltered.

The last layer of the neural network is a dense layer with *softmax* activation function, to assign a single-class label to each region under consideration. To train our model, we selected the *categorical crossentropy* loss function, with five different categories (four events and no-event). Each model was individually trained for 10 epochs with a batch size of 8 using *Stochastic Gradient Descend* with a learning rate of 0.001 and momentum of 0.9 [43]. The validation loss was monitored after each epoch to retain the overall best training weights.

For the classification of real samples, all trained models are ensembled. The final classification is performed using a *soft-voting* procedure, in which each network emits its confidence score for every class and the softmax of the sum of all models is then used for the final prediction.

In our initial experiments, we tried three kind of architectures: EfficientNet, InceptionNet, and ResNet, not finding particular differences in their outcomes and scalability. In this work, we choose ResNet for its simplicity, ease of use, and scalability.
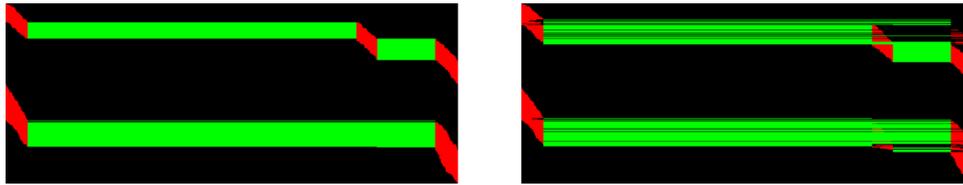
Fig. 5 Cross-feed example: (left) each sample expresses a single transcript, (right) same region with a cross-feed of 20% in both directions. Images generated by the *Squish* encoding method

2.4 Training the model

Each model has been trained on its respective image representations starting from the same set of annotated regions. Each dataset of images consisted of approx 57000 images which was then randomly split into training and test set with a ratio of 70:30.

Firstly, we performed a fivefold cross-validation by randomly splitting the training set into training and validation and evaluating each fold against the entire test set, with the goal of testing the absence of overfitting and bias selection and to ensure the ability of classify new data. Table 1 shows the average and standard deviation of precision ($\frac{tp}{tp+fp}$), recall ($\frac{tp}{tp+fn}$) and F1-score ($\frac{2tp}{2tp+fp+fn}$) across the fivefolds, where deepSpecas clearly express extremely good classification results, with all measures having a score greater than 99%.

We then retrained the models on the entire (sub-sampled) training set for further analysis; the results of the entire test set for each model are shown in Fig. 6 as confusion matrices, and all the models achieve an F1-score greater than 99%.

3 Results

Dataset. To assess the efficacy of deepSpecas in classifying AS events, we performed an experimental evaluation on a real RNA-Seq dataset. To this aim, we considered the dataset provided in [44] (SRA BioProject ID: PRJNA255099), which have been extensively used to evaluate tools for the analysis of AS events [20, 23]. In the study, control samples are compared with samples obtained after the double knockdown of the TRA2 splicing regulatory proteins (TRA2A and TRA2B). The dataset consists of six Illumina Hiseq2000 samples (three replicates for two conditions) of 101bp-long paired-end reads. In this investigation, a total of 81 exon skipping events have been RT-PCR validated in laboratory. Since we wanted to assess the ability of deepSpecas in detecting other AS events, we have extended the ground truth set of events besides the RT-PCR validated exon skipping events, by adding all AS events that are reported and considered significant (*i.e.*, events with reported *p*-value < 0.05) by both rMATS [17] and SUPPA2[23]—two state-of-the-art tools for differential analysis of AS events. Overall, the ground truth set of AS events has 62 alternative acceptor (A3), 54 alternative donor (A5), and 41 intron retention events, in addition to the 81 RT-PCR validated exon skipping events.

We have manually inspected the ground truth set and we have found several cases where the consensus of rMATS [17] and SUPPA2[23] is likely to be incorrect or uncertain—we detail in Figs. 7 and 8 two such cases and the rationale for our decision on whether to exclude each of them from the ground truth. We will first consider Fig. 7. It is evident that the second condition (in green) contains a high amount of reads expressing an exon skipping compared to the first condition (in red). Although there are different annotated transcripts and it is affected by spuriously aligned reads, there clearly is a different event expressed by the two conditions. On the other hand, the region in Fig. 8 does not contain a gene annotation that supports an exon skipping event, but more importantly the

Table 1 Accuracy scores for fivefold cross-validation show an elevated accuracy of the model on all the image representations, achieving $> 99\%$ on all measures

Method	Precision (fivefold)	Recall (fivefold)	F1-score (fivefold)
Cov	0.9964 ± 0.0015	0.9906 ± 0.0058	0.9934 ± 0.0036
Squish	0.9975 ± 0.0006	0.9945 ± 0.0022	0.9960 ± 0.0013
Comb	0.9955 ± 0.0016	0.9852 ± 0.0023	0.9902 ± 0.0011
Cov-4D	0.9990 ± 0.0004	0.9989 ± 0.0004	0.9989 ± 0.0004
Squish-4D	0.9987 ± 0.0009	0.9968 ± 0.0019	0.9978 ± 0.0013
Comb-4D	0.9994 ± 0.0006	0.9994 ± 0.0004	0.9994 ± 0.0005

(a) Cov   (b) Squish   (c) Comb

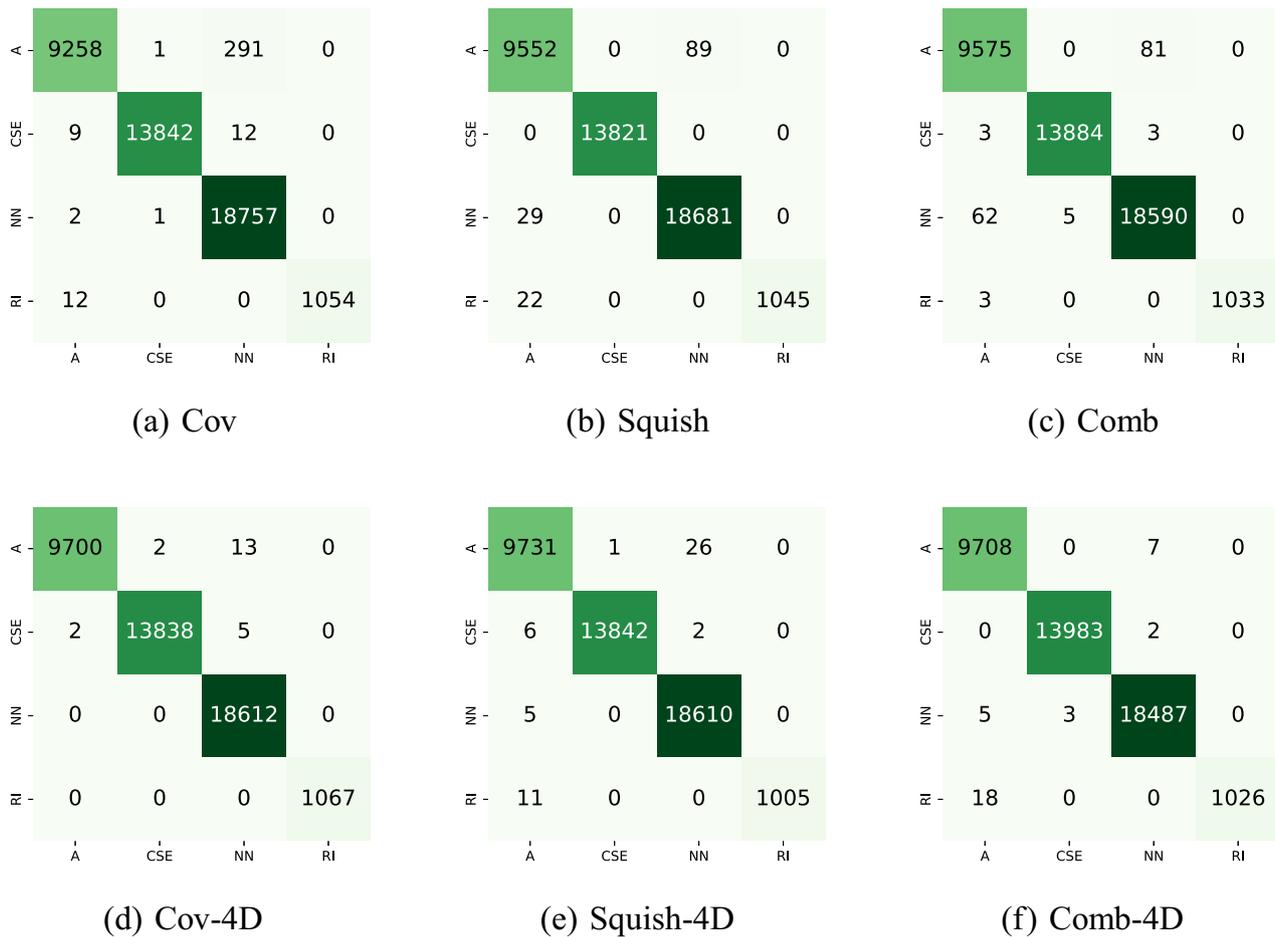(d) Cov-4D   (e) Squish-4D   (f) Comb-4D

Fig. 6 Confusion matrices for all the models obtained on the entire test set. All the models achieve an F1-score $> 99\%$

two conditions do not express two different events; indeed, they contain roughly the same amount of reads that cover and skip the exon in the middle.

In our curation, we do not consider the transcript annotations, as they might not be complete, and we greatly prioritize the different levels of expression present in the two conditions. With this assumption, we then filtered out events that did not respect such criterion, even if RT-PCR validated (in the case of CSE events) or if rMATS and SUPPA2 agreed on the call. After a manual inspection, we removed all regions in which (i) the event was supported by a very small number of reads, (ii) the region contained multiple AS events, and (iii) there was no significant difference between the two samples, hence no differential expression between samples. Overall, we removed 16/81 (20%) CSE events, 34/62 (54%) A3 events, 24/54 (44%) A5 events, and 24/41 (58%) RI events.

In the following, we refer to the ground truth set before our curation as the *non-curated* dataset, while the subset obtained after our manual inspection as the *curated* dataset. The complete collection of dataset and scripts is available

in the repository and the entire experimental analysis is reproducible via a provided Snakemake file. Furthermore, the curated dataset is available at https://github.com/scic colella/deepSpecas/tree/master/benchmark with the regions and the subset of reads belonging to them divided by conditions, for replication of our results and as possible benchmark dataset.

AS analysis We performed the AS classification on such datasets using deepSpecas with the *soft-voting* ensemble of all the models obtaining the results shown in Table 2 ("Non-curated" row), where deepSpecas correctly classify between 70% and 80% across the different events. When we perform the same classification on the curated dataset, we see a considerable increase in performance of our method—CSE +2.4%, A3 +6.25%, A5 +19.2%, RI +14.5%—with respect to the non-curated as shown in Table 2 ("Curated" row). It is interesting to notice that while we filtered all the events that we considered to be invalid, we retained regions that, even if not absolutely evident, still showed some level of differential expression across the conditions, thus showing the ability of
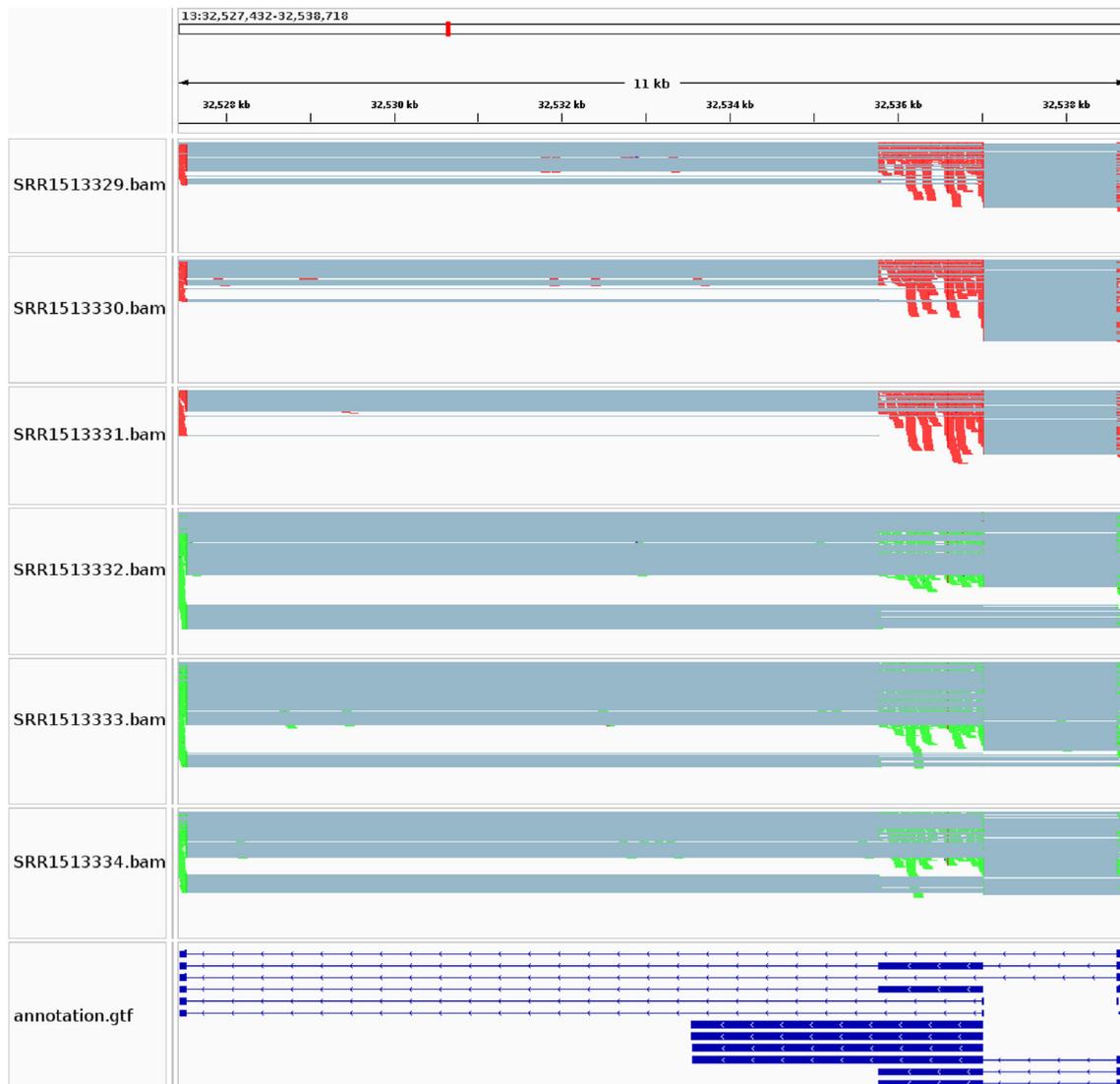
Fig. 7 Example of an exon skipping event reported by both SUPPA2 and rMATS as differentially expressed in the two samples (indicated in red and green). Even if spurious reads are present, it is evident that green condition expresses an exon skipping compared with the red one. This event was included in the manually curated subset of events

deepSpecas to classify noisy data. Moreover, it is important to consider that rMATS and SUPPA2 do call all the considered regions, but they also perform many debatable calls, implying high recall at the expense of lower precision. Furthermore, they do rely on the gene annotation and use it to perform the classification, whereas deepSpecas does not use such information and shows considerable performance while avoiding possible biases due to the use of incomplete annotation. Lastly, Fig. 9 shows the confusion matrices for the real datasets. With respect to both non-curated (a) and curated (b) dataset, we can see a high accuracy of results with no particular class that overshadow the others. In both datasets, no *non-event* (NN) is present as input, given their construction. Furthermore, it is also clear that a number of events that were correctly classified by deepSpecas in the non-curated dataset are discarded by the manual curation, thus demonstrating that no bias toward deepSpecas was employed when filtering the original dataset.

4 Conclusions

In this paper, we devised some novel image representations of differential alternative splicing events and developed a deep neural network model trained on these dataset to accurately classify AS events. We trained a different network for each image type, achieving high accuracy levels. Furthermore, we developed a soft-voting approach that
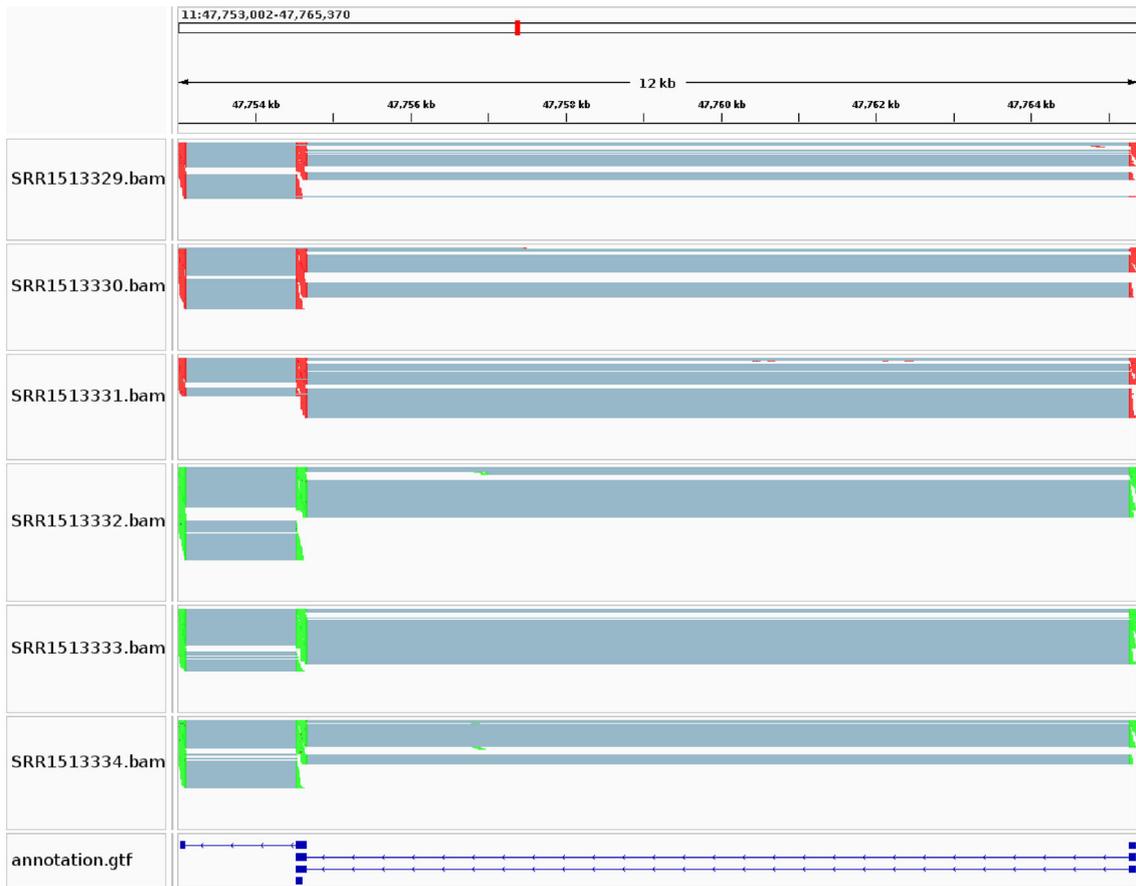
Fig. 8 Another example of an exon skipping event reported by both `SUPPA2` and `rMATS` as differentially expressed in the two samples (indicated in red and green). However, in this case, we believe that there is not enough evidence supporting the differential expression of the splicing event. Indeed, the annotation does not include a transcript with the exon skipping event and the levels of expression in the two conditions are very similar. This event was *not* included in the manually curated subset of events

Table 2 Results of `deepSpecas` classification on the real AS events datasets obtained by consensus of `rMATS` and `SUPPA2`, where we show the number of events that are correctly classified and the total number of events, for each type of AS event

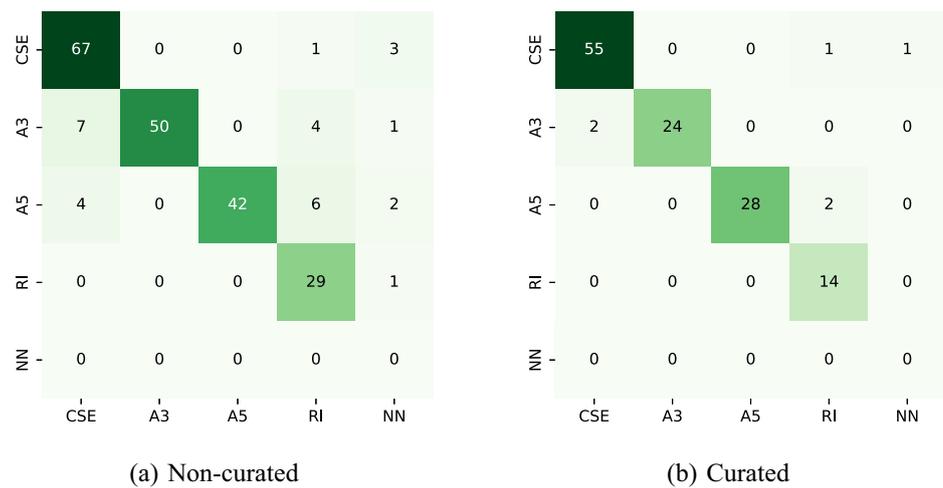|	CSE	A3	A5	RI
Non-curated	67/81 (83%)	50/62 (81%)	42/54 (78%)	29/41 (70%)
Curated	55/65 (85%)	24/28 (86%)	28/30 (93%)	14/17 (82%)

The manually curated dataset shows a significant improvement over the non-curated events. Furthermore, it is interesting to notice that even in very noisy settings, `deepSpecas` is still able to perform an acceptable classification

leverages the different information provided by each image model, to provide a final classification by ensemble of the previous networks. We showed that `deepSpecas` also performs very well on real datasets, even in the presence of highly noisy data. More importantly, `deepSpecas` does not use gene and transcript information to perform its classification step, thus avoiding any bias derived from such annotations. This lack of bias makes `deepSpecas` an interesting tool to use when the annotation cannot be completely trusted, such as in cancerous cells, or when we only have incomplete information. To the best of our knowledge, `deepSpecas` is the first deep learning tool that can reliably infer AS events without relying on a gene annotation on noisy data. On a negative note, this novelty means that we have not been able to assess the quality of `deepSpecas` by comparing it with different tools.

Crucially, we also created a manually curated dataset of AS events that can be used to evaluate the performance of future studies. Indeed, while there are several tools to predict AS events, the field does not have a benchmark to

Fig. 9 Confusion matrices of `deepSpecas` classification on the real AS events dataset obtained by consensus of `rMATS` and `SUPPA2`. No *non-event* (NN) region is present in the dataset since they are consensus of calls of the other tools (**a**) and furthermore a manually curated subset of them (**b**)

	CSE	A3	A5	RI	NN
CSE	67	0	0	1	3
A3	7	50	0	4	1
A5	4	0	42	6	2
RI	0	0	0	29	1
NN	0	0	0	0	0

(a) Non-curated

	CSE	A3	A5	RI	NN
CSE	55	0	0	1	1
A3	2	24	0	0	0
A5	0	0	28	2	0
RI	0	0	0	14	0
NN	0	0	0	0	0

(b) Curated

compare different tools. We have manually inspected a large set of predicted events and we built a set of supported events which, although relatively small, can be used for driving future improvements of `deepSpecas` and other tools for AS events detectiong and quantification. Alongside `deepSpecas`, we also released such a benchmark of regions with the set of aligned reads, already divided by event type and ready to be used in future studies.

The current version of `deepSpecas` is designed to work on specific regions that are provided by the user, since it is not fast enough to be used in a genome-scale investigation. Future work will be devoted to extend `deepSpecas` so that it is able to predict AS events on a much larger scale, such as on an entire transcriptome.

Declarations

Conflict of interest The authors declare no Conflict of interest.

References

1. Graveley Brenton R (2001) Alternative splicing: increasing diversity in the proteomic world. Trends Genet 17(2):100–107
2. Wang Y, Liu J, Huang L-F et al (2015) Mechanism of alternative splicing and its regulation. Biomed Rep 3(2):152–158
3. Yeo G, Holste D, Kreiman G, Burge CB (2004) Variation in alternative splicing across human tissues. Genome Biol 5:1–15
4. Biamonti G, Amato A, Belloni E, Di Matteo A, Infantino L, Pradella D, Ghigna C (2021) Alternative splicing in alzheimer's disease. Aging Clin Exp Res 33:747–758
5. Sveen A, Kilpinen S, Ruusulehto A, Lothe RA, Skotheim RI (2016) Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. Oncogene 35(19):2413–2427
6. Sophie CB, Irene LÃ³pez-Oreja, Juan V (2020) Roles and mechanisms of alternative splicing in cancer–implications for care. Nature Rev Clin Oncol 17(8):457–474
7. Bonizzoni P, Rizzi R, Pesole G (2006) Computational methods for alternative splicing prediction. Brief Funct Genom Proteomic 5(1):46–51
8. Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, Rawool S, Thakare RP, Banday S, Mishra AK, Das G (2023) Next-generation sequencing technology: current trends and advancements. Biology 12(7):997
9. Qun P, Ofer S, Leo JL, Brendan JF, Benjamin JB (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nature Genet 40(12):1413–1415
10. Liu J, Lin CX, Zhang X, Li Z, Huang W, Liu J, Guan Y, Li HD (2023) Computational approaches for detecting disease-associated alternative splicing events. Briefin Bioinform 24(3):bbad106
11. Mihaela P, Geo MP, Antonescu CM, Tsung-Cheng C, Mendell JT, Steven LS (2015) StringTie enables improved reconstruction of a transcriptome from rna-seq reads. Nature Biotechnol 33(3):290–5
12. Cole T, Brian W, Pertea G et al (2010) Transcript assembly and quantification by rna-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnol 28:511–515

13. Mitchell G, Manuel G, Levin J et al (2010) Ab initio reconstruction of transcriptomes of pluripotent and lineage committed cells reveals gene structures of thousands of lincrnas. Nature Biotechnol 28(5):503–510

14. Li W, Feng J, Jiang T (2011) Isolasso: a lasso regression approach to rna-seq based transcriptome assembly. J Comput Biol 8(11):1693–707

15. Bray N, Pimentel H, Melsted P et al (2016) Near-optimal probabilistic rna-seq quantification. Nature Biotechnol 34:525–527

16. Rob P, Geet D, Michael L, Rafael I, Carl K (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods 14:417–419

17. Shihao S, Won PJ, Zhi-xiang L, Lan L, Henry Michael D, Nian WY, Qing Z, Yi X (2014) rmats: robust and flexible detection of differential alternative splicing from replicate rna-seq data. Proc Natl Acad Sci 111(51):E5593–E5601

18. Michael R, James C, RyangGuk K, Chung WW, John W (2012) Spliceseq: a resource for analysis and visualization of rna-seq data on alternative splicing and its functional impacts. Bioinformatics 8(18):2385–2387

19. Kahles A, Ong CS, Zhong Y, Ratsch G (2016) SplAdder: identification, quantification and testing of alternative splicing events from RNA-Seq data. Bioinformatics 32(12):1840–1847

20. Luca D, Raffaella R, Stefano B, Della VG, Marco P, Paola B (2018) ASGAL: aligning rna-seq data to a splicing graph to detect novel alternative splicing events. BMC Bioinform 19(1):1–21

21. Cozzi D, Bonizzoni P, Denti L (2023) Esgq: alternative splicing events quantification across conditions based on event splicing graphs. bioRxiv, p 2023–07

22. Ciccolella S, Cozzi D, Della Vedova G, Kuria SN, Bonizzoni P, Denti L (2024) Differential quantification of alternative splicing events on spliced pangenome graphs. PLOS Comput Biol 20(12):e1012665

23. Trincado Juan L, Entizne Juan C, Gerald H, Babita S, Miha S, Elliott David J, Eduardo E (2018) Suppa2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. Genome Biol 19:1–11

24. Hartley S, Mullikin J (2016) Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq. Nucl Acids Res 44(15):e127–e127

25. Vaquero-Garcia J, Barrera A, Gazzara MR, Gonzalez-Vallinas J, Lahens NF, Hogenesch JB, Lynch KW, Barash Y (2016) A new view of transcriptome complexity and regulation through the lens of local splicing variations. eLife 5:e11752

26. Li Yang I, Knowles David A, Jack H, Barbeira Alvaro N, Dickinson Scott P, Kyung IH, Pritchard Jonathan K (2018) Annotation-free quantification of rna splicing using leafcutter. Nat Genet 50(1):151–158

27. Simon A, Alejandro R, Wolfgang H (2012) Detecting differential usage of exons from rna-seq data. Genome Res 22(10):2008–2017

28. Alharbi WS, Rashid M (2022) A review of deep learning applications in human genomics using next-generation sequencing data. Hum Genom 16(1):26

29. Shen F, Hu C, Huang X, He H, Yang D, Zhao J, Yang X (2023) Advances in alternative splicing identification: deep learning and pantranscriptome. Front Plant Sci 14:1232466

30. Li Jia T, Dingyuan LS, Zhifu G, Xiaowei S (2024) Identification of alternative splicing regulatory patterns and characteristic splicing factors in heart failure using rna-seq data and machine learning. Heliyon 10(15):e35408

31. Benegas G, Fischer J, Song YS (2022) Robust and annotation-free analysis of alternative splicing across diverse cell types in mice. Elife 11:e73520

32. Ryan P, Pi-Chuan C, Alexander D et al (2018) A universal SNP and small-indel variant caller using deep neural networks. Nat Biotechnol 36(10):983–987

33. Jaganathan K, Panagiotopoulou SK, McRae JF et al (2019) Predicting splicing from primary sequence with deep learning author links open overlay panel. Cell 176(3):535–548

34. Jasper Z, Godin F, Kim M, Soete A, Saeys Y, De Neve W (2018) Splicerover: interpretable convolutional neural networks for improved splice site prediction. Bioinformatics 34(24):4180–4188

35. Akpokiro V, Oluwadare O, Kalita J (2021) Deepsplicer: an improved method of splice sites prediction using deep learning. In: 2021 20th IEEE international conference on machine learning and applications (ICMLA), pp 606–609

36. Yaron S, Jenny L, Mahesan N, Diana B (2022) CI-SpliceAI-Improving machine learning predictions of disease causing splicing variants using curated alternative splice sites. PLoS ONE 17(6):e0269159

37. Dondi A, Lischetti U, Jacob F, Singer F, Borgsmuller N, Coelho R, Heinzelmann-Schwarz V, Beisel C, Beerenwinkel N (2023) Detection of isoforms and genomic alterations by high-throughput full-length single-cell RNA sequencing in ovarian cancer. Nature Commun 14(1):7780

38. Helga T, James R, Mesirov J (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14(2):178–192

39. Wei J, Liang C (2021) Alternative splicing: Human disease and quantitative analysis from high-throughput sequencing. Comput Struct Biotechnol J 19:183–195

40. Frazee AC, Jaffe AE, Langmead B, Leek JT (2015) Polyester: simulating RNA-seq datasets with differential transcript expression. Bioinformatics 31(17):2778–2784

41. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29(1):15–21

42. He K, Zhang X, Ren S, Sun J (2015) Deep residual learning for image recognition. CoRR, arXiv:1512.03385

43. Sutskever I, Martens J, Dahl G, Hinton G (2013) On the importance of initialization and momentum in deep learning. In: International conference on machine learning, pp 1139–1147. PMLR

44. Best A, James K, Dalgliesh C, Hong E, Kheirolahi-Kouhestani M, Curk T, Yaobo X, Danilenko M, Hussain R, Keavney B et al (2014) Human tra2 proteins jointly control a chek1 splicing switch among alternative and constitutive target exons. Nat Commun 5(1):4760

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.