Review

Jianbo Wang, Zhenqing Ye, Tim H.-M. Huang, Huidong Shi and Victor Jin*

A survey of computational methods in transcriptome-wide alternative splicing analysis

Abstract: Alternative splicing is widely recognized for its roles in regulating genes and creating gene diversity. Consequently the identification and quantification of differentially spliced transcripts is pivotal for transcriptome analysis. Here, we review the currently available computational approaches for the analysis of RNA-sequencing data with a focus on exon-skipping events of alternative splicing and discuss the novelties as well as challenges faced to perform differential splicing analyses. In accordance with operational needs we have classified the software tools, which may be instrumental for a specific analysis based on the experimental objectives and expected outcomes. In addition, we also propose a framework for future directions by pinpointing more extensive experimental validation to assess the accuracy of the software predictions and improvements that would facilitate visualizations, data processing, and downstream analyses along with their associated software implementations.

Keywords: alternative splicing (AS); differential splicing; exon-skipping; graph-based exon-skipping scanner (GESS); RNA-sequencing (RNA-seq).

Introduction

In higher eukaryotes, alternative splicing (AS), a post-transcriptional regulatory mechanism can generate

*Corresponding author: Victor Jin,** Department of Molecular Medicine, University of Texas Health Science Center, 8403 Floyd Curl, San Antonio, TX 78229, USA, e-mail: jinv@uthscsa.edu
Jianbo Wang, Zhenqing Ye and Tim H.-M. Huang: Department of Molecular Medicine, University of Texas Health Science Center, 8403 Floyd Curl, San Antonio, TX 78229, USA
Huidong Shi: Department of Biochemistry and Molecular Biology, Georgia Regents University, Augusta, GA 30901, USA

multiple transcripts from the same gene to yield functionally diverse proteins (1, 2). More than 90% of human genes are alternatively spliced (3), and variation in the relative ratio of alternatively spliced isoforms of a single gene can lead to various diseases (4). Mechanistically, different types of AS events can be categorized into five main subgroups: (i) exon-skipping, where the exon can be spliced out of the transcript together with its flanking introns; (ii) alternative 5′ splice site (5′ss) and (iii) 3′ splice site (3′ss), which are the results of the recognition of two or more splice sites at one end of an exon; and (iv) intron retention, in which an intron can remain in the mature mRNA molecule; and finally, a rare subtype of AS is represented by (v) mutually exclusive exon (MXE) splicing. The MXEs are characterized by splicing of exons in a coordinated manner such that two or more splicing events are not independent. In addition, other complex events albeit less frequent can also give rise to alternative transcript variants that include alternative transcription start sites, and multiple polyadenylation sites (5).

Analysis of AS by RNA-sequencing (RNA-seq) has been performed on many human tissues and cell lines (6, 7). Indeed, the ability to globally detect every possible spliced isoform has enabled the discovery of a much larger amount of AS in human tissues than previously estimated. Considering that more than 90% of the human multi-exon genes have been found to undergo AS, exon-skipping could be one of the most important ways to steer transcriptional regulation (8). RNA-seq enables better characterization and quantification of individual transcriptomes because it can detect both gene expression and AS. Current RNA-seq protocols depend on high-throughput short-read sequencing of cDNA. With increased availability of next-generation sequencing instruments, RNA-seq has become a frequently used tool to study AS under multiple conditions (9), which rapidly yields a vast amount of data and significantly increases the complexity of the information to be analyzed. In the past years, a wide variety of tools have been developed to allow researchers to process RNA-seq data and to study the expression of isoforms and splicing events, and their changes under different

conditions. In this review article, we provide an overview of the computational methods available for studying exon-skipping events in AS from short-read RNA-seq data. Precisely, we have categorized these methods according to the different aspects they cover:

i. Summary of novel computational methods for detecting exon-skipping events in AS.
ii. Recapitulation of novel computational methods for the identification and quantification of differentially spliced transcripts in transcriptome analysis.
iii. Visualization of AS.
iv. Review of the correlation of epigenetic marks with exon-skipping events.

Summary of novel computational methods for detecting exon-skipping events in alternative splicing

Exon-skipping is the most common AS mechanism known in mammals, and is a major contributor to protein diversity in mammals (10). Exon-skipping results in the loss of an exon in the alternatively spliced mRNA. In this mode, the middle exon in three consecutive exons may be included in mature mRNA under some conditions or in particular tissues, and excluded from the mature mRNA in others. Several computational methods (Table 1) have been developed to detect exon-skipping events. We briefly summarize some of the methods in this section.

Our group has developed a novel computational method, graph-based exon-skipping scanner (GESS) (detection scheme summarized in Figure 1), to detect *de novo* exon-skipping events (Table 1) directly from raw RNA-seq data without the prior knowledge of gene annotation information. Firstly, we build a splicing-site-linking graph from the splicing-aware aligned reads, and then iteratively scan this linking graph to obtain those patterns conforming to skipping events by a greedy algorithm. Finally, we apply the Mixture of Isoforms (MISO) model (15) to calculate the ratio of skipping isoform versus inclusion isoform and determine the dominated isoform at these skipping event sites.

Alt Event Finder deploys a related approach based on deriving AS events from RNA-seq data. This method analyzes the transcripts built by genome-guided construction tools such as Cufflinks or Scripture and outputs annotated AS event which can be processed by MISO. Alt Event Finder enables users to identify novel AS events in the human genome. It can also be used for species of which known AS event annotations are not available.

ASprofile consists of several programs for extracting, quantifying, and comparing AS events from RNA-seq data. It reveals differences among transcriptomes caused by the interplay between constitutive transcription and AS and builds up a complex and dynamic picture of AS across tissue types. It allows us to compile the whole set of putative exon-skipping and other classes of AS events in normal human tissues detected from RNA-seq data, which will be a valuable resource with regard to the studies on transcriptional regulation and biomarkers of diseases.

ASTALAVISTA (alternative splicing transcriptional landscape visualization tool) has been designed to extract AS events from a given genomic annotation of exon-intron coordinates, and then categorize them into groups of common types and visualize a comprehensive picture of the resulting AS landscape. By comparing these transcripts, ASTALAVISTA identifies the difference in their splicing pattern and detects all AS events. Therefore, ASTALAVISTA can characterize AS in RNA-seq data from reference annotations (GENCODE, REFSEQ, ENSEMBL) as

Table 1 Summary of the available software packages for alternative splicing analysis.

| Method	Splice model	Pair end	Uses annotation	Visualization	URL	References
GESS	Exon-skipping	No	No	No	http://motif.bmi.ohio-state.edu/GESS_Web/	(1)
Alt Event Finder	Exon-skipping	Yes	No	No	http://compbio.iupui.edu/group/6/pages/alteventfinder	(11)
ASprofile	Exon-skipping	Yes	Yes	No	http://ccb.jhu.edu/software/ASprofile/	(12)
AStalavista	Exon-skipping	No	Yes	Yes	http://sammeth.net/confluence/display/ASTA/2+−+Download	(13)
AltAnalyze	Exon-skipping	No	Yes	Yes	http://www.altanalyze.org/	(14)
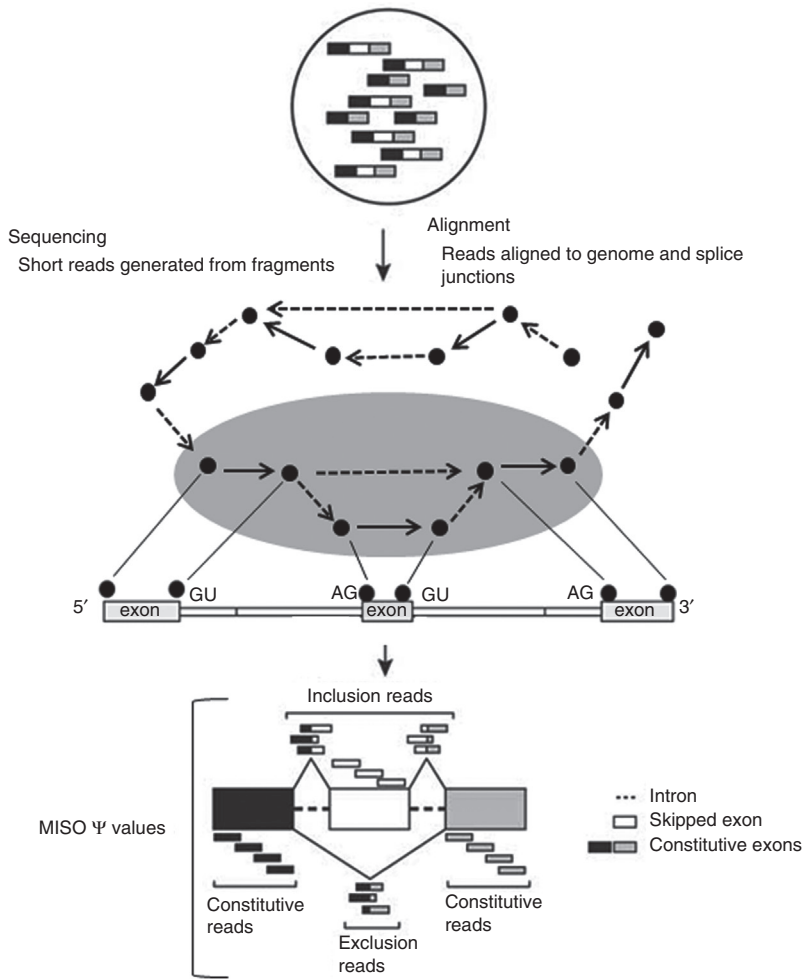
Figure 1 Scheme of the exon-skipping event detection pipeline (GESS).

well as for the genes selected according to structural and functional aspects of investigation.

AltAnalyze provides a comprehensive software workflow for the statistical and visual analysis of exon expression data. This workflow explores the functional impact of AS from Affymetrix Exon and Gene Arrays at the level of proteins, domains, microRNA binding sites, molecular interactions, and pathways. It includes rigorous statistical methods (FIRMA, MiDAS, and DABG filtering) and does not require prior knowledge of exon array analysis or programming. Moreover, these data can be visualized together with affected pathways and gene or protein interaction networks, allowing a straightforward identification of potential biological effects due to AS.

As transcriptional regulation in a cell is complex and dynamic, resulting in diverse outcomes under different physiological conditions, many current approaches for the identification of exon-skipping event rely on annotated exon information. Such approaches may not be able to

capture the complete landscape of gene expression *in situ*, sometimes leading to errors in interpretation of results. A notable advantage of the GESS and Alt Event Finder methods is that both the approaches empower investigators to capture *de novo* exon-skipping events. However, the current version of Alt Event Finder supports skipped exon events only. Interestingly, our GESS method is not only capable of capturing *de novo* exon-skipping events by building a splice-site-link graph and implementing an iteratively walking strategy on this graph, but also can provide a more accurate and thorough picture of the skipping events associated with a particular physiological condition within a cell. Furthermore, we integrated the MISO model into our method to determine which isoform, skipping- or inclusion-isoform, is the dominant transcript produced from a skipping-event site, where maintenance of the subtle balance between the two mRNA molecules is vital for proper cellular function and dynamics. Although our tool does not integrate visualization function, it can

be supplemented by tools such as ASTALAVISTA and AltAnalyze for better understanding of the complexity of AS and biological effects aroused by exon-skipping. Most of the methods discussed earlier analyze exon-skipping events within one sample, which need to combine differential splicing and gene expression in the analysis of normal and diseased individuals.

Recapitulation of novel computational methods for the identification and quantification of differentially spliced transcripts in transcriptome analysis

Differential splicing analysis reveals the differences in AS site between two samples or more. This is critical for understanding the mechanisms implicated with AS and its regulation. In addition, differential splicing analysis detects functional diversity that is not covered by differential gene expression analysis. Available software tools and packages undertake different approaches to identify differential splicing at the level of the gene. The choice of methods for a given analysis depends on the experimental objectives and expected outcomes. Table 2 summarizes the available software packages that are frequently used.

DiffSplice shows good characteristics for discovering and quantitating AS variants present in an RNA-seq dataset, without relying on annotated transcriptome or splice pattern. For two groups of samples, DiffSplice further utilizes a nonparametric permutation test to identify significant differences in expression at both gene level and transcription level. DiffSplice takes the sequence alignment/map format (SAM) files as input that supply the alignment of the RNA-seq reads on the reference genome, obtained from an RNA-seq aligner like MapSplice. The results of DiffSplice are summarized as a decomposition of the genome and can be visualized using the University of California Santa Cruz (UCSC) genome browser.

DSGseq focuses on exon usage by equating differential usage of nonterminal exons with differential splicing. Read densities are calculated for every annotated exon in the genome, data dispersion is modeled, and junction information is ignored. DSGseq uses the negative binomial distribution to model read-counts on all exons of a gene, considers over-dispersion in read-count distribution and borrows information across samples to get better estimation of the signal of each exon, and includes the mean read density across the gene (coverage). However, DSGseq reports no quantification (e.g., fold change) of the differential exon usages.

GLiMMPS is a robust statistical method for detecting splicing quantitative trait loci from RNA-seq data. GLiMMPS takes into account the individual genetic variation in sequencing coverage and the noise prevalent in RNA-seq data. Analyses of simulated and real RNA-seq datasets demonstrate that GLiMMPS is good at competing statistical models. As population-scale RNA-seq studies become increasingly affordable and popular, GLiMMPS provides a useful tool for elucidating the genetic variation of AS in humans and model organisms.

GPSeq proposes a two-parameter generalized Poisson (GP) model to the position-level read counts.

Table 2 Summary of major differential splicing analysis software.

Method	Splice model	Pair end	Uses annotation	Visualization	URL	References
DiffSplice	Junction + exon	No	No	Yes	http://www.netlab.uky.edu/p/bioinfo/DiffSplice	(16)
DSGseq	Exon	No	Yes	No	http://bioinfo.au.tsinghua.edu.cn/software/DSGseq/	(17)
GLiMMPS	Exon	Yes	Yes	No	http://www.mimg.ucla.edu/faculty/xing/glimmps/	(18)
GPSeq	Exon	No	Yes	No	http://www-rcf.usc.edu/~liangche/software.html	(19)
MATS	Junction + exon	Yes	No	No	http://rnaseq-mats.sourceforge.net/	(20)
MISO	Exon	Yes	Yes	No	http://genes.mit.edu/burgelab/miso/	(15)
JuncBASE	Exon junction	No	No	No	http://compbio.berkeley.edu/proj/juncbase/Home.html	(14)
FineSplice	Exon junction	No	Yes	No	http://sourceforge.net/projects/finesplice/	(21)
SpliceTrap	Exon inclusion	Yes	No	No	http://rulai.cshl.edu/splicetrap/	(22)

GP model fits the data much better than the traditional Poisson model. Based on the GP model, we can better estimate gene or exon expression, perform a more reasonable normalization across different samples, and improve the identification of differentially expressed genes as well as differentially spliced exons. The usefulness of the GP model is demonstrated by applications to multiple RNA-seq data sets.

MATS provides a computational tool for the detection of exon inclusion [percent spliced in (PSI)] from junction and exon reads for the exon and its two flanking exons. The statistical model of MATS calculates the p-value and false discovery rate that the difference in the isoform ratio of a gene between two conditions exceeds a given user-defined threshold. From the RNA-seq data, MATS can automatically detect and analyze AS events corresponding to all major types of AS patterns. MATS recognizes the type of AS (e.g., skipped exon and retained intron) and returns both p-value and magnitude (ΔPSI) for each alternative splice in the dataset.

MISO (Mixture of Isoforms) is a probabilistic framework that quantitates the expression level of alternatively spliced genes from RNA-seq data, and identifies differentially regulated isoforms or exons across samples. Through modeling the splicing process by which reads are produced from isoforms in RNA-seq, the MISO model uses Bayesian inference to compute the probability that a read originated from a particular isoform. MISO uses the inferred assignment of reads to isoforms to quantitate the abundances of the underlying set of alternative mRNA isoforms. Confidence intervals over estimates are accrued to quantitate the reliability of the estimates.

JuncBASE is used to identify and classify AS events from RNA-seq data. Alternative splicing events are identified from splice junction reads from RNA-seq read alignments and annotated exon coordinates. In addition to the identification of AS events, JuncBASE also uses read counts to quantify the relative expression of each isoform and identifies splice events that are significantly differentially expressed across two or more samples. JuncBASE was initially developed to characterize annotated and novel AS events throughout Drosophila development as well as splice events that are altered upon knockdown of splice factors. Consequently, it has been further utilized for AS analysis in cancer RNA-seq studies.

FineSplice takes advantage of novel pipeline based on TopHat2 combined with a splice junction detection algorithm to dissect the impact of design and method on the mapping, detection, and quantification of splice junctions from multi-exon reads from RNA-seq data, at enhanced detection precision with small loss in sensitivity. Following alignment with TopHat2 using known transcript annotations, FineSplice takes the input as resulting BAM file and outputs a confident set of expressed splice junctions with the corresponding read counts. Potential false positives arising from spurious alignments are filtered out through a semi-supervised anomaly detection strategy based on logistic regression. Multiple mapping reads with a unique location after filtering are rescued and reallocated to the most reliable candidate location.

SpliceTrap determines exon inclusion ratios in paired-end RNA-seq data, with broad applications for the study of AS. SpliceTrap approaches exon inclusion level estimation as a Bayesian inference problem by enumerating each tri-exon combination generated by shuffling explicitly known, annotated exons. For every exon it quantifies the extent to which it is included, skipped, or subjected to size variations due to alternative 3′/5′ splice sites or Intron Retention. In addition, SpliceTrap can quantify AS within a single cellular condition, with no need of a background set of reads.

As a gene normally carries several splice isoforms, splicing analysis for specific one requires more input data than gene expression analysis. A read that arises from sequencing workflow can map anywhere along a gene and count towards expression, but a read must include the AS region to count towards splicing analysis. In a recent human study, 100–150 M reads (100 nt, paired end) were required to detect 80% AS events, and more than 400 M reads were needed to detect 80% AS differences between two conditions (23). Given current sequencing costs and computational approaches, statistically robust detection of differential splicing is biased towards the more abundant transcripts. Nonetheless, the problem is complex because many genes have multiple AS choices, and each isoform may have biologically unique properties and some of them work together to achieve same biological function. To reconstruct which full-length isoforms are present in what amounts requires either full-length sequencing of transcripts or full knowledge of all possible isoforms as well as considerable modeling based on the shorter reads. In future, when high-throughput full-length transcriptome sequencing becomes available, splicing analysis will undoubtedly be based on comparison of isoforms. However, current approaches ignore isoform reconstruction in favor of local analysis. The challenges associated with differential splicing analysis apply equally to mammalian systems and to organisms with more compact genomes. The primary difference is that better performance is expected for all approaches in compact genomes, due to their lower complexity (24).

Visualizing alternative splicing

Primarily, the ability to visualize the complexity of AS is an important aspect of this analysis. Visualization for RNA-seq requires specialized tools that can efficiently process large amount of data from multiple samples. This has triggered the development of specialized tools to visualize alternative isoforms and events from RNA-seq data. Perhaps the simplest way to visualize isoforms and events is to generate track files for a genome browser. For instance, RSEM produces WIG files that can be viewed as tracks in the UCSC browser. Similarly, SpliceGrapher (25) and DiffSplice produce files in GFF-like formats (http://gmod.org/wiki/GFF), which can be uploaded into visualization tools. SpliceGrapher and Alexa-Seq (26) have their own visualization utilities. Other tools have been developed independently from the analysis method. For instance, the Sashimi plots are made using alignments (stored in the SAM/BAM format) and gene model annotations. Sashimi plots can be used to quickly screen differentially spliced exons along genomic regions of interest. The Integrative Genomics Viewer that was developed by and is available from the Broad Institute is one of the most widely used tools for visualizing short-read sequence data in the context of rich annotations such as known gene models. Finally, SpliceSeq and SplicingViewer (27) are stand-alone tools that also provide visualization of results besides mapping reads, quantifying events, and differential splicing.

Correlation of epigenetic marks with exon-skipping events

It has been widely accepted that chromatin state plays essential roles in regulating gene expression. DNA methylation, nucleosome occupancy, and modifications of histone are involved in determining the chromatin state, while some transcription factors can bind to specific regulatory regions to interact with chromatin and regulate gene expression (28). All these factors can be considered as epigenetic features that regulate gene expression from a broad perspective (29). Though epigenetic signatures are mainly found to be enriched in promoters, it has become increasingly clear that they are also present in exon regions, indicating a potential link of epigenetic regulation to splicing (30). For example, DNA methylation level shows distinctive differential patterns between intron and exon (31). Histone modifications such as H3K36me3, H3K79me1, H2BK5me1, H3K27me1,

H3K27me2, and H3K27me3 were found to be related with exon expression (32).

As epigenetic features have been implied to have a connection with splicing, we wonder whether they may also be involved in AS. Indeed, nucleosome occupancy level was found to be lower in cassette exons than in constitutively spliced exons. H3K36me3 and H3K9ac were found to be related to the exon-skipping event of *NCAM* (33). The level of H3K36me3 was found to be different in mutually exclusive exons of *FGFR2*, *TPM1, TPM2*, and *PKM2* between PNT2 and hMSC cell types. H3K36me3 was also found to be depleted in skipped exon in a genome-wide study across different species (34), though this finding is still controversial. Other histone methylations, such as H3K4me1, H3K4me3, H3K27me3 and H3K9me1, were found to be associated with the AS events of *FGFR2* (35), while H3K4me3 was suggested to affect the AS events of *CHD1* (36), and H3K9me3 was found to be associated with the multiple exon-skipping of *CD44* (37).

With respect to AS, evidence implicates certain histone modification in helping to regulate the choice of splicing junction by altering the rates of transcription, nucleosome positioning, or direct interactions with proteins that mark exon-intron junction of pre-mRNA. The changes in physiology condition can alter the chromatin modifications at these junctions and concomitantly modulate exon-skipping. A recent study on DNA methylation/chromatin epigenetic interrelationships with exon-skipping found that intron-exon junctions are enriched in sharp transition in DNA methylation levels (31). A relevant report (38) shows that malignant prostate cancer cells bear enrichment of DNA hypermethylation at exon-intron junctions suggesting that the cancer-linked alteration of DNA methylation levels may affect AS.

Programmed changes in DNA methylation in intra- and intergenic regions are not restricted to differentiation-related events. For example, electroconvulsive stimulation of mouse neuronal cells *in vivo* was recently demonstrated to cause rapid decrease and increase in DNA methylation in a substantial minority of CpG sites, especially at CpG-poor regions (39). The physiologically linked DNA demethylation included rapid demethylation of exons and introns in various positions of the genes. Importantly, there was enrichment in these DNA epigenetic changes in the vicinity of brain-related genes. Though the accumulating evidence suggests a potential connection between epigenetics and AS, to date no systematic studies have been carried out by any group to establish the association of epigenetic features with different types of AS.

Conclusions and perspectives

Previously, initial estimation suggested that at least 60% of human genes might undergo AS. In fact, current application of next generation sequencing technologies revealed that almost all genes (95%) in the human genome will undergo AS, highlighting its important role in the complexity of eukaryotic gene expression. However, although much research work has been dedicated to the understanding of the AS phenomenon, the regulatory mechanism of AS is still not fully understood.

An array of studies on the genomes of human and many other vertebrates, invertebrates, yeast, protozoa and plants, as well as sequences of large numbers of transcripts generated from each organism provides the data for comparative genomics and transcriptomics analyses. As a matter of fact, further knowledge acquired by sequencing of genomes and transcripts from more organisms of a broader range of lineages should advance our understanding on the mechanism of AS.

Notably, quite often the skipping event is not confined to the exon carrying the splicing mutation, such as the majority of 5′SS, 3′SS, and regulatory element mutation lead to skipping of the affected exon. However, it can also be extended to neighboring exon, either upstream or downstream. Precisely, it is conceivable that significant amount of studies are still needed to reveal the mystery of the genomic milieu. Undoubtedly, the increase in knowledge on the basic mechanisms of splicing provided by functional experiments such as *in silico* assays will help to provide new insights into identifying splicing biomarkers. Concurrently it is equally important that bioinformatics predictions must be validated experimentally.

Extending *de novo* transcriptome assembly methods to calculate differential expression of isoforms between two or more conditions could facilitate the analysis of isoform expression for nonmodel organisms. Although under present circumstances, this may be accomplished by utilizing a combination of methods, a tool integrating all of these approaches would render a powerful approach to study expression and splicing in tumors expected to undergo potential multiple genome rearrangements and copy number alterations. In the case of a different scenario, given that a reference genome sequence does not represent all DNA which can be possibly transcribed in a cell, unmapped RNA reads may emerge from functional RNAs not represented in the genome annotation. The tools that map reads to a genome reference and simultaneously attempt to perform transcript assembly will be also quite useful to perform systematic analyses of RNA in cancer tissues as well as in partially assembled genomes.

There are new technologies for single-molecule sequencing that can be used to probe the transcriptome. This may preclude the need to perform reconstruction of isoforms. Currently, RNA-seq may still be necessary for efficient quantification. Single-molecule sequencing (40) technologies will open up a whole new set of problems, like that of reconciling new cell-specific RNA sequences with the information available for the genome sequence and its annotation. With the advent of enhanced knowledge on the complexity of DNA and transcription machinery in tumor cells, it is quite prudent to anticipate that a full catalog of likely disease-causing variants will be divulged. In conclusion, modern cutting edge bioinformatics and genomic approaches together should be instrumental to improve our understanding of dysregulated pathways in each patient and to identify markers for prediction of diseases.

References

1. Ye Z, Chen Z, Lan X, Hara S, Sunkel B, Huang TH-M, Elnitski L, Wang Q, Jin VX. Computational analysis reveals a correlation of exon-skipping events with splicing, transcription and epigenetic factors. Nucleic Acids Res 2014; 42: 2856–69.
2. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. Alternative isoform regulation in human tissue transcriptomes. Nature 2008; 456: 470–6.
3. Matlin AJ, Clark F, Smith CWJ. Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol 2005; 6: 386–98.
4. Tazi J, Bakkour N, Stamm S. Alternative splicing and disease. Biochim Biophys Acta 2009; 1792: 14–26.
5. Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet 2010; 11: 345–55.
6. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 2009; 6: 377–82.
7. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, Schmidt D, O'Keeffe S, Haas S, Vingron M, Lehrach H, Yaspo M-L. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 2008; 321: 956–60.
8. 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.
9. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 2008; 18: 1509–17.
10. Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol 2009; 10: 741–54.

11. Zhou A, Breese MR, Hao Y, Edenberg HJ, Li L, Skaar TC, Liu Y. Alt Event Finder: a tool for extracting alternative splicing events from RNA-seq data. BMC Genomics 2012; 13: S10.

12. Florea L, Song L, Salzberg SL. Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. F1000Res 2013; 2: 188.

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

14. Brooks AN, Yang L, Duff MO, Hansen KD, Park JW, Dudoit S, Brenner SB, Graveley BR. Conservation of an RNA regulatory map between *Drosophila* and mammals. Genome Res 2011; 21: 193–202.

15. 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.

16. Hu Y, Huang Y, Du Y, Orellana CF, Singh D, Johnson AR, Monroy A, Kuan PF, 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 2013; 41: e39.

17. Wang W, Qin Z, Feng Z, Wang X, Zhang X. Identifying differen-tially spliced genes from two groups of RNA-seq samples. Gene 2013; 518: 164–70.

18. Zhao K, Lu ZX, Park JW, Zhou Q, Xing Y. GLiMMPS: robust statisti-cal model for regulatory variation of alternative splicing using RNA-seq data. Genome Biol 2013; 14: R74.

19. Srivastava S, Chen L. A two-parameter generalized Poisson model to improve the analysis of RNA-seq data. Nucleic Acids Res 2010; 38: e170.

20. Shen S, Park JW, Huang J, Dittmar KA, Lu Z-X, 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.

21. Gatto A, Torroja-Fungairiño C, Mazzarotto F, Cook SA, Barton PJR, Sánchez-Cabo F, Lara-Pezzi E. FineSplice, enhanced splice junction detection and quantification: a novel pipeline based on the assessment of diverse RNA-Seq alignment solu-tions. Nucleic Acids Res 2014; 42: e71.

22. Wu J, Akerman M, Sun S, McCombie WR, Krainer AR, Zhang MQ. SpliceTrap: a method to quantify alternative splicing under single cellular conditions. Bioinformatics 2011; 27: 3010–6.

23. Liu Y, Ferguson JF, Xue C, Silverman IM, Gregory B, Reilly MP, Li M. Evaluating the impact of sequencing depth on transcrip-tome profiling in human adipose. PLoS One 2013; 8: e66883.

24. Steijger T, Abril JF, Engstrom PG, Kokocinski F, The RC, Hubbard TJ, Guigó R, Harrow J, Bertone P. Assessment of tran-script reconstruction methods for RNA-seq. Nat Methods 2013; 10: 1177–84.

25. 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.

26. Griffith M, Griffith OL, Mwenifumbo J, Goya R, Morrissy AS, Morin RD, Corbett R, Tang MJ, Hou Y-C, Pugh TJ, Robertson G, Chittaranjan S, Ally A, Asano JK, Chan SY, Li HI, McDonald H,

Teague K, Zhao Y, Zeng T, Delaney A, Hirst M, Morin GB, Jones SJM, Tai IT. Alternative expression analysis by RNA sequencing. Nat Methods 2010; 7: 843–7.

27. 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.

28. Goodrich JA, Tjian R. Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. Nat Rev Genet 2010; 11: 549–58.

29. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. Cell 2007; 129: 823–37.

30. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M, Edsal L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 2009; 462: 315–22.

31. Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J, Wei CL. Dynamic changes in the human methylome during differentiation. Genome Res 2010; 20: 320–31.

32. Andersson R, Enroth S, Rada-Iglesias A, Wadelius C, Komorowski J. Nucleosomes are well positioned in exons and carry characteristic histone modifications. Genome Res 2009; 19: 1732–41.

33. Schor IE, Rascovan N, Pelisch F, Allo M, Kornblihtt AR. Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. Proc Natl Acad Sci USA 2009; 106: 4325–30.

34. Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. Differential chromatin marking of introns and expressed exons by H3K36me3. Nat Genet 2009; 41: 376–81.

35. Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. Regulation of alternative splicing by histone modifica-tions. Science 2010; 327: 996–1000.

36. Sims RJ, 3rd, Millhouse S, Chen CF, Lewis BA, Erdjument-Bromage H, Tempst P, Manley JL, Reinberg D. Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. Mol Cell 2007; 28: 665–76.

37. Saint-Andre V, Batsche E, Rachez C, Muchardt C. Histone H3 lysine 9 trimethylation and HP1gamma favor inclusion of alter-native exons. Nat Struct Mol Biol 2011; 18: 337–44.

38. Yegnasubramanian S, Wu Z, Haffner MC, Esopi D, Aryee MJ, Badrinath R, He TL, Morgan JD, Carvalho B, Zheng Q, De Marzo AM, Irizarry RA, Nelson WG. Chromosome-wide map-ping of DNA methylation patterns in normal and malignant pros-tate cells reveals pervasive methylation of gene-associated and conserved intergenic sequences. BMC Genomics 2011; 12: 313.

39. Guo JU, Ma DK, Mo H, Ball MP, Jang M-H, Bonaguidi MA, Balazer JA, Eaves HL, Xie B, Ford E, Zhang K, Ming G-L, Gao Y, Song H. Neuronal activity modifies the DNA methylation land-scape in the adult brain. Nat Neurosci 2011; 14: 1345–51.

40. Pushkarev D, Neff NF, Quake SR. Single-molecule sequencing of an individual human genome. Nat Biotechnol 2009; 27: 847–50.