



# Intron Retention as a Mode for RNA-Seq Data Analysis

Jian-Tao Zheng<sup>1</sup>, Cui-Xiang Lin<sup>1\*</sup>, Zhao-Yu Fang<sup>2</sup> and Hong-Dong Li<sup>1\*</sup>

<sup>1</sup> Hunan Provincial Key Lab on Bioinformatics, Center for Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha, China, <sup>2</sup> School of Mathematics and Statistics, Central South University, Changsha, China

Intron retention (IR) is an alternative splicing mode whereby introns, rather than being spliced out as usual, are retained in mature mRNAs. It was previously considered a consequence of mis-splicing and received very limited attention. Only recently has IR become of interest for transcriptomic data analysis owing to its recognized roles in gene expression regulation and associations with complex diseases. In this article, we first review the function of IR in regulating gene expression in a number of biological processes, such as neuron differentiation and activation of CD4<sup>+</sup> T cells. Next, we briefly review its association with diseases, such as Alzheimer's disease and cancers. Then, we describe state-of-the-art methods for IR detection, including RNA-seq analysis tools IRFinder and iREAD, highlighting their underlying principles and discussing their advantages and limitations. Finally, we discuss the challenges for IR detection and potential ways in which IR detection methods could be improved.

**Keywords:** alternative splicing, intron retention, gene regulation, disease association, RNA-seq

## OPEN ACCESS

### Edited by:

Raya Khanin,  
Cornell University, United States

### Reviewed by:

Younghee Lee,  
University of Utah, United States  
Yuriy L. Orlov,  
First Moscow State Medical  
University, Russia

### \*Correspondence:

Cui-Xiang Lin  
lincxsu@csu.edu.cn  
Hong-Dong Li  
hongdong@csu.edu.cn

### Specialty section:

This article was submitted to  
Computational Genomics,  
a section of the journal  
Frontiers in Genetics

**Received:** 13 December 2019

**Accepted:** 14 May 2020

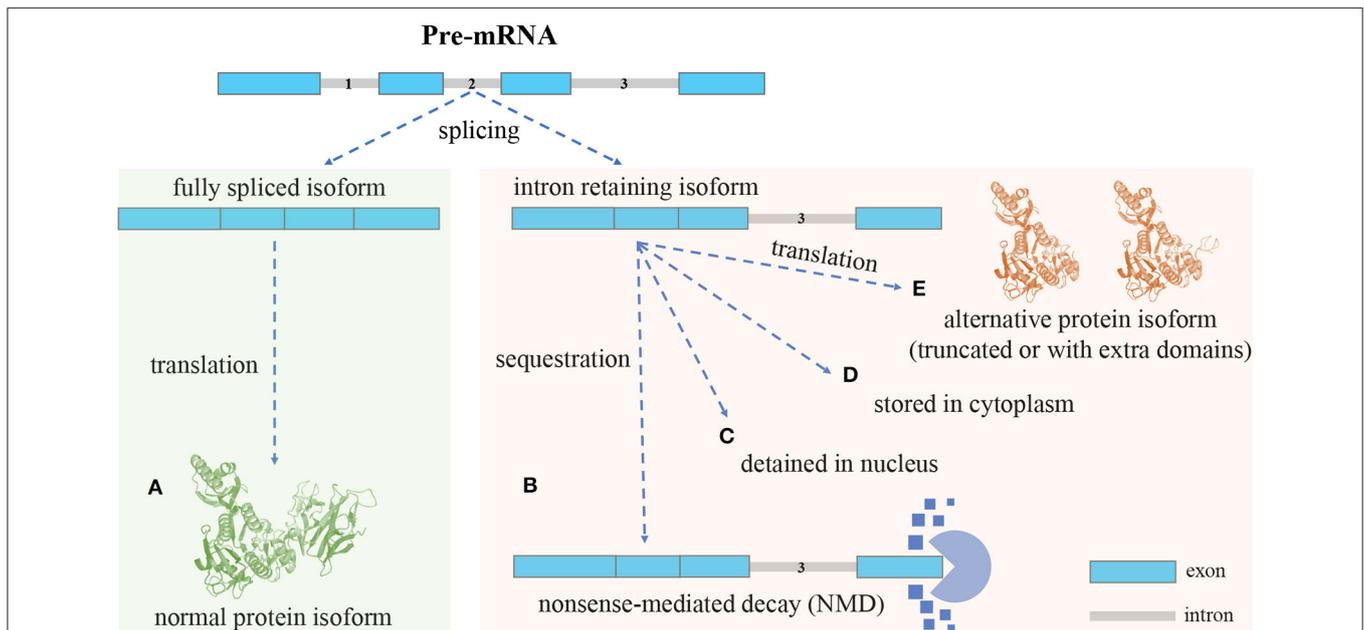
**Published:** 07 July 2020

### Citation:

Zheng J-T, Lin C-X, Fang Z-Y and  
Li H-D (2020) Intron Retention as a  
Mode for RNA-Seq Data Analysis.  
*Front. Genet.* 11:586.  
doi: 10.3389/fgene.2020.00586

## 1. INTRODUCTION

Different mRNA splicing isoforms can be produced from pre-mRNA by skipping or joining coding/non-coding gene fragments, referred to as alternative splicing (AS) (Ner-Gaon et al., 2004). AS includes five major forms: exon skipping, intron retention (IR), mutually exclusive exons, alternative 5' splice sites, and alternative 3' splice sites. IR is the least understood form in mammalian cells (Sznajder et al., 2018; Monteuuis et al., 2019; Broseus and Ritchie, 2020). The process of IR is illustrated in **Figure 1**. In most cases, mature mRNA isoforms with introns fully spliced are exported out of the nucleus for translation (Cuenca-Bono et al., 2011; Palazzo et al., 2013). Because introns often contain premature termination codons (PTCs), intron-retaining isoforms (IRIs) are often rapidly degraded by the nonsense-mediated decay (NMD) pathway that is triggered by PTCs (Ge and Porse, 2014). IRIs may be retained in the nucleus or cytoplasm and be subject to further splicing in response to stimuli or stress (Naro et al., 2017). IRIs may also escape from the NMD pathway (Lykke-Andersen and Jensen, 2015) and be translated into protein isoforms that are often truncated (Lindeboom et al., 2016; Ottens and Gehring, 2016) and harmful to cells (Brady et al., 2017; Kanagasabai et al., 2017; Uzor et al., 2018; Mukherjee et al., 2019; Wang et al., 2019). Regarding the proportion of IRIs escaping from the NMD pathway, it has been shown that ~10% of human alternatively spliced nonsense-mediated decay (AS-NMD) transcripts are translated into truncated proteins (de Lima Morais and Harrison, 2010). Studies have shown that the truncated protein isoform may be shorter (i.e., have fewer domains) than or include extra domains over the normal protein isoform (Gontijo et al., 2011; Rekossh and Hammarskjold, 2018). As for the frequency of truncation, to the best of our knowledge, no estimates of the percentage of truncated proteins translated from IRIs seem to be available in the existing literature.



**FIGURE 1** | An overview of the intron retention (IR) mechanism: different isoforms can be produced from a single gene through AS. **(A)**, Isoforms with introns fully spliced are sent out of the nucleus for translation. Intron-retaining isoforms (IRIs) can be generated through IR (no intron retention): **(B)**, In most cases, the IRIs are degraded by the nonsense-mediated decay (NMD) pathway, the reason being that retained introns often contain premature termination codons (PTCs) that can trigger NMD (with intron retention): **(C)**, In some cases, the IRIs are detained in the nucleus, and in response to stimuli these IRIs can undergo further splicing to remove the retained intron, before being exported out of nucleus for translation (with intron retention): **(D)**, In the case of cytoplasmic splicing, IRIs are shuttled to the cytoplasm for preservation and may be subject to further splicing (with intron retention): **(E)**, In yet another case, IRIs escape from the NMD pathway and are translated into protein isoforms, which, compared with normal protein isoforms, are often truncated and may lose domains; however, it could also be that the alternative protein isoforms include extra domains formed by the amino acid sequences translated from retained introns (with intron retention).

AS is a regulated process during gene expression (Koch, 2017). Since introns do not encode proteins, historically they were considered junk DNA as well as a burden on transcription and splicing (Wong et al., 2000; Roy and Irimia, 2008; Morris and Mattick, 2014; Parenteau and Elela, 2019). IR refers to an ineffective or inefficient splicing of introns that may have a negative impact on cells (Lim et al., 2011; Singh and Cooper, 2012; Wong et al., 2016). For example, IR of the *globin* gene will trigger NMD, which in turn affects red blood cell differentiation (Reimer and Neugebauer, 2018); IR generated an *Id3* isoform that limits the growth of smooth muscle cells during the formation of vascular disease (Forrest et al., 2004). IR is also associated with the development and maintenance of complex diseases. For example, many introns that are preferentially retained in primary cancers can be detected in the cytoplasm of cancer cells, and the abundant IRIs in cancer cells can increase the diversity of cancer cell transcriptomes (Dvinge and Bradley, 2015). In recent years, the transcriptome analysis of IR has received increasing attention.

Currently, the detection of IR is based on computational analysis of high-throughput RNA-seq data. In recent years, tools dedicated to IR detection have been developed, such as IRCall and IR classifier (Bai et al., 2015), Keep Me Around (KMA) (Pimentel et al., 2015a), intron Retention Analysis and Detector (iREAD) (Li et al., 2020), and IRFinder (Middleton et al., 2017). In addition, some tools originally designed to detect AS events can also be used to detect IR, such as mixture-of-isoforms (MISO) (Katz et al., 2010), multivariate analysis of transcript splicing

(MATS) (Shen et al., 2012), replicate MATS (rMATS) (Shen et al., 2014), comprehensive alternative splicing hunting (CASH) (Wu et al., 2017), and DEXSeq (Anders et al., 2012). In recent years, deep learning-based AS detection methods have been developed, such as deep learning augmented RNA-seq analysis of transcript splicing (DARTS) (Zhang et al., 2019) and SpliceAI (Jaganathan et al., 2019).

In the following sections, we will review the association of IR with gene expression regulation and complex diseases. Last but not least, we will describe current computational approaches to IR detection and discuss their advantages and limitations.

## 2. INTRON RETENTION IN GENE EXPRESSION REGULATION

IR plays an important role in regulating gene expression through triggering NMD (Wong et al., 2013; Ge and Porse, 2014). IRIs often contain PTCs (Braunschweig et al., 2014). The signal of a PTC can be recognized by the protein factors in the NMD pathway, and IRIs can thus be degraded by NMD. Consequently, IR leads to down-regulation of the isoform and of the protein products if translated (Ge and Porse, 2014). In this section we review some studies exploring the relationship between IR and the regulation of gene expression in different cell types, as well as studies investigating the relationship between IR and cell differentiation.

Some studies have found that IR is related to gene expression regulation in different types of cells. For example, Kienzle et al. (1999) suggested that retained introns can introduce a stop codon in an open reading frame or frameshift, which can contribute to gene expression regulation via premature termination of translation without changing the transcriptional activity. Taking the *EBNA-3* gene as an example, the presence of introns would effectively disrupt the translation process and thereby affect the expression of the *EBNA-3* protein, suggesting that IR may provide a means of fine-tuning the expression of the *EBNA-3* family gene in human B lymphocytes. Ni et al. (2016) found that the up-regulation of most genes in activated T cells was accompanied by a significant decrease in the level of IR. In their human and mouse CD4<sup>+</sup> T cell validation experiments, 185 of 1,583 genes were mainly regulated by IR and were highly enriched in the proteasome pathway, revealing a novel post-transcriptional regulatory mechanism. This mechanism can help cells coordinate and respond quickly to extracellular stimuli, such as acute infections. Forrest et al. (2004) found that during the formation of vascular lesions in rats, an IRI called helix-loop-helix transcription factor Id3 (*Id3a*) was abnormally expressed in the early stage of lesion formation. Using the *Id3a*-specific antibody they developed, they found that the *Id3a* protein was induced to be translated in vascular lesions. This protein does not promote the growth of smooth muscle cells but stimulates their apoptosis and inhibits the production of endogenous *Id3a* isoforms.

Other studies have found IR to be associated with cell differentiation. By analyzing high-coverage poly(A)<sup>+</sup> RNA-seq data, Braunschweig et al. (2014) found that the increase of IR during neuronal differentiation plays a major role in down-regulating gene expression. First, genes containing introns have higher retention rates in differentiated neurons than in murine embryonic stem cells and are significantly enriched in multiple Gene Ontology (GO) terms associated with the cell cycle. Second, the increase of IR reduces the mRNA expression of the *Ssrp1* gene during neuronal differentiation. Pimentel et al. (2015b) observed a dynamic increase of IR in late erythroblasts, indicating that IR explicitly regulates the differentiation process of erythroblasts. They also discovered many unique and extensive IR events during the differentiation of red blood cells. They inferred that IR is a multidimensional process that post-transcriptionally regulates multiple gene groups during normal erythropoiesis, and that its misregulation may be the cause of human disease. In the late phases of mammalian germ cell differentiation, the required transcripts must be synthesized and stored in advance (Paronetto and Sette, 2010). From observing the accumulation of the *ADAM3* protein, Naro et al. (2017) found that IRs detained in the nucleus can regulate the use of transcripts.

In summary, there are a number of studies that show various ways in which IR can regulate gene/protein isoform production (Nilsen and Graveley, 2010; Floor and Doudna, 2016; Jacob and Smith, 2017), RNA stability, and translation efficiency (Thiele et al., 2006; Sterne-Weiler et al., 2013). These studies suggest that IR as a post-transcriptional splicing pattern plays an essential role in fine-tuning gene expression. (Mauger et al., 2016).

### 3. INTRON RETENTION IS ASSOCIATED WITH COMPLEX DISEASES

IR has been shown to be associated with complex diseases. For example, IR represents a mechanism (Mauger et al., 2016) and provides a sensitive and disease-specific diagnostic biomarker for neurodegenerative diseases (Jeromin and Bowser, 2017; Sznajder et al., 2018). In addition, IR was found to be widespread across a series of cancer transcriptomes (Dvinge and Bradley, 2015) and was thought to be related to tumor suppressor inactivation (Jung et al., 2015). We performed a literature survey and identified 60 papers (published after 2016) on the association of IR with diseases: 28 are on neurodegenerative diseases, 23 are on cancers, and the remaining are about other diseases, such as Duchenne muscular dystrophy (DMD), chronic lymphocytic leukemia (CLL), and myelodysplastic syndromes (MDS). Below, we will briefly review these studies on the association of IR with neurodegenerative diseases, cancers, and the other less-studied diseases.

With regard to neurodegenerative diseases, Xu et al. (2008) studied Alzheimer's disease (AD). A number of studies have shown that the apolipoprotein E4 (*apoE4*) isoform is associated with AD. In the primary neuron transfection experiment of Xu et al., they found that neuronal expression of the *apoE4* isoform was significantly higher when intron-3 was deleted from the genomic DNA structure and, conversely, significantly lower when intron-3 was inserted into the cDNA. This finding suggests that the retention/splicing of intron-3 controls the expression of the *apoE4* isoform in neurons, implying an association between IR and AD. Over-expression of the *peripherin* gene may lead to degeneration of motor neurons in transgenic mice. Xiao et al. (2008) identified the normal splicing variants of peripheral proteins and a novel transcript of peripheral proteins retaining introns 3 and 4. The IRI of the *peripherin* gene was found to be expressed at a low stoichiometric level. When the expression of IRI is up-regulated, it will lead to the aggregation of *peripherin*. This observation suggests that the abnormal splicing of peripheral protein in amyotrophic lateral sclerosis produces a splice isoform that is prone to aggregation.

Several studies have identified an association between IR and cancers. Zhang et al. (2014) used whole transcriptome sequencing data from five lung adenocarcinoma tissues and matched normal tissues to detect IR. A large number of IR events were found in both the tumor and the normal tissues, and 2,340 and 1,422 genes contained only tumor-specific and normal tissue-specific retention events, respectively. Subsequent functional analysis indicated that genes with tumor-specific retention include known lung cancer driver genes, such as *EGFR*, *ROS1*, and *RUNX1*, and are enriched in pathways that are important in carcinogenesis. IR in these genes causes frameshift, which generally invokes NMD and reduces the expression levels of mRNAs. These over-expressed or highly mutable driver genes may have a protective effect on patients. The work of Jung et al. (2015) demonstrated that IR is a mechanism leading to the inactivation of tumor suppressor genes. By analyzing the RNA sequencing and exome data from 1,812 cancer patients, they determined that at least

163 of the 900 splice-disrupted somatic exon single-nucleotide variants caused IR in an allele-specific manner and were enriched in tumor suppressor genes.

In particular, Dvinge and Bradley (2015) performed extensive experiments to analyze the association between IR and 16 cancers. By analyzing the genome-wide RNA splicing patterns of 805 matched tumor and control samples from 16 cancers, they found that abnormal RNA splicing occurs in the form of IR in cancers. The most common spliceosomal mutations, such as the specific missense changes of the *SF3B1*, *SRSF2*, and *U2AF1* proteins, are abundant in a variety of diseases, including MDS, lymphoid leukemia, and solid tumors of the lung, breast, pancreas, and eyes. They also used the transcriptomic data generated by the Cancer Genome Atlas project to identify large-scale differences in RNA splicing between the cancerous and control samples. In all the cancers except breast cancer, IRs are up-regulated in the cancer samples, with the increase ranging from 2-fold (acute myeloid leukemia) to 40-fold (colon cancer) compared with control samples. Many introns that are preferentially retained in primary cancers are detectable in the cytoplasmic fractions of cancer cell lines. This finding suggests that IR is a common factor associated with tumorigenesis. Abundant IRs in cancer cells may increase the diversity of cancer transcriptomes. Finally, through genome-wide quantitative analysis and unsupervised clustering analysis, Dvinge et al. confirmed that although some retained introns are shared by most cancer types, most are either present at a low frequency in multiple cancers or unique to primary cancers. For example, two adjacent introns in *FUS* were recurrently retained in multiple types of cancers, such as breast and colon cancers. Most introns (1,205 out of 1,767) were retained in a few samples of a particular cancer. The retained intron in *CDK10*, for example, was specific to and frequently retained in the colon cancer samples. Clustering results showed that cancers originating from similar tissues, such as the colon and rectum, have similar patterns of IR.

There are also studies on the association of IR with other less-studied diseases, such as DMD, CLL, and MDS. For example, the high-level retention of introns 40, 58, and 70 in DMD transcripts may be responsible for the lack of dystrophin expression in CRL-2061 cells, resulting in the elimination of the tumor suppressor activity of dystrophin (Niba et al., 2017). It was found that the *SF3B1* modulator sudemycin D6 (SD6) can effectively inhibit the growth of CLL cells and that IR in SD6-treated CLL cells increased significantly (Han et al., 2019). In the gene subset represented by *SF3B1*, the non-productive interaction between intron-terminal splice sites and decoy exons could prevent the excision of introns and was found to regulate a pivotal subset of IR events during erythroblast differentiation (Parra et al., 2018).

In recent years, splicing regulation therapy strategies have been developed and are currently being tested in clinical trials for a range of diseases (Scotti and Swanson, 2016; Di et al., 2019), including muscular dystrophy and motor neuron diseases. It is therefore increasingly important to understand the relationship between IR and diseases.

## 4. METHODS FOR INTRON RETENTION DETECTION

An important component of high-throughput sequencing, transcriptome sequencing technology (RNA-seq) is a useful tool in transcriptomics analysis (Conesa et al., 2016; Hrdlickova et al., 2017). RNA-seq data can be used to analyze transcriptome information, such as gene expression and splice sites (Vanichkina et al., 2018).

Currently, tools dedicated to IR detection are available (Bai et al., 2015; Pimentel et al., 2015a; Middleton et al., 2017; Li et al., 2020). Bai et al. (2015) developed IRcall (a ranking strategy) and IRclassifier (random forest classifiers) to detect IR events. IRcall integrates seven features—including gene expression, intron read counts, flanking exon read counts, and splice junctions—to calculate a joint score for IR events. The joint score can help to reduce false positive identification to a certain extent. IRclassifier constructs a training set based on the prediction of other IR detection methods and uses 21 features to characterize each intron; it then builds a random forest classifier to predict IR events. A limitation of these two methods is that the features used for model construction depend on the quality of the junction alignment tool (Bai et al., 2015). KMA (Pimentel et al., 2015a) is an IR detection pipeline that leverages existing isoform expression quantification tools. It can combine biological replicates to reduce the number of false positives. The isoform quantification and analysis of IR are performed in different software environments, which may be inconvenient (Li et al., 2020). IRFinder (Middleton et al., 2017) provides a complete pipeline for identifying IR events, including genome preparation, data preparation and quality control, IR quantification, and differential analysis. IRFinder quantifies IR in terms of splicing level and intronic abundance, where the *IR ratio* metric indicates the proportion of transcripts containing the intron of interest. IRFinder identifies IR candidates based on *IR ratio* and the number of intronic reads. iREAD (Li et al., 2020) uses the Shannon entropy to quantify the uniformity of the distribution of reads across the intron. To avoid ambiguity, only the independent intron that does not overlap with any exon of any gene is considered in iREAD. One limitation of iREAD is that it does not provide differential analysis.

In addition to the techniques above, methods for detecting AS can also be used to detect IR. For example, MISO (Katz et al., 2010) is a method for inferring isoform regulation from RNA-seq data. It models the generation process of reads in isoforms, considers all isoform expression levels in genes as random variables, and uses Markov chain Monte Carlo sampling to estimate the distribution of the variables. Therefore, MISO can estimate expression at both the AS event level and the whole mRNA isoform level. For differential analysis, MISO can perform comparison on only two samples (e.g., samples without replicates) (Wang et al., 2018). DEXSeq (Anders et al., 2012) is a method that was originally developed to detect exon usage. It can be used to detect IR if introns rather than exons are used as the genomic feature for calculating usage. DEXSeq integrates

different methods to detect the deviation of reads of each exon, and the result is robust. Differential expression is performed for exon usage. MATS (Shen et al., 2012) uses a Bayesian statistical framework to detect differential AS from RNA-seq samples without replicates. rMATS (Shen et al., 2014) is an extended version of MATS and is capable of processing replicate samples. rMATS uses read counts that uniquely map to isoforms to estimate the exon inclusion level, and it takes into account both the uncertainty in individual samples and the variability between replicates. It is worth mentioning that rMATS adopts a flexible likelihood-ratio test, allowing users to define the threshold of inclusion level differences between groups. A limitation of rMATS is that it relies on known annotations of transcripts and has insufficient detection ability for novel AS events (Denti et al., 2018). In consideration of the currently incomplete transcript annotation, CASH (Wu et al., 2017) combines the annotated exon sites in the reference transcriptome and the novel splice site detected in RNA-seq data to reconstruct all splice sites for each gene. In this way, CASH has the potential to detect novel AS events (Carazo et al., 2019). Conventional quantification of exons (exon inclusion level) and isoforms (isoform ratio) depends on transcript models or predefined splicing events, which may be incomplete, partly because of the limitation of disease-specific abnormal transcripts (Li et al., 2018), for example. LeafCutter is an annotation-free method for quantifying both known and novel AS events based on exon-exon junction reads. It focuses on the intron removal rate rather than the exon inclusion rate (Li et al., 2018). The benefit of focusing on intron excision is that transcript annotation is not necessary and there is no need to estimate isoform or exon usage in complex splicing events. In brief, LeafCutter first defines introns that overlap and share the acceptor or donor splice site as intron clusters. The intron removal rate difference, quantified as  $\Delta$ PSI of the intron cluster between samples, is used to find differentially excised introns. Although IR is not explicitly modeled by LeafCutter (Vaquero-Garcia et al., 2018),  $\Delta$ PSI could reflect the possibility of IR.

Deep learning is a machine learning approach that has the ability to extract useful information or patterns from large numbers of samples. In recent years, deep learning has been introduced for AS analysis. For example, DARTS (Zhang et al., 2019) uses a large amount of sequencing data from public databases, such as ENCODE (Chi, 2016) and Roadmap (Romanoski et al., 2015) as input to the model. Then, the Bayesian hypothesis statistical test (BHT) is applied to obtain the training labels for each AS event. A deep neural network (DNN) is used to train all AS events. The prediction result of the DNN is fed into the BHT again to get the final prediction label for AS events including IR. SpliceAI (Jaganathan et al., 2019) uses a deep residual network to predict the splice sites of any pre-mRNA transcript sequence. The resulting splice sites can be used to infer whether IR has occurred. SpliceAI also explores the effects of gene mutations and exon and intron lengths on the splicing strength of splice sites.

## 5. CONCLUSION

The rapid development of high-throughput sequencing technology has enabled genome-wide detection of IR. Although significant progress has been made in this area, there are still challenges at present. First, current methods detect retained introns at the gene level instead of the isoform level. Identifying isoforms in which introns are retained is a question that remains to be resolved. Third-generation sequencing technologies, such as the PacBio single-molecule real time (SMRT) technology (Edge and Bansal, 2019), which can sequence the entire transcript, may help to address this challenge (Wu et al., 2017). Second, introns that are enriched in low-complexity and repetitive sequences may restrict the unique mapping of sequencing data (Broeseus and Ritchie, 2020), and such introns if retained may be more difficult to detect. Third, there are currently no benchmark data available on retained introns, making it difficult to evaluate IR detection methods.

Current IR detection methods could be improved through integrating prior knowledge, selecting suitable thresholds for parameters, and so on. For prior knowledge, features, such as intron length, the distribution of the splicing regulatory elements, canonical or non-canonical status of splice sites, and splicing strength could be used as prior knowledge to improve IR detection (Mao et al., 2014; Cui et al., 2017; Kim et al., 2018; Zhang et al., 2018). For parameter thresholds, designing methods to incorporate sequence features and read coverage variations of introns to adaptively determine individual intron-specific optimal thresholds of parameters could be helpful for IR detection (Broeseus and Ritchie, 2020). It is worth noting that any rigid thresholding may cause downstream analysis, such as GO enrichment to be heavily skewed toward genes with high expression (Young et al., 2010; Timmons et al., 2015). As an important mode of alternative splicing, IR is expected to advance our understanding of gene expression regulation and diseases from a new perspective (Wong et al., 2016; Jacob and Smith, 2017; Vanichkina et al., 2018; Monteuis et al., 2019).

## AUTHOR CONTRIBUTIONS

H-DL and C-XL conceived and wrote the manuscript. J-TZ wrote the manuscript. Z-YF wrote part of and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the National Natural Science Foundation of China (Nos. 61702556, 61772557, and 61702555), the 111 Project (No. B18059), and the Hunan Provincial Science and Technology Program (2018WK4001).

## REFERENCES

- Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. *Genome Res.* 22, 2008–2017. doi: 10.1101/gr.133744.111
- Bai, Y., Ji, S., and Wang, Y. (2015). IRcall and IRclassifier: two methods for flexible detection of intron retention events from RNA-Seq data. *BMC Genomics* 16:S9. doi: 10.1186/1471-2164-16-S2-S9
- Brady, L. K., Wang, H., Radens, C. M., Bi, Y., Radovich, M., Maity, A., et al. (2017). Transcriptome analysis of hypoxic cancer cells uncovers intron retention in EIF2B5 as a mechanism to inhibit translation. *PLoS Biol.* 15, 1–29. doi: 10.1371/journal.pbio.2002623
- Braunschweig, U., Barbosa-Morais, N. L., Pan, Q., Nachman, E. N., Alipanahi, B., Gonatopoulos-Pournatzis, T., et al. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* 24, 1774–1786. doi: 10.1101/gr.177790.114
- Broseus, L., and Ritchie, W. (2020). Challenges in detecting and quantifying intron retention from next generation sequencing data. *Comput. Struct. Biotechnol. J.* 18, 501–508. doi: 10.1016/j.csbj.2020.02.010
- Carazo, F., Romero, J. P., and Rubio, A. (2019). Upstream analysis of alternative splicing: a review of computational approaches to predict context-dependent splicing factors. *Brief. Bioinformatics* 20, 1358–1375. doi: 10.1093/bib/bby005
- Chi, K. R. (2016). The dark side of the human genome. *Nature* 538, 275–277. doi: 10.1038/538275a
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17:13. doi: 10.1186/s13059-016-0881-8
- Cuenca-Bono, B., García-Moliner, V., Pascual-García, P., Dopazo, H., Llopis, A., Vilardell, J., et al. (2011). SUS1 introns are required for efficient mRNA nuclear export in yeast. *Nucleic Acids Res.* 39, 8599–8611. doi: 10.1093/nar/gkr496
- Cui, Y., Zhang, C., and Cai, M. (2017). Prediction and feature analysis of intron retention events in plant genome. *Comput. Biol. Chem.* 68, 219–223. doi: 10.1016/j.compbiolchem.2017.04.004
- de Lima Morais, D. A., and Harrison, P. M. (2010). Large-scale evidence for conservation of NMD candidature across mammals. *PLoS ONE* 5:e11695. doi: 10.1371/journal.pone.0011695
- Denti, L., Rizzi, R., Beretta, S., Della Vedova, G., Previtali, M., and Bonizzoni, P. (2018). ASGAL: aligning RNA-Seq data to a splicing graph to detect novel alternative splicing events. *BMC Bioinformatics* 19:444. doi: 10.1186/s12859-018-2436-3
- Di, C., Zhang, Q., Chen, Y., Wang, Y., Zhang, X., Liu, Y., et al. (2019). Function, clinical application, and strategies of Pre-mRNA splicing in cancer. *Cell Death Differ.* 26, 1181–1194. doi: 10.1038/s41418-018-0231-3
- Dvinge, H., and Bradley, R. K. (2015). Widespread intron retention diversifies most cancer transcriptomes. *Genome Med.* 7:45. doi: 10.1186/s13073-015-0168-9
- Edge, P., and Bansal, V. (2019). Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. *Nat. Commun.* 10, 1–10. doi: 10.1038/s41467-019-12493-y
- Floor, S. N., and Doudna, J. A. (2016). Tunable protein synthesis by transcript isoforms in human cells. *eLife* 5:e10921. doi: 10.7554/eLife.10921
- Forrest, S. T., Barringhaus, K. G., Perlegas, D., Hammarskjöld, M.-L., and McNamara, C. A. (2004). Intron retention generates a novel Id3 isoform that inhibits vascular lesion formation. *J. Biol. Chem.* 279, 32897–32903. doi: 10.1074/jbc.M404882200
- Ge, Y., and Porse, B. T. (2014). The functional consequences of intron retention: alternative splicing coupled to NMD as a regulator of gene expression. *Bioessays* 36, 236–243. doi: 10.1002/bies.201300156
- Gontijo, A. M., Miguela, V., Whiting, M. F., Woodruff, R., and Dominguez, M. (2011). Intron retention in the *Drosophila melanogaster* Rieske Iron Sulphur Protein gene generated a new protein. *Nat. Commun.* 2:323. doi: 10.1038/ncomms1328
- Han, Q., Wang, J., Shull, A. Y., Shi, F., Deng, L., Choi, J.-H., et al. (2019). Modulation of SF3B1 causes global intron retention and downregulation of the B-cell receptor pathway in chronic lymphocytic leukemia. *Cancer Res.* 79, 5230–5230. doi: 10.1158/1538-7445.AM2019-5230
- Hrdlickova, R., Toloue, M., and Tian, B. (2017). RNA-Seq methods for transcriptome analysis. *WIREs RNA* 8:e1364. doi: 10.1002/wrna.1364
- Jacob, A. G., and Smith, C. W. (2017). Intron retention as a component of regulated gene expression programs. *Hum. Genet.* 136, 1043–1057. doi: 10.1007/s00439-017-1791-x
- Jaganathan, K., Panagiotopoulou, S. K., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., et al. (2019). Predicting splicing from primary sequence with deep learning. *Cell* 176, 535–548. doi: 10.1016/j.cell.2018.12.015
- Jeromin, A., and Bowser, R. (2017). Biomarkers in neurodegenerative diseases. *Neurodegen. Dis.* 15, 491–528. doi: 10.1007/978-3-319-57193-5\_20
- Jung, H., Lee, D., Lee, J., Park, D., Kim, Y. J., Park, W.-Y., et al. (2015). Intron retention is a widespread mechanism of tumor-suppressor inactivation. *Nat. Genet.* 47, 1242–1248. doi: 10.1038/ng.3414
- Kanagasabai, R., Serdar, L., Karmahapatra, S., Kientz, C. A., Ellis, J., Ritke, M. K., et al. (2017). Alternative RNA processing of topoisomerase II $\beta$  in etoposide-resistant human leukemia K562 cells: intron retention results in a novel C-terminal truncated 90-kDa isoform. *J. Pharmacol. Exp. Ther.* 360, 152–163. doi: 10.1124/jpet.116.237107
- Katz, Y., Wang, E. T., Airoidi, E. M., and Burge, C. B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* 7, 1009–1015. doi: 10.1038/nmeth.1528
- Kienzle, N., Young, D. B., Liaskou, D., Buck, M., Greco, S., and Sculley, T. B. (1999). Intron retention may regulate expression of Epstein-Barr virus nuclear antigen 3 family genes. *J. Virol.* 73, 1195–1204. doi: 10.1128/JVI.73.2.1195-1204.1999
- Kim, D., Shivakumar, M., Han, S., Sinclair, M. S., Lee, Y.-J., Zheng, Y., et al. (2018). Population-dependent intron retention and DNA methylation in breast cancer. *Mol. Cancer Res.* 16, 461–469. doi: 10.1158/1541-7786.MCR-17-0227
- Koch, L. (2017). Alternative splicing: a thermometer controlling gene expression. *Nat. Rev. Genet.* 18:515. doi: 10.1038/nrg.2017.61
- Li, H.-D., Funk, C. C., and Price, N. D. (2020). iREAD: a tool for intron retention detection from RNA-seq data. *BMC Genomics* 21:128. doi: 10.1186/s12864-020-6541-0
- Li, Y. I., Knowles, D. A., Humphrey, J., Barbeira, A. N., Dickinson, S. P., Im, H. K., et al. (2018). Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* 50, 151–158. doi: 10.1038/s41588-017-0004-9
- Lim, K. H., Ferraris, L., Filloux, M. E., Raphael, B. J., and Fairbrother, W. G. (2011). Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc. Natl. Acad. Sci. U.S.A.* 108, 11093–11098. doi: 10.1073/pnas.1101135108
- Lindeboom, R. G., Supek, F., and Lehner, B. (2016). The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat. Genet.* 48:1112. doi: 10.1038/ng.3664
- Lykke-Andersen, S., and Jensen, T. H. (2015). Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat. Rev. Mol. Cell Biol.* 16, 665–677. doi: 10.1038/nrm4063
- Mao, R., Kumar, P. K. R., Guo, C., Zhang, Y., and Liang, C. (2014). Comparative analyses between retained introns and constitutively spliced introns in *Arabidopsis thaliana* using random forest and support vector machine. *PLoS ONE* 9:e104049. doi: 10.1371/journal.pone.0104049
- Mauger, O., Lemoine, F., and Scheiffle, P. (2016). Targeted intron retention and excision for rapid gene regulation in response to neuronal activity. *Neuron* 92, 1266–1278. doi: 10.1016/j.neuron.2016.11.032
- Middleton, R., Gao, D., Thomas, A., Singh, B., Au, A., Wong, J. J., et al. (2017). IRFinder: assessing the impact of intron retention on mammalian gene expression. *Genome Biol.* 18:51. doi: 10.1186/s13059-017-1184-4
- Monteuuis, G., Wong, J. J. L., Bailey, C. G., Schmitz, U., and Rasko, J. E. J. (2019). The changing paradigm of intron retention: regulation, ramifications and recipes. *Nucleic Acids Res.* 47, 11497–11513. doi: 10.1093/nar/gkz1068
- Morris, K. V., and Mattick, J. S. (2014). The rise of regulatory RNA. *Nat. Rev. Genet.* 15, 423–437. doi: 10.1038/nrg3722
- Mukherjee, S., Sengupta, S., Mukherjee, A., Basak, P., and Majumder, A. L. (2019). Abiotic stress regulates expression of galactinol synthase genes post-transcriptionally through intron retention in rice. *Planta* 249, 891–912. doi: 10.1007/s00425-018-3046-z
- Naro, C., Jolly, A., Di Persio, S., Bielli, P., Setterblad, N., Alberdi, A. J., et al. (2017). An orchestrated intron retention program in meiosis controls timely usage of transcripts during germ cell differentiation. *Dev. Cell* 41, 82–93. doi: 10.1016/j.devcel.2017.03.003

- Ner-Gaon, H., Halachmi, R., Savaldi-Goldstein, S., Rubin, E., Ophir, R., and Fluhr, R. (2004). Intron retention is a major phenomenon in alternative splicing in Arabidopsis. *Plant J.* 39, 877–885. doi: 10.1111/j.1365-313X.2004.02172.x
- Ni, T., Yang, W., Han, M., Zhang, Y., Shen, T., Nie, H., et al. (2016). Global intron retention mediated gene regulation during CD4<sup>+</sup> T cell activation. *Nucleic Acids Res.* 44, 6817–6829. doi: 10.1093/nar/gkw591
- Niba, E. T. E., Yamanaka, R., Rani, A. Q. M., Awano, H., Matsumoto, M., Nishio, H., et al. (2017). DMD transcripts in CRL-2061 rhabdomyosarcoma cells show high levels of intron retention by intron-specific PCR amplification. *Cancer Cell Int.* 17:58. doi: 10.1186/s12935-017-0428-4
- Nilsen, T. W., and Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457–463. doi: 10.1038/nature08909
- Ottens, F., and Gehring, N. H. (2016). Physiological and pathophysiological role of nonsense-mediated mRNA decay. *Pflügers Archiv Eur. J. Physiol.* 468, 1013–1028. doi: 10.1007/s00424-016-1826-5
- Palazzo, A. F., Mahadevan, K., and Tarnawsky, S. P. (2013). ALREX-elements and introns: two identity elements that promote mRNA nuclear export. *Wiley Interdiscipl. Rev. RNA* 4, 523–533. doi: 10.1002/wrna.1176
- Parenteau, J., and Elela, S. A. (2019). Introns: good day junk is bad day treasure. *Trends Genet.* 35, 923–934. doi: 10.1016/j.tig.2019.09.010
- Paronetto, M. P., and Sette, C. (2010). Role of RNA-binding proteins in mammalian spermatogenesis. *Int. J. Androl.* 33, 2–12. doi: 10.1111/j.1365-2605.2009.00959.x
- Parra, M., Booth, B. W., Weiszmann, R., Yee, B., Yeo, G. W., Brown, J. B., et al. (2018). An important class of intron retention events in human erythroblasts is regulated by cryptic exons proposed to function as splicing decoys. *RNA* 24, 1255–1265. doi: 10.1261/rna.066951.118
- Pimentel, H., Conboy, J. G., and Pachter, L. (2015a). Keep me around: intron retention detection and analysis. *arXiv [Preprint]*. arXiv:1510.00696.
- Pimentel, H., Parra, M., Gee, S. L., Mohandas, N., Pachter, L., and Conboy, J. G. (2015b). A dynamic intron retention program enriched in RNA processing genes regulates gene expression during terminal erythropoiesis. *Nucleic Acids Res.* 44, 838–851. doi: 10.1093/nar/gkv1168
- Reimer, K., and Neugebauer, K. (2018). Blood relatives: splicing mechanisms underlying erythropoiesis in health and disease. *F1000Research* 7:F1000 Faculty Rev-1364. doi: 10.12688/f1000research.15442.1
- Rekosh, D., and Hammarskjöld, M.-L. (2018). Intron retention in viruses and cellular genes: detention, border controls and passports. *WIREs RNA* 9:e1470. doi: 10.1002/wrna.1470
- Romanoski, C. E., Glass, C. K., Stunnenberg, H. G., Wilson, L., and Almouzni, G. (2015). Epigenomics: roadmap for regulation. *Nature* 518, 314–316. doi: 10.1038/518314a
- Roy, S. W., and Irimia, M. (2008). Intron mis-splicing: no alternative? *Genome Biol.* 9:208. doi: 10.1186/gb-2008-9-2-208
- Scotti, M. M., and Swanson, M. S. (2016). RNA mis-splicing in disease. *Nat. Rev. Genet.* 17, 19–23. doi: 10.1038/nrg.2015.3
- Shen, S., Park, J. W., Huang, J., Dittmar, K. A., Lu, Z., Zhou, Q., et al. (2012). MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. *Nucleic Acids Res.* 40:e61. doi: 10.1093/nar/gkr1291
- Shen, S., Park, J. W., Lu, Z., Lin, L., Henry, M. D., Wu, Y. N., et al. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U.S.A.* 111, E5593–E5601. doi: 10.1073/pnas.1419161111
- Singh, R. K., and Cooper, T. A. (2012). Pre-mRNA splicing in disease and therapeutics. *Trends Mol. Med.* 18, 472–482. doi: 10.1016/j.molmed.2012.06.006
- Sterne-Weiler, T., Martinez-Nunez, R. T., Howard, J. M., Cvitovik, I., Katzman, S., Tariq, M. A., et al. (2013). Frac-seq reveals isoform-specific recruitment to polyribosomes. *Genome Res.* 23, 1615–1623. doi: 10.1101/gr.148585.112
- Sznajder, Ł., Thomas, J. D., Carrell, E. M., Reid, T., McFarland, K. N., Cleary, J. D., et al. (2018). Intron retention induced by microsatellite expansions as a disease biomarker. *Proc. Natl. Acad. Sci. U.S.A.* 115, 4234–4239. doi: 10.1073/pnas.1716617115
- Thiele, A., Nagamine, Y., Hauschildt, S., and Clevers, H. (2006). AU-rich elements and alternative splicing in the  $\beta$ -catenin 3' UTR can influence the human  $\beta$ -catenin mRNA stability. *Exp. Cell Res.* 312, 2367–2378. doi: 10.1016/j.yexcr.2006.03.029
- Timmons, J. A., Szkop, K. J., and Gallagher, I. J. (2015). Multiple sources of bias confound functional enrichment analysis of global-omics data. *Genome Biol.* 16:186. doi: 10.1186/s13059-015-0761-7
- Uzor, S., Zorzou, P., Bowler, E., Porazinski, S., Wilson, I., and Ladomery, M. (2018). Autoregulation of the human splice factor kinase CLK1 through exon skipping and intron retention. *Gene* 670, 46–54. doi: 10.1016/j.gene.2018.05.095
- Vanichkina, D. P., Schmitz, U., Wong, J. J.-L., and Rasko, J. E. (2018). Challenges in defining the role of intron retention in normal biology and disease. *Semin. Cell Dev. Biol.* 75, 40–49. doi: 10.1016/j.semcdb.2017.07.030
- Vaquero-Garcia, J., Norton, S., and Barash, Y. (2018). LeafCutter vs. MAJIQ and comparing software in the fast moving field of genomics. *bioRxiv*. doi: 10.1101/463927
- Wang, Y., Bernhardt, A. J., Nacson, J., Kraiss, J. J., Tan, Y.-F., Slifker, M., et al. (2019). “AP30: BRCA1 intron retention generates truncated proteins that avoid BRCT mutation misfolding and promote PART inhibitor resistance,” in *Proceedings of the 12th Biennial Ovarian Cancer Research Symposium* (Seattle, WA; Philadelphia, PA: AACR; Clin Cancer Res).
- Wang, Z., Ballut, L., Barbosa, I., and Le Hir, H. (2018). Exon Junction Complexes can have distinct functional flavours to regulate specific splicing events. *Sci. Rep.* 8, 1–8. doi: 10.1038/s41598-018-27826-y
- Wong, G. K.-S., Passey, D. A., Huang, Y., Yang, Z., and Yu, J. (2000). Is “junk” DNA mostly intron DNA? *Genome Res.* 10, 1672–1678. doi: 10.1101/gr.148900
- Wong, J. J.-L., Au, A. Y., Ritchie, W., and Rasko, J. E. (2016). Intron retention in mRNA: no longer nonsense: known and putative roles of intron retention in normal and disease biology. *Bioessays* 38, 41–49. doi: 10.1002/bies.201500117
- Wong, J. J.-L., Ritchie, W., Ebner, O. A., Selbach, M., Wong, J. W., Huang, Y., et al. (2013). Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* 154, 583–595. doi: 10.1016/j.cell.2013.06.052
- Wu, W., Zong, J., Wei, N., Cheng, J., Zhou, X., Cheng, Y., et al. (2017). CASH: a constructing comprehensive splice site method for detecting alternative splicing events. *Brief. Bioinformatics* 19, 905–917. doi: 10.1093/bib/bbx034
- Xiao, S., Tjostheim, S., Sanelli, T., McLean, J. R., Horne, P., Fan, Y., et al. (2008). An aggregate-inducing peripherin isoform generated through intron retention is upregulated in amyotrophic lateral sclerosis and associated with disease pathology. *J. Neurosci.* 28, 1833–1840. doi: 10.1523/JNEUROSCI.3222-07.2008
- Xu, Q., Walker, D., Bernardo, A., Brodbeck, J., Balestra, M. E., and Huang, Y. (2008). Intron-3 retention/splicing controls neuronal expression of apolipoprotein E in the CNS. *J. Neurosci.* 28, 1452–1459. doi: 10.1523/JNEUROSCI.3253-07.2008
- Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11:R14. doi: 10.1186/gb-2010-11-2-r14
- Zhang, A. Y., Su, S., Ng, A. P., Holik, A. Z., Asselin-Labat, M.-L., Ritchie, M. E., et al. (2018). A data-driven approach to characterising intron signal in RNA-seq data. *bioRxiv [Preprint]*. doi: 10.1101/352823
- Zhang, Q., Li, H., Jin, H., Tan, H., Zhang, J., and Sheng, S. (2014). The global landscape of intron retentions in lung adenocarcinoma. *BMC Med. Genomics* 7:15. doi: 10.1186/1755-8794-7-15
- Zhang, Z., Pan, Z., Ying, Y., Xie, Z., Adhikari, S., Phillips, J., et al. (2019). Deep-learning augmented RNA-seq analysis of transcript splicing. *Nat. Methods* 16, 307–310. doi: 10.1038/s41592-019-0351-9

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Zheng, Lin, Fang and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.