



Single RNA-Seq Transcriptome Data Used for Retrieving Diverse Molecular Functions

Amrendra Kumar^{1,2}, Adhitthan Shanmugam¹, Minu Kesheri³, R. Sanjeevi², Annamalai Arunachalam⁴ and PTV Lakshmi^{1*}

¹Department of Bioinformatics, Pondicherry University, Kalapet, India

²Department of Bioinformatics, NIMS Institute of Allied Medical Sciences and Technology, NIMS University, Jaipur, India

³Department of Bioinformatics, Boise State University, Idaho, United States

⁴Department of Food Science and Technology, Pondicherry University, Kalapet, India

*Corresponding author: PTV Lakshmi, Department of Bioinformatics, Pondicherry University, Kalapet, Pondicherry, India; E-mail: lakanna@bicpu.edu.in

Received date: 20 October, 2024, Manuscript No. JABCB-24-150508;

Editor assigned date: 23 October, 2024, PreQC No. JABCB-24-150508 (PQ);

Reviewed date: 08 November, 2024, QC No. JABCB-24-150508;

Revised date: 09 July, 2025, Manuscript No. JABCB-24-150508 (R);

Published date: 16 July, 2025, DOI: 10.4172/2329-9533.1000298.

Abstract

RNA sequencing and microarray methods, which produce RNA from plant or animal tissue and transform it into complementary DNA (cDNA) to construct a sequencing library, are the methods used to produce transcriptome data. Transcriptome (RNA-seq) data is used for various molecular functions. Many bioinformatics tools and techniques have been developed for transcriptome data analysis and retrieving more than ten functional information. This review paper describes almost all possible tools that are used to gain scientific knowledge from a single transcriptome (RNA-seq) data analysis, including data collection from the public domain, quality control, read alignment and count, differential gene or transcript isoform expression, hub genes, specific gene identification like transcription factor, CAZyme, resistance genes, alternatively spliced genes, protein-protein interaction, pathway analysis, or functional analysis. Furthermore, non-coding RNA identifications, alternative polyadenylation, and transposable elements can be identified from transcriptome data. We highlighted all challenges and came to proceed with the steps of tools. In conclusion, we discuss the viewpoint for new analysis and technologies that are fluctuating to retrieve information from single transcriptome data.

Keywords: Transcriptome data; RNA-seq; Non-coding; Transposable element

Introduction

RNA sequencing methods are powerful in generating transcriptome data that enables comprehensive or complex analysis of genes [1]. By May 2024, the National Centre for Biotechnology Information (NCBI) will have roughly 207,892 transcriptome data in Sequence Read Archive (SRA) records from 50,461 BioProject samples of different organisms that were generated using different sequencing platforms.

Other databases are also available for transcriptome data access, like the TCGA database (<https://portal.gdc.cancer.gov/>) and the Science Data Bank (ScienceDB) (<https://www.scidb.cn/en>). Transcriptome data started to be generated after 24 years; the double helix model was proposed by Watson and Crick in 1953 [2]. At the current time, 12 sequencing techniques are available in the scientific field [3], which is divided into four generations. The first-generation sequencing technology developed by Sanger, as well as Maxam and Gilbert, chain termination, and chemical degradation methods in 1977 [4]. The most popular second-generation (next-generation) sequencing platform is Illumina in the world; its work on the Sequencing by Synthesis (SBS) approach produces millions of DNA fragments, which gives highly sensitive and accurate expression of known and unknown genes as well as characterized non-coding RNAs. These single excrement RNA-seq data can be used to retrieve various information (Figure 1). Third-generation, which also belongs to next-generation sequencing, is known as long-read sequencing [5]. It's very different compared to second generations of sequencing because it reads nucleotide sequences at a single molecule level and produces much longer reads, while the second generation breaks the longer strand sequences into multiple small fragments. These generations' sequencing methods have their benefit because it is sequencing nucleotides without clonal amplification or transcript assembly to full-length complementary DNA (cDNA). Third-generation sequencing platforms Pacific Biosciences (PacBio) [6] and Oxford nanopore [7], in which Single-Molecule Real-Time (SMRT) and nanopore sequencing technology are fit for all types of sequencing projects under turnaround time, budgets, and satisfied accurate sequencing data. In addition to the updated Oxford nanopore technology, known as fourth-generation sequencing technology, which works fast whole-genome scans within 15 minutes, and low-cost sequencing technology, the National Institutes of Health of the United States goal set may achieve \$1000 per human genome [8].

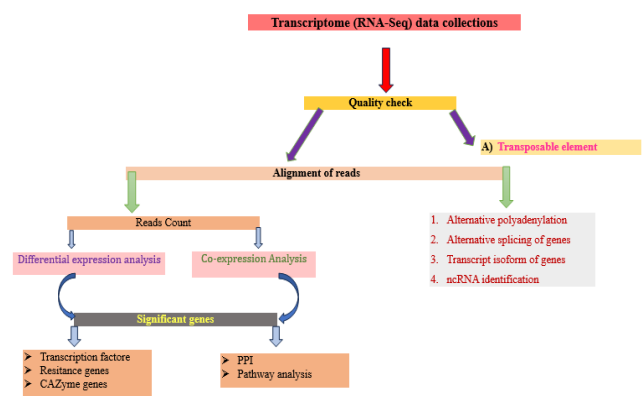


Figure 1: Describes almost all possible tools to be used to gain scientific knowledge from a single transcriptome (RNA-seq) data analysis.

Bioinformatics is a multi-disciplinary subject used to analyze large biological data and understand the molecular mechanisms of plants and animals [9]. This bioinformatics developed as tools to apply for identified biological information. Current era 2032 tools and software developed (till now), which are used to analysis of sequencing data (<https://bioinformatics.com/tools/tools-main.html>), where 497 free tools are available for RNA-seq data analysis. In this review paper, we are going to point out multiple bioinformatics analysis tools

that are used in single transcriptome (RNA-seq) data analysis and identify various information, including alternative polyadenylation; alternative splicing of genes; annotation 3 prime UTR; CAZyme genes; genes fusion; intron-exon retention; known and unknown genes; ncRNA identification; resistance genes; somatic mutation; transcript isoform of genes; transcription factor; translational efficiency of genes; transposable elements.

Materials and Methods

RNA-seq data collections

Many publicly available databases, but these three SRA, TCGA, and ScienceDB databases are very friendly using retrieving RNA-seq data directly, and sometimes big-size transcriptome data is downloaded using the SRA-toolkit inbuild program prefetch for SRA format and fastq-dump for fastq format. Researchers store data, but they face problems with analysis because primer users upload their transcriptome data to the public repository after analysis and publication of the paper. So, one question in the secondary data user's mind is, what can I do now on this transcriptome data? This review gives multiple opportunities to secondary users for different analyses and results since the primary users only cover four to five analyses in one transcriptome data, but we have multiple opportunities by using

different tools and algorithms of bioinformatics to make various types of results from single transcriptome data.

Quality check of RNA-Seq data

Bioinformaticians developed various tools for quality control and assessment of RNA-Seq data. Some tools work with filtering and trimming sequences, but some tools have only function sequences: trim, filter, and visualization (Table 1). Raw sequences generated by sequencing technologies have the possibility of some unwanted sequence adaptors and duplicate read sequences, which decrease the quality of sequences. The quality of base pair sequences is measured by the Phred quality score (Q score) that is calculated through logarithmically related to the base calling error probabilities (P).

$$Q = -10 \log_{10} P$$

The 90% base call accuracy means that one base is incorrect out of ten base sequences, and the Phred quality score is 10 (Q10), while Q20 represents one base incorrect out of 100, and base call accuracy is 99%. In addition, base call accuracy increases according to a numeric number of Phred quality scores; Q30 means 99.9% accuracy, so Q5 means one base is incorrect in 100,000 bases, and accuracy is 99.999% of sequences.

S. no.	Filtering and trimming	Trimming	Filtering and visualization
1	PRINSEQ	CUTADAPT	TRAPR
2	TRIMMOMATIC	CUTADAPT-IP	UMIS
	FASTP	SOLEXAQA	SCATER
	AFTERQC	FASTQPURI	FASTQC
	RNA-QC-CHAIN	NGSHORT	SOLEXAQA
	CLINQC	RBOWTIE2	FQSTAT
	QUASR	SKEWER	HTSEQTOOLS

Table 1: RNA-seq pre-analysis tools.

GC content bias mentions systemic variations at library preparations by PCRA amplification and sequencing that can distribute GC content across the read sequences, which affects the expression value of genes. The adaptor, a short DNA sequence called an adapter, is ligated onto the ends of a DNA insert, also known as a target sequence, to prepare libraries. Uncertain bases (Ns) can have a substantial influence on the accuracy of downstream studies and are a good sign of low-quality sequences.

Alignment of read sequences

Several alignment tools, based on different algorithms, are available for read sequence alignment; these tools have their significance. These two, Hierarchical Indexing for Spliced Alignment of Transcripts version 2 (HISAT2) and Spliced Transcripts Alignment to a Reference (STAR), are used for read alignment to identify alternative splicing genes. Its extensive alignment method allows it to align reads spanning huge introns and complex splicing patterns. It also works well in recognising novel splice junctions and alternative splicing events. Working with big datasets requires HISAT2 to be especially efficient in terms of speed and memory utilization. The Rail-RNA is also a splice-aware tool to fit introns spliced out of mature mRNA

transcripts, as well as creating exon-exon junctions in large groups (hundreds) of biological replicate samples. Another splice-unaware reference-based alignment tool, Bowtie 2, is flexible and effective for a wider range of sequencing applications because it allows local alignment, paired-end reads, and gapped alignment using a Burrows-Wheeler Transform (BWT) and FM-index. Furthermore, these tools have eschewed backtracking in favor of a dynamic programming process accelerated by single instruction multiple data. In addition, in the world plant and animal species, approximately 8 million have been identified, which includes 71886 genomes that have been assembled and/or partially sequenced and reported in the NCBI of various species, including mammals (676), birds (590), fish (865), insects (1796), fungi (3747), plants (1025), bacteria (33724), viruses (26004), archaea (2040), and others, so lots of species have not references information that time use *de novo* (references-free) assembly tools to identify prior knowledge such as new genes, transcript variant genes fusion, etc. There are many *de novo* assembly tools that have been built based on greedy, overlap layout consensus, de Bruijn graph, etc. algorithms like Trinity, SOAPdenovo-Trans, IDBA, Trans-ABYSS used for sort sequences, and rnaSPAdes. These tools can handle paired-end reads, multiple isoforms, and insert sizes to make a

consolidated output in contigs (transcripts) and are used backed to identified expression values of sort reads.

Expression quantification of genes

Genes are materials of heredity whose function is the transfer of information from one cell to another through the expression (activation) of those genes. gene expression values, we need to identify to understand the effect level of biological and molecular functions of cells. Nowadays, there are multiple tools available for quantifying the expression value of genes from transcriptome (RNA-seq) data with Gene Transfer Format (GTF) files, such as feature counts, HTSeq, and Salmon. These tools are very popular and suitable for finding the expression (read count) value; they have multi-thread support, so they quantify results fast as well, and they are suitable for single-cell RNA-seq (scRNA-seq) data. The PennDiff tool finds the expression on the exon level and transcript isoform only.

Hub genes retrieve from multiple samples

Pearson and Spearman correlations statistical base algorithms tool, Weighted Gene Co-Expression Network Analysis (WGCNA), is commonly used for identifying intersectional expressed genes in the multiple samples, it's called hub genes. Nowadays, lots of different versions of the same plant or animal transcriptome (RNA-Seq) data are available in the public domain, like NCBI. Researchers are re-analyzing and finding some novel information, but sometimes they need to analyze co-relationships between various samples and identify clusters of genes as well as disperse genes. The WGCNA tool defines co-expression networks, meaning that network nodes corresponding to gene expression (read count) and edge interaction among genes are determined by the pairwise correlations between gene expression. Furthermore, these tools' eigengene significance is calculated by the eigenvector of expressed genes of samples (control or treated) to refer to module membership and also known as eigengene connectivity of inter-individual variation in gene expression. Finding potential candidate (hub) genes that are co-expressed consistently, associated with the traits (sample), and potentially implicated in related biological pathways is essential.

Comparative analysis from different samples

This is a very common and popular analysis of transcriptome (RNA-Seq) data because most of the project analyses of control vs. treated, drought vs. normal, disease vs. normal, etc., to retrieve significant genes. Multiple tools are used to identify differential expression genes from the read count of samples. The DESeq2, ABSSeq, and DEXUS algorithms use negative binomial distribution, and DEXUS works on a finite mixture of unknown conditions, while other tools EdgeR, Limma, and maSigPro use a Generalized Linear Model (GLM). Additional SARTools tools used both algorithms, negative binomial and general linear models, to identify and generate HTML reports. Furthermore, multiple web browsers are available, like integrated Differential Expression and Pathway (iDEP) analysis (<http://bioinformatics.sdstate.edu/idep11/>), gene expression normalization analysis and visualization (GENAVi), <https://junkdnalab.shinyapps.io/GENAVi/>, DEAPP <https://yanli.shinyapps.io/DEApp/> and ideal: interactive differential expression analysis (iDEAL) <http://shiny.imbei.uni-mainz.de:3838/ideal/tools>, all of which used DESeq2, EdgeR, and Limma-Voom algorithms to identify differentially expressed genes in multiple replicate samples.

Transcriptome isoform analysis

Based on the total size of isoforms, the DESeq and edgeR programs are primarily utilized to identify Differentially Expressed Genes (DEGs), i.e., the total number of copies for every isoform, and fail to recognize the structures of the isoforms. Furthermore, the raw count approaches were reliable in capturing true changes in gene expression when all isoforms of a gene were consistently up or down-regulated between two conditions; however, this has also become unreliable. For the accurate and statistically trustworthy identification of changes in gene expression at a higher resolution, transcript-level calculations are therefore essential. So, research developed a huge number of tools for identifying differential expression analysis based on isoforms, such as RSEM-EBSeq (RNA-Seq by Expectation Maximization), PennSeq, Mixture-of-Isoforms (MISO), and RDiff. These tools used the following method to classify and count isoform expressions, including, The RSEM-EBSeq pipeline clusters the isoforms into K uncertainty groups based on the unmappability scores of the reads using the K-means algorithm; PennSeq used state-of-the-art algorithms for isoform expression estimation; MISO worked by Markov Chain Monte Carlo (MCMC) and Metropolis-Hastings (MH) algorithms; and rDiff applied the simplified variant of the parametric test, called rDiff. Poisson is based on the Poisson distribution instead of the negative binomial distribution.

Alternative Splicing (AS) events identification

Alternative Splicing (AS) is a process that occurs in many human, plant, and other metazoan genes that produces multiple mRNA isoforms from a single gene locus. Because Genetically Modulated Alternative Splicing (GMAS) processes are essentially independent of cell type, genetic variations that modify splicing may have a wide range of roles. Genetic variants present in various genomic regions (exons or introns) can affect alternative splicing, which is involved in almost all aspects of gene regulation, ranging from transcription to post-translation. In the current era, there has been a lot of research done on variants that are found close to the splicing sites in either the exon or the intron, and play directly impacts the splicing events, which leads to aberrant transcript isoform abundance. The most five common alternative splicing events occur, which are recognized by rMATS tools. They work on different methods and algorithms, such as replicate Multivariate Analysis of Transcript Splicing (rMATS), which is a robust method and works in various experimental factors to find differential alternative splicing by using the likelihood ratio test and its calculated p-value. SUPPA2 tools work faster than rMATS and have multiple testing options for occurring AS events under the same genes, which work on the Benjamini-Hochberg method. Modeling Alternative Junction Inclusion Quantification (MAJIQ) applies a combination of read rate modeling, Bayesian modeling, and bootstrapping to quantified subsequently estimated Percent Splice In (PSI) or dPSI for Local Splice Variation (LSV) detected by the builder. The AStool tool was developed only for alternative splicing (Intron Retention (IR), Exon Skipping (ES), Alternative 5 Splice Sites (A5SS), and Alternative 3' Splice Sites (A3SS)) identified in transcriptome data of plants. The IR event is most predominant in plants, while the ES event highly occurs in animals.

Furthermore, Bioconductor R also has the package, which is used to quantify alternative splicing. SpliceR is an R package that is used to predict the Nonsense-Mediated Decay (NMD) sensitivity of each coding potential transcript; it also calculates the Untranslated Region (UTR) and Open Reading Frame (ORF). Other newly developed R

package splice Wiz for alternative splicing analysis, these tool differential AS find by Generalized Linear Model (GLM) of complex experiment designs, and compared to other tools, its need less computation power for transcriptome (RNA-seq) data analysis.

Results and Discussion

Non-coding RNA identifications

DNA sequences consist of both coding and noncoding regions. Non-coding RNAs are classified into two groups based on their biological function and sequence length: housekeeping non-coding RNAs and regulatory non-coding RNAs. Housekeeping non-coding RNAs include small nuclear RNAs (snRNAs), ribosomal RNAs (rRNAs), and small nucleolar RNAs (snoRNAs), which are essential for processes such as intron splicing, RNA processing, and mRNA translation. Regulatory non-coding RNAs encompass long non-coding RNAs (lncRNAs) and small RNAs (sRNAs), which play a role in regulating gene expression through various mechanisms, including transcriptional, post-transcriptional, and epigenetic processes. Many regulatory non-coding RNA have been identified from transcriptome data, which interact with transcription factors to activate or repress gene expression. More numbering tools are available for the identification of lncRNAs; we are highlighting only the most popular and latest ones: CNCI (Coding-Non-Coding Index) works on Adjoining Nucleotide Triplets (ANT), which successfully extract coding and non-coding RNA sequences from all species transcriptome data. Coding Potential Calculator V2 (CPC2) tools also provide a web portal (<http://cpc2.cbi.pku.edu.cn>) for easy analysis coding and no coding sequences. This tool has a novel discriminative Support Vector Machine (SVM) model, which is used for four sequence intrinsic features as Fickett TESTCODE score, Open Reading Frame (ORF) length, ORF integrity, and isoelectric point (pI). Flnc also incorporates a type of feature: genomic location, multiple exons, promoter signature, and transcript length archive state-of-art prediction power. FEELnc (FIExible Extraction of LncRNAs) is an alignment-free program that accurately annotates lncRNA by the random forest model trained with general features also included like K-mer frequency and open reading frame, FEELnc filtering of coding potential score, and non-lnc RNA transcript through an all-in-one solution to formalize the lncRNA classes. Long non-coding RNA detection (lncDC) tool is user-friendly to customize the model with its data, lncDC alignment-free machine learning-based tool, so this tool automatically balances the training data of large variance mRNA and lncRNA from RNA-seq data by the three feature categories Secondary Structure Features (SSFs), Sequence Intrinsic Features (SIFs), and Protein Features (PFs) with the XGBoost algorithm. One more tool, PLEK (Predictor of long non-coding RNAs and messenger RNAs based on k-mer scheme), works alignment-free based on k-mer frequency and Support Vector Machine (SVM) of the sequences, and the k-mer usage is calibrated according to the size of k-mer strings from mRNA sequences.

Alternative polyadenylation analysis

Alternative Polyadenylation (APA) is a pivotal process regulating mRNA function. Its misregulation can profoundly impact cellular and organismal phenotypes, particularly in the transcriptomes of diseased cells, including tumors. The polyadenylation machinery that recognizes poly(A) sites operates through interactions between multiple transacting factors and cis-regulatory elements. APA's

influence on mRNA abundance is well recognized, but its broader effects on mRNA metabolism remain largely unexplored. The APA can occur within two positions in the intron and 3'UTR sequence of the entire gene length, which can affect the protein-coding and noncoding. However, it lacks a PolyA-seq platform for APA analysis and fills this gap in the scientific community; nowadays, bioinformatics tools are available to retrospectively analyze the APA from existing RNA-seq data in the public domain. Dynamic analysis of Alternative Polyadenylation from RNA-seq (DaPars) tool is first for analysis of APA at standard RNA-seq to effectively analyze the dynamic regulation of APA by comparing RNA-seq read coverage across different experimental conditions. The above mentions many tools that work on gene/isoform expression analysis, but among the few tools, DaPars is one tool that analyzes 3'UTR from RNA-Seq data (<https://github.com/ZhengXia/DaPars>). The other tool, especially for humans and mice, is RNA-Seq data analysis. Quantification of Alternative Polyadenylation (QAPA), which works based on gene models to extract and annotate 3'UTR sequences (<https://github.com/morrislab/qapa>). APATrap (<https://sourceforge.net/projects/apatrap/>), which is based on a mean squared error model, is designed to identify and quantify Alternative Polyadenylation (APA) sites from RNA-seq data. APATrap excels at discovering novel 3' UTRs and 3' UTR extensions, which enhances the identification of potential poly(A) sites in previously overlooked regions, thereby improving genome annotations. Additionally, APATrap aims to catalog all potential poly(A) sites and detect genes with differential APA site usage between conditions. It is also capable of handling batch samples (more than two) simultaneously, facilitating large-scale data analysis and avoiding inconsistencies that may arise from different pair-wise comparisons. Furthermore, APATrap utilizes standard file formats (e.g., SAM format) from read mapping, allowing researchers to easily integrate it into their analysis pipelines for detecting poly(A) sites or refining 3' UTRs. When used in combination with other tools, such as Cufflinks, MISO, or Kleat, APATrap can provide complementary or consensus evidence for 3' UTR annotations or APA events identified by different approaches. Another tool, Roar and REPAC, is implemented in the Bioconductor R package, which uses an alignment (BAM) file and annotation (gtf) for the identification of APA (<https://github.com/vodkatad/roar/>).

Transposable element identification

Transposable elements, normally known as jumping genes, can move within genomes from one location to another. This rotation plays a positive or negative role in the gene promoters, enhancers, and other cis-regulatory sequences. Two or three bioinformatics tools were developed that use transcriptome data and detect Transposable Elements (TE). ChimeraTE is one of the tools that uses RNA-Seq data to identify the TEs based on two different chimeric transcript modes. Mode 1 utilizes a reference-guided strategy that relies on canonical genome alignment, whereas mode 2 detects chimeras originating from fixed or internationally polymorphic Transposable Elements (TEs) in the absence of a reference genome. One more LION tool has been used for TE identification from transcriptome align data, which annotates the intersection between the assembly, a reference gene set, and a repeat set.

Essential gene and pathway identification

A Protein-Protein Interactions (PPI) network was constructed for the more significant common gene accession using the STRING

database (v11) Szklarczyk by applying an interaction filter score >0.4 (medium confidence) and further visualized through Cytoscape. Within the networks, the parameters K-mean cluster score=2, degree cutoff=2, and max. depth=100 were set. Subnetwork analysis was performed using Molecular Complex Detection (MCODE) for clustering connected. For the identification of key/essential genes from the PPI network, the CytoHubba plugin of Cytoscape was used, which extracted the top 100 genes with the selected four scoring methods of CytoHubba, namely Maximal Clique Centrality (MCC), Maximum Neighborhood Component (MNC), Edge Percolated Component (EPC), and node-connect degree, respectively.

Transcription factor, resistance genes, and CAZyme genes identifications

To recognize and identify the functions of the significantly expressed genes, Biomart, ShinyGo, and UniProt were employed; the information was generated and further verified through BLASTx against PlantTFdb and the PRGdb version 3.0 for the recognition of genes encoding for resistance, transcription factors, and specific pathways related to them. The DRAGO (PRGdb tool) pipeline also classified the R-gene classes and domain. In contrast, dbCAN2 with default parameters integrated with automated annotation tools of Diamond, HMMER, and Hotpep enabled the identification of the polysaccharide degradation enzymes, also known as CAZymes.

Conclusion

Transcriptome data analysis has become an indispensable tool for researchers in various fields, including biology, medicine, and agriculture. By leveraging the power of bioinformatics tools and techniques, scientists can gain a deeper understanding of gene expression, cellular processes, and the underlying mechanisms of diseases.

Funding

Not applicable.

Ethical Approval

Not applicable.

Conflict of Interest

Not applicable.

Informed Consent

Not applicable.

Acknowledgments

The authors thank the Department of Bioinformatics, Pondicherry University, and NIMS University for providing all the necessary infrastructure for writing the review.

References

1. Anders S, Pyl PT, Huber W (2015) HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31: 166–169.
2. Athanasopoulou K, Boti MA, Adamopoulos PG, Skourou PC, Scorilas A (2021) Third-generation sequencing: The spearhead towards the radical transformation of modern genomics. *Life* 12: 30.
3. Bahaian A, Thompson IR, Lever J, Gagnier L, Karimi MM, et al. (2019) LIONS: analysis suite for detecting and quantifying transposable element initiated transcription from RNA-seq. *Bioinformatics* 35: 3839–3841.
4. Bayley H (2015) Nanopore sequencing: FROM imagination to reality. *Clin Chem* 61: 25–31.
5. Benovoy D, Kwan T, Majewski J (2008) Effect of polymorphisms within probe–target sequences on oligonucleotide microarray experiments. *Nucleic Acids Res* 36: 4417–4423.
6. Bushmanova E, Antipov D, Lapidus A, Prjibelski AD (2019) RNA spades: a de novo transcriptome assembler and its application to RNA-Seq data. *Gigascience* 8: giz100.
7. Chaudhary S, Khokhar W, Jabre I, Reddy AS, Byrne LJ, et al. (2019) Alternative splicing and protein diversity: plants versus animals. *Front Plant Sci* 10: 708.
8. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21.
9. Drewe P, Stegle O, Hartmann L, Kahles A, Bohnert R, et al. (2013) Accurate detection of differential RNA processing. *Nucleic Acids Res* 41: 5189–5198.