



Analysis of Transcriptome and Epitranscriptome in Plants Using PacBio Iso-Seq and Nanopore-Based Direct RNA Sequencing

Liangzhen Zhao¹, Hangxiao Zhang¹, Markus V. Kohnen¹, Kasavajhala V. S. K. Prasad², Lianfeng Gu^{1*} and Anireddy S. N. Reddy^{2*}

OPEN ACCESS

Edited by:

Jiannis Ragoussis,
McGill University, Canada

Reviewed by:

Milind Ratnaparkhe,
ICAR Indian Institute of Soybean
Research, India
Pawel P. Labaj,
University of Natural Resources
and Life Sciences Vienna, Austria

*Correspondence:

Lianfeng Gu
lfgu@fafu.edu.cn
Anireddy S. N. Reddy
Anireddy.Reddy@colostate.edu;
reddy@colostate.edu

Specialty section:

This article was submitted to
Genomic Assay Technology,
a section of the journal
Frontiers in Genetics

Received: 15 October 2018

Accepted: 06 March 2019

Published: 21 March 2019

Citation:

Zhao L, Zhang H, Kohnen MV,
Prasad KVS, Gu L and Reddy ASN
(2019) Analysis of Transcriptome
and Epitranscriptome in Plants Using
PacBio Iso-Seq and Nanopore-Based
Direct RNA Sequencing.
Front. Genet. 10:253.
doi: 10.3389/fgene.2019.00253

¹ Basic Forestry and Proteomics Research Center, College of Forestry, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Fujian Agriculture and Forestry University, Fuzhou, China, ² Program in Cell and Molecular Biology, Department of Biology, Colorado State University, Fort Collins, CO, United States

Nanopore sequencing from Oxford Nanopore Technologies (ONT) and Pacific BioSciences (PacBio) single-molecule real-time (SMRT) long-read isoform sequencing (Iso-Seq) are revolutionizing the way transcriptomes are analyzed. These methods offer many advantages over most widely used high-throughput short-read RNA sequencing (RNA-Seq) approaches and allow a comprehensive analysis of transcriptomes in identifying full-length splice isoforms and several other post-transcriptional events. In addition, direct RNA-Seq provides valuable information about RNA modifications, which are lost during the PCR amplification step in other methods. Here, we present a comprehensive summary of important applications of these technologies in plants, including identification of complex alternative splicing (AS), full-length splice variants, fusion transcripts, and alternative polyadenylation (APA) events. Furthermore, we discuss the impact of the newly developed nanopore direct RNA-Seq in advancing epitranscriptome research in plants. Additionally, we summarize computational tools for identifying and quantifying full-length isoforms and other co/post-transcriptional events and discussed some of the limitations with these methods. Sequencing of transcriptomes using these new single-molecule long-read methods will unravel many aspects of transcriptome complexity in unprecedented ways as compared to previous short-read sequencing approaches. Analysis of plant transcriptomes with these new powerful methods that require minimum sample processing is likely to become the norm and is expected to uncover novel co/post-transcriptional gene regulatory mechanisms that control biological outcomes during plant development and in response to various stresses.

Keywords: SMRT isoform sequencing, nanopore direct RNA sequencing, RNA modification, alternative splicing, alternative polyadenylation, epitranscriptome

INTRODUCTION

Analysis of transcriptomes, which represent the activity of genes in the genome, is vital for understanding the relationship between genotype and phenotype. The dynamics and complexity of transcriptome regulate all aspects of plant growth, development, and responses to various external biotic and abiotic cues. Different methods such as expressed sequence tag (EST) sequencing (Wu et al., 2002), serial analysis of gene expression (SAGE) (Matsumura et al., 1999), DNA microarray (Hihara et al., 2001), and recently RNA sequencing (RNA-Seq) using next-generation sequencing (NGS) technologies (Mortazavi et al., 2008) have been developed to analyze transcriptomes. Since 2005, second-generation short-read sequencing platforms quickly replaced first-generation Sanger sequencing technology for various high-throughput applications due to lower costs and greater sequencing depth (Sedlazeck et al., 2018). However, the read length is the major limitation in second-generation short-read sequencing, which made it harder to analyze several aspects of co/post-transcriptional processing events. To overcome this limitation, in the past few years, researchers are sequencing full-length transcripts mostly using two platforms, Pacific BioSciences (PacBio) (Rhoads and Au, 2015) and Oxford Nanopore Technologies (ONT) (Bayega et al., 2018), which are referred to as “third” and “fourth” generation sequencing technologies, respectively (Slatko et al., 2018). These two platforms increased read length considerably as compared to other NGS methods and can, therefore, be used to address a larger variety of research questions. Single-molecule real-time (SMRT) isoform sequencing (Iso-Seq) using PacBio platform captures the full length of transcripts (Gonzalez-Garay, 2016) and thereby presents easier and more accurate ways for different applications, such as gene annotation (Zhao et al., 2018), isoform identification (Abdel-Ghany et al., 2016; Wang T. et al., 2017), identification of fusion transcripts (Weirather et al., 2015), and long non-coding RNA (lncRNA) discovery (Li et al., 2016). Here, we discuss applications and broader utility of PacBio and ONT in transcriptome studies. Recently developed direct RNA-Seq using nanopore can avoid amplification biases (Garalde et al., 2018). Furthermore, this technology has the potential to provide a complete view of RNA modifications such as N⁶-methyladenosine, 5-methylcytidine, and 5-hydroxymethylcytidine (Li X. et al., 2017), which are collectively referred to as the “epitranscriptome.”

Parts of the core algorithm for PacBio and ONT long-read analyses are similar to short-read analysis strategies used in second-generation sequencing approaches. Nevertheless, specific new bioinformatics tools have been designed for several of the applications, which have not been part of second-generation sequencing pipelines. These tools are needed to provide greater flexibility to achieve different goals as well as to address new issues, such as higher error rates and low throughput. We present currently available bioinformatics methods for PacBio and ONT read analysis, including reads-of-interest (ROI) extraction, error correction (Au et al., 2012), mapping (Wu and Watanabe, 2005), isoform clustering (Fu et al., 2012), and identification of multiple transcript isoforms (Abdel-Ghany et al., 2016). Improvements in

these new methods and computational pipelines will expand the landscape of transcriptome complexity at the transcript isoform and epitranscriptome level with higher throughput and higher accuracy. Here, we discussed PacBio Iso-Seq and ONT direct RNA-Seq methodologies, the current status of bioinformatics tools used to analyze the long-reads and highlighted various applications of these methods.

LIBRARY PREPARATION AND EXTRACTION OF READ-OF-INSERT FROM PACBIO ISO-SEQ

Generally, high-quality RNA is poly(A) selected to construct PacBio long-read sequencing libraries using, e.g., Clontech SMARTer PCR kit (Ramsköld et al., 2012; Li et al., 2016). The length of sequencing reads is dependent on the quality of RNA and generation of full-length cDNAs. To enrich for full-length cDNAs in the library, cap-dependent linker ligation method has been used (Cartolano et al., 2016). Alternatively, full-length RNAs can be enriched by combining poly(A)⁺ RNA selection with capturing of 5' capped mRNAs using a cap-binding protein (Blower et al., 2013). Full-length mRNA is then used for first-strand cDNA synthesis with oligo (dT) primer followed by second-strand cDNA synthesis with a size selection of full-length cDNA in several different sizes (Xu et al., 2015). With the new Sequel system, cDNAs can be sequenced without size selection. By ligating hairpin adaptors to double-stranded cDNA, SMRTbellTM libraries are generated which can be subsequently sequenced on either the RSII or Sequel platform (Xu et al., 2017). Comparison of 5' ends with annotated transcript start sites shown that this protocol enables full-length cDNA sequencing with little loss of 5' or 3' ends (Ramsköld et al., 2012).

At present, PacBio offers two fourth-generation sequencers: the RSII was the first commercially available sequencing instrument and the recently improved Sequel device provides much higher throughput (up to 20 Gb per SMRT Cell). PacBio's sequencing strategy is based on the usage of zero-mode waveguide (ZMW) technology, which consists of tiny nano-wells initially described in 2003 (Levene et al., 2003). The ZMWs allow the immobilization of sequencing templates through the interaction with the sequencing engine, a polymerase enzyme complex, which is affixed at the bottom of ZMWs (Rhoads and Au, 2015). Then the incorporation of fluorescent-labeled DNA bases emits fluorescent signals that are captured by a detector in real time (McCarthy, 2010). Hairpin adaptors that are added to both ends of double-stranded DNA during library preparation generate a closed circular DNA template, which could be repeatedly traversed by long lifetime polymerase to improve the accuracy. In this way, PacBio platform could generate multiple subreads including adapter sequences in a single ZMW and yield a continuous long read (CLR), which can generate more accurate circular consensus sequence (CCS) reads (Weirather et al., 2017).

Subsequently, the RSII system and the Sequel system store the base-call data and associated quality metrics in HDF5 and BAM

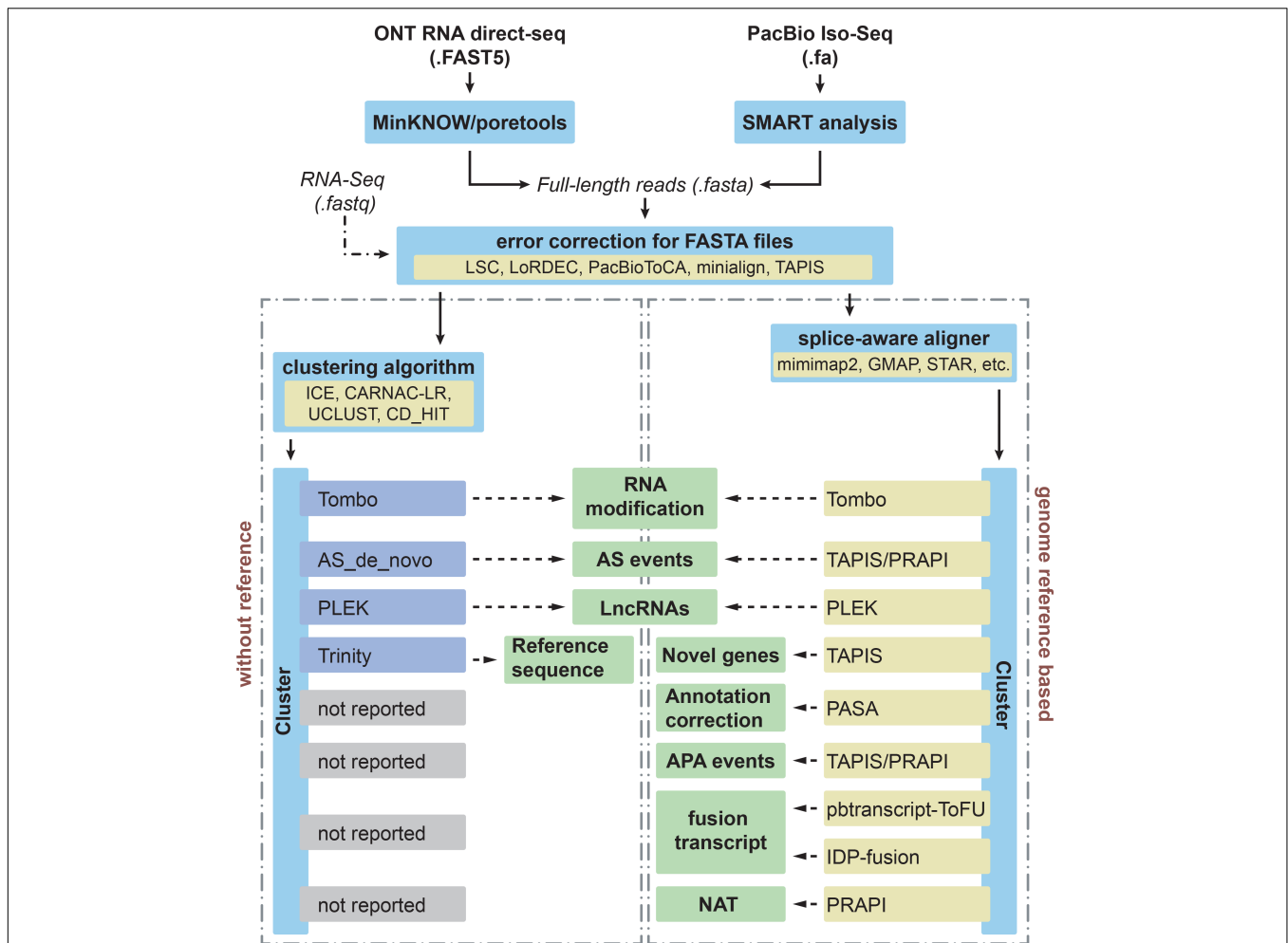


FIGURE 1 | Different applications and bioinformatics solutions for PacBio Iso-Seq and Nanopore direct RNA sequencing in plants. Iso-seq and direct RNA sequencing can be performed using the SMRT analysis from PacBio and MinKNOW from ONT, respectively. Then these long reads with FASTA format will undergo error correction step before downstream analysis. In order to deal with different applications (middle boxes), computational tools used to process and analyze long-reads for each application are indicated for species without and with reference genomes in left and right boxes, respectively.

files format, respectively. The bax2bam tool can convert HDF5 file format into BAM format¹.

The SMRT Analysis module from SMRT Link from PacBio is adopted for obtaining effective subreads (Figure 1). Then extraction of ROI for each ZMW is the second step in PacBio Iso-Seq bioinformatics analysis workflow. This step is performed with the SMRT Link pipeline, which includes steps for trimming adapters and generating CCSs. Then ROIs are cleaned of polyA/T tails, primers, artificial concatemers, and transcript strand direction is identified (Bayega et al., 2018). ToFu Pacbio pipeline from SMRT Analysis package can be used to search for sequencing adapters for extracting ROI and full-length non-chimeric (FLNC) reads (Wang T. et al., 2017; Xu et al., 2017). Afterward, the FLNC reads, which contain both 5' and 3' primers and poly-A tail, can be analyzed using iterative clustering for error correction (ICE) to build consensus clusters to improve

consensus accuracy. Subsequently, PacBio RS II and Sequel use Quiver and Arrow to polish consensus sequences, respectively (Bayega et al., 2018).

LIBRARY CONSTRUCTION AND BASE-CALLING FOR NANOPORE DIRECT RNA SEQUENCING

The starter pack for direct RNA-Seq costs only \$1000 (pricing as of January 2019), which includes one MinION sequencer, two flow cells, one sequencing kit, and a wash kit². Compared to NGS or PacBio, the MinION is portable (weighs 90 g), real-time, long-read, and low-cost device. It is also possible to use the SMARTer protocol for full-length cDNA synthesis (Ramsköld et al., 2012), which includes end-repair, dA-Tailing,

¹<https://github.com/PacificBiosciences/PacBioFileFormats/wiki/BAM-recipes>

²<https://nanoporetech.com>

and adapter ligation. However, this kind of library construction will remove all RNA modification information during double-strand cDNA synthesis. The Nanopore direct RNA library construction workflow uses poly(dT) adapter and SuperScript III Reverse Transcriptase to generate RNA–DNA hybrids, which are subsequently ligated to nanopore sequencing adapters using T4 DNA ligase prior to sequencing. Then Agencourt RNAClean XP magnetic beads are used to purify RNA–DNA duplexes³. After estimating the sample concentration, the Nanopore direct RNA library can be loaded into flow cells using MinION, GridION, or PromethION sequencer. Compared to the MinION, the GridION and PromethION sequencers provide higher throughput. The motor protein pulls the 3' end of the RNA strand inside the nanopore channel (Bayega et al., 2018). Then changes in the ionic current are detected at each pore by a sensor.

Prior to sequencing, the computer hardware should be checked to meet the minimum requirement. The minimal requirements for MinION are: CPU: i7 or Xeon with 4+ cores, memory: 16 GB RAM, storage: 1 TB internal SSD, ports: USB3⁴. Any computer with above minimal requirements can run a MinION without deterioration of performance during sequencing. Desktop or laptop computer with a MinKNOW and EPI2ME Desktop Agent installation provided by Oxford Nanopore and Metrichor Ltd., respectively, need to be connected with MinION (Figure 1). MinKNOW controls the MinION device, tests hardware, checks flow cells, and performs sequencing runs. EPI2ME further analyzes the raw electrical signals generated and stores in FAST5 files, which serve as input for Metrichor for base-calling. Then FASTQ and FASTA sequences can be extracted from FAST5 files using poretools (Loman and Quinlan, 2014). However, detection tools to identify base modifications are currently not available. The majority of the applications using Nanopore direct RNA-Seq have been focused on mammals. At present, Nanopore direct RNA-Seq has not been reported for studies on plants. However, it is anticipated that more and more laboratories will use this tool to study RNA modifications in plants.

LONG-READ PREPROCESSING: ERROR CORRECTION, MAPPING, AND CLUSTERING OF LONG-READS

Although the length of PacBio and ONT reads is longer than NGS, one common concern regarding these technologies is high error rates (Koren et al., 2012). Thus, it is necessary to reduce the error rate before subsequent utilization. At present, correcting PacBio and ONT reads fall into three distinct categories: hybrid error correction strategy, self-correction method, and reference-based error correction.

Hybrid error correction strategy uses short reads from NGS to correct long reads. LSC (Au et al., 2012), LoRDEC (Salmela and Rivals, 2014), and PacBioToCA (Koren et al., 2012) are three widely used methods for error correction (Figure 1). Unlike LSC

and PacBioToCA, LoRDEC avoids mapping of short reads by building short reads De Bruijn graph (DBG) of order k and threads the long reads through this short reads DBG to correct. Thus LoRDEC requires less time/memory and less disk space (Salmela and Rivals, 2014). Recently, Nanocorr was developed specifically to correct Nanopore long reads using high-quality short reads (Goodwin et al., 2015).

Alternatively, self-correction software is distinct from the above hybrid error correction strategy, which depends on short-reads. Long-read multiple aligner (LoRMA) is one of the methods for error correction that relies only on long reads (Salmela et al., 2016). Compared to another self-correction method PacBio corrected reads (PBcR) algorithm (Koren et al., 2012), LoRMA achieved higher throughput and lower error rate. However, self-correction method needs a high coverage in order to obtain accurate correction, which limits its application.

The third method provides reference-based error correction during alignment of long reads to reference genome and some tools that do this type of error correction are minimap2 (Li, 2018) and minialign⁵. These are fast and accurate alignment tools for PacBio and Nanopore long reads with high insertion and deletion error rate. Transcriptome Analysis Pipeline from Isoform Sequencing (TAPIS) (Abdel-Ghany et al., 2016) also performs reference-based error correction. In addition to minimap2, GMAP (Wu and Watanabe, 2005) and STAR (Dobin et al., 2013) are two splice-aware aligners, which can be used for mapping full-length reads to reference genome for downstream analysis. However, GMAP and STAR do not perform error correction during mapping. In addition to canonical splice sites, GMAP and STAR capture non-canonical splice sites, hence should be cautious during downstream AS analysis. Compared to GMAP, minimap2 is more consistent with existing annotation and works well with noisy reads (Li, 2018).

Highly expressed genes could generate multiple identical isoforms, which would take more time for downstream processing/analyses and are hard to visualize without collapsing redundant reads. Clustering step could group full-length reads into a cluster, which is a necessary step to further improve quality and identify unique splicing isoforms. After mapping Iso-Seq to reference genome, Cupcake ToFU could be used to collapse redundant isoforms and obtain unique isoforms⁶. The majority of clustering strategies used for species without reference genome have been developed for ESTs, which appeared before the age of PacBio and ONT. Clustering programs designed for ESTs, such as UCLUST (Edgar, 2010) and CD-HIT (Fu et al., 2012), are widely used to group and collapse redundant sequences. However, these methods were not designed for full-length sequences with high error rates as compared to ESTs or short reads from NGS. At present, there are two *de novo* algorithms for clustering of long reads by genes: the ICE algorithm (Gordon et al., 2015) can cluster FLNC reads from PacBio sequencing to generate consensus isoforms and the CARNAC-LR algorithm designed for ONT long-read sequencing data (Marchet et al., 2018). After collapsing the redundant isoforms, the read count information

³<https://nanoporetech.com>

⁴<https://community.nanoporetech.com>

⁵<https://github.com/ocxtal/minialign>

⁶https://github.com/Magdoll/cDNA_Cupcake

for expression levels would be lost. If expression level analysis needs to be performed, one can go back and retrieve the read counts from the original sequencing files.

APPLICATIONS AND BIOINFORMATICS TOOLS FOR ISO-SEQ AND NANOPORE DIRECT RNA SEQUENCING IN PLANTS

At present, PacBio and ONT deep sequencing are increasingly used for genome annotation, identification of co/post-transcriptional events and fusion transcripts. Recently, several studies collected and reanalyzed long reads from Iso-Seq into comprehensive databases such as Plant ISOform sequencing database (PISO) (Feng et al., 2019) and ISOdb (Xie et al., 2018). ISOdb and PISO deposited 8 and 19 species, respectively. Since the new technology has a higher resolution than second-generation sequencing and detects modified RNA bases, additional aspects of transcriptional and post-transcriptional regulation can be studied more comprehensively. Therefore, we highlight bioinformatics solutions and various applications that are difficult to investigate using NGS.

DE NOVO GENOME ANNOTATION, NEW LOCUS IDENTIFICATION, AND GENE MODEL CORRECTION

For species without an available reference genome, such as *Drynaria roosii* (Sun et al., 2018) and *Asparagus officinalis* (Kakrana et al., 2018), Iso-Seq was successfully used recently to capture the complete and full-length transcriptome. Due to the longer reads from PacBio and ONT, Iso-Seq has proven to be more advantageous in resolving many complex features in transcriptomes when compared to short-read RNA-Seq, which depends on software for reconstructing transcript sequences (Haas et al., 2013; Steijger et al., 2013). Thus, one key advantage of long-reads from PacBio and ONT was to accurately infer gene models by generating full-length transcripts without further assembly, which is challenging for complex isoforms (Gordon et al., 2015). The utility of long-read transcripts in inferring gene models has been reported in medicinal herb *Panax ginseng* (Jo et al., 2017; Kim et al., 2018), allohexaploid wheat (Clavijo et al., 2017), bread wheat (Cartolano et al., 2016), sugar beet (Minoche et al., 2015), the coffee bean (Cheng et al., 2017), and Para rubber tree (Pootakham et al., 2017). Full-length transcripts generated by Iso-Seq are ideal for improving gene model prediction and identification of novel genes, which do not map to annotated gene loci. For example, recent studies revealed 2171 novel genes in *Sorghum bicolor* (Abdel-Ghany et al., 2016), 8091 in *Phyllostachys edulis* (Wang T. et al., 2017), and 3026 in *Triticum aestivum* (Gordon et al., 2015). Also in *Populus trichocarpa* (Filichkin et al., 2018), allopolyploid cotton (Wang et al., 2018), and *Populus "Nanlin 895"* (Chao Q. et al., 2018), 15,087, 13,551, and 1575 novel transcribed regions, respectively, were recently identified. In addition to isoform and new locus identification, Iso-Seq has been used to refine gene models in *Vitis vinifera*

cv. Cabernet Sauvignon (Minio et al., 2018) and allopolyploid cotton (Wang et al., 2018). Furthermore, recent studies corrected 178 and 2241 annotated genes, which covered more than one transcript assemblies in *S. bicolor* (Abdel-Ghany et al., 2016) and *P. edulis* (Wang T. et al., 2017), respectively. Program to Assemble Spliced Alignments (PASA) is one bioinformatics tool that corrects such gene annotations (Haas et al., 2008). Recently, long-read annotation (LoReAn) pipeline used a combination of PacBio SMRT or MinION long-reads and other information such as protein evidence for gene annotation (Cook et al., 2019).

CHARACTERIZATION OF ALTERNATIVE TRANSCRIPTION INITIATION, ALTERNATIVE POLYADENYLATION, AND ALTERNATIVE SPLICING

Alternative transcription initiation (ATI), alternative cleavage and alternative polyadenylation (APA), and alternative splicing (AS) events are three major processes that contribute to transcriptome diversity. AS of precursor mRNAs (pre-mRNAs) can potentially increase the number of protein isoforms produced from multiexon genes and regulate gene expression through multiple mechanisms such as altered translational efficiency of splice isoforms, non-sense-mediate decay, and miRNA-mediated mRNA degradation (Reddy et al., 2013). Though individual AS events can be quantified and annotated using NGS with great accuracy, it is hard to deduce full-length splicing isoforms that contain a combination of these individual AS events (Steijger et al., 2013). Long-read sequencing provides the possibility to obtain full-length sequences and thus identify complex splice isoforms, which are hard to detect and reconstruct by NGS. Iso-Seq has allowed identification of over 110,000 non-redundant isoforms in *Zea mays* (Wang et al., 2016), >42,000 in *P. edulis* (Wang T. et al., 2017), and >16,000 in *Salvia miltiorrhiza* (Xu et al., 2015). Additionally, Iso-Seq identified 29,730 novel isoforms in *Trifolium pratense* L., 2501 new alternative transcripts in *V. vinifera* cv. Cabernet Sauvignon (Minio et al., 2018), and over 11,000 novel splice isoforms in *S. bicolor* L. Moench (Abdel-Ghany et al., 2016). For 35.74% of the unigenes of bermudagrass, three or more distinct isoforms were identified using Iso-Seq (Zhang B. et al., 2018). In the wild strawberry *Fragaria vesca*, Iso-Seq revealed that pre-mRNAs from ~58% of multiexon genes are alternatively spliced (Li Y. et al., 2017).

In addition to the full-length isoform detection, AS events can be classified into five different types: retained intron (RI), skipped exon (SE), alternative 5' splicing site (A5SS), alternative 3' splicing site (A3SS), and mutually exclusive exons (Shen et al., 2014). In addition to above five common categories, many other complex types, such as alternative position, i.e., alternative 3' and 5' site (Wang and Brendel, 2006), AS and transcriptional initiation (ASTI) (Nagasaki et al., 2006) alternative first exons (Chen et al., 2007), and composite patterns (Wang and Rio, 2018), can occur. Although NGS can detect these AS events, long reads from PacBio and ONT

provide an advantage on detecting AS events because long-read sequencing could avoid any possible issues during transcriptome reconstruction. For example, Iso-Seq revealed 10,053, 172,743, 133,229, and 21,154 AS events in *S. bicolor* (Abdel-Ghany et al., 2016), *Z. mays* (Wang et al., 2016), allopolyploid cotton (Wang et al., 2018), and *P. edulis* (Wang T. et al., 2017), respectively.

Alternative polyadenylation has multiple regulatory roles in RNA transportation, localization, stability, and translation by producing isoforms with different 3' cleavage sites, which generates transcript diversity and complexity (Tilgner et al., 2015; Abdel-Ghany et al., 2016; Wang T. et al., 2017). For APA identification using NGS, Poly(A) Site Sequencing (PAS-Seq) libraries can be constructed using degenerate nucleotides in combination with oligo(T) primers (Shepard et al., 2011; Zhang et al., 2015). Internal priming issue was defined as cDNA primers hybridizing to internal continuous As instead of the actual poly(A) tail (Beaudoing et al., 2000). If six continuous As or more than seven As existed in a 10 nt window, it was internal priming candidate (Tian et al., 2005). PAS-Seq based on NGS methods could not avoid the internal priming because internal A-rich sequences could prime the oligo(dT) (Nam et al., 2002; Sherstnev et al., 2012). Both Iso-Seq and Nanopore direct RNA-Seq methods could avoid internal priming. Using Iso-Seq, 7700 genes containing two or more polyadenylation sites have recently been detected in *S. bicolor* (Abdel-Ghany et al., 2016). In allopolyploid cotton, 6935 genes have at least five poly(A) sites (Wang et al., 2018). At present, quantification analysis of APA still depends on NGS due to the low sequence depth of Iso-Seq and Nanopore direct RNA-Seq. A recent study in *P. edulis* used a method that combined NGS with Iso-Seq to identify 1224 differential APA sites (Wang T. et al., 2017). In the future, it is expected that both Iso-Seq and Nanopore direct RNA-Seq can be used for quantification analysis once the throughput increases.

Alternative transcription initiation is another key mechanism to generate diverse transcripts (Tanaka et al., 2009). Alternative usage of transcription start sites attracted little attention in plants as compared to the studies on AS and APA. Paired-end analysis of transcription start sites (PEAT) strategy, which requires complex library construction, following NGS has been used for monitoring global transcription start site usage (Ni et al., 2010). Using the PEAT protocol, millions of transcription start sites that fall into three categories have been identified in Arabidopsis roots (Morton et al., 2014). Since PacBio Iso-Seq and Nanopore direct RNA-Seq can sequence full-length transcripts from 5' ends to polyadenylated tails, it would be a perfect tool to detect ATI.

For traditional RNA-Seq, the identification of the major AS events, including exon skipping events, intron retention, alternative 5' donor, and alternative 3' donor usage is quite simple by using several tools, including rMATS (Shen et al., 2014), JUM (Wang and Rio, 2018), PASA pipeline (Campbell et al., 2006), and ASTALAVISTA (Foissac and Sammeth, 2007). For the analysis of post-transcriptional regulation based on long-read sequencing, TAPIS pipeline (Abdel-Ghany et al., 2016) and PRAPI (Gao et al., 2017) are two main bioinformatics tools that use Iso-Seq reads to

identify AS and APA (Figure 1). In addition, PRAPI (Gao et al., 2017) can also identify several other events/processes, such as ATI, and production of circular RNAs (circRNAs).

IDENTIFICATION OF FUSION TRANSCRIPTS

Fusion transcripts are the result of a trans-splicing event (Li et al., 2008) that joins two separately encoded pre-RNAs into one transcript. Fusion transcripts have been identified in diverse plant species (Zhang et al., 2010; Wang et al., 2016). Paired-end RNA-Seq datasets based on NGS have been successfully analyzed for fusion transcript (Maher et al., 2009). Recently, Iso-Seq provided a more reliable way to identify fusion transcripts. In total, 1430 fusion transcripts had been detected in *Z. mays* using Iso-Seq (Wang et al., 2016). Furthermore, 3762 and 222 fusion transcripts were identified in *T. pratense* L (Chao Y. et al., 2018) and allopolyploid cotton (Wang et al., 2018), respectively.

The standard for fusion transcript identification is based on the simple idea that two or more fragments from one transcript can be mapped to several loci (Wang et al., 2016). Multiple fusion transcript detection algorithms based on NGS have been developed (Liu S. et al., 2015). However, these algorithms were specially designed for paired-end RNA-Seq data. PacBio pbtranscript-ToFU package provides a script to detect fusion transcripts (fusion_finder.py)⁷, which is specially designed for reads from Iso-Seq. Isoform Detection and Prediction (IDP) fusion (Figure 1) also presents another algorithm to detect fusion events using both PacBio long-read sequencing and NGS (Weirather et al., 2015).

lncRNA IDENTIFICATION

Long ncRNAs are defined as RNAs with more than 200 nt and have no discernable coding potential (Jin et al., 2013). In plants, lncRNAs can be generated from intergenic, intronic, or coding regions and play an important role in gene regulation (Wang and Chekanova, 2017). The majority of lncRNAs are polyadenylated in plants, thus RNA-Seq on Illumina platforms can also detect the expression of lncRNAs. However, recent studies showed that lncRNAs undergo complex post-transcriptional regulation (Liu J. et al., 2015). Thus, full-length sequencing provides a great advantage in identifying gene model of lncRNAs. Recently, several studies reported the identification of lncRNAs using Iso-Seq in plants. For example, PacBio Iso-Seq revealed 1187 and 4333 lncRNAs in poplar "Nanlin 895" (Chao Q. et al., 2018) and *T. pratense* L. (Chao Y. et al., 2018), respectively. These studies suggested that Iso-Seq is a well-suited method for identification of lncRNAs. GREEEN and CANTATAdb are two resources to search for sequence homology of lncRNAs from long reads, which have been reported in *P. edulis* (Wang T. et al., 2017). Also, long reads containing sequence

⁷<https://github.com/PacificBiosciences>

homology to miRNAs could also be regarded as non-coding RNA, as has been reported in *S. bicolor* (Abdel-Ghany et al., 2016). In *Z. mays*, lncRNAs were identified using PLEK, a classification model trained on known high-confidence lncRNAs (Wang et al., 2016).

NATURAL ANTISENSE TRANSCRIPTS IDENTIFICATION

Natural antisense transcripts (NATs) including head-to-head, tail-to-tail, and fully overlapping types have been shown to function in transcriptional and post-transcriptional gene regulation (Faghihi and Wahlestedt, 2009). In total, 932 *cis*-NATs were identified using a strand-specific PacBio SMRT dataset by performing pair-wise comparisons of overlapping coordinates from oppositely oriented full-length transcripts (Zhang H. et al., 2018). Furthermore, PRAPI was developed to identify NAT based on PacBio/ONT long reads (Gao et al., 2017). At the same time, PRAPI can also quantify the expression of NAT by combining NGS reads using strand-specific library construction (**Figure 1**).

ANALYSIS OF LONG-READS IN THE ABSENCE OF A REFERENCE GENOME

Due to recent developments in long-read sequencing, more and more genome sequencing studies are using long-read sequencing platforms to obtain longer reads than N50, such as *de novo* assembling of grass *Oropetium thomaeum* (VanBuren et al., 2015), sunflower (Badouin et al., 2017), and citrus (Wang X. et al., 2017). However, there are still many species without available genome sequences. Thus, it will be valuable to develop reference-free analyses for transcription annotation using Trinity (Haas et al., 2013) and other tools for post-transcriptional analysis. Recent studies have shown that it is feasible to reconstruct full-length transcript models for species without a reference genome, such as *Astragalus membranaceus* (Li J. et al., 2017), *Arabidopsis pumila* (Yang et al., 2018), and *Zanthoxylum bungeanum* Maxim (Tian et al., 2018) using long reads.

Recently, AS_*de_novo*⁸ has reported AS identification based on Iso-Seq without reference genomes (Liu et al., 2017). The basic idea originated from searching for the deletion or insertion in the clustering units (Ner-Gaon et al., 2007; Zhou et al., 2011; Wu et al., 2014; Liu et al., 2017). Thus, clustering long reads from PacBio Iso-Seq or ONT should be the first step before AS identification. Several clustering programs, such as the widely used CD-HIT, can be used for this analysis (Fu et al., 2012). Recently, one clustering approach designed for Oxford Nanopore long reads has been released (Marchet et al., 2018). After the clustering step, all-vs-all BLAT comparison can be used for the identification of insertion or deletion segmentation caused by AS events (Liu et al., 2017). Hybrid sequencing and map finding (HySeMaFi) combined PacBio Iso-Seq and NGS to identify

splicing and quantify the isoforms abundance (Ning et al., 2017). AStrap adopted machine-learning model to identify AS events by integrating more than 500 assembled features (Ji et al., 2018).

THE APPLICATION OF NANOPORE DIRECT RNA SEQUENCING

Since full-length native RNA-Seq (nRNA-Seq) of ONT provides multiple benefits compared to NGS, this method has been applied for detecting viral transcriptomes (Moldován et al., 2018a), 16S rRNA base modifications (Smith et al., 2017), viral pathogen (Depledge et al., 2018), and identification of artifactual splice isoforms during reverse transcription due to the template switching (Moldován et al., 2018b). Finally, a significant advantage of direct RNA-Seq is that it allows detection of co/post-transcriptional base modifications in RNA since it does not require reverse transcription and PCR amplification steps. Many reversible chemical modifications of bases occur in mRNAs, which are collectively referred to as the “epitranscriptome” (Gilbert et al., 2016). These covalent reversible chemical modifications of nucleotides regulate many aspects of gene expression. Recent studies indicate that epitranscriptomic modifications are key players in regulating pre-mRNA splicing, nuclear export, mRNA stability and localization, and translation efficiency (Gilbert et al., 2016; Xiao et al., 2016; Roundtree et al., 2017; Slobodin et al., 2017) and also several developmental processes in plants (Fray and Simpson, 2015; Vandivier and Gregory, 2018). There is no simple high-throughput tool to detect mRNA modifications and their dynamics in plants. A widely used method for transcriptome-wide analysis of RNA modifications is challenging as it requires specific antibodies for each modification. These antibodies are then used to precipitate RNA with modifications, which is then subjected to high-throughput sequencing (**Figure 2**). This method has been used to identify transcriptome-wide m⁶A localization and abundance in animals (Dominissini et al., 2012; Meyer et al., 2012). In *Arabidopsis thaliana*, a transcriptome-wide 6-methyladenine (m⁶A) and 5-methylcytosine (m⁵C) profiles were reported using the m⁶A- or m⁵C-targeted antibodies, respectively, for RNA immunoprecipitation (RIP) followed by high-throughput sequencing (m⁶A-Seq/m⁵C-Seq) (Luo et al., 2014; Cui et al., 2017). This RIP-Seq approach has several limitations including the need for specific antibody for each modification. It is also time-consuming and laborious. Furthermore, it is difficult to obtain a sufficient amount of immunoprecipitated RNA. More importantly, this method does not provide the precise location of the modified base. Recently, it has been shown that RNA modifications can be detected using Oxford Nanopore direct RNA-Seq (Garalde et al., 2018).

The library construction protocol for direct RNA-Seq was designed for poly(A) transcripts (Garalde et al., 2018). Steps involved in native RNA-Seq are illustrated in **Figure 3**. Although single-stranded RNA is depicted in this figure, RNA-DNA hybrid can be used for direct RNA-Seq where only the RNA strand in the hybrid is sequenced. The use of RNA-DNA hybrids may alleviate some issues associated with RNA secondary structures and

⁸https://github.com/liuxiaoxian/IsoSeq_AS_de_novo

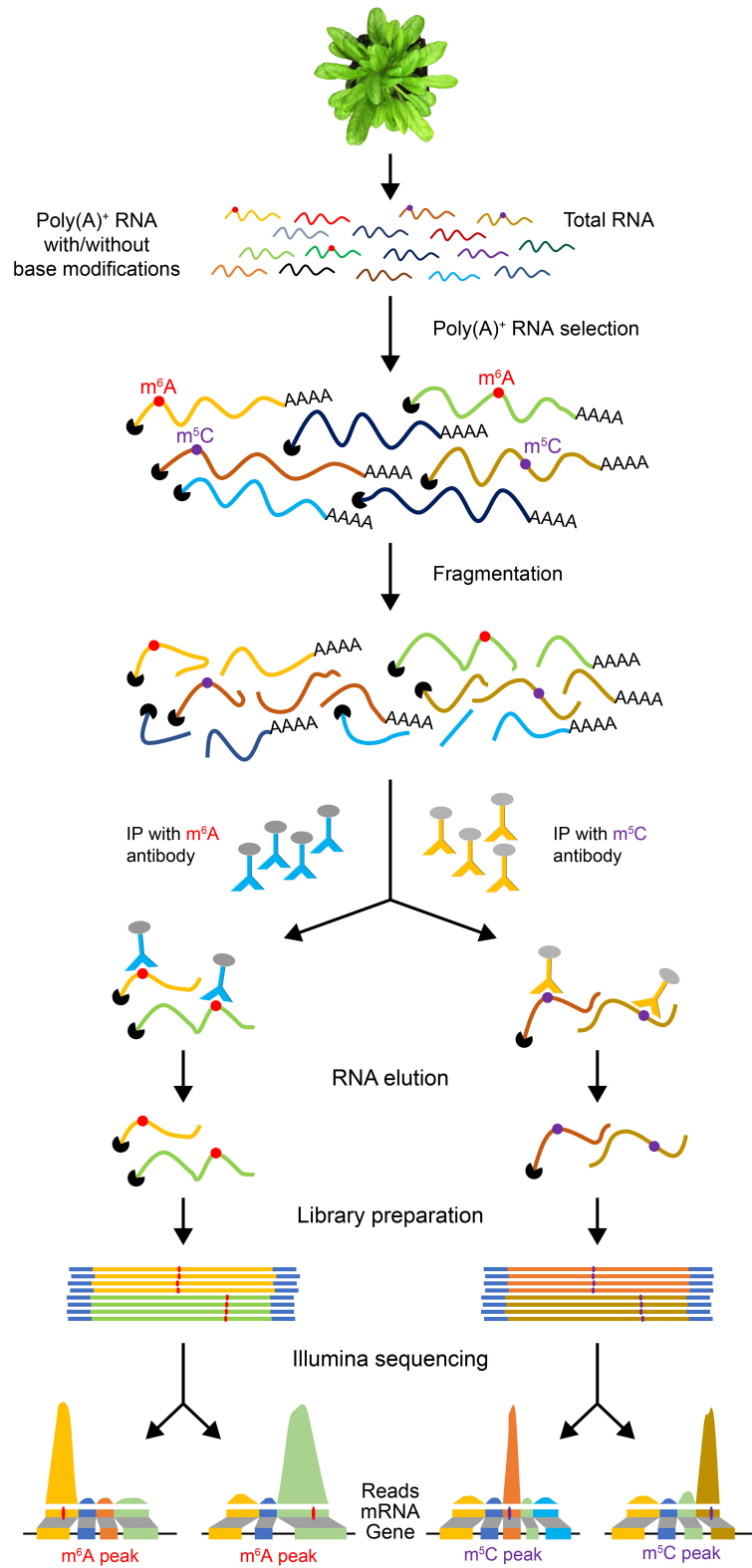


FIGURE 2 | An illustration of epitranscriptome analysis using antibodies to identify RNAs with base modifications. Poly(A)⁺ mRNA is used for RNA immunoprecipitation with antibodies specific to a base modification (e.g., m⁶A or m⁵C). The IP'ed RNA is then used to generate a cDNA library for high-throughput sequencing. The reads are then aligned to the reference genome.

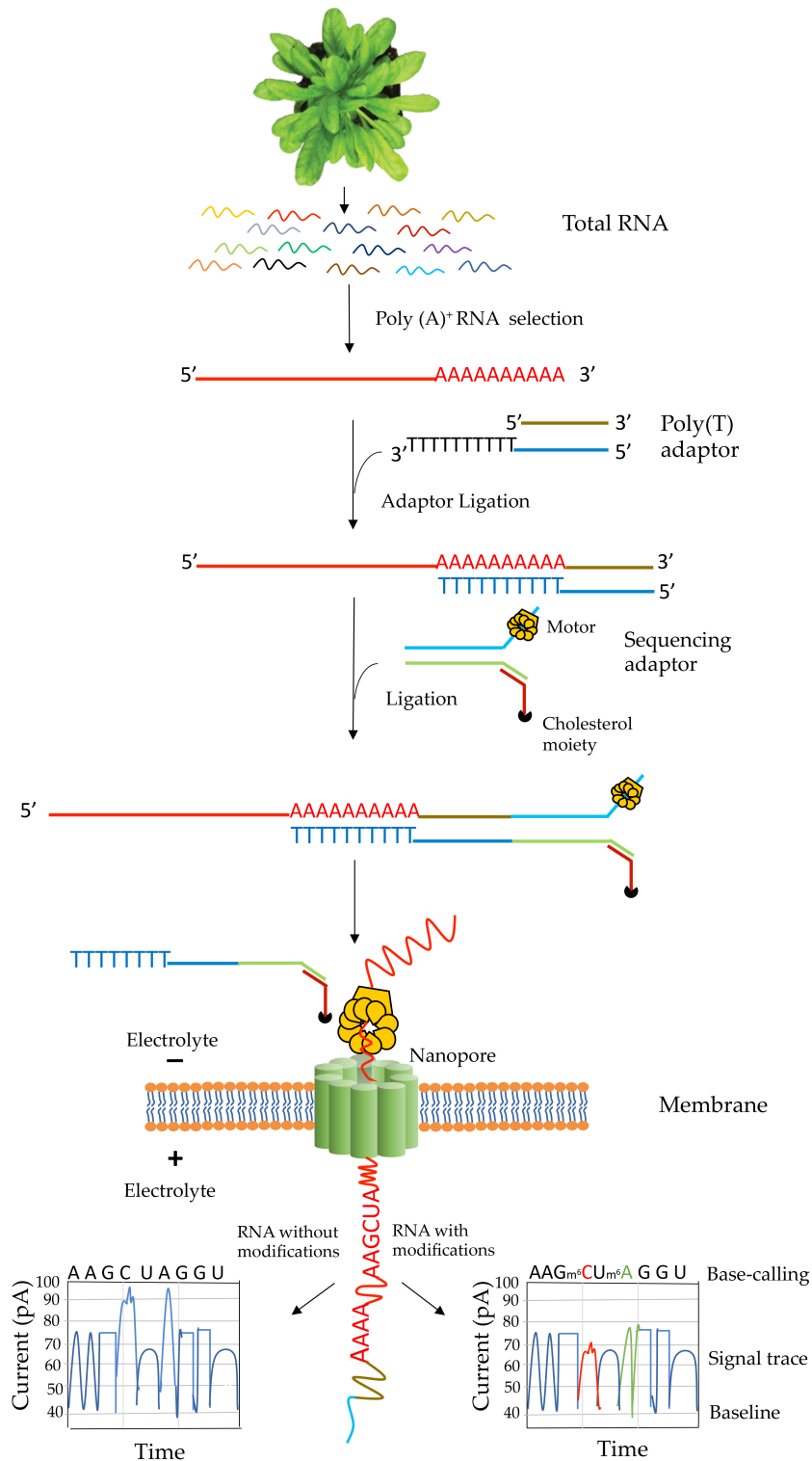


FIGURE 3 | Schematic illustration of direct RNA sequencing using the Oxford Nanopore Technology. Poly(A)⁺ mRNA from total RNA is isolated, then a poly(T) adaptor and a sequencing adaptor with a motor enzyme are added to the 3' end of poly(A)⁺ mRNA. It is then subject to sequencing on a membrane with thousands of nanopore channels, each of which is coupled to ammeters that measure current passing through the pore. The motor enzyme interacts with a nanopore on an electrically resistant synthetic membrane and the RNA strand is fed through the nanopore. A voltage across the membrane is applied and as the RNA moves through the nanopore nucleotide bases cause a characteristic change in current through the pore that is unique to each normal and modified base. The current output is then used in base-calling. An example of current output when RNA with (right box) or without modified RNA bases (left box) move through a pore is shown.

improve sequence throughput and quality (Garalde et al., 2018). In characterizing the yeast transcriptome using direct RNA-Seq, single-stranded RNA was used (Garalde et al., 2018) whereas in analyzing the human transcriptome, RNA–DNA hybrids were used (Workman et al., 2018). Transcripts without poly(A) tail can also be sequenced by enzymatically adding a 3′ poly (A) tail. One of the limitations for direct RNA-Seq is about the truncated reads. Studies in both pseudorabies virus (Moldován et al., 2018b) and *Saccharomyces cerevisiae* (Jenjaroenpun et al., 2018) revealed truncated reads, especially missing nucleotides at the 5′ end of the transcripts. It was speculated that it might be due to the premature release of the sequencing transcripts by the motor protein (Moldován et al., 2018b). However, longer transcripts over 5 kb could be generated using direct RNA-Seq (Jenjaroenpun et al., 2018). Thus, the motor protein might not be the major reason for the truncated reads. Another limitation is that at present bioinformatics tools for identification of RNA modification are rare. Tombo is the only reported tool to identify modified nucleotides from ONT (Stoiber et al., 2016). Also, base-calling algorithms for most RNA modifications are yet to be developed. Recently, soybean (*Glycine max*) seed transcriptome has been sequenced using MinION sequencing. However, this study adopted cDNA sequencing method, which could not be used for characterization of RNA modifications (Fleming et al., 2018). So far, only two direct RNA-sequencing studies – one with yeast poly(A)⁺ RNA (Garalde et al., 2018) and one with human poly(A)⁺ RNA (Workman et al., 2018) – have been performed with eukaryotic mRNAs. Interestingly, native sequencing of human poly(A)⁺ RNA uncovered a large number of novel isoforms (over 65% of all detected isoforms are novel) (Workman et al., 2018). The authors of the human transcriptome study were able to assess poly(A)⁺ length, allele-specific expression, base modifications (N⁶-methyladenine and inosine) in mRNA from direct RNA-Seq data (Workman et al., 2018).

FUTURE DIRECTIONS

From the Iso-Seq library construction step, it becomes apparent that the RNA modification information will be removed. Thus, common Iso-Seq libraries cannot be used for detecting RNA modification. Beside direct RNA-Seq, the PacBio reads from genome sequencing without any PCR amplification step can be used to detect DNA methylation marks, such as m⁶A, m⁵C, 5-hydroxymethylcytosine (Flusberg et al., 2010; Fang et al., 2012), and 4-methylcytosine (4mC) (Ye et al., 2016), respectively. Bisulfite sequencing (BS-Seq) using NGS can also detect m⁵C in a genome-wide manner (Krueger et al., 2012). However, long reads without PCR amplification provide new opportunities to detect additional modifications, which present distinct kinetic profiles and cannot be detected using NGS technologies. In *A. thaliana*, global profiling of m⁶A residues has been investigated using this method at single-nucleotide resolution (Liang et al., 2018). ONT sequencing can detect native genomic methylation, which has been reported in *Escherichia coli* (Rand et al., 2017) and humans (Simpson et al., 2017). It can be expected that both PacBio and ONT with enough coverage can replace present methylation

detecting methods, such as bisulfite-treated DNA following NGS for m⁵C identification (Frommer et al., 1992). By using a reverse transcriptase, instead of DNA polymerase, in ZMWs, cDNA synthesis has been observed in real time (Saletore et al., 2012). Furthermore, the presence of a modified (e.g., m⁶A) in RNA has been shown to alter the kinetics of nucleotide incorporation at the modified site. Based on this, it was suggested that by monitoring cDNA synthesis in real time in ZMWs modifications in RNA can be identified using the altered kinetic signature (Saletore et al., 2012).

Previous studies have shown that it is difficult to reconstruct splice isoforms and quantify differential expression of isoforms using short reads obtained with second-generation sequencing (Steijger et al., 2013; Kratz and Carninci, 2014). In comparison with Illumina, the read length is the great advantage in Iso-Seq cDNA transcript sequencing and Oxford Nanopore direct RNA-Seq, which can capture entire transcripts (Wang et al., 2016). Comparison of the gene expression between Illumina datasets and MinION revealed high correlation coefficient (Seki et al., 2018), which suggests that MinION is a useful platform to calculate expression level of transcripts by read count, or relative abundance of an RNA as transcripts per million transcripts (TPM) (Marinov, 2017). Indeed, single-molecule long-read sequencing in maize revealed tissue-specific isoforms (Wang et al., 2016). These new technologies provide great strengths and new avenues to explore complex transcriptomes. A combination of different techniques can offer solutions to overcome weaknesses of NGS and PacBio/ONT (Rhoads and Au, 2015). At present, IDP (Au et al., 2013) was developed to use long reads for identification of complex transcript structure and next-generation short reads for quantification. This hybrid method can solve the limitation for both technologies. A recent study showed a high correlation between ONT and Illumina on quantifying gene expression (Byrne et al., 2017). With improvements in sequencers (from MinION, GridION to PromethION), Oxford Nanopore direct RNA-Seq with sufficient throughput and accuracy can possibly be used to perform quantitative analyses of full-length isoforms on a whole transcriptome level.

AUTHOR CONTRIBUTIONS

LZ, HZ, MK, KP, LG, and AR wrote, discussed, and edited the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 31570674), the National Key R&D Program of China (2016YFD0600106 and 2018YFD0600101), the International Science and Technology Cooperation and Exchange Fund from Fujian Agriculture and Forestry University (KXGH17016), Natural Science Foundation of Fujian Province of China (Grant No. 2018J01608), and the Department of Energy Office of Science, Office of Biological and Environmental Research (Grant No. DE-SC0010733).

REFERENCES

- Abdel-Ghany, S. E., Hamilton, M., Jacobi, J. L., Ngam, P., Devitt, N., Schilkey, F., et al. (2016). A survey of the sorghum transcriptome using single-molecule long reads. *Nat. Commun.* 7:11706. doi: 10.1038/ncomms11706
- Au, K. F., Sebastiano, V., Afshar, P. T., Durruthy, J. D., Lee, L., Williams, B. A., et al. (2013). Characterization of the human ESC transcriptome by hybrid sequencing. *Proc. Natl. Acad. Sci. U.S.A.* 110, E4821–E4830. doi: 10.1073/pnas.1320101110
- Au, K. F., Underwood, J. G., Lee, L., and Wong, W. H. (2012). Improving PacBio long read accuracy by short read alignment. *PLoS One* 7:e46679. doi: 10.1371/journal.pone.0046679
- Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., et al. (2017). The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* 546, 148–152. doi: 10.1038/nature22380
- Bayega, A., Fahiminiya, S., Oikonomopoulos, S., and Ragoussis, J. (2018). “Current and future methods for mrna analysis: a drive toward single molecule sequencing,” in *Gene Expression Analysis: Methods in Molecular Biology*, Vol. 1783, eds N. Raghavachari and N. Garcia-Reyero (New York, NY: Springer), 209–241.
- Beaudoing, E., Freier, S., Wyatt, J. R., Claverie, J. M., and Gautheret, D. (2000). Patterns of variant polyadenylation signal usage in human genes. *Genome Res.* 10, 1001–1010. doi: 10.1101/gr.10.7.1001
- Blower, M. D., Jambhekar, A., Schwarz, D. S., and Toombs, J. A. (2013). Combining different mRNA capture methods to analyze the transcriptome: analysis of the *Xenopus laevis* transcriptome. *PLoS One* 8:e77700. doi: 10.1371/journal.pone.0077700
- Byrne, A., Beaudin, A. E., Olsen, H. E., Jain, M., Cole, C., Palmer, T., et al. (2017). Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nat. Commun.* 8:16027. doi: 10.1038/ncomms16027
- Campbell, M. A., Haas, B. J., Hamilton, J. P., Mount, S. M., and Buell, C. R. (2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC Genomics* 7:327. doi: 10.1186/1471-2164-7-327
- Cartolano, M., Huettel, B., Hartwig, B., Reinhardt, R., and Schneeberger, K. (2016). cDNA library enrichment of full length transcripts for smrt long read sequencing. *PLoS One* 11:e0157779. doi: 10.1371/journal.pone.0157779
- Chao, Q., Gao, Z. F., Zhang, D., Zhao, B. G., Dong, F. Q., Fu, C. X., et al. (2018). The developmental dynamics of the Populus stem transcriptome. *Plant Biotech. J.* 17, 206–219. doi: 10.1111/pbi.12958
- Chao, Y., Yuan, J., Li, S., Jia, S., Han, L., and Xu, L. (2018). Analysis of Transcripts and splice isoforms in Red Clover (*Trifolium pratense* L.) by single-molecule long-read sequencing. *bioRxiv*. doi: 10.1101/330977330977
- Chen, W. H., Lv, G., Lv, C., Zeng, C., and Hu, S. (2007). Systematic analysis of alternative first exons in plant genomes. *BMC Plant Biol.* 7:55. doi: 10.1186/1471-2229-7-55
- Cheng, B., Furtado, A., and Henry, R. J. (2017). Long-read sequencing of the coffee bean transcriptome reveals the diversity of full-length transcripts. *Gigascience* 6, 1–13. doi: 10.1093/gigascience/gix086
- Clavijo, B. J., Venturini, L., Schudoma, C., Accinelli, G. G., Kaithakottil, G., Wright, J., et al. (2017). An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Res.* 27, 885–896. doi: 10.1101/gr.217117.116
- Cook, D. E., Valle-Inclan, J. E., Pajoro, A., Rovenich, H., Thomma, B., and Faino, L. (2019). Long-read annotation: automated eukaryotic genome annotation based on long-read cDNA sequencing. *Plant Physiol.* 179, 38–54. doi: 10.1104/pp.18.00848
- Cui, X., Liang, Z., Shen, L., Zhang, Q., Bao, S., Geng, Y., et al. (2017). 5-Methylcytosine RNA methylation in *Arabidopsis thaliana*. *Mol. Plant* 10, 1387–1399. doi: 10.1016/j.molp.2017.09.013
- Depledge, D. P., Puthankalam, S. K., Sadaoka, T., Beady, D., Mori, Y., Placantonakis, D., et al. (2018). Native RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *bioRxiv* doi: 10.1101/373522
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485, 201–206. doi: 10.1038/nature11112
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461
- Faghihi, M. A., and Wahlestedt, C. N. R. M. C. B. (2009). Regulatory roles of natural antisense transcripts. *Nat. Rev. Mol. Cell Biol.* 10, 637–643. doi: 10.1038/nrm2738
- Fang, G., Munera, D., Friedman, D. I., Mandlik, A., Chao, M. C., Banerjee, O., et al. (2012). Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nat. Biotechnol.* 30, 1232–1239. doi: 10.1038/nbt.2432
- Feng, J. W., Huang, S., Guo, Y. X., Liu, D., Song, J. M., Gao, J., et al. (2019). Plant ISOform sequencing database (PISO): a comprehensive repertory of full-length transcripts in plants. *Plant Biotechnol. J.* doi: 10.1111/pbi.13076 [Epub ahead of print].
- Filichkin, S. A., Hamilton, M., Dharmawardhana, P. D., Singh, S. K., Sullivan, C., Ben-Hur, A., et al. (2018). Abiotic Stresses modulate landscape of poplar transcriptome via alternative splicing, differential intron retention, and isoform ratio switching. *Front. Plant Sci.* 9:5. doi: 10.3389/fpls.2018.00005
- Fleming, M. B., Patterson, E. L., Reeves, P. A., Richards, C. M., Gaines, T. A., and Walters, C. (2018). Exploring the fate of mRNA in aging seeds: protection, destruction, or slow decay? *J. Exp. Bot.* 69, 4309–4321. doi: 10.1093/jxb/ery215
- Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., et al. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7, 461–465. doi: 10.1038/nmeth.1459
- Foissac, S., and Sasmeth, M. (2007). ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. *Nucleic Acids Res.* 35, W297–W299. doi: 10.1093/nar/gkm311
- Fray, R. G., and Simpson, G. G. (2015). The Arabidopsis epitranscriptome. *Curr. Opin. Plant Biol.* 27, 17–21. doi: 10.1016/j.pbi.2015.05.015
- Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., et al. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. U.S.A.* 89, 1827–1831. doi: 10.1073/pnas.89.5.1827
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565
- Gao, Y., Wang, H., Zhang, H., Wang, Y., Chen, J., and Gu, L. (2017). PRAPI: post-transcriptional regulation analysis pipeline for Iso-Seq. *Bioinformatics* 34, 1580–1582. doi: 10.1093/bioinformatics/btx830
- Garalde, D. R., Snell, E. A., Jachimowicz, D., Sipos, B., Lloyd, J. H., Bruce, M., et al. (2018). Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* 15, 201–206. doi: 10.1038/nmeth.4577
- Gilbert, W. V., Bell, T. A., and Schaening, C. (2016). Messenger RNA modifications: form, distribution, and function. *Science* 352, 1408–1412. doi: 10.1126/science.aad8711
- Gonzalez-Garay, M. L. (2016). “Introduction to isoform sequencing using pacific biosciences technology (Iso-Seq),” in *Transcriptomics and Gene Regulation*, ed. J. Wu (Dordrecht: Springer), 141–160.
- Goodwin, S., Gurtowski, J., Ethe-Sayers, S., Deshpande, P., Schatz, M. C., and McCombie, W. R. (2015). Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res.* 25, 1750–1756. doi: 10.1101/gr.191395.115
- Gordon, S. P., Tseng, E., Salamov, A., Zhang, J., Meng, X., Zhao, Z., et al. (2015). Widespread polycistronic transcripts in fungi revealed by single-molecule mRNA sequencing. *PLoS One* 10:e0132628. doi: 10.1371/journal.pone.0132628
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protocols* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9:R7. doi: 10.1186/gb-2008-9-1-r7

- Hihara, Y., Kamei, A., Kanehisa, M., Kaplan, A., and Ikeuchi, M. (2001). DNA microarray analysis of cyanobacterial gene expression during acclimation to high light. *Plant Cell* 13, 793–806. doi: 10.1105/tpc.13.4.793
- Jenjaroenpun, P., Wongsurawat, T., Pereira, R., Patumcharoenpol, P., Ussery, D. W., Nielsen, J., et al. (2018). Complete genomic and transcriptional landscape analysis using third-generation sequencing: a case study of *Saccharomyces cerevisiae* CEN. PK113-7D. *Nucleic Acids Res.* 46:e38. doi: 10.1093/nar/gky014
- Ji, G., Ye, W., Su, Y., Chen, M., Huang, G., and Wu, X. (2018). AStrap: identification of alternative splicing from transcript sequences without a reference genome. *Bioinformatics* doi: 10.1093/bioinformatics/bty1008 [Epub ahead of print].
- Jin, J., Liu, J., Wang, H., Wong, L., and Chua, N.-H. (2013). PLncDB: plant long non-coding RNA database. *Bioinformatics* 29, 1068–1071. doi: 10.1093/bioinformatics/btt107
- Jo, I. H., Lee, J., Hong, C. E., Lee, D. J., Bae, W., Park, S. G., et al. (2017). Isoform sequencing provides a more comprehensive view of the panax ginseng transcriptome. *Genes* 8:E228. doi: 10.3390/genes8090228
- Kakrana, A., Mathioni, S. M., Huang, K., Hammond, R., Vandivier, L., Patel, P., et al. (2018). Plant 24-nt reproductive phasiRNAs from intramolecular duplex mRNAs in diverse monocots. *Geome Res.* 28, 1333–1344. doi: 10.1101/gr.228163.117
- Kim, N. H., Jayakodi, M., Lee, S. C., Choi, B. S., Jang, W., Lee, J., et al. (2018). Genome and evolution of the shade-requiring medicinal herb Panax ginseng. *Plant Biotechnol. J.* 16, 1904–1917. doi: 10.1111/pbi.12926
- Koren, S., Schatz, M. C., Walenz, B. P., Martin, J., Howard, J. T., Ganapathy, G., et al. (2012). Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat. Biotechnol.* 30, 693–700. doi: 10.1038/nbt.2280
- Kratz, A., and Carninci, P. (2014). The devil in the details of RNA-seq. *Nat. Biotechnol.* 32, 882–884. doi: 10.1038/nbt.3015
- Krueger, F., Kreck, B., Franke, A., and Andrews, S. R. (2012). DNA methylome analysis using short bisulfite sequencing data. *Nat. Methods* 9:145. doi: 10.1038/nmeth.1828
- Levene, M. J., Korch, J., Turner, S. W., Foquet, M., Craighead, H. G., and Webb, W. W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* 299, 682–686. doi: 10.1126/science.1079700
- Li, H. (2018). Minimap2: versatile pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191
- Li, H., Wang, J., Mor, G., and Sklar, J. (2008). A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells. *Science* 321, 1357–1361. doi: 10.1126/science.1156725
- Li, J., Harata-Lee, Y., Denton, M. D., Feng, Q., Rathjen, J. R., Qu, Z., et al. (2017). Long read reference genome-free reconstruction of a full-length transcriptome from *Astragalus membranaceus* reveals transcript variants involved in bioactive compound biosynthesis. *Cell Discov.* 3:17031. doi: 10.1038/celldisc.2017.31
- Li, S., Yamada, M., Han, X., Ohler, U., and Benfey, P. N. (2016). High-resolution expression map of the *Arabidopsis* root reveals alternative splicing and lincRNA regulation. *Dev. Cell* 39, 508–522. doi: 10.1016/j.devcel.2016.10.012
- Li, X., Xiong, X., and Yi, C. (2017). Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat. methods* 14:23. doi: 10.1038/nmeth.4110
- Li, Y., Dai, C., Hu, C., Liu, Z., and Kang, C. (2017). Global identification of alternative splicing via comparative analysis of SMRT- and Illumina-based RNA-seq in strawberry. *Plant J.* 90, 164–176. doi: 10.1111/tpj.13462
- Liang, Z., Shen, L., Cui, X., Bao, S., Geng, Y., Yu, G., et al. (2018). DNA N6-Adenine Methylation in *Arabidopsis thaliana*. *Dev. Cell* 45, 406–416.e3. doi: 10.1016/j.devcel.2018.03.012
- Liu, J., Wang, H., and Chua, N. H. (2015). Long noncoding RNA transcriptome of plants. *Plant Biotechnol. J.* 13, 319–328. doi: 10.1111/pbi.12336
- Liu, S., Tsai, W.-H., Ding, Y., Chen, R., Fang, Z., Huo, Z., et al. (2015). Comprehensive evaluation of fusion transcript detection algorithms and a meta-caller to combine top performing methods in paired-end RNA-seq data. *Nucleic Acids Res.* 44:e47. doi: 10.1093/nar/gkv1234
- Liu, X., Mei, W., Soltis, P. S., Soltis, D. E., and Barbazuk, W. B. (2017). Detecting alternatively spliced transcript isoforms from single-molecule long-read sequences without a reference genome. *Mol. Ecol. Resour.* 17, 1243–1256. doi: 10.1111/1755-0998.12670
- Loman, N. J., and Quinlan, A. R. (2014). Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics* 30, 3399–3401. doi: 10.1093/bioinformatics/btu555
- Luo, G.-Z., Macqueen, A., Zheng, G., Duan, H., Dore, L. C., Lu, Z., et al. (2014). Unique features of the m6A methylome in *Arabidopsis thaliana*. *Nat. Commun.* 5:5630. doi: 10.1038/ncomms6630
- Maier, C. A., Palanisamy, N., Brenner, J. C., Cao, X., Kalyana-Sundaram, S., Luo, S., et al. (2009). Chimeric transcript discovery by paired-end transcriptome sequencing. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12353–12358. doi: 10.1073/pnas.0904720106
- Marchet, C., Lecompte, L., Da Silva, C., Cruaud, C., Aury, J. M., Nicolas, J., et al. (2018). Clustering de novo by gene of long reads from transcriptomics data. *Nucleic Acids Res.* 42:e2. doi: 10.1093/nar/gky1834
- Marinov, G. K. (2017). On the design and prospects of direct RNA sequencing. *Brief. Funct. Genom.* 16, 326–335. doi: 10.1093/bfpg/elw043
- Matsumura, H., Nirasawa, S., and Terauchi, R. (1999). Transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE). *Plant J.* 20, 719–726. doi: 10.1046/j.1365-313X.1999.00640.x
- McCarthy, A. (2010). Third generation DNA sequencing: pacific biosciences' single molecule real time technology. *Chem. Biol.* 17, 675–676. doi: 10.1016/j.chembiol.2010.07.004
- Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E., and Jaffrey, S. R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 149, 1635–1646. doi: 10.1016/j.cell.2012.05.003
- Minio, A., Massonnet, M., Vondras, A., Figueroa-Balderas, R., Blanco-Ulate, B., and Cantu, D. (2018). Isoform-scale annotation and expression profiling of the Cabernet Sauvignon transcriptome using single-molecule sequencing of full-length cDNA. *bioRxiv* doi: 10.1534/g3.118.201008
- Minoche, A. E., Dohm, J. C., Schneider, J., Holtgrawe, D., Viehover, P., Montfort, M., et al. (2015). Exploiting single-molecule transcript sequencing for eukaryotic gene prediction. *Genome Biol.* 16:184. doi: 10.1186/s13059-015-0729-7
- Moldován, N., Tombác, D., Szűcs, A., Csabai, Z., Balázs, Z., Kis, E., et al. (2018a). Third-generation sequencing reveals extensive polycistronism and transcriptional overlapping in a baculovirus. *Sci. Rep.* 8:8604. doi: 10.1038/s41598-018-26955-8
- Moldován, N., Tombác, D., Szűcs, A., Csabai, Z., Snyder, M., and Boldogkői, Z. (2018b). Multi-platform sequencing approach reveals a novel transcriptome profile in pseudorabies virus. *Front. Microbiol.* 8:2708. doi: 10.3389/fmicb.2017.02708
- Mortazavi, A., Williams, B. A., Mccue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5:621. doi: 10.1038/nmeth.1226
- Morton, T., Petricka, J., Corcoran, D. L., Li, S., Winter, C. M., Carda, A., et al. (2014). Paired-end analysis of transcription start sites in *Arabidopsis* reveals plant-specific promoter signatures. *Plant Cell* 26, 2746–2760. doi: 10.1105/tpc.114.125617
- Nagasaki, H., Arita, M., Nishizawa, T., Suwa, M., and Gotoh, O. (2006). Automated classification of alternative splicing and transcriptional initiation and construction of visual database of classified patterns. *Bioinformatics* 22, 1211–1216. doi: 10.1093/bioinformatics/btl067
- Nam, D. K., Lee, S., Zhou, G., Cao, X., Wang, C., Clark, T., et al. (2002). Oligo (dT) primer generates a high frequency of truncated cDNAs through internal poly (A) priming during reverse transcription. *Proc. Natl. Acad. Sci. U.S.A.* 99, 6152–6156. doi: 10.1073/pnas.092140899
- Ner-Gaon, H., Leviatan, N., Rubin, E., and Fluhr, R. (2007). Comparative cross-species alternative splicing in plants. *Plant Physiol.* 144, 1632–1641. doi: 10.1104/pp.107.098640
- Ni, T., Corcoran, D. L., Rach, E. A., Song, S., Spana, E. P., Gao, Y., et al. (2010). A paired-end sequencing strategy to map the complex landscape of transcription initiation. *Nat. Methods* 7:521. doi: 10.1038/nmeth.1464
- Ning, G., Cheng, X., Luo, P., Liang, F., Wang, Z., Yu, G., et al. (2017). Hybrid sequencing and map finding (HySeMaFi): optional strategies for extensively deciphering gene splicing and expression in organisms without reference genome. *Sci. Rep.* 7:43793. doi: 10.1038/srep43793
- Pootakham, W., Sonthirod, C., Naktang, C., Ruang-Areerate, P., Yoocha, T., Sangsrakru, D., et al. (2017). De novo hybrid assembly of the rubber tree

- genome reveals evidence of paleotetraploidy in *Hevea* species. *Sci. Rep.* 7:41457. doi: 10.1038/srep41457
- Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O. R., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* 30, 777–782. doi: 10.1038/nbt.2282
- Rand, A. C., Jain, M., Eizenga, J. M., Musselman-Brown, A., Olsen, H. E., Akeson, M., et al. (2017). Mapping DNA methylation with high-throughput nanopore sequencing. *Nat. Methods* 14, 411–413. doi: 10.1038/nmeth.4189
- Reddy, A. S., Marquez, Y., Kalyana, M., and Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *Plant Cell* 25, 3657–3683. doi: 10.1105/tpc.113.117523
- Rhoads, A., and Au, K. F. (2015). PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* 13, 278–289. doi: 10.1016/j.gpb.2015.08.002
- Roundtree, I. A., Evans, M. E., Pan, T., and He, C. (2017). Dynamic RNA modifications in gene expression regulation. *Cell* 169, 1187–1200. doi: 10.1016/j.cell.2017.05.045
- Saletore, Y., Meyer, K., Korlach, J., Vilfan, I. D., Jaffrey, S., and Mason, C. E. (2012). The birth of the Epitranscriptome: deciphering the function of RNA modifications. *Genome Biol.* 13:175. doi: 10.1186/gb-2012-13-10-175
- Salmela, L., and Rivals, E. (2014). LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30, 3506–3514. doi: 10.1093/bioinformatics/btu538
- Salmela, L., Walve, R., Rivals, E., and Ukkonen, E. (2016). Accurate self-correction of errors in long reads using de Bruijn graphs. *Bioinformatics* 33, 799–806. doi: 10.1093/bioinformatics/bt1321
- Sedlazeck, F. J., Lee, H., Darby, C. A., and Schatz, M. C. (2018). Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nat. Rev. Genet.* 19, 329–346. doi: 10.1038/s41576-018-0003-4
- Seki, M., Katsumata, E., Suzuki, A., Sereewattanawoot, S., Sakamoto, Y., Mizushima-Sugano, J., et al. (2018). Evaluation and application of RNA-Seq by MinION. *DNA Res.* 26, 55–65. doi: 10.1093/dnares/dsy038
- Shen, S., Park, J. W., Lu, Z.-X., 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
- Shepard, P. J., Choi, E.-A., Lu, J., Flanagan, L. A., Hertel, K. J., and Shi, Y. (2011). Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* 17, 761–772. doi: 10.1261/rna.2581711
- Sherstnev, A., Duc, C. L., Cole, C., Zacharaki, V., Hornyik, C., Oszolac, F., et al. (2012). Direct sequencing of *Arabidopsis thaliana* RNA reveals patterns of cleavage and polyadenylation. *Nat. Struct. Mol. Biol.* 19, 845–852. doi: 10.1038/nsmb.2345
- Simpson, J. T., Workman, R. E., Zuzarte, P., David, M., Dursi, L., and Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* 14, 407–410. doi: 10.1038/nmeth.4184
- Slatko, B. E., Gardner, A. F., and Ausubel, F. M. (2018). Overview of next-generation sequencing technologies. *Curr. Protoc. Mol. Biol.* 122:e59. doi: 10.1002/cpmb.59
- Slobodin, B., Han, R., Calderone, V., Vrieliink, J., Loayza-Puch, F., Elkon, R., et al. (2017). Transcription impacts the efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. *Cell* 169:e312. doi: 10.1016/j.cell.2017.03.031
- Smith, A. M., Jain, M., Mulroney, L., Garalde, D. R., and Akeson, M. (2017). Reading canonical and modified nucleotides in 16S ribosomal RNA using nanopore direct RNA sequencing. *bioRxiv*. doi: 10.1101/132274
- Steijger, T., Abril, J. F., Engstrom, P. G., Kokocinski, F., Hubbard, T. J., Guigo, R., et al. (2013). Assessment of transcript reconstruction methods for RNA-seq. *Nat. Methods* 10, 1177–1184. doi: 10.1038/nmeth.2714
- Stoiber, M. H., Quick, J., Egan, R., Lee, J. E., Celniker, S. E., Neely, R., et al. (2016). De novo identification of DNA modifications enabled by genome-guided nanopore signal processing. *bioRxiv*. doi: 10.1101/094672
- Sun, M.-Y., Li, J.-Y., Li, D., Huang, F.-J., Wang, D., Li, H., et al. (2018). Full-length transcriptome sequencing and modular organization analysis of naringin/neoeriocitrin related gene expression pattern in *Drynaria roosii*. *Plant Cell Physiol.* 59, 1398–1414. doi: 10.1093/pcp/pcy072
- Tanaka, T., Koyanagi, K. O., and Itoh, T. (2009). Highly diversified molecular evolution of downstream transcription start sites in rice and *Arabidopsis*. *Plant Physiol.* 149, 1316–1324. doi: 10.1104/pp.108.131656
- Tian, B., Hu, J., Zhang, H., and Lutz, C. S. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res* 33, 201–212. doi: 10.1093/nar/gki158
- Tian, J., Feng, S., Liu, Y., Zhao, L., Tian, L., Hu, Y., et al. (2018). Single-molecule long-read sequencing of *Zanthoxylum bungeanum* Maxim. Transcriptome: Identification of Aroma-Related Genes. *Forest* 9:765. doi: 10.3390/f9120765
- Tilgner, H., Jahanbani, F., Blauwkamp, T., Moshrefi, A., Jaeger, E., Chen, F., et al. (2015). Comprehensive transcriptome analysis using synthetic long-read sequencing reveals molecular co-association of distant splicing events. *Nat. Biotechnol.* 33, 736–742. doi: 10.1038/nbt.3242
- VanBuren, R., Bryant, D., Edger, P. P., Tang, H., Burgess, D., Challabathula, D., et al. (2015). Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. *Nature* 527, 508–511. doi: 10.1038/nature15714
- Vandivier, L. E., and Gregory, B. D. (2018). New insights into the plant epitranscriptome. *J. Exp. Bot.* 69, 4659–4665. doi: 10.1093/jxb/ery262
- Wang, B., Tseng, E., Regulski, M., Clark, T. A., Hon, T., Jiao, Y., et al. (2016). Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat. Commun.* 7:11708. doi: 10.1038/ncomms11708
- Wang, B. B., and Brendel, V. (2006). Genomewide comparative analysis of alternative splicing in plants. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7175–7180. doi: 10.1073/pnas.0602039103
- Wang, H.-L. V., and Chekanova, J. A. (2017). “Long noncoding RNAs in plants,” in *Long Non Coding RNA Biology: Advances in Experimental Medicine and Biolog.* ed. M. Rao (Singapore: Springer), 133–154. doi: 10.1007/978-981-10-5203-3_5
- Wang, M., Wang, P., Liang, F., Ye, Z., Li, J., Shen, C., et al. (2018). A global survey of alternative splicing in allopolyploid cotton: landscape, complexity and regulation. *New Phytol.* 217, 163–178. doi: 10.1111/nph.14762
- Wang, Q., and Rio, D. C. (2018). JUM is a computational method for comprehensive annotation-free analysis of alternative pre-mRNA splicing patterns. *Proc. Natl. Acad. Sci. U.S.A.* 115, E8181–E8190. doi: 10.1073/pnas.1806018115
- Wang, T., Wang, H., Cai, D., Gao, Y., Zhang, H., Wang, Y., et al. (2017). Comprehensive profiling of rhizome associated alternative splicing and alternative polyadenylation in moso bamboo (*Phyllostachys edulis*). *Plant J.* 91, 684–699. doi: 10.1111/tpj.13597
- Wang, X., Xu, Y., Zhang, S., Cao, L., Huang, Y., Cheng, J., et al. (2017). Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nat. Genet.* 49, 765–772. doi: 10.1038/ng.3839
- Weirather, J. L., Afshar, P. T., Clark, T. A., Tseng, E., Powers, L. S., Underwood, J. G., et al. (2015). Characterization of fusion genes and the significantly expressed fusion isoforms in breast cancer by hybrid sequencing. *Nucleic Acids Res.* 43:e116. doi: 10.1093/nar/gkv562
- Weirather, J. L., De Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X. J., et al. (2017). Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *PLoS One* 12:e0170000. doi: 10.1371/journal.pone.0170000
- Workman, R. E., Tang, A. D., Tang, P. S., Jain, M., Tyson, J. R., Zuzarte, P. C., et al. (2018). Nanopore native RNA sequencing of a human poly(A) transcriptome. *bioRxiv*. doi: 10.1101/459529
- Wu, B., Suo, F., Lei, W., and Gu, L. (2014). Comprehensive analysis of alternative splicing in *Digitalis purpurea* by strand-specific RNA-Seq. *PLoS One* 9:e106001. doi: 10.1371/journal.pone.0106001
- Wu, J., Maehara, T., Shimokawa, T., Yamamoto, S., Harada, C., Takazaki, Y., et al. (2002). A comprehensive rice transcript map containing 6591 expressed sequence tag sites. *Plant Cell* 14, 525–535. doi: 10.1105/tpc.010274
- Wu, T. D., and Watanabe, C. K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21, 1859–1875. doi: 10.1093/bioinformatics/bti310
- Xiao, W., Adhikari, S., Dahal, U., Chen, Y. S., Hao, Y. J., Sun, B. F., et al. (2016). Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. *Mol. Cell* 61, 507–519. doi: 10.1016/j.molcel.2016.01.012
- Xie, S. Q., Han, Y., Chen, X. Z., Cao, T. Y., Ji, K. K., Zhu, J., et al. (2018). ISOdb: a comprehensive database of full-length isoforms generated by Iso-Seq. *Int. J. Genomics* 2018:9207637. doi: 10.1155/2018/9207637

- Xu, Q., Zhu, J., Zhao, S., Hou, Y., Li, F., Tai, Y., et al. (2017). Transcriptome profiling using single-molecule direct RNA sequencing approach for in-depth understanding of genes in secondary metabolism pathways of *Camellia sinensis*. *Front. Plant Sci.* 8:1205. doi: 10.3389/fpls.2017.01205
- Xu, Z., Peters, R. J., Weirather, J., Luo, H., Liao, B., Zhang, X., et al. (2015). Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis. *Plant J.* 82, 951–961. doi: 10.1111/tpj.12865
- Yang, L., Jin, Y., Huang, W., Sun, Q., Liu, F., and Huang, X. (2018). Full-length transcriptome sequences of ephemeral plant *Arabidopsis pumila* provides insight into gene expression dynamics during continuous salt stress. *BMC Genomics* 19:717. doi: 10.1186/s12864-018-5106-y
- Ye, P., Luan, Y., Chen, K., Liu, Y., Xiao, C., and Xie, Z. (2016). MethSMRT: an integrative database for DNA N6-methyladenine and N4-methylcytosine generated by single-molecular real-time sequencing. *Nucleic Acids Res.* 45, D85–D89. doi: 10.1093/nar/gkw950
- Zhang, B., Liu, J., Wang, X., and Wei, Z. (2018). Full-length RNA sequencing reveals unique transcriptome composition in bermudagrass. *Plant Physiol.* 132, 95–103. doi: 10.1016/j.plaphy.2018.08.039
- Zhang, H., Wang, H., Zhu, Q., Gao, Y., Wang, H., Zhao, L., et al. (2018). Transcriptome characterization of moso bamboo (*Phyllostachys edulis*) seedlings in response to exogenous gibberellin applications. *BMC Plant Biol.* 18:125. doi: 10.1186/s12870-018-1336-z
- Zhang, G., Guo, G., Hu, X., Zhang, Y., Li, Q., Li, R., et al. (2010). Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res.* 20, 646–654. doi: 10.1101/gr.100677.109
- Zhang, Y., Gu, L., Hou, Y., Wang, L., Deng, X., Hang, R., et al. (2015). Integrative genome-wide analysis reveals HLP1, a novel RNA-binding protein, regulates plant flowering by targeting alternative polyadenylation. *Cell Res.* 25, 864–876. doi: 10.1038/cr.2015.77
- Zhao, H., Gao, Z., Wang, L., Wang, J., Wang, S., Fei, B., et al. (2018). Chromosome-level reference genome and alternative splicing atlas of moso bamboo (*Phyllostachys edulis*). *Gigascience* 7:giy115. doi: 10.1093/gigascience/giy115
- Zhou, R., Moshgabadi, N., and Adams, K. L. (2011). Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16122–16127. doi: 10.1073/pnas.1109551108

Conflict of Interest Statement: 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 © 2019 Zhao, Zhang, Kohmen, Prasad, Gu and Reddy. 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.