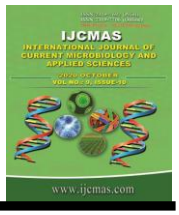
Int.J.Curr.Microbiol.App.Sci (2020) *9*(10): 891-905

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Volume 9 Number 10 (2020)**
Journal homepage: http://www.ijcmas.com



Review Article https://doi.org/10.20546/ijcmas.2020.910.107

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*

Keywords

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

Article Info

Accepted:
10 September 2020
Available Online:
10 October 2020

A B S T R A C T

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.

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 behavr 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

et al., 2017). 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 reversed-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 analysed 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

Platform	Sequencing chemistry	Method of amplification	Sequencing yield per run (Gb)	Run Type	Run time	Read length (bp)	Observed raw error rate (%)
Roche (454 GS FLX+)	Pyrosequencing	Emulsion PCR	0.7	SE	20-23 hours	700	1
Illumina GAIIx	synthesis	Bridge amplifications clusters on the flow cell surface	30	PE	10 days	150	0.76
Illumina HiSeq 2000/2500			600	PE	11 days	150	0.26
Illumina MiSeq V2			15	PE	27 hours	150	0.80
SOLiD-550xI	Ligation	Emulsion PCR on beads catalysed by DNA ligase	300	SE & PE	10 days (SE) 14 days (PE)	85	<0.1
Helicos	SMRT	No amplification	15	SE	10 days	30	5
Ion torrent PGM	Synthesis	Emulsion PCR	1.5-2	PE	2-5 hours	200-400	1
PacBio RS	SMRT	No amplification	0.1	SE	2 hours	15,000	10-15

Fig.1 The basic workflow of RNAseq experiment

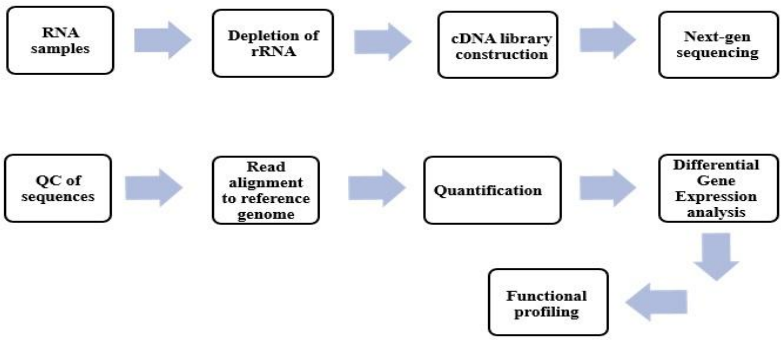
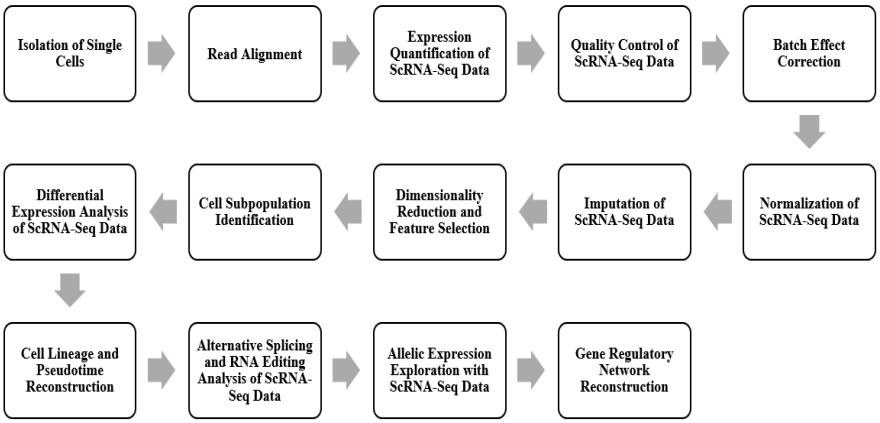


Fig.2 General steps involved in scRNA sequencing



In general, the greater the genome coverage, 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.,* 2016a; 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

Adams, M. D. *et al.,* 1992. Sequence identification of 2,375 human brain genes. Nature 1992, 355:632–634.

Adiconis X, D. Borges-Rivera, R. Satija, D. S. DeLuca, M. A. Busby, A. M. Berlin, A. Sivachenko, D. A. Thompson, A. Wysoker, T. Fennell, *et al.,* 2013. Comparative analysis of RNA sequencing methods for degraded or low input samples. Nat Methods, 10:623–629. doi:10.1038/nmeth.2483.

Amarasinghe, S. L., S. Su, X. Dong, L. Zappia, M. E. Ritchie and Gouill, Q. 2020. Opportunities and challenges in long-read sequencing data analysis. Genome Biology 21:30 https://doi.org/10.1186/s13059-020-1935-5

Anamika, K., S. Verma, A. Jere and Desai, A. 2015. Transcriptomic Profiling Using Next Generation Sequencing - Advances, Advantages, and Challenges. http://dx.doi.org/10.5772/61789.

Arnold, C. D., D. Gerlach, C. Stelzer, L. M. Boryn, M. Rath and Stark, A. 2013. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science. 339(6123):1074–7.

Auer, P. L. and Doerge, R. W. 2010. Statistical design and analysis of RNA sequencing data. Genetics 185, 405–416.

Berger, M. F., J. Z. Levin, K. Vijayendran, A. Sivachenko, X. Adiconis, J. Maguire, L. A. Johnson, J. Robinson, R. G. Verhaak, C. Sougnez, R. C. Onofrio, L. Ziaugra, K. Cibulskis, E. Laine, J. Barretina, W. Winckler, D. E. Fisher, G. Getz, *et al.,* 2010. Integrative analysis of the melanoma transcriptome. Genome Res, 20:413-27.

Bennett S. 2004. Solexa Ltd, Pharmacogenomics 5, 433-438.

Blignaut, M. 2012. Review of non-coding

RNAs and the epigenetic regulation of gene expression. Epigenetics. 7(6):664–666. DOI: 10.4161/epi.20170.

Buermans, H. P. J. and den Dunnen, J. T. 2014. Next generation sequencing technology: Advances and applications. Biochimica et Biophysica Acta 1842, 1932–1941.

Buniello, A., J. A. L. MacArthur, M. Cerezo, L. W. Harris, J. Hayhurst, Malangone, C. *et al.,* 2019. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Research. 47(D1): D1005{D1012. doi:10.1093/nar/gky1120.

Cánovas, A., G. Rincon, A. Islas-Trejo, R. Jimenez-Flores, A. Laubscher and Medrano J, F. 2013. RNA sequencing to study gene expression and single nucleotide polymorphism variation associated with citrate content in cow milk. J. Dairy Sci. 96, 2637–2648.

Cao, J., J. S. Packer, V. Ramani, D. A. Cusanovich, C. Huynh, Daza, R. *et al.*, 2017. Comprehensive single-cell transcriptional profiling of a multicellular organism. Science 357, 661–667. doi: 10.1126/science. aam8940

Chen, Q., G. He, W. Zhang, T. Xu, H. Qi, J. Li, Y. Zhang and Gao, M. Q. 2016a. Stromal fibroblasts derived from mammary gland of bovine with mastitis display inflammation-specific changes. Sci. Rep. 6, 27462. (doi:10.1038/srep27462)

Chen, G., J. P. Schell, J. A. Benitez, S. Petropoulos, M. Yilmaz, Reinius, B. *et al.,* 2016b. Single-cell analyses of X Chromosome inactivation dynamics and pluripotency during differentiation. Genome Res. 26, 1342–1354. doi: 10.1101/gr.201954.115

Chen, G., B. Ning and Shi, T. 2019. Single-cell RNA-seq technologies and related computational data analysis. Front. Genet. 10:317. doi:10.3389/fgene.2019.00317

Chu, Y. and Corey, D. R. 2012. RNA Sequencing: Platform Selection, Experimental Design, and Data Interpretation. NUCLEIC ACID THERAPEUTICS 22 (4). Mary Ann Liebert, Inc., DOI:10.1089/nat.2012.0367

Cie´slik, M. and Chinnaiyan, A. M. 2017. Cancer transcriptome profiling at the juncture of clinical translation. Nat. Rev. Genet. 19:93–109

Cloonan, N. and Grimmond, S. 2008. Transcriptome content and dynamics at single-nucleotide resolution. Genome Biol. 9, 234.

Cloonan, N. *et al.,* 2008. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature Methods* 5, 613–619.

Corney, D. C. and Basturea, G. N. 2013. RNA-seq. *Mater methods*, 3:203.

Costa-Silva, J., D. Domingues and Lopes, F. M. 2017. RNAseq differential expression analysis: an extended review and a software tool. PLoS ONE 12, e0190152. doi:10.1371/journal.pone. 0190152.

Dal Molin, A., G. Baruzzo and Di Camillo, B. 2017. Single-Cell RNA-Sequencing: Assessment of Differential Expression Analysis Methods. Front. Genet. 8:62. doi: 10.3389/fgene.2017.00062

De Maio, N., L. P. Shaw, A. Hubbard, S. George, N. D. Sanderson, J. Swann, *et al.,* 2019. Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. Microb Genom. 5(9): https://doi.org/10.1099/mgen.0.000294.

Delseny, M., B. Han and Hsing, Y. 2010. High throughput DNA sequencing: The new sequencing revolution. *Plant. Sci*. 179, 407–422.

Denoeud, F., J. -M. Aury, C. Da Silva, B. Noel, O. Rogier, M. Delledonne, M. Morgante, G. Valle, P. Wincker, C. Scarpelli, O. Jaillon and Artigue- nave, F. 2008. Annotating genomes with massive-scale RNA sequencing. Genome Biol. 9, R175.

Esteve-Codina, A. 2018. RNA-Seq Data Analysis, Applications and Challenges. *Comprehensive Analytical Chemistry*. https://doi.org/10.1016/bs.coac.2018.06.001.

Farkas, M. H., E. D. Au, M. E. Sousa and Pierce, E. A. 2015. RNAseq: improving our understanding of retinal biology and disease. Cold Spring Harbor Perspect. Med. 5, a017152. doi:10.1101/cshperspect.a017152.

Gallagher, M. D. and Chen-Plotkin, A. S. 2018. The post-GWAS era: from association to function. Am. J. Hum. Genet 102, 717–730 (2018). [PubMed: 29727686].

Gamazon, E. R., H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E. Eyler, J. C. Denny, D. L. Nicolae, Cox, N. J. *et al.,* 2015. A gene-based association method for mapping traits using reference transcriptome data. Nat. Genet., 47, 1091–1098.

Garber, M., M. G. Grabherr, M. Guttman and Trapnell, C. 2011. Computational methods for transcriptome annotation and quantification using RNASeq. Nat Methods., 8:469–477. DOI: 10.1038/nmeth.1613

Gonzalez-Garay, M. L. 2016. Introduction to isoform sequencing using Pacific Biosciences Technology (Iso-Seq). In Transcriptomics and Gene Regulation, ed. J Wu, pp. 141–60. Dordrecht, Neth.: Springer Neth.

Gosalbes, M. J., A. Durban, M. Pignatelli, J. J. Abellan, N. Jimenez-Hernandez, A. E. Perez-Cobas, A. Latorre, A. Moya. 2011. Metatranscriptomic approach to analyze the functional human gut microbiota. PLoS ONE. 6(3):e17447. DOI: 10.1371/journal.pone.0017447

Grinberg, N. F. and Wallace, C. 2020. Multi-tissue transcriptome-wide association studies. bioRxiv preprint doi: https://doi.org/10.1101/2020.07.13.201111.

Grun, D., A. Lyubimova, L. Kester, K. Wiebrands, O. Basak, Sasaki, N. *et al.,* 2015. Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature 525, 251–255. doi: 10.1038/nature14966

Gupta, J. P., B. Bhushan, M. Panigrahi, *et al.,* 2016. Study on genetic variation of short tandem repeats (STR) markers and their association with somatic cell scores (SCS) in crossbred cows. : Indian Journal of Animal Research. 50 (4) pp.450-454.

Gusev, A., A. Ko, H. Shi, G. Bhatia, W. Chung, B.W.J.H. Penninx, R. Jansen, E. J. de Geus, D. I. Boomsma, Wright, F. A., *et al.,* 2016. Integrative approaches for large-scale transcriptome wide association studies. Nat. Genet., 48, 245–252.

Han, Y., S. Gao, K. Muegge, W. Zhang and Zhou, B. 2015. Advanced applications of RNA sequencing and challenges. Bioinform Biol Insights, 9:29–46. doi:10.4137/BBI.S28991.

Han, Z., Y. Fan, Z. Yang, J. J. Loor and Yang, Y. 2020. Mammary Transcriptome Profile during Peak and Late Lactation Reveals Differentially Expression Genes Related to Inflammation and Immunity in Chinese Holstein. Animals, 10, 510; doi:10.3390/ani10030510.

Hansen, K. D., S.E. Brenner and Dudoit, S. 2010. Biases in Illumina tran- scriptome sequencing caused by random hexamer priming, Nucleic Acids Res. 38 (2010)

e131.

Hrdlickova, R., M. Toloue and Tian, B. 2016. RNAseq methods for transcriptome analysis. Wiley Interdiscip. Rev. RNA 8, e1364. (doi:10.1002/wrna.1364).

Koren, S. and Phillippy, A. M. 2015. One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Curr Opin Microbiol*. 23:110–20.

Kovaka, S., A. V. Zimin, G. M. Pertea, R. Razaghi, S. L. Salzberg and Pertea, M. 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biology 20:278 https://doi.org/10.1186/s13059-019-1910-1

Kuosmanen, A., T. Norri and Mäkinen, V. 2018. Evaluating approaches to find exon chains based on long reads. Brief Bioinform., 19:404–14.

Langmead, B., K. Hansen and Leek, J. 2010. Cloud-scale RNA-sequencing differential expression analysis with Myrna. Genome Biol. 11, R83.

Levin, J. Z., M. Yassour, X. Adiconis, C. Nusbaum, D. A. Thompson, N. Friedman, A. Gnirke and Regev, A. 2010. Comprehensive comparative analysis of strand-specific RNA sequencing methods, Nat. Methods 7:709–715.

Liu, L., Y. Li, S. Li, N. Hu, Y. He, R. Pong, D. Lin, L. Lu and Law, M. 2012. Comparison of Next-Generation Sequencing Systems. Journal of Biomedicine and Biotechnology, Article ID 251364, 11 pages doi:10.1155/2012/251364.

Lu, B., Z. Zeng and Shi T. 2013. Comparative study of de novo assembly and genome-guided assembly strategies for transcriptome reconstruction based on RNASeq. Sci China Life Sci. 56(2):143–155. DOI: 10.1007/s11427-013-4442-z

Mackenzie, R.J. 2018. RNAseq: Basics, Applications and Protocol. Science Writer for Technology Networks. Accessed on 21st Aug. 2020. https://www.technologynetworks.com/genomics/articles/RNAseq-basics-applications-and-protocol-299461

Maher, C. A., C. Kumar-Sinha, Cao, X., *et al.*, 2009. Transcriptome sequencing to detect gene fusions in cancer. *Nature*. 458(7234):97–101.

Mardis, E. R. 2008. The impact of next-generation sequencing technology on genetics. Trends in Genetics, 24(3), pp. 133-141.

Medrano, J. F., G. Rincon and Islas-Trejo, A. 2010. Comparative analysis of bovine milk and mammary gland transcriptome using RNAseq. Department of Animal Science, University of California, Davis, California 95616, USA. https://www.researchgate.net/publication/262485617

Mellmann, A., D. Harmsen, Cummings, C. A. *et al.*, 2011. "Prospective genomic characterization of the german enterohemorrhagic Escherichia coli O104:H4 outbreak by rapid next generation sequencing technology," PLoS ONE, vol. 6, no. 7, Article ID e22751.

Minoche, A., J. Dohm and Himmelbauer, H. 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. Genome Biol. 12, R112.

Morozova, O., M. Hirst and Marra, M. A. 2009. Applications of new sequencing technologies for transcriptome analysis. Annu. Rev. Genomics Hum. Genet. 10,135–151.

Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer and Wold, B. 2008: Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat

Methods, 5:621-8.

Nagalakshmi, U. *et al.*2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320, 1344–1349.

Pan, Q., O. Shai, L. J. Lee, B. J. Frey and Blencowe, B. J. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*. 40(12):1413–5.

Panigrahi, M., Kumar, H., Sah, V., Dillipkumar Verma, A., Bhushan, B., and Parida, S. 2020. Transcriptome profiling of buffalo endometrium reveals molecular signature distinct to early pregnancy. *Gene, 743*, 144614. https://doi.org/10.1016/j.gene.2020.144614

Park, E., B. Williams, B. J. Wold and Mortazavi, A. 2012. RNA editing in the human ENCODE RNA-seq data. Genome Res. 22:1626–33.

Reinartz, J., E. Bruyns, J. -Z. Lin, T. Burcham, S. Brenner, B. Bowen, M. Kramer and Woychik, R. 2002. Massively parallel signature sequencing (MPSS) as a tool for indepth quantitative gene expression profiling in all organisms. Brief. Funct. Genomics Proteomics 1,95–104.

Robertson, G., J. Schein, Chiu, R. *et al.*, 2010. De novo assembly and analysis of RNA-seq data. *Nat Methods*., 7(11):909–12.

Robles, J.A., S. E. Qureshi, S. J. Stephen, S. R. Wilson, C. J. Burden and Taylor, J.M. 2012. Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. BMC Genomics 13, 484.

Rosenberg, A.B., C. M. Roco, R. A. Muscat, A. Kuchina, P. Sample, Yao, Z., *et al.*, 2018. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science 360, 176-182. doi:10.1126/science.aam8999.

Shabalina, S. A and Spiridonov, N. A. 2004. The mammalian transcriptome and the function of non-coding DNA sequences. Genome Biol. 5(4):105. DOI: 10.1186/gb-2004-5-4-105

Shukla, S. K., S. Shukla, A. Chauhan, Sarvjeet, R. Khan, A. Ahuja, L. V. Singh, N. Sharma, C. Prakash, A. V. Singh and Panigrahi, M. 2017. Differential gene expression in Mycobacterium bovis challenged monocyte-derived macrophages of cattle. Microbial pathogenesis, 113, 480–489. https://doi.org/10.1016/j.micpath.2017.11.030

Soneson, C. and Delorenzi, M. A. 2013. comparison of methods for differential expression analysis of RNA-se data. BMC Bioinformatics.14:91. DOI: 10.1186/1471-2105-14-91.

Strunz, T., S. Lauwen, K. Christina, International AMD Genomics Consortium (IAMDGC), A. den Hollander and Weber, B. H. F. 2020. A transcriptome-wide association study based on 27 tissues identifies 106 genes potentially relevant for disease pathology in age-related macular degeneration Scientific Reports | *10:1584* | https://doi.org/10.1038/s41598-020-58510-9.

Sulabh, S. M. Panigrahi, S. Kumar, R. Varshney, A. Verma, N. A. Baba, J. P. Gupta, A. Chauhan, P. Kumar, T. Dutt and Bhushan, B. 2019. Differential cytokine response of E coli LPS stimulated PBMCs in crossbred, Tharparkar cattle and Murrah buffalo - An in vitro study. Spanish Journal of Agricultural Research, Volume 17, Issue 1, e0501. https://doi.org/10.5424/sjar/2019171-1.

Trapnell, C., B. A. Williams, Pertea, G. *et al.*,

2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*., 28(5):511–5.

Usman, T., Y. Wang, C. Liu, Y. He, X. Wang, Y. Dong, H. Wu, A. Liu and Yu, Y. 2017. Novel SNPs in IL-17F and IL-17A genes associated with somatic cell count in Chinese Holstein and Inner-Mongolia Sanhe cattle. *Journal of Animal Science and Biotechnology*, 8:5. DOI 10.1186/s40104-016-0137-1.

Van den Berge, K., K. M. Hembach, C. Soneson, S. Tiberi, L. Clement, M. I. Love, R. Patro and Robinson, M. D. 2019. RNA Sequencing Data: Hitchhiker's Guide to Expression Analysis. *Annu. Rev. Biomed. Data Sci*., 2:139–73

Van Dijk, E. L., Y. Jaszczyszyn and Thermes, C. 2014. Library preparation methods for next-generation sequencing: tone down the bias. Exp Cell Res 2014, 322:12–20. doi:10.1016/j.yexcr.2014.01.008.

Velculescu, V. E., L. Zhang, B. Vogelstein and Kinzler, K. W. 1995. Serial analysis of gene expression. Science, 270:484–487.

Wang, X., Q. Sun, S, D. McGrath, E. R. Mardis, P. D. Soloway and Clark, A.G. 2008. Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. PLoS One 3, e3839.

Wang, Z., M. Gerstein and Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10, 57–63. https://doi.org/10.1038/nrg2484

Wickramasinghe, S., A. Cánovas, G. Rincón and Medrano, J. F. 2014. RNA-Sequencing: A tool to explore new frontiers in animal genetics. Livestock Science 166, 206–216. http://dx.doi.org/10.1016/j.livsci.2014.06.015.

Wilhelm, B.T. and Landry, J. R. 2009. RNA-Seq—quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* 48, 249–257.

Wilson, N. K., Kent, D. G., Buettner, F., *et al.*, 2015. Combined single-cell functional and gene expression analysis resolves heterogeneity within stem cell populations. *Cell Stem Cell*. 16(6):712–24.

Xu, T., R. Deng, X. Li, Y. Zhang and Gao, M. Q. 2019. RNAseq analysis of different inflammatory reactions induced by lipopolysaccharide and lipoteichoic acid in bovine mammary epithelial cells. *Microbial Pathogenesis.* 130, pp. 169-177. https://doi.org/10.1016/j.micpath.2019.03.015.

Yan, Z. *et al.*, 2020. Integrating RNAseq with GWAS reveals novel insights into the molecular mechanism underpinning ketosis in cattle. BMC Genomics 21:489 https://doi.org/10.1186/s12864-020-06909-z

Ziegenhain, C., B. Vieth, S. Parekh, B. Reinius, A. Guillaumet-Adkins, Smets, M. *et al.*, 2017. Comparative analysis of single-cell RNA sequencing methods. Mol. Cell 65, 631–643.e4. doi: 10.1016/j.molcel.2017.01.023

How to cite this article:

Kaiho Kaisa, Harshit Kumar, Manjit Panigrahi, Triveni Dutt and Bharat Bhushan. 2020. RNA Sequencing: A Potent Transcription Profiling Tool. *Int.J.Curr.Microbiol.App.Sci.* 9(10): 891-905. doi: https://doi.org/10.20546/ijcmas.2020.910.107