RNA Sequencing: A Potent Transcription Profiling Tool

Kaiho Kaisa, Harshit Kumar, Manjit Panigrahi, Triveni Dutt and Bharat Bhushan*

Division of Animal Genetics, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly-243122, Uttar Pradesh, India

*Corresponding author

Abstract

The RNAseq approach is currently commonly used in genome-wide transcription profiling that stimulates biological applications through deep-sequencing technologies. RNAseq offers greater advantages over microarray-based analysis as it provides in-depth and more detailed knowledge on the characterization and quantification of transcriptomes. Although RNAseq still raises problems that need to be solved, there are various advances in innovative approaches and innovations that overcome the challenges that are continuously provided in the modern field of science as an indispensable tool. We are therefore presenting a summary of how RNAseq operates, summarizing its related difficulties and advantages over other methods, and explaining the variety of its applications that has pioneered transcriptome studies and influenced modern genetic science.

Keywords
RNAseq, transcriptome, Next Generation Sequencing, Gene expression, microarray

Article Info
Accepted: 10 September 2020
Available Online: 10 October 2020

Introduction

The transcriptome is comprised of multiple types of the RNA molecule or RNA transcript. RNA molecules are key components of any living cell. RNA molecules play a key role in the physiological response and understanding how these molecules are regulated is extremely crucial to understand the functional genome. The primary goal of RNA research is to understand under a specific condition the nature and quantity of each RNA molecule within a cell or tissue. There are already several development ventures underway on transcriptome instead of just genome and proteome because only 1-2% among the genes are coding and perhaps most transcribed genes were not translated into proteins and are expected to be involved in controlling epigenetic regulation and gene expression (Blignaut, 2012; Shabalina et al., 2004). The detailed profiling of RNA molecules presents the possibility of obtaining insight into the biological behavior of a cell or tissue at any point in time and to relate the knowledge to the phenotypic variation that affects almost every area of biological sciences and is now widely adopted for clinical use (Berger et al., 2010; Costa-Silva...
The advancement of Massively Parallel Signature Sequencing (MPSS) (Reinartz et al., 2002) and Solexa technology (Bennett, 2004) has led in recent years to the advancement of the ground-breaking RNA-sequencing (RNAseq) (Delseny et al., 2010). The Next Generation Sequencing (NGS) technology i.e. RNAseq is a technology that allows the sequencing of millions of nucleotide fragments in parallel, has emerged as a superior and effective method for studying entire transcriptome characterization and profiling (Wickramasinghe et al., 2014; Anamika et al., 2015; Hrdlickova et al., 2016) because of its ground-breaking effect and quick declining costs (Esteve-Codina, 2018) on transcriptome study laying the ground for modern genetic research. Despite earlier advances in various technologies such as microarray-based deduction and quantification of the transcriptome, RNAseq technology has evolved to be the best choice. With development, numerous problems and challenges are tackled, whilst RNAseq technology is still an indispensable tool constantly in the study both known and novel organism transcripts by providing comprehensive light on the role of gene expression in development, differential expression between conditions, gene expression changes in disease progression, alternative splicing incidents, RNA editing, gene fusion, allele-specific expression, etc.

**RNA sequencing (RNAseq)**

RNAseq is a central technique capable of evaluating the quantity and sequences of RNA in a sample using a combination of next-generation sequencing (NGS) or deep sequencing. It is a revolutionary tool used to map and measure the transcriptome (Chu and Corey, 2012). The RNAseq study gives us a snapshot of the transcriptome, the entire cellular composition of RNAs such as mRNA, rRNA, and tRNA to name a few, while the gene expression profiles are simultaneously evaluated on a genome basis (Mortazavi et al., 2008). It is important to understand the transcriptome (Wang et al., 2009) to interpret the underlying functional elements of the whole genome for the protein expression and to disclose the genetic constitutions of the different tissues and cells and its correlation with the various development and disease. RNAseq has been largely used since the outset of its use as an effective approach in characterising transcripts, gene expression profiling and detecting RNA biogenesis and metabolism, and providing powerful tools to recognize molecular pathways in growth, differentiation, and disease (Costa-Silva et al., 2017).

RNAseq has overcome many of the drawbacks imposed by previously evolved technologies, including the expressed sequence tag (EST) technique (Adams et al., 1992) analyses in which the higher cost of sequencing has restricted its use in expression analysis. While the Serial Analysis of Gene Expression (SAGE) approach arises to reduce the cost of expression analysis per gene (Velculescu et al., 1995), the emergence of superior DNA microarray technology has outperformed EST and SAGE gene expression analysis approaches, predominantly because of their much efficient coverage for large-scale studies (Farkas et al., 2015). The major drawback that inhibits its widespread applicability and growth, however, is that microarrays only enable the relative quantification of transcripts and rely on foreknowledge of the adequately annotated genome to pick a probe without the ability to recognize alternate splice sites or new exons, and other limitations such as cross-hybridization history and signal saturation-induced detection (Han et al., 2015). Subsequently, the innovative Next Generation Sequencing (NGS) or deep sequencing, i.e. RNAseq, has influenced a paradigm shift in
biology and medicine with its ability to obtain a comparatively broad and diverse collection of data in a brief period of time and to quickly turn RNA studies than ever before. In comparison, the RNAseq technique enables an accurate analysis of transcriptomes (Adiconis et al., 2013; Hrdlickova et al., 2016). RNAseq was pivotal in capturing the range of species with a novel transcript including long non-coding RNA, siRNA, miRNA, snRNA, etc. involved in RNA stability control, protein translation, or chromatin modulation (Robertson et al., 2010; Trapnell et al., 2010). RNAseq has recently been used to research biological issues, including the accurate location of regulatory elements (Arnold et al., 2013). RNAseq information can also detect allele-specific expression, disease-associated single nucleotide polymorphisms (SNP), and gene fusions that lead to our understanding of disease-causing variants (Maher et al., 2009; Berger et al., 2010; Usman et al., 2017). Currently, scRNA-seq has rapidly emerged as an alternative for individual cell transcriptome analysis (Van den Berge et al., 2019) to research core biological issues of cell heterogeneity and diversity in stem cell biology and neuroscience (Wilson et al., 2015), which significantly enhances transcriptomic studies (Gupta et al., 2016; Chen et al., 2019) revealing significant cell-to-cell gene expression differences (Rosenberg et al., 2018).

**RNAseq vs microarrays**

Since the mid-1900s, microarray analysis has been used for the study of gene expression evaluation. RNAseq does however have strong advantages over microarray approaches. RNAseq can be aligned to particular genome regions with fewer background signals, while cross-hybridization and background signals often result in low precision or poor sensitivity for some genes and restrict the diverse use of the microarray (Hrdlickova et al., 2016). RNAseq encompasses a broader spectrum of levels of expression across which transcripts can be identified (> 9,000 folds) (Nagalakshmi et al., 2008) and can be used to evaluate certain organisms for which the whole reference genome hasn't yet been constructed, unlike microarray-based methods that are limited to genomic sequence prior knowledge for species-specific samples sequencing (Hrdlickova et al., 2016). RNAseq is more quantitative since it has no upper bound for quantification but offers an open architecture and is free from significant problems associated with microarrays in identifying expression levels that are extremely high or extremely poor. RNAseq specifically uncovers sequence identity, essential for studying unknown genes and novel transcript isoforms, SNPs, or other modifications, and with its high resolution and specificity has mapped 5' and 3' borders for several genes (Wang et al., 2009; Mackenzie, 2018). The RNAseq findings also indicate a high degree of data reproducibility for scientific and biological replicates (Cloonan et al., 2008). RNAseq is the first sequencing-based strategy that enables very high-throughput and quantitative surveying of the entire genome context transcriptome. This approach provides both single-base annotation resolution and 'digital' gene expression levels on the scale of the genome, and at a reduced cost than previous techniques (Wang et al., 2009). Therefore, RNAseq technology is favored over array-based transcriptome profiling approaches.

**RNAseq workflow**

Several mature protocols were adapted over the years from those of the original RNAseq protocols, released over a decade (Mortazavi et al., 2008). The basic workflow (Figure 1) of the primary steps involved in a typical
RNAseq experiment (Mackenzie, 2018) initiates with extraction and purification of the RNA from the given sample, followed by the enrichment of the target RNAs or the depletion of the rRNA (Van Dijk et al., 2014). RNAs were further chemically or enzymatically fragmented into suitable size molecules. The technological advantage of instruments developed for DNA-based sequencing has confined most RNAseq experiments to be performed on instruments that sequence DNA molecules (Han et al., 2015), so the preparation of cDNA libraries from RNA is a fundamental step for RNAseq. Therefore, the target RNAs are reverse-transcribed to cDNA (first strand), then the RNA is degraded, and the first-strand cDNA is then complemented forming a double strand. Adapters are either linked to the double-stranded cDNA ends of 3' and 5'. cDNA libraries can be prepared in any one of the two ways: single-end (only one end of the cDNA insert is sequenced) or paired-end (both ends are sequenced, providing two reads in opposite direction). Each molecule is then sequenced in a high-throughput fashion so-called next-generation sequencing with or without amplification, to obtain short sequence readings from single or both ends (Wang et al., 2009). Most RNAseq studies contain between 10 million and 100 million reads, with a bias over time towards deeper sequencing. Although the number of samples per project has remained stable over the years, the median number of samples was about eight (Van den Berge et al., 2019). The resulting sequence reads are mapped with the reference genome or transcriptome or integrated de novo assembly without the genomic sequence to generate a genome-scale transcription map consisting of both the transcription structure and/or expression level for either gene. The resulting sequence data will then be analyzed for evaluating a differential expression, identifying variants, annotation of genomes, identification of new transcripts, editing of RNA, functional profiling, etc.

**RNAseq technologies**

The RNAseq data generation is a continuous phase of progress that involves constant development of sequencing technologies, experiment design, and algorithm creation. Several next-generation sequencing (NGS) platforms are available for RNAseq such as Roche, Illumina, Strong, PacBio and Ion Torrent, etc. and all of these techniques have different sequencing chemistries and yields that vary accordingly, proportionately influencing the final analysis of the experiments (Han et al., 2015). One of the most essential prerequisites during the preparation of the RNAseq experiment is the selection of the correct sequencing platform for a given application towards experimental success (Anamika et al., 2015). In 2004, Roche released the first commercially available RNAseq NGS technology, based on "pyrosequencing" technology (Mardis, 2008). The new improved Roche 454 GS FLX + device can generate approximately 1000bp average length sequence reads. The Illumina Genome Analyzer (GA) developed in 2006 (Mardis, 2008) based on Sequencing by Synthesis covering a sequence reading of about 36-100 bp (Wickramasinghe et al., 2014), offering higher sequencing capability at a low false-positive rate, even in repeated sequence regions. A sequencing platform based on ligation chemistry was also developed in 2007, known as sequencing by oligo ligation and detection (SOLiD) technology, with an average reading length of 85 bases approximately (Cloonan et al., 2008). In 2010, Illumina launched the upgraded HiSeq1000 and HiSeq2000 systems (Minoche et al., 2011) designed to deliver far higher sequencing reads of 100–150 bp at a much cheaper price (Nagalakshmi et al., 2008). Further, to study small genomes in a
limited of time, Illumina also launched the MiSeq method. Illumina has newly presented the HiSeq 2500 instrument, which is capable of much quicker sequencing than ever before. The reduced sequencing error margin (< 1%) of Illumina or SOLiD is now a significant tool for extremely smaller microRNA sequencing (Cloonan et al., 2008; Mardis, 2008). In 2010, Helicos technology implemented a single molecule sequencing method independent of an amplification step, generating sequence readings with a length of 55bp at an average (Morozova et al., 2009) that directly quantify RNA expression levels. On the counter, this sequencer is marked by an intrinsically high error rate (5 percent) (Chu and Corey, 2012). Similarly, another single-molecule real-time sequencing technology (SMRT) was also launched and named PacBio RS (Pacific Biosciences) which was commercially available first in 2011 and subsequently RS II and then the sequel sequencer (upgraded version) were also built (Van den Berge et al., 2019) which can perform much longer readings and offer reduced sequencing costs (Gonzalez-Garay, 2016). These PacBio sequencers are also used to decode the mystery of alternate splicing and detect gene fusion isoforms (Hrdlickova et al., 2016). The paired-end (PE) sequencing semi-conductor technology-based Ion Torrent Personal Genome Machine (PGM) was developed at the end of 2010, as the first commercial sequencing system that is independent of fluorescence and camera scanning, resulting in higher efficiency, lower cost, and smaller sample size, is mainly intended for therapeutic applications and small laboratories (Mellmann et al., 2011). The Paired-end sequencing is much more insightful than single-end sequencing notably if the objective is to research alternate splicing, spot gene fusions, or recreate isoforms de novo (Esteve-Codina, 2018). Very recently, Oxford Nanopore Technologies (ONT) has built portable devices such as MinION and PromethION powered by nanopore technology (Han et al., 2015). Thus, PacBio and ONT's third-generation sequencing technologies, which can achieve read lengths exceeding 10,000 bp, provide the potential for significant advances in isoform detection and discovery accuracy (Kuosmanen et al., 2018; Kovaka et al., 2019). In general, long-read sequencing: Pacific Biosciences' (PacBio) based on single-molecule real-time (SMRT) sequencing and the Oxford Nanopore Technologies' (ONT) nanopore sequencing has acquired attention currently, for improved transcriptome construction due to its potential to produce long reads (Amarasinghe et al., 2020), although they have less accuracy per read than short-read sequencing in comparison with short-read sequencing. Few of the NGS platforms commonly used for RNAseq are mentioned in table 1 (Cloonan et al., 2008; Nagalakshmi et al., 2008; Liu et al., 2012; Buermans and den Dunnen, 2014).

**Post Genome-Wide Association Studies (GWASs) period of transcriptome-wide association studies (TWASs)**

GWASs had successfully identified thousands of SNP-trait associations across the genome, connecting widespread genetic diversity to various complex traits and diseases (Buniello et al., 2019). Conversely, so many of these reported genetic markers reside well outside the protein-coding domains, intronic or intergenic genomic regions, making it nearly impossible to comprehend the biological mechanisms behind these established associations (Gusev et al., 2016). GWAS usually seldom refers to genetic variations with a direct functional effect on the integrity of cells. In several ways, to name a few, this lack of interpretability about; which are those causal variants, what are their molecular roles, which genes does the causal variants influenced, how alterations in
the role or control of the causal genes contribute to an altered probability of disease, have contributed to critique of GWASs (Gallagher and Chen-Plotkin, 2018; Strunz et al., 2020). A fruitful approach was developed to resolve the constraints and is known as the TWAS. TWASs have recently been broadly extended to prioritize genetic variants whose genetically controlled expression is correlated with diseases and complex traits (Gamazon et al., 2015). The TWAS is indeed a way to incorporate expression profile and GWAS that facilitates gene research correlated with interesting traits (Grinberg and Wallace, 2020). Gallagher and Chen-Plotkin (2018), in their report, they hypothesized that a greater focus on the subsequent functional analysis of the already established GWAS loci, instead of the quest for more and more GWAS loci, will most probably gain pathophysiological awareness.

**Single-cell RNA sequencing**

Single-cell RNA sequencing (scRNA-seq) approaches allow the examination of single-cell gene expression that dramatically revolutionizes transcriptomic studies. A variety of scRNA-seq techniques was developed and these approaches have unique characteristics with distinct strengths and weaknesses rendering to technological constraints and biological considerations, scRNA-seq data is not clean and more complex than bulk RNA-seq data. In recent years scRNA-seq has been extended to numerous species, in particular to multiple human tissues (including normal and cancer) and these experimentations have established considerable cell-to-cell gene expression heterogeneity (Grun et al., 2015; Chen et al., 2016b; Cao et al., 2017; Rosenberg et al., 2018). Every scRNA-seq method has its own merits and demerits, resulting in various scRNA-seq methods having distinct features and differential performance (Ziegenhain et al., 2017). Unique scRNA-seq technologies may need to be used to perform a single-cell transcriptomic analysis, considering the balance between the research purpose and the sequencing expense. The general steps involved in scRNA-seq is elaborated in Figure 2.

**Challenges**

For a decade and a half now, so much has been understood about RNAseq and its continuously upgraded methods that offer an open framework for transcriptional output profiling on a wide scale and hence have a wide array of uses. On the other hand, several challenges must also be addressed. In this summary, we primarily discuss the problems associated with the generation and analysis of the data.

In species lacking an annotated reference genome and with low mRNA characterization, the RNAseq study is still difficult and requires more improved computational tools for de novo gene assembly (Cloonan and Grimmond, 2008).

The typical Illumina technique depends on randomly priming double-stranded cDNA synthesis. A significant downside of this strategy is that the fragment's directionality cannot be established, and the exact orientation of the fragment to the genome strand is lacking (Hansen et al., 2010). For this cause, several strand-specific latest protocols to prepare RNAseq library has been designed (Levin et al., 2010) although strand-specific libraries demand more labor to generate (Cloonan et al., 2008).

Another complex challenge is the need for more computer space for data storage, without which there are problems of inconsistency in image processing and simple and low-quality readings (Wang et al., 2009). As with NGS
technological development, it causes the processing and storing of the vast volumes of data and images generated during the study. Cloud computing has been introduced as a way of solving this problem and tools are being developed for storing and analyzing RNAseq data in cloud applications. In addition, NCBI and EMBL have already started storing processed NGS sequence data in MINSEQE (minimum information about a high-throughput nucleotide sequencing experiment) format (Wilhelm and Landry, 2009; Langmead et al., 2010).

The mapping of millions of short reads produced from RNAseq to the reference sequence can take considerable computational time; hence, it is important to build bioinformatics tools specifically suited for a massive volume of short sequence readings. SOAP, MAQ, SSAH, ELAND, and BOWTIE are some of the new alignment technologies available for this phase (Wilhelm and Landry, 2009).

Another difficulty is the alleviation of the multi-match crisis. And larger is the difficulty for short reads that have higher copy numbers (> 100) and lengthy stretches of repeated regions. Longer sequence readings platforms, including 454 and PacBio, are usually used to address the problem (Cloonan et al., 2008; Mortazavi et al., 2008).

Microarray technology has developed various standardisation or normalisation methods to address the bias in data counts and statistical analysis, and the same has been adopted by the RNAseq but due to the particular variations between the distribution, dynamic range, and form of data produced from RNAseq and array technology, it is important to evolve methods specific to RNAseq analysis (Wang et al., 2009).

### Table 1 RNA sequencing platforms

<table>
<thead>
<tr>
<th>Platform</th>
<th>Sequencing chemistry</th>
<th>Method of amplification</th>
<th>Sequencing yield per run (Gb)</th>
<th>Run Type</th>
<th>Run time</th>
<th>Read length (bp)</th>
<th>Observed raw error rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roche GS FLX+</td>
<td>Pyrosequencing</td>
<td>Emulsion PCR</td>
<td>0.7</td>
<td>SE</td>
<td>20-23 hours</td>
<td>700</td>
<td>1</td>
</tr>
<tr>
<td>Illumina GAII</td>
<td>synthesis</td>
<td>Bridge amplifications</td>
<td>30</td>
<td>PE</td>
<td>10 days</td>
<td>150</td>
<td>0.76</td>
</tr>
<tr>
<td>Illumina HiSeq 2000/2500</td>
<td></td>
<td>clusters on the flow cell surface</td>
<td>600</td>
<td>PE</td>
<td>11 days</td>
<td>150</td>
<td>0.26</td>
</tr>
<tr>
<td>Illumina MiSeq V2</td>
<td></td>
<td></td>
<td>15</td>
<td>PE</td>
<td>27 hours</td>
<td>150</td>
<td>0.80</td>
</tr>
<tr>
<td>SOLiD 550x1</td>
<td>Ligation</td>
<td>Emulsion PCR on beads catalysed by DNA ligase</td>
<td>300</td>
<td>SE &amp; PE</td>
<td>10 days (SE)</td>
<td>14 days (PE)</td>
<td>85 &lt;0.1</td>
</tr>
<tr>
<td>Helicos</td>
<td>SMRT</td>
<td>No amplification</td>
<td>15</td>
<td>SE</td>
<td>10 days</td>
<td>30</td>
<td>5</td>
</tr>
<tr>
<td>Ion torrent PGM</td>
<td>Synthesis</td>
<td>Emulsion PCR</td>
<td>1.5-2</td>
<td>PE</td>
<td>2-5 hours</td>
<td>200-400</td>
<td>1</td>
</tr>
<tr>
<td>PacBio RS</td>
<td>SMRT</td>
<td>No amplification</td>
<td>0.1</td>
<td>SE</td>
<td>2 hours</td>
<td>15,000</td>
<td>10-15</td>
</tr>
</tbody>
</table>
Fig. 1 The basic workflow of RNAseq experiment

- RNA samples
- Depletion of rRNA
- cDNA library construction
- Next-gen sequencing
- QCs of sequences
- Read alignment to reference genome
- Quantification
- Differential gene expression analysis
- Functional profiling

Fig. 2 General steps involved in scRNA sequencing

- Isolation of single cells
- Read alignment
- Expression quantification of scRNA-Seq data
- Quality control of scRNA-Seq data
- Batch effect correction
- Differential expression analysis of scRNA-Seq data
- Cell subpopulation identification
- Dimensionality reduction and feature selection
- Imputation of scRNA-Seq data
- Normalization of scRNA-Seq data
- Cell lineage and pseudotime reconstruction
- Alternative splicing and RNA editing analysis of scRNA-Seq data
- Allelic expression exploration with scRNA-Seq data
- Gene regulatory network reconstruction

In general, the greater the genome coverage, the more and more complex the transcriptome gets, the more depth of sequencing is needed, and the cost increases proportionately. Similarly, is the condition when the analysis is carried out for a lesser abundant transcript. Thus, RNAseq studies should be tailored with due consideration of the research goals and the budget available. To reduce the sequencing depth, approaches such as running experimental multiplex designs can be implemented (Auer and Doerge, 2010). In their report, Robbles et al., (2012) indicated that the incorporation with biological replicates is more successful as compared to mechanical replicates and that the depth of sequencing could be decreased to 15% without a substantial impact on false-positive rates.

It remains a great challenge to investigate more complicated transcriptomes and to describe the expression profile of less abundant RNA isoforms. Indeed, the scope of long-read transcriptomics is expanding rapidly lately. Transcriptome assemblies largely based on long reads (such as PacBio, developed from 454) are favored as generally they produce extremely complete and consistent genomes (Koren et al., 2015), nevertheless, there are also cases in which shorter readings (Illumina) further enhance the results (De Maio et al., 2019). Therefore, in their report, Amarasinghe et al., (2020) suggested combining subsets of technologies such as nanopore / SMRT (generate strong contigs) with shorter-read (ensure base-level accuracy) sequencing or generating data from different platforms and combining the research (Anamika et al., 2015) to overcome these challenges to validate findings and exclude false positive.
Notably, there is no conclusion about the best-suited instrument for all laboratory settings, as several experiments have found performance differences in a method in varying circumstances (Soneson et al., 2013).

Depending on the availability of the reference genome, transcriptome assembly may be directed by reference or de novo, de novo transcriptome assembly typically takes longer and is more procedural than reference assembly (Garber et al., 2011). In addition, the continuity and completeness of the de novo assembled transcriptome are lower than that of the reference-based assembly particularly for data with lower sequencing depth (Lu et al., 2013).

Applications

Using cost-effective high-throughput sequencing techniques, it has resulted in the development of millions of readings at an unprecedented scale than the traditional techniques could, as advanced technology will give crucial information to the different applications that provide the central context for any current genetic analysis. One of the key areas of research is gene regulation, to locate genes that alter their expression in order to fully understand the molecular mechanisms used or altered or the regulatory components that are used in various stages of development, in a disease state relative to normal cells, or particular experimental stimuli relative to physiological or pathological conditions (Medrano et al., 2010; Shukla et al., 2017; Sulabh et al., 2019; Xu et al., 2019; Han et al., 2020; Panigrahi et al., 2020). Genome annotation is another application domain. Transcriptomic studies show novel phenomena involving RNA base modifications, such as micro exons, obscure exons, enhancer RNAs, fusion genes, epitranscriptome, etc. (Van den Berge et al., 2019). Several facets of the existing gene annotation can be modified with RNaseq's single-base resolution, including gene limits and introns for known genes and the identification of new transcribed regions (Nagalakshmi et al., 2008). RNaseq technology enables one to estimate alternate splicing events on an unbiased and genome-wide scale (Pan et al., 2008). There is substantial interest in using RNaseq in clinical applications (Chen et al., 2016; Cie'slik and Chinnaiyan, 2017) to increase genome sequencing knowledge. RNaseq also offers an approach to detect genetic variation, such as the single nucleotide polymorphism (SNP) variation associated with it (Cánovas et al., 2013). RNaseq offers a rare ability to classify allele-specific expression at high throughput in hundreds of loci in the study of complex traits (Wang, 2008). The incorporation of RNaseq technology with other high-throughput techniques such as Chromatin Immunoprecipitation Sequencing (ChIP-Seq) and SNP chip genome-wide association research will solve many long-standing animal genetics problems (Wickramasinghe et al., 2014; Yan et al., 2020). Another emerging application is Metatranscriptomics-NGS technology for microbial transcriptome analysis (Gosalbes et al., 2011). Other uses provide an effective means of finding novel non-coding RNAs and of researching in depth the non-coding RNA variants and the RNA editing mechanism (Morozova et al., 2009; Park et al., 2012). Using RNaseq technology, a researcher can study not only the transcribed field of the genome, but the non-translated regions and introns of 3' and 5' (Wickramasinghe et al., 2014). Thanks to its cost-effectiveness and less background noise, RNaseq sequence reads may be used as an efficient alternative to the cDNA microarray technique to create de novo gene models in animals (Denoeud et al., 2008; Nagalakshmi et al., 2008). Having discovered that gene expression levels differ considerably from cell to cell, researchers concentrate on finding new cell types or gene
expression dissection at single-cell resolution to investigate cell heterogeneity and diversity, scRNA-seq (Single-cell RNA-Seq) (Dal Molin et al., 2017; Chen et al., 2019). The research, however, suggested the need for developments in modern methods that can efficiently resolve the higher technological noise compared to the bulk RNAseq analysis.

In conclusions, RNAseq has helped us to produce an outstanding global view of the transcriptome and its detailed structure for many organisms and types of cells. RNAseq has strong advantages over Transcriptomic methods previously developed. Technologies such as pair-end sequencing, strand-specific sequencing, and the use of longer readings to expand coverage and depth will overcome the difficulties and advance the RNAseq objectives. RNAseq data are indeed an intermediary phase in exploration in which the molecular modifications observed constitute candidates for further applications. Continuous innovation in RNAseq technology and parallel developments in bioinformatics techniques to further develop and mitigate challenges will significantly accelerate biological and clinical research and provide in-depth insights into gene expression while the future study will continue, it is not untrue to suggest that substantial progress has been accomplished so far, resulting in better quality more accurate NGS expression. In the end, RNAseq's success lies in its large variety of uses and lower sequencing costs as well as lower error rates and correctly resolving the due limitations and obstacles.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

This review-manuscript preparation did not require any sort of funding as such.

**References**


Blignaut, M. 2012. Review of non-coding


Methods, 5:621-8.
Trapnell, C., B. A. Williams, Pertea, G. et al.,


**How to cite this article:**