We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the
TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

Assessing Host-Pathogen Interaction Networks via RNA-Seq Profiling: A Systems Biology Approach

Sudhesh Dev Sareshma and Bhassu Subha

Abstract

RNA sequencing is a valuable tool brought about by advances in next generation sequencing (NGS) technology. Initially used for transcriptome mapping, it has grown to become one of the 'gold standards' for studying molecular changes that occur in niche environments or within and across infections. It employs high-throughput sequencing with many advantages over previous methods. In this chapter, we review the experimental approaches of RNA sequencing from isolating samples all the way to data analysis methods. We focus on a number of NGS platforms that offer RNA sequencing with each having their own strengths and drawbacks. The focus will also be on how RNA sequencing has led to developments in the field of host-pathogen interactions using the dual RNA sequencing technique. Besides dual RNA sequencing, this review also explores the application of other RNA sequencing techniques such as single cell RNA sequencing as well as the potential use of newer techniques like 'spatialomics' and ribosome-profiling in host-pathogen interaction studies. Finally, we examine the common challenges faced when using RNA sequencing and possible ways to overcome these challenges.

Keywords: RNA-Seq, transcriptome, next generation sequencing, systems biology, host-pathogen interactions

1. Introduction

1.1 RNA sequence profiling

RNA sequencing (most commonly abbreviated as RNA-Seq) is an advanced sequencing approach that has transformed the way we look at the intricacies that exist within complex biological systems. Using high-throughput next generation sequencing (NGS) technology, RNA-Seq allows the detection and quantification of RNA transcripts in a biological sample with high accuracy [1]. Further analysis of RNA-Seq data can reveal a dynamic scale of information ranging from alternative spliced transcripts, gene fusions, single nucleotide polymorphisms (SNPs), post-translational modifications, temporal fluctuations in RNA expression during infection across cells [2–5]. This extensive capability of RNA-Seq has also recently found its way into studies investigating host-pathogen interaction networks with hopes of further elucidating this multi-faceted system [6, 7].

One of the earliest papers describing the term 'RNA-Seq' successfully mapped the transcriptome of the yeast genome using a high-throughput sequencing platform [8]. In fact, a handful of studies had already started using the RNA-Seq method even before the term was coined [9–13]. Commonly referred to as 'transcriptome sequencing', these studies mainly adopted the massively parallel pyro-sequencing technology which was one of the newer sequencing technologies at the time [14]. While DNA sequencing and genomic studies have led to many breakthroughs, RNA-Seq brings forth a more functional, integrated view of expressed genes with distinct advantages over previous methods. Different aspects of RNA-Seq will be discussed in the following sections leading to its role in unravelling host-pathogen interaction networks.

2. Introduction to RNA Seq approaches in biology and medicine

Transcriptomics is an area that is being continuously developed especially with the recent advances in technology that make it easier to carry out large-scale analysis of RNA. Prior to the use of RNA-Seq, traditional methods used to study transcriptomes include hybridization-based, sequence-based and tag-based approaches [15]. A popular hybridization-based approach is the use of microarrays. The main principle behind microarrays is complementary binding of nucleotides. A microarray or 'gene chip' is prepared containing thousands of different oligonucleotides or cDNA molecules [16]. Extracted RNA samples converted into cDNA are fluorescently labelled and allowed to hybridise on the microarray [17]. This approach has proven to be useful in studies looking to compare the levels of gene expression but it does not generate quantitative values and can only be used for known genes [18]. A related method called genome tiling array, however, has the ability to examine genomic regions without prior knowledge of its expression [19]. Like any other method scrutinised over time, the pitfall of microarrays stem from inconsistent protocols, high background noise due to cross-hybridization, low technical reproducibility as well as other technical issues [20, 21].

As for sequence-based approaches, a method used for gene discovery early on was expressed sequence tags (ESTs), which are single-pass sequence reads selected from cDNA libraries [22]. Aside from being expensive, the single-pass reads produced using this method are more prone to error and likely to have redundancies in large datasets [23]. On the other hand, tag-based approaches like serial analysis of gene expression (SAGE) and massively parallel signature sequencing (MPSS) employ the principle of generating short 'tags' (9–20 base pairs) which are then sequenced and quantified on a large scale [24, 25]. Both methods make use of bead-based technology and produce accurate quantitative levels of gene expression but mostly focusing on the 3′-ends [26]. Cap analysis of gene expression (CAGE) was then introduced to examine 5′-end short tag sequences revealing more information about promoters and transcription start sites [27]. Altogether, these relatively costly methods were common during the Sanger sequencing era and could only be optimally used in conjunction with already known genome or EST databases. In addition, these approaches had limitations such as cloning biases, technical challenges and general lack of strength to be solid stand-alone approaches for transcriptome analysis [28, 29].

After decades of utilising Sanger sequencing, the development of Next Generation Sequencing (NGS) was a giant leap for researchers everywhere. There has been constant development in NGS technologies hence they can be more distinctly categorised as second-, third- and even fourth generation sequencing.

Second generation sequencing mainly consists of two methods which are sequencing by hybridization (SBH) and sequencing by synthesis (SBS) [30]. SBH was the main principle behind microarray technology using known DNA sequences as explained previously. Meanwhile, SBS is different from Sanger sequencing because dideoxy terminators are not used. In addition, it employs repeated cycles of nucleotide incorporation and also tiny-volume reactions that are massively run in parallel. Most second generation methods commonly rely on sequencing reactions that take place in micro wells or channels [30]. One of the most common second generation sequencing technology is developed by Illumina, producing short read lengths. On the other hand, third- and fourth generation sequencing technologies are more focused on producing longer read lengths. These technologies have creatively exploited the principle of sequencing reactions occurring in millions of tiny wells either by specially engineered chambers or biological nanopores [30]. The front runners of third- and fourth generation sequencing are currently Pacific Biosciences and Oxford Nanopore Technologies. Their technologies will be discussed in the coming sections. Also known as deep sequencing, these high-throughput sequencing technologies eventually led to the development of next generation RNA-Seq. Originally described by Nagalakshmi et al. [8], preliminary RNA-Seq studies focused on improving genomic annotation by examining novel untranslated regions, promoter regions, intergenic transcripts, alternative gene splicing events and single nucleotide polymorphisms (SNPs) among others [31–35]. Advances in next generation RNA-Seq has allowed diverse studies spanning areas like diagnosis of genetic conditions, characterisation of immune microenvironments, understanding cellular frameworks and viral genetics [36–40]. **Table 1** shows a comparison of RNA-Seq with some of the main methods used to study the transcriptome.

	Microarray	SAGE*	Next-Gen* RNA-Seq
Type of method	Hybrid-based	Tag-based	cDNA library preparation & high throughput sequencing
Amount of input material	High	High	Low
Probes	Yes	No	No
Cost	Medium	Low	High
Data analysis	Based on relative intensity	Based on amplified SAGE tag counts	Based on amplified & sequenced cDNA fragments producing raw read counts
Detection of novel genes/transcripts	No	Limited	Yes
Detection of alternatively spliced isoforms	Limited	No	Yes
Detection of single nucleotide polymorphisms	No	No	Yes
Detection of non-coding transcripts	Limited	Limited	Yes
Prior knowledge of gene sequence	Yes	Limited	No

SAGE – Serial Analysis of Gene Expression.
Next-Gen – Next Generation.

Table 1.
Comparison of commonly used methods for gene and transcriptome analysis.

2.1 Experimental flow in approaches

The flow chart in **Figure 1** shows the initial steps involved when carrying out an RNA-Seq experiment.

The first step in an RNA-Seq experiment is to isolate RNA from any biological sample (e.g. cell or tissue populations). As a quality control step, the integrity of extracted RNA samples is commonly measured using an Agilent Bioanalyzer. Based on electrophoretic separation of RNA and a built-in software algorithm, it produces an RNA Integrity Number (RIN) depicting levels of RNA degradation [43]. The next step involves either an enriching or depleting procedure to select specific RNA species. In any given total RNA sample, a variety of RNA species would be present including messenger RNAs, ribosomal RNAs, precursor RNAs, non-coding RNAs,
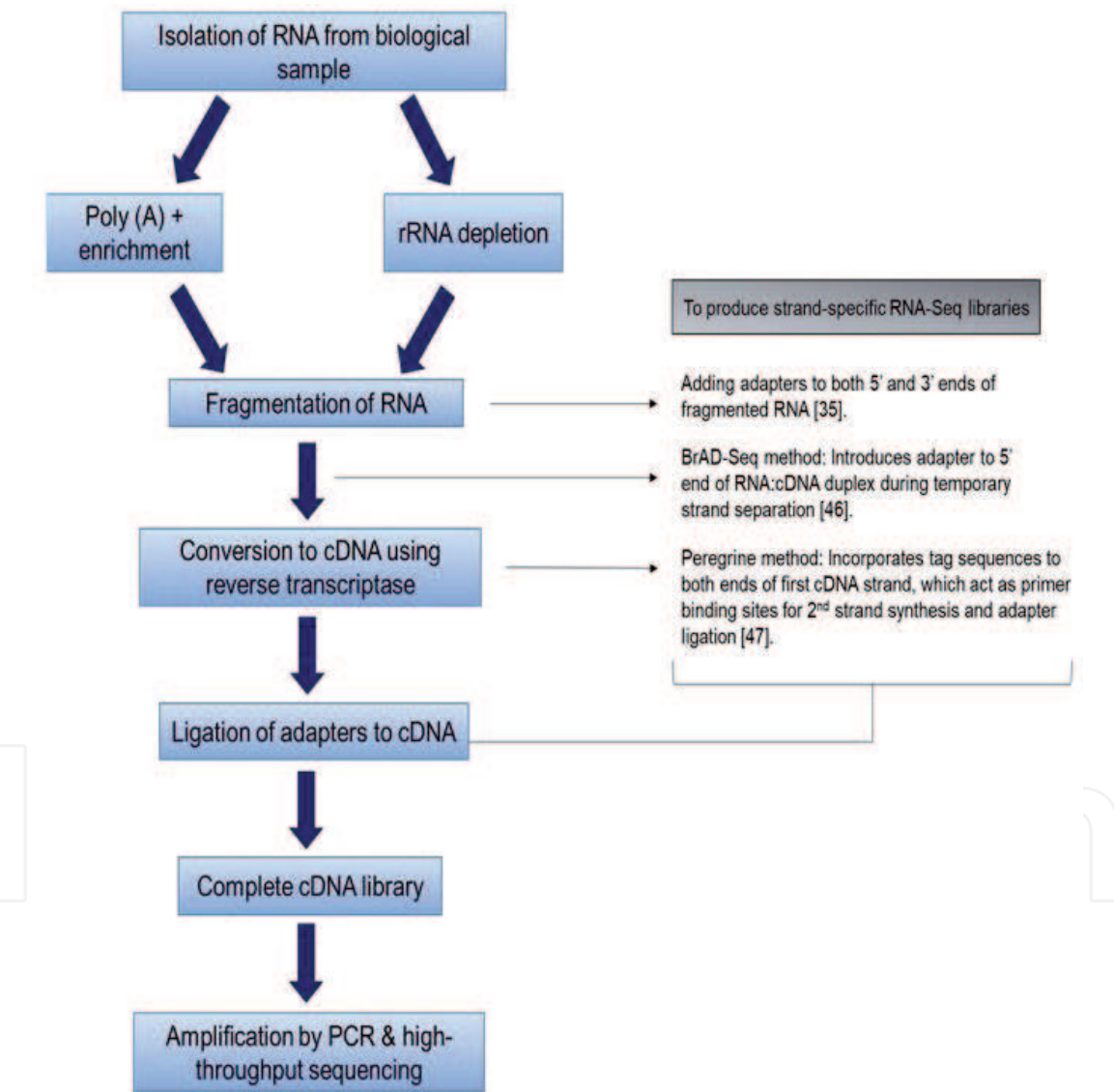

Figure 1.
Overview of second generation RNA-Seq workflow. Firstly, RNA samples are extracted from biological samples. Selection of specific RNA species is carried out either by enriching transcripts expressing poly-adenylated (poly-A) tails (usually mRNA) or by removing the abundant ribosomal RNAs (rRNAs). Next, the enriched or depleted RNA samples are fragmented followed by reverse transcription to generate cDNA. The next step is ligation of adapters, however, standard adapter ligation loses information about RNA strand-specificity hence a few methods have been developed to prevent this. These include adding adapters directly to the 5′ and 3′ ends of fragmented RNA [31], the BrAD-Seq method which adds an adapter to 5′ end of the RNA:cDNA duplex during reverse transcription [41], and lastly the Peregrine method which incorporates tag sequences to 5′ and 3′ ends of the first cDNA strand [42]. Once library preparation is completed, samples are amplified by PCR and RNA-Seq library is now ready to be sequenced.

etc. A bulk of the RNA portion (~95%) in most cells comprises of rRNA which if not removed, would make up a large part of the sequencing reads. Since this would largely restrict the study of less-abundant RNA species, protocols were created to circumvent this issue.

One such protocol is the enrichment of polyadenylated (poly-A) RNAs. This procedure selects for poly (A) + RNA mainly mRNA and exploits the fact that rRNAs generally lack this structure. A particular study however did find the presence of rRNA polyadenylation but only in small amounts [44]. This selection step can be carried out by using magnetic beads coated with oligo-dT or reverse transcription (RT) using oligo-dT primers [45]. An alternative step is rRNA depletion which serves to eliminate them from total RNA samples. There are various approaches used by different researchers for this method. One such approach uses probes like biotinylated DNA or locked nucleic acid which are allowed to hybridise to rRNAs. This is followed by a depleting step using streptavidin beads [46]. Another method that can be used for rRNA depletion is known as probe-directed degradation (PDD). This method involves obtaining cDNA:RNA duplexes, circularising them and then hybridising them with rRNA-specific probes. The final step involves digestion with Duplex-Specific Nuclease (DSN) which renders the hybridised-sequences unusable [47]. Some researchers also use not-so-random (NSR) primers that bind to specific RNA molecules during RT, excluding rRNAs [48]. In essence, the variety of methods that exist for rRNA depletion focuses on unique features of rRNA that can be singled out and developed into an eliminating step. The choice of using either poly (A) + selection or rRNA depletion ultimately depends on the aims of the experiment. Evaluation of these two methods showed that while rRNA depletion could record more unique characteristics of the transcriptome, poly(A) + selection was more accurate in terms of gene quantification [49].

Following poly (A) + enrichment or rRNA depletion, RNA samples need to be fragmented to shorter sequences according to the size restrictions of sequencing platforms. RNAs are usually fragmented chemically using alkaline solutions, divalent cations or enzymes [45]. Alternatively, RNA can be reverse transcribed (RT) first followed by cDNA fragmentation. Similarly, enzymes like DNAses can be used to fragment cDNA with recent advances including a transposon-based approach [50]. Next, either fragmented RNAs or cDNAs are ligated with adapters that are specific to the sequencing platform to be used. This step however overlooks RNA directionality whereby there is lack of information about DNA strands and their corresponding sense RNA strands. This may impede the identification of novel RNA species and also make it harder to accurately measure sense RNA expression [45]. Methods have been developed to preserve this directionality and they can be carried out either directly on fragmented RNA, cDNA or even on RNA:cDNA hybrids that are formed during RT. One of these approaches include adding distinct adapters to the 5′ and 3′ ends of fragmented RNA [31]. This difference in sequences at both ends preserve the strandedness of RNA. Other methods to preserve strand-specificity of RNA are BrAD-Seq [41] and the Peregrine method [42]. The BrAD-Seq method exploits the transient strand separation or 'breathing' of RNA:cDNA hybrid during reverse transcription to add an adapter to the 5′ end of the duplex. This is followed by incorporation of nucleotides by *E.coli* DNA Polymerase I to form the second strand and eventually a complete strand-specific cDNA library. Meanwhile, the Peregrine method incorporates short unidentical tag sequences to the ends of cDNA during first strand synthesis. These then serve as primer binding sites for subsequent adaptor ligation during second strand synthesis.

Finally, after cDNA synthesis and adapter ligation, cDNA libraries need to be amplified using PCR. Once amplified, they are ready for sequencing using a chosen NGS sequencing technology.

3. Next-generation sequencing technologies

3.1 Illumina, second generation sequencing technology

In 2005, Solexa released the Genome Analyser which established a quality standard for the transformation of sequencing platforms that came after. Solexa was bought over by Illumina in 2007 and continued developing second-generation sequencing platforms for specific aims [51]. The strategy behind Illumina's sequencing process is a four-colour reversible termination sequencing method. After clonal amplification of DNA, sequencing occurs through base incorporation onto the template strand successively, followed by washing, imaging and cleavage. In this method, the polymerisation reaction is halted using fluorescently-labelled dNTPs and unincorporated bases are removed. Final analysis is carried out on the obtained four-colour images to ascertain base composition [52]. Currently, Illumina provides an impressive number of sequencing platforms which include MiniSeq, MiSeq, NextSeq 550, NovaSeq 6000, etc. NextSeq 500 was discontinued with the introduction of NextSeq 550 which has more flexible features of microarray scanning and sequencing. Their newest sequencing systems, NextSeq 1000 and 2000, boasts an integrated cartridge containing fluidics, waste compartment and reagents. It also possesses a novel system taking advantage of super resolution optics resulting in higher sensitivity and increased accuracy of imaging data [53].

3.2 Pacific Biosciences, third generation sequencing technology

The single-molecule real-time sequencing (SMRT) method is a third-generation sequencing approach developed by Pacific Biosciences (PacBio). This method directly observes DNA or cDNA synthesis by DNA polymerase as it occurs in real time [54]. The principle behind this method is the use of zero-mode waveguide (ZMW) technology. A ZMW is essentially a tiny, zeptoliter-sized hole deposited slightly above a glass surface [54]. Within each ZMW is a chamber containing a single DNA polymerase molecule affixed to the bottom glass surface using a biotin/streptavidin system. Fluorophore-labelled nucleotides are added to the compartment above an array of ZMWs. Diffusion then occurs whereby labelled nucleotides travel downwards through the ZMW to reach DNA polymerase for incorporation onto the DNA strand. The ZMW system is sufficiently sensitive to detect incorporations against background nucleotides. In addition, one of the first commercially available sequencing system employing SMRT contains an assembly of ~75000 ZMWs [54]. Therefore, single-molecule sequencing can be carried out massively in parallel. As of now, PacBio also has an Iso-Seq method used to analyse long reads produced by SMRT to examine novel transcripts, gene fusion, alternative splicing, etc. Their newest system release is the sequel IIe system that promotes higher quality data, shorter analysis time and cheaper costs [55].

3.3 Oxford Nanopore Technologies, fourth generation sequencing technology

As suggested by their name, Oxford Nanopore Technologies (ONT) developed and commercialised nanopore-based sequencing. The idea behind this strategy is that each nucleotide can induce a unique fluctuation in ionic current while passing through a tiny channel [56]. An α-hemolysin pore secreted by *Staphylococcus aureus* was used to form single transmembrane channels through which nucleic acid polymers would pass through [56]. This study aimed to determine the length of nucleic acid polymers but also proposed that if each nucleotide could provide a characteristic current change based on their chemical or molecular properties, it could very

well be used to determine nucleotide sequences as well. The current technology employed by ONT consists of a group of tiny wells contained in a sequencing flow cell. Within each well is a synthetic bilayer fabricated with biologic nanopores. As described earlier, sequencing is achieved by assessing the distinct current changes induced during base incorporation carried out by a molecular motor protein [57]. Presently, the devices provided by ONT include the Flongle, MinION, GridION and PromethION. Flongle and MinION are more for smaller scale experiments while GridION generates high-throughput data up to 150GB. PromethION, on the other hand, provides ultra- high-throughput data of up to a remarkable scale of 8 TB [58].

3.4 Other genome analysers

Roche 454 pyrosequencing was the first commercially successful 2nd generation sequencing platform, initially developed by 454 Life Sciences and later acquired by Roche. Sequencing by this platform depended on the detection of visible light produced by a group of enzymes correlating to the pyrophosphate release during nucleotide incorporation [59]. Roche however stopped supplying the 454 sequencing machines and any accompanying reagents since 2016 [51]. Another NGS instrument is Sequencing by Oligonucleotide Ligation and Detection (SOLiD) released by Applied Biosystems Instruments (ABI). This technology uses sequencing by ligation. It involves cycles of annealing and ligation of primers and probes. Four-colour imaging is also carried out after which ligated probes are cleaved to allow another cycle of ligation [60]. Despite being quite accurate, it has a long run time and requires experts to analyse raw data [51]. Furthermore, another sequencing approach called DNA nanoball sequencing was developed by Complete Genomics and later acquired by Beijing Genomics Institute (BGI) [51]. This approach combines the principles of hybridization and ligation. DNA nanoballs are produced by amplifying DNA or cDNA using rolling-circle replication. They are then added onto a flow cell with an array of wells and each nanoball in each well are sequenced at high density. This process only yields short reads however and takes a long time. Meanwhile, Ion Torrent technology introduced by the team behind the 454 sequencer is based on the electronic detection of pH changes as opposed to detection of light as previously used [61]. Each incorporated nucleotide generates an electronic signal detected by electronic sensors placed at the bottom of each flow cell [51]. Lastly, a third generation sequencing platform called Helicos sequencing employs the principle of single-molecule fluorescent sequencing [62]. The Helicos sequencer, Heliscope, does not require clonal amplification and uses a very sensitive fluorescence detection system [60]. This method merges sequencing by synthesis and hybridization.

3.5 NGS advantages

All NGS platforms have significant advantages over previously used methods however, each platform has their own strengths and unique features. The four major sequencing platforms being used currently are Illumina, Pacific Biosciences, Oxford Nanopore Technologies (ONT) and Ion Torrent. Both Illumina and Ion Torrent are highly accurate but they are relatively more costly and have short reads (≤ 400). The problem with short read lengths is that it prevents researchers from performing de novo assembly and impedes the detection of structural variations [63]. On the other hand, PacBio and ONT platforms produce long reads (≥ 500) but they have variable accuracies. Although, both ONT and PacBio have similar read lengths, ONT specifically the MinION device, has higher error rates of up to 38.2% [64]. ONT also produces a higher yield but PacBio has better data quality overall [65].

All these platforms have a similar disadvantage which is a long turnaround time except for ONT. In addition, ONT also has lower capital costs compared to the others [66].

Illumina sequencing has error rates of <1% and one of their systems called the NextSeq 550, employs the use of two-channel sequencing strategies instead of the four-channel strategy used by previous systems. This method only needs two images to detect nucleotides which makes data processing much faster [67]. However, a few studies found that PacBio sequence data produced better results than Illumina datasets specifically when used for *de novo* assembly purposes in addition to improved resolution [68–70]. Meanwhile, when comparing Illumina against ONT, ONT proved to have a significantly shorter turn-around time of <15 hours while Illumina analysis took around 3 to 6 days. Therefore, ONT sequencing was deemed more suitable for urgent, smaller scale sequencing requirements especially during public health emergencies [71]. Lastly, the Ion Torrent Personal Genome Machine (PGM) has a unique plus point which is the ability to identify single nucleotide polymorphisms (SNPs) better than Illumina and PacBio [72]. Lahens et al. [73] did however conclude from his experiments that both Illumina and Ion Torrent are equally capable in detecting differential gene expressions. There are a large number of studies that have found certain platforms to perform better than others, however it ultimately depends on the aims of the experiment. Another useful method is combining datasets from more than one platform to acquire a more complete genome assembly [74–79].

NGS technologies are also capable of producing either single-end or paired-end reads during sequencing. The question that normally arises is which type of sequencing to perform. Single-end sequencing in RNA-Seq is when a cDNA fragment is sequenced from only one end whereas paired-end sequencing is when both ends of a fragment are sequenced [80]. Paired-end sequencing produces twice the amount of data which increases the accuracy of read alignment. It also more sensitive and allows the detection of events like gene fusions and new splice isoforms. On the other hand, single-end sequencing is much cheaper than paired-end sequencing. It is also more suitable for some methods such as ChIP-Seq and small RNA-Seq [80]. Although it is the more economical choice, it has drawbacks such as lower read counts per RNA feature and a weaker ability to assign reads to features. In the context of functional profiling, single- and paired-end reads in an RNA-Seq experiment only showed a 65% agreement in the top 20 gene ontology (GO) terms obtained. However, when looking at the top 300 GO terms, both led to similar broad conclusions [81]. Since the cost of sequencing is an important consideration to make, Corley et al. [81] suggested that single-end sequencing could be carried out with more biological replicates as they found that it was comparable to the results obtained using paired-end sequencing if functional analysis is done cautiously. As mentioned before, the utility of single- or paired-end sequencing ultimately comes down to the research question. For instance, if the main objective of the experiment is transcriptome assembly, then paired-end sequencing would be the more suitable choice.

4. Application of systems biology in understanding host-pathogen interactions

Systems biology is the comprehensive study of a biological system encompassing molecular- level interactions, sub-cellular dynamics and overall physiological functions of cells, tissues and organs [82]. A systems biology approach aims to looks at

the larger picture involved in a given system or condition. For a long time, research had centred on the molecular understanding of genes and proteins. Current illustrations or diagrams of interconnecting pathways are just not enough to completely understand a system. Kitano et al. [83] aptly describes these diagrams as mere static roadmaps, whereas what we seek to understand leans more toward patterns, their causes and regulatory dynamics. In the context of host-pathogen interactions, a systems biology view is examining components from both the host and pathogen as well as their interactions with one another. Some of the approaches used in systems biology include identification of key molecules or biomarkers, inference between networks and disease module discovery [84]. The advancement of -omics technologies supported by high throughput sequencing has increased the whole-system analyses focusing on host-pathogen interaction between genes, proteins and small ligands [85]. This is accomplished by carrying out dual RNA sequencing whereby both host and pathogen transcriptomes are profiled during the course of an infection. Multiple cascades of events are triggered by an infection and dual RNA-Seq allows the monitoring of host and pathogen in parallel. Knowledge gained from comprehensive host-pathogen interaction studies especially with the use of dual RNA-Seq can guide efforts toward better therapeutics against infection. Dual RNA-Seq was first described by Westermann et al. [86] however it only started gaining attention recently resulting in a surge of studies utilising this method.

4.1 Bacteria-host interaction

Interaction between bacteria and hosts usually begin with a compulsory attachment or adherence of bacteria to host cells followed by subsequent internalisation which may involve direct or indirect receptor binding [87]. Entry into the host may seem like a straightforward step but it involves a drastic change in environment for the pathogen. Hence, entry and any subsequent mechanism employed are bound to involve a complex interplay between the host and pathogen. Previous methods were limited in the sense that they only allow the analysis of mRNA in either infected host cells or bacteria [88]. Dual RNA-Seq has provided researchers everywhere an access to the complete story. Some of the host-bacteria interaction studies utilising dual RNA-Seq have looked at bacteria infecting humans, such as *Salmonella enterica* [89], *Haemophilus influenza* [90], *Streptococcus pneumonieae* [91, 92], *Mycobacterium tuberculosis* [93, 94] and *Mycobacterium leprae* [95]. Despite the diversity of these bacteria-host dual RNA-Seq studies, one similarity is that all their findings encompass several aspects or levels of a biological system instead of mere isolated observations. For instance, in the study by Baddal et al. [90], not only did they characterise preferential binding of nontypeable *H.influenzae* (NTHi) to ciliated bronchial epithelial cells, they also observed differential expression of various bacterial virulence factors, alteration of host cell adherence junctions, host-dependent modulation of NTHi metabolic machinery and rearrangement of host extracellular matrix and cytoskeletons. In addition, they discovered small RNA regulatory elements that were differentially expressed including novel snoRNAs that have never been associated with NTHi before. Meanwhile, Aprianto et al. [91] observed the generation of reactive oxygen species (ROS) by *S. pneumoniae*, the glutathione-dependent detoxification of ROS as a counteraction by the host, expression of chemokine IL-8 for immune response repression and also the activation of bacterial sugar transporters sensitive to host-derived non-glucose carbohydrates. Lastly, Yimthin et al. [96] analysed the whole blood transcriptome of 29 patients with melioidosis which is the infection caused by *B. pseudomallei* often leading to mortality in endemic areas. Using RNA-Seq, they managed to identify survivor- and non-survivor-specific

expressions related to cell lineage processes and immune activation pathways with the potential to be biomarkers against melioidosis. These findings further reiterate the importance of a systems biology-based view when analysing RNA-Seq data spanning multiple gene networks and pathways.

4.2 Virus-host interaction

Viruses are obligate intracellular parasites manipulating various machinery and components of the host cell. The human body has developed efficient responses against viruses particularly the interferon system. An antiviral state is induced by the family of interferon proteins and other effectors upon viral infection. However, over time, certain viruses have evolved mechanisms to dodge these immune responses [97]. Given the complex nature of viral infections, it most certainly involves multi-level interactions and a method like dual RNA-Seq can help us understand these elaborate interactions networks. One of the first studies examining host-virus interactions using dual RNA-Seq was carried out using a murine infection model for cytomegalovirus (CMV) [98]. This study found some unexpected results such as highly abundant viral transcripts with unknown functions and also a viral transcript bearing functions of both non-coding RNA and mRNA. From the host perspective, expected upregulation of genes involved in inflammation and immunity were observed. Certain unforeseen results include upregulation of genes associated with development and differentiation. More importantly, this study found many differentially expressed genes within specific biological pathways including certain networks with unknown relevance to infection, providing new insights into CMV pathogenesis. The use of dual RNA-Seq has been applied to a range of studies analysing host-virus interactions which include infections by avian influenza (H5N8) [99], varicella zoster virus [100], Crimean-Congo hemorrhagic fever virus (CCHFV) [101], influenza A (H3N2) [102], and Zika virus [103]. Similar to host-bacterial studies, a wide range of findings were uncovered including variable alternative gene splicing events, association between clinical phenotypes and viral gene induction, remodelling of host epidermal environment, inhibition of functional pathways, host metabolic regulation and many more. Michlmayr et al. [103] successfully identified CD169 (Siglec-1) on CD14+ monocytes as a potential biomarker against acute infections of Zika virus while also providing evidence that dengue-immune patients did not necessarily have an upper hand when faced with Zika virus. Another interesting study by Wesolowska-Andersen et al. [104] using dual RNA-Seq found that transcriptionally active respiratory viruses were present in children even in the absence of any observable respiratory illness. These viral carriers also displayed alterations in their nasal transcriptomes. This shows that underlying host-virus interaction networks are still being engaged 'silently' and not necessarily in cases where the illness clearly manifests itself. In due time, these studies will hopefully reveal horizontal inter-study patterns which will point toward the discovery of common disease modules or host-pathogen interaction networks. Furthermore, the discovery of a novel coronavirus in Hong Kong was achieved through a series of eliminating laboratory tests and eventually genome sequencing [105]. In addition to discovery of novel pathogens, RNA-Seq analysis can provide information relating to genome sequence, gene expression, pathogen abundance and a myriad of information that will provide useful insight regarding the pathogen and how it causes disease [106]. Currently, most RNA-Seq studies examining novel viruses are focused on plant viruses [107, 108]. The rapid detection of novel viruses in humans by RNA-Seq is an area that should be further investigated and optimised as it can help us take precautionary steps before the wide spread of disease.

4.3 Fungi-host interaction

There are at least 712 000 existing fungal species around the world however the total number of fungal species is estimated to be more than 1.5 million [109]. The proportion of fungal species causing human diseases are quite small comparatively [110]. Some of the most common opportunistic fungal pathogens are *Aspergillus fumigatus* and *Candida albicans*. Previous studies have elucidated certain interactions of these fungi with their host including interference of host phagolysosome mechanisms, activation of complement system, morphological switches and formation of neutrophil extracellular traps (NETs) [111–113]. These studies mainly use assay- and imaging-based techniques to study interaction and are mostly focused on specific pathways or components. From a systems biology perspective, pathogenic fungi often co-evolve with the host and commensals resulting in an equilibrium shift within the host leading to a myriad of changes affecting many networks [114]. The use of RNA-Seq has allowed a more comprehensive study of host-fungal interactions. Initially, a number of studies used RNA-Seq to delineate transcriptional landscapes for fungi like *Candida albicans* and *Candida glabrata* [115, 116]. In terms of host-fungal interaction, RNA-Seq has shed light on alternative splicing events during host invasion, gene expression profiles in mice models of fungal keratitis and also differences in regulatory networks between *Candia albicans* and *Mus musculus* [117–119]. Dual RNA-Seq analysis of *Trichophyton rubrum*-infected human keratinocytes also demonstrated the upregulation of genes increasing the efficiency of nutrient uptake, production of keratinolytic proteases as well as host-derived antimicrobial proteins [120].

4.4 Combination of pathogens and host interactions

Aside from the pathogens discussed above, some other pathogens that exist are parasites, prions and in rare cases, algae [121–123]. Parasites in particular have extremely complex life cycles involving different hosts at different life stages [124]. A clear comprehension of parasitic life cycles will undoubtedly require a systems biology approach and RNA-Seq has provided an avenue for that. RNA-Seq studies have allowed inter-sex, inter-stage and inter-host studies involving parasites like *Plasmodium falciparum* [125], *Trypanosoma vivax* [126], *Brugia malayi* [127], *Trichuris trichiura* [128] and *Schistosoma mansoni* [129]. A dual RNA-Seq study examining the interactions between murine hosts and the parasite *Toxoplasma gondii* also provided many insights into acute and chronic infection stages by this parasite that is prevalent in humans [130]. Prions, which are misfolded proteins, cause several neurodegenerative diseases in humans including Jakob-Creutzfeldt disease, kuru and fatal familial insomnia [122]. Despite being a protein-only infection, it involves extensive processes occurring simultaneously in the brain including synaptic alterations, inflammation, neural cell death and protein aggregation [131]. RNA-Seq has revealed unique miRNA profiles produced by components of prion-infected cells, mechanisms of prion-induced neurotoxicity and signature glial gene expressions among others [132–134]. Meanwhile, algal infections in humans are quite rare however they have been documented such as human protothecosis caused by the *Prototheca* species [121]. Genome sequencing studies have been carried out to study the sequence and expression of these species, however the use of RNA-Seq in this area is still scarce [135, 136]. There are also cases of co-infections whereby more than one pathogen infects a host simultaneously. Transcriptomic profiling studies of co-infections have shed some light on disease mechanism, molecular phenotypes and inter-disease relationships. One example of a complex co-infection is when HIV-infected patients develop cryptococcal meningitis which is a fungal infection.

Some patients undergoing treatment for these infections also start to develop paradoxical cryptococcosis-associated immune reconstitution inflammatory syndrome (C-IRIS) characterised by various clinical deteriorations. By assessing the whole blood transcriptome of infected patients, Vlasova-St. Louis et al. [137] identified novel and unique biomarkers for both early and late stages of C-IRIS which are difficult to distinguish due to their similar clinical manifestations. Moreover, an ambitious study by Seelbinder et al. [138] managed to carry out a triple RNA-Seq analysis in host monocyte-derived dendritic cells infected by the fungus, *Aspergillus fumigatus* and human cytomegalovirus (CMV). These two pathogens are commonly co-occurring pulmonary pathogens. A highlight from their comprehensive study is that host expression levels that were upregulated during single infection by either pathogen were downregulated instead during co-infection. This implied interference or opposing effects of the two distinct host responses induced and also a possible synergistic relationship between *A. fumigatus* and CMV.

5. Bioinformatics and statistical approaches in analysing RNA-Seq data

The initial experimental workflow of RNA-Seq has been described earlier which briefly include depletion of rRNA or enrichment of mRNA, fragmentation of samples and subsequent reverse transcription to form a cDNA library. These cDNA fragments are then sequenced using a high-throughput sequencing platform. This section will describe the data analysis of RNA-Seq data including statistical approaches taken to analyses differentially expressed genes. The whole process is simplified in **Figure 2**, covering all the important analytical steps involved.

Once sequencing data is obtained in the form of raw reads, quality control and sequence filtering need to be carried. This is a key pre-processing step because next-generation sequencing data may contain unexpected artefacts, poor quality reads, low-complexity regions, high GC content and sequencing errors [139, 140]. The presence of these low-quality sequences will further effect downstream analysis leading to inaccuracies in overall RNA-Seq data interpretation. There are a variety of tools that can be used to perform data pre-processing. Two important pre-processing concepts are the quality assessment of reads and also processing/filtering to remove contaminants, adapter sequences, low-quality sequences [141]. Some of the methods developed include FastQC [142], RSeQC [143], NGSQC [144], Trimmomatic [145] and CutAdapt [146]. Weaknesses of these tools include the inability to carry out both data quality control and processing steps, slow run times and single-platform services [147, 148]. Recently developed tools are more comprehensive, encompassing all steps required in raw reads processing. Some of these include FastProNGS [147], FastqPuri [149], Zseq [140], RNA-QC-Chain [150] and fastp [151].

The next step is mapping or aligning the quality-assessed reads onto a genome or transcriptome. Reads can be mapped either uniquely to a single position or multiple positions (multi-reads) in the reference genome. Some of the mapping software or algorithms available are STAR [152], TopHat2 [153], MapSplice [154], BowTie2 [155] and Magic-BLAST [156] among others. A range of bench-marking studies have compared the efficiencies of various RNA-Seq aligners. Baruzzo et al. [157] examined 14 common RNA-Seq aligners, whereas Schaarschmidt et al. [158] evaluated 7 alignment tools. In addition, Engstrom et al. [159] carried out comprehensive analysis on a total of 26 alignment protocols. A similarity across these three studies is that they all found STAR to be one of the more reliable aligners, although other aligners do have their own strengths. After alignment, transcript identification
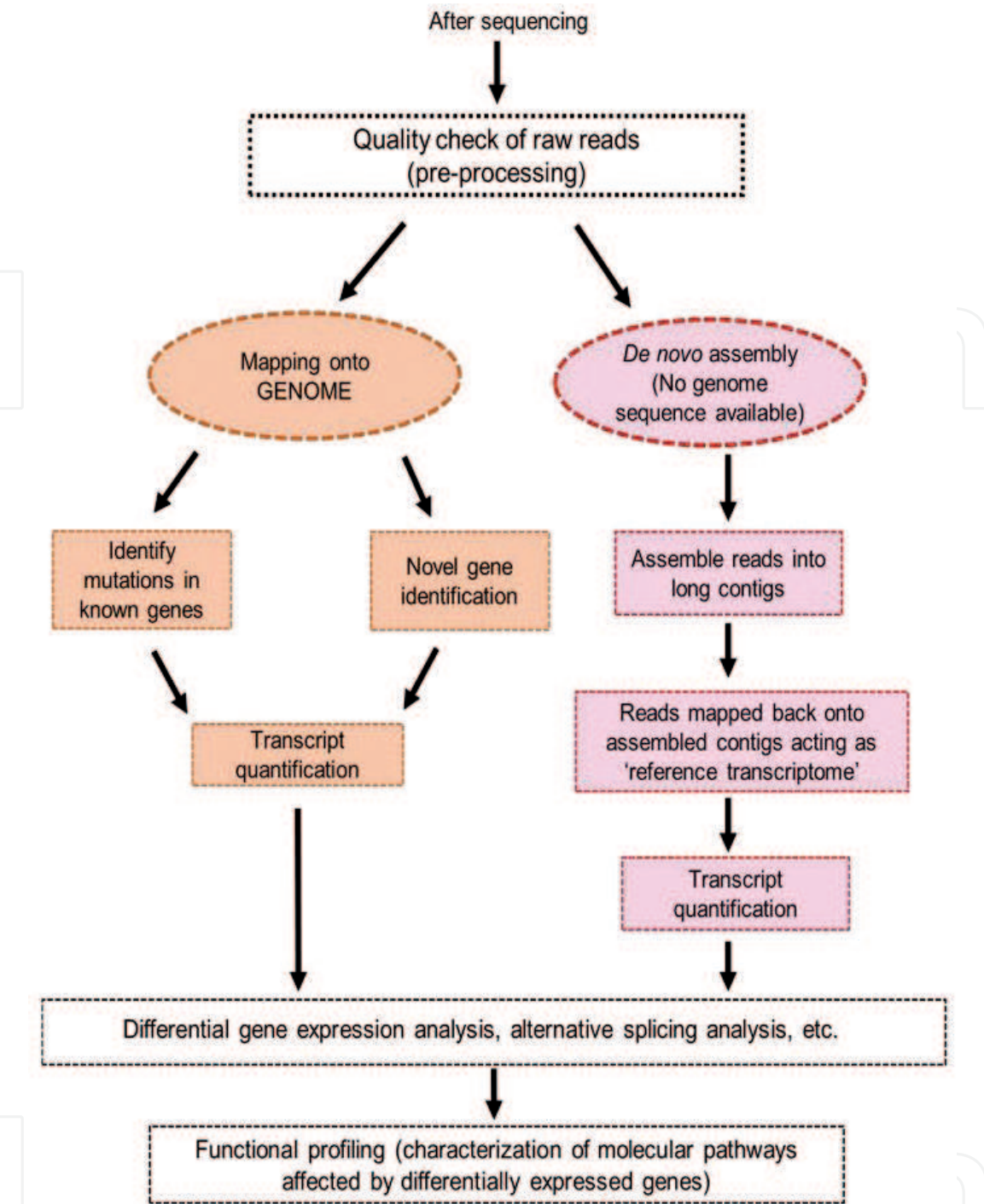
Figure 2.
General RNA-Seq data analysis workflow. The first step after sequencing is pre-processing the sequence reads to obtain data with higher quality. Reads can be either mapped onto a reference genome (e.g., GRCh38) or in cases where a reference genome is unavailable, de novo assembly is carried out. When using a reference genome, novel transcript discovery is possible. After identification of relevant transcripts, quantification or counting is carried out. When the genome sequence is unavailable, de novo assembly is used to assemble reads into long contigs. Reads are then mapped back onto assembled transcriptome followed by quantification. In both cases, differential gene expression and alternative splicing analysis can be carried out in addition to other methods depending on the experiment. Finally, functional profiling is done to characterise molecular pathways and interactions.

is carried out. Reads that are mapped onto known reference transcriptomes can only focus on quantification and not novel transcript discovery. Meanwhile, reads mapped onto a reference genome can either be identified as known transcripts or alternative transcripts [139]. For rapid discovery of novel transcripts, a popular programme called Cufflinks utilises existing annotated genomes as a reference to assist in transcript assembly [160]. Other methods focusing on novel transcript

identification are SLIDE [161], iReckon [162] and StringTie [163]. In the case where a reference genome is absent or incomplete, de novo transcript reconstruction is carried out. Reads are first assembled into longer contigs, then this is treated as the 'reference transcriptome' to which the reads are mapped back onto for quantification purposes. Some of the tools available for de novo transcript assembly include Trinity [164], SOAPdenovo-Trans [165], TransABySS [166] and Oases [167]. Depending on the experiment, transcript identification and quantification can be carried out either simultaneously or sequentially. One of the most frequent applications of RNA-Seq is estimating the abundance of gene or transcript expressions. HTSeq-count and featureCounts are two gene-level quantification approaches with HTSeq-count being specially designed for downstream differential expression analysis [168, 169]. These are 'union exon'-based approaches whereby exons that overlap are merged to form a union-exon. This method can assign reads to respective genes with high confidence however, difficulty arises when dealing with alternatively spliced transcripts [170]. Due to biases related to transcript length and number of reads, within-sample normalisation methods are used to standardise reads with some common measures like RPKM (reads per kilobase of exon model per million reads), FPKM (fragments per kilobase of exon model per million mapped reads) and TPK (transcripts per million) [34, 139]. Besides union exon-based methods, several transcript-level statistical quantification methods also exist such as RSEM [171], eXpress [172] and TIGAR2 [173]. Recently, alignment-free methods have also been developed like Salmon [174], kallisto [175] and Sailfish [176].

A crucial step before carrying out differential gene expression (DGE) analysis is data normalisation. The within-sample normalisation approaches during quantification are not sufficient in cases where high numbers of differentially expressed transcripts exist [139]. The current software that exist for RNA-Seq differential gene expression analysis can be mainly categorised into four groups based on the statistical methods employed [177]. These include (1) Poisson or negative binomial model-based methods – baySeq [178], DESeq [179], DESeq2 [180], EBSeq [181], edgeR [182], NBPSeq [183], PoissonSeq [184], TSPM [185], (2) *t*-test analogical methods – Cuffdiff [186], Cuffdiff2 [187], (3) non-parametric methods – NOIseq [188] and SAMseq [189], (4) linear models – limma [190] and voom [191]. Other methods have also been developed including a hybrid full Bayes-empirical Bayes method (ShrinkSeq) and also and binomial distribution-based method called DEGSeq [192, 193]. There are also specific methods that have been developed to study differential gene expression using de novo transcriptome assemblies [194]. There is still no consensus as to which methods are significantly superior however many studies have done comparative analyses of these methods. **Table 2** summarises past studies that have compared the ability of various statistical methods.

A common finding across these studies is that no single method is superior in all circumstances. Each method has their own strengths and weaknesses. Out of the seven studies mentioned in **Table 2**, edgeR and DESeq were commonly found to perform better than other softwares however, a few studies did find contrasting results. Ultimately, the choice of statistical approach largely depends on the nature of study, type of biological sample, number of replicates, budget of study and many other factors that need to be matched to the strengths of any particular approach.

The next step usually examines differential gene expression at a transcript level which is alternative splicing (AS) events. Many computational tools exist that can infer AS events including some of the previously mentioned methods [202]. These include exon-based methods like DEXSeq [203] and JunctionSeq [204], event-based methods like MAJIQ [205], dSpliceType [206] and SUPPA2 [207] and lastly isoform-based methods like Cuffidiff2 [187] and DiffSplice [208]. The final step is a pathway enrichment analysis. The list of DEGs obtained are further analysed

Author (Year)	Statistical methods compared	Data used	Main Findings
Robles et al. [195]	DESeq, edgeR, NBPSeq	Simulations using statistical models derived from real RNA-Seq data	• DESeq performs more conservatively • More biological replicates result in higher quality and reliability of DEG detection
Soneson & Delorenzi [196]	baySeq, DESeq, EBSeq, edgeR, NBPSeq, NOIseq, SAMseq, ShrinkSeq, TSPM, voom+limma, vst + limma	Simulations using statistical models derived from real RNA-Seq data	• voom+limma and vst-limma performed well under many conditions like detection of DEGs, gene ranking and detection of true positives. • SAMseq did well with large sample sizes • TSPM most affected by sample size
Rapaport et al. [197]	baySeq, Cuffdiff, DESeq, edgeR, limma, PoissonSeq	Used benchmark datasets: SEQC dataset & ENCODE project data	• Negative binomial methods (baySeq, DESeq & edgeR) have better specificity, sensitivity & good control of false positive errors • Cuffdiff had low specificity, sensitivity & high false positives • Number of sample replicates greatly affect DEG detection accuracy.
Zhang et al. [198]	Cuffdiff2, DESeq, edgeR	Real RNA-Seq & simulated datasets: MAQC dataset (human), K_N dataset (mouse), LCL dataset (human)	• edgeR performs better than Cuffdiff2 & DESeq in uncovering true positives • Cuffdiff2 more sensitive to sequencing depth, DESeq more sensitive to unbalanced sequencing depths between groups • All three perform better with biological/technical replicates
Seyednasrollah et al. [199]	baySeq, Cuffdiff2, DESeq, EBSeq, edgeR, limma, NOIseq, SAMseq	Real mouse RNA-Seq and human RNA-Seq data	• DESeq & limma most reliable choices • edgeR had large variability, SAMseq had low power • Cuffdiff2 & NOIseq did not do well with large replicates
Rajkumar et al. [200]	Cuffdiff2, DESeq2, edgeR, TSPM	Real RNA-Seq data from mice amygdalae micro-punches	• edgeR had relatively high sensitivity & specificity • Cuffdiff2 had high false positive rates • DESeq2 & TSPM had high false negative rates • RNA sample pooling is discouraged due to low positive predictive values
Costa-Silva et al. [201]	baySeq, DESeq, DESeq2, EBSeq, edgeR, limma+voom, NOIseq, SAMseq	Real RNA-Seq dataset produced for MAQC project	• DESeq2, limma+voom & NOIseq produced most consistent results in terms of accuracy, precision & sensitivity

Abbreviations: TSPM: Two-stage Poisson Model, DEG: Differentially expressed genes, SEQC: Sequencing Quality Control, ENCODE: Encyclopaedia of DNA Elements, MAQC: MicroArray Quality Control, LCL: Lymphoblastoid cell line.

Table 2.
A compilation of numerous studies that have compared common statistical methods used for differential gene expression analysis in RNA-Seq.

to characterise their molecular involvement in biological pathways. Some of the RNA-Seq-specific tools developed for this aim are GOSeq [209], Gene Set Variation Analysis (GSVA) [210] and SeqGSEA [211]. Annotation databases such as KEGG [212], Gene Ontology [213] and Bioconductor [214] also complement functional profiling of DEGs. This is an important step particularly in host-pathogen interactions to unravel the interaction networks that exist. Common databases and softwares used by dual RNA-Seq studies examining host-pathogen interactions are Gene Ontology and KOBAS (KEGG Orthology-based Annotation System) [215–217]. Novel transcripts detected based on de novo assembly can be functionally annotated by finding orthologous proteins in protein databases. Challenges arise when annotating non-protein coding transcripts like long non-coding RNAs which still lack proper functional-annotation procedures [139].

6. Other applications of RNA-Seq in host-pathogen interaction studies

RNA-Seq can be applied in very innovative ways to answer many of the questions and mysteries posed by biology and disease. Initially, it was used for simpler research goals like profiling transcriptomes and monitoring gene expression. Over time, RNA-Seq technology has developed rapidly and one of its vital uses is characterising host-pathogen interaction networks. Dual RNA-Seq in particular has been applied to many infection models ranging from bacteria, virus, fungi and parasites as described in previous sections. Understanding the mechanics of infection induced by pathogens and subsequent host response is a crucial step required before proceeding to figure out clinical treatment strategies. Besides utilising dual RNA-Seq, as extensively detailed earlier, another application of RNA-Seq is single cell RNA sequencing (scRNA-Seq). The difference between bulk RNA-Seq and scRNA-Seq is that the latter allows transcriptional comparison of single-cell populations and has the ability to capture cellular heterogeneity that is normally obscured by bulk RNA-Seq [218]. In the context of host-pathogen interaction studies, dual scRNA-Seq is commonly utilised. ScRNA-Seq involves an extra step which is isolating single cells from tissue samples using techniques like fluorescence-activated cell sorting (FACS), micro-dissection and droplet-based methods instead of bulk sequencing various cell populations [218]. While dual RNA-Seq provides insight about the bigger picture, dual scRNA-Seq can elucidate the smaller scale interactions that sum up to produce the host outcome during infection [219].

It is common for bacteria to have distinct co-existing subpopulations due to their dynamic adaptability. This heterogeneity can lead to phenotypic variations in infection and scRNA-Seq is capable of characterising these variabilities [220]. Avraham et al. [220] examined individual macrophages infected with Salmonella typhimurium and found molecular variations despite what seemed to be identical infections in these cells. They discovered that the type I interferon response pathway is influenced by PhoPQ activity levels in the bacterium. Host cells infected with a bacterium expressing high levels of PhoPQ had an increased type I interferon response. Another similar study also examined bone marrow-derived macrophages exposed to Salmonella with their method called scDual-Seq [221]. From their time-dependent analysis of macrophage single-cell transcriptomes, they found that within infected cells, some had fully induced immune responses while others only had 'partially induced' immune responses. They also found two intracellular classes of Salmonella having unique transcriptional signatures. One of their interesting findings is how the infection progresses from partially induced to fully induced immune responses which also involve changes in Salmonella subpopulations [221].

Meanwhile, scRNA-Seq has also been applied to host-viral interaction studies. In HIV infections, the virus has the ability to persist in latent reservoirs where they are not completely eradicated by treatments like antiretroviral therapy (ART). Golumbeanu et al. [222] used scRNA-Seq to characterise the transcriptomes of latent and reactivated HIV-infected cells. They identified two main subpopulations with one cell cluster being more predisposed to HIV reactivation. Their results provide interesting insights for the identification of potential latency reversing agents and biomarkers for susceptible cells. However, the use of scRNA-Seq in host-pathogen interactions studies are still in its infant stages. Many more questions can be answered using scRNA-Seq such as the mechanism behind selective infections of host cells, antibiotic tolerance of certain bacteria, the switch between active and latent infection in viruses and the list goes on [219].

Furthermore, scRNA-Seq has also played a role in the development of human organoids from stem cells by assessing the similarity between these organoids and primary tissue counterparts [223]. In addition, scRNA-Seq can be used to properly characterise the development and maturation stages of stem cells to specific organ tissue or even used as a blueprint to direct the recreation of actual human organs [224, 225]. Moreover, scRNA-Seq can be used in conjunction with the well-known CRISPR-based gene editing tool to provide confirmation of target gene activation/repression [226]. Advancements in the application of scRNA-Seq in these research areas can provide valuable tools for host-pathogen interaction studies in the future. For instance, the successful creation of human organoids which are highly accurate to real organs can be used as infection models to study disease mechanisms.

Innovations of RNA-Seq methods based on experimental needs have led to its application in various settings. Two of these methods are spatially resolved RNA-Seq known as 'spatialomics' and ribosome-profiling using RNA-Seq to understand the translatome [227]. Spatial information is not provided when using bulk RNA-Seq or scRNA-Seq and this information could be crucial to comprehend cellular processes and how they relate to gene expression. The main concept behind spatialomics is in situ transcriptomics which produce data within tissue sections either using sequencing or imaging [227]. Some of the approaches that have been used in spatial transcriptomics are fluorescent in situ RNA sequencing (FISSEQ) and also a combination of scRNA-Seq data with single molecule fluorescence in situ hybridization method (smFISH) to examine spatial division of genes along liver lobules and investigate gene expression as well as post-transcriptional modifications while preserving spatial information [228, 229]. The smFISH method however had limitations in the number of RNA species that could be imaged at once in single cells. Hence, another method called multiplexed error-robust FISH (MERFISH) was developed which allows thousands of RNA species to be imaged in individual cells with spatial distribution information as well [230]. The use of spatialomics in host pathogen interaction studies shows great promise as many infections by pathogens induce alterations in specific subcellular compartments [231]. Understanding both temporal and spatial changes that occur during the course of an infection can improve our comprehension of host-pathogen interplay. As for ribosome-profiling, the highly regulated process of mRNA translation by ribosomes inspired this translatome-based analysis with an assumption that protein synthesis is proportional to the density of mRNA ribosomes [227]. By sequencing the ribosome-protected mRNAs, studies have gained insight on translational control in yeast, codon usage biases and unannotated translational events [232–234]. Ribosome profiling coupled with RNA-Seq has been carried out as well to study infections by pathogens like *Toxoplasma gondii* and the vaccinia virus. Holmes et al. [235] found open reading frames that may be involved in selective stress-induced translation of parasitic mRNA while Dai et al. [236] found that mRNAs involved in cellular energy

production were increased which supported vaccinia virus replication. The applications of RNA-Seq and its combinations with existing methods are increasingly being advanced and modified to suit specific experimental needs.

7. Challenges in RNA-Seq

The rapid surge of RNA-Seq technology has led to many new discoveries and is currently the go-to method for transcriptomic analysis. Although significant advancements have resulted from the use of RNA-Seq, it is still continuously evolving with many aspects that need to be improved. The drawbacks of short-read sequencing platforms as mentioned before have been mostly solved with the advent of long-read technology. While long-read technology has its own strengths, analysing long-read datasets still poses a challenge. Aside from lower accuracies per read compared to short-read platforms, most of the long-read transcriptomic tools do not take into account factors like coverage bias and high error rates [237]. Several studies have found beneficial effects of combining short- and long-read technologies, however integrating different tools are often laborious hence it still needs to be improved [238, 239]. There are certain challenges faced with library preparations as well. In this process, cDNA is generated from fragmented RNAs followed by adapter ligation, amplification and finally sequencing. Linsen et al. [240] compared three different library preparation methods and found that each method had large differences in the frequency of miRNAs captured. Other biases include PCR amplification bias which might be introduced due to variations in template length and base composition during parallel amplification of multiple templates [241, 242]. Yet another issue faced in library preparation is the influence of batch effects. Batch effects may arise from various factors including experimental conditions, quality of reagents, pipetting abilities and also the individual/technician in charge on a particular day [243]. Careful considerations should be made by researchers in order to reduce the effects of these confounding variables.

A recent discovery was the abundance of circular RNAs in various eukaryotic organisms including humans [244]. Previous RNA-Seq protocols were mostly biased against circular RNAs (circRNAs) whereby the poly (A) enrichment step would efficiently deplete all circRNAs since they lack poly (A) tails. The development of alternate protocols more suited to non-coding transcripts like rRNA depletion improved detection of circRNAs. However, these approaches are not entirely efficient for circRNAs and further research is required to improve the detection sensitivity of circRNA and possibly other non-coding RNA transcripts [245]. There are several technical challenges associated with scRNA-Seq as well. With regard to host-bacterial studies, the bacterial lysing protocols employed, whether physical or chemical, are not very compatible with further downstream steps in RNA-Seq like amplification and library preparation. These steps also do not preserve the RNA effectively. Another problem is the accurate identification of minority transcripts in bacteria. ScRNA-Seq protocols commonly employ poly (A) enriching strategies which are useful for eukaryotes however, prokaryotic mRNAs are not poly-adenylated. Analysis of non-polyadenylated RNAs have been attempted however, they involve complex and specialised protocols which need to be simplified [218, 246]. This problem is also faced when analysing viral infections in host cells because certain viruses like dengue virus and hepatitis C virus have non-polyadenylated mRNAs. There needs to be a more optimum procedure to accurately quantify bacterial and viral transcripts. Furthermore, scRNA-Seq examines individual cells leading to very low input material. This results in high levels of technical noise which can be confused with biological variability [247]. A few statistical models have

been proposed which are capable of quantifying this technical noise but additional research is required to assess the validity of these models [247, 248].

The development of more complex tools for RNA-Seq analysis are quite possible and challenges may arise in the comprehension or use of such approaches. Efforts should be made to increase the practicality of approaches to avoid methods that are only manageable for those with very high expertise. While many tools exist for the analysis of RNA-Seq data, they seem to be more than we can handle. There are a multitude of pipelines incorporating many different tools with multiple versions and licences [249]. This is a major challenge especially in the context of translating RNA-Seq into clinic. Bringing a laboratory test into clinic involves an important step that is demonstration of analytical validity. One aspect of analytical validity is accuracy that is commonly measured by comparing obtained values to a reference standard [249]. The development of a reference standard especially for NGS data can reduce method- and platform-specific biases [250]. One of the first reference standards that existed for RNA-Seq was developed by the External RNA Controls Consortium (ERCC) using synthetic RNA spike-in controls [251]. Other projects like the Sequencing Quality Control (SEQC) [252], Association of Biomolecular Resource Facilities (ABRF) [253] and GEUVADIS [254] carried out extensive studies investigating the accuracy of RNA-Seq data across many platforms, protocols and laboratory sites, providing a guide for other researchers. The continuous technological advancements occurring in the field of sequencing technologies have to be accompanied by more reference standards [250]. The constant development and assessment of reference standards are required to reduce the variability that arises from the emergence of numerous tools. Conquering this challenge will also allow improved translation of RNA-Seq into clinic and ensure the smooth transition of NGS technologies into clinical settings.

8. Summary

RNA-Seq has revolutionised the approach taken by researchers in exploring host-pathogen interactions. From scRNA-Seq to bulk RNA-Seq, the vast amount of information derived from these studies provide novel insights into the exact mechanisms of disease and host counter- reactions in combating the disease. RNA-Seq has allowed us to examine the mechanisms of gene expression, differentially expressed genes in development or disease, alternative splicing events, gene fusion events, transcriptional regulation and many more. The use of dual RNA-Seq has changed our current perspectives of host-pathogen interactions. It is clear that systems-level alterations are induced by infection all the way from immune responses to metabolic processes. These studies are laying the foundation for more complex interrogations of our immune system and eventually its translation into clinical settings. Other creative innovations to RNA-Seq are also bound to occur as long as the determination to answer biological questions are present. The use of spatialomics seems very promising as it allows the known transcripts to be assessed while preserving the three dimensional suurounding of the tissue. This has major implications especially in studies investigating the influence of cellular architecture on infection progression. Single-cell RNA-Seq is also slowly gaining momentum in the field of host-pathogen interaction studies namely due to its ability to elucidate pathogen subpopulations. This is a key factor that will provide further information about their pathogenesis, host cell susceptibility and potential targeted treatment strategies. The current discrepancies and biases that exist within RNA-Seq protocols are challenges that need to be met in order to ensure its upward trajectory. The next few years will be a period of concurrent growth for RNA-Seq technology and biomedical research. A new biological discovery phase has just begun and RNA-Seq has proved to be a valuable tool to guide us through this phase.

Author details

Sudhesh Dev Sareshma[1,2] and Bhassu Subha[1,2]*

1 Genetics and Molecular Biology, Institute of Biological Sciences, Faculty of Science, University Malaya, Kuala Lumpur, Malaysia

2 Center of Biotechnology for Agriculture, University of Malaya, Kuala Lumpur, Malaysia

*Address all correspondence to: subhabhassu@um.edu.my

IntechOpen

References

[1] Denoeud F, Aury JM, Da Silva C, Noel B, Rogier O, Delledonne M, et al. Annotating genomes with massive-scale RNA sequencing. Genome biology. 2008;9(12):R175.

[2] Ren S, Peng Z, Mao J-H, Yu Y, Yin C, Gao X, et al. RNA-seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicings. Cell Research. 2012;22(5):806-21.

[3] Gaidatzis D, Burger L, Florescu M, Stadler MB. Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. Nature Biotechnology. 2015;33(7):722-9.

[4] Pareek CS, Błaszczyk P, Dziuba P, Czarnik U, Fraser L, Sobiech P, et al. Single nucleotide polymorphism discovery in bovine liver using RNA-seq technology. PLOS ONE. 2017;12(2):e0172687.

[5] Zhao H, Chen M, Tellgren-Roth C, Pettersson U. Fluctuating expression of microRNAs in adenovirus infected cells. Virology. 2015;478:99-111.

[6] Rao R, Bing Zhu Y, Alinejad T, Tiruvayipati S, Lin Thong K, Wang J, et al. RNA-seq analysis of *Macrobrachium rosenbergii* hepatopancreas in response to *Vibrio parahaemolyticus* infection. Gut Pathogens. 2015;7(1):6.

[7] Westermann AJ, Barquist L, Vogel J. Resolving host–pathogen interactions by dual RNA-seq. PLOS Pathogens. 2017;13(2):e1006033.

[8] Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing.

Science (New York, NY). 2008;320(5881):1344-9.

[9] Bainbridge MN, Warren RL, Hirst M, Romanuik T, Zeng T, Go A, et al. Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. BMC genomics. 2006;7:246.

[10] Cheung F, Haas BJ, Goldberg SM, May GD, Xiao Y, Town CD. Sequencing *Medicago truncatula* expressed sequenced tags using 454 Life Sciences technology. BMC genomics. 2006;7:272.

[11] Emrich SJ, Barbazuk WB, Li L, Schnable PS. Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome research. 2007;17(1):69-73.

[12] Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS. SNP discovery via 454 transcriptome sequencing. The Plant journal : for cell and molecular biology. 2007;51(5):910-8.

[13] Weber AP, Weber KL, Carr K, Wilkerson C, Ohlrogge JB. Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. Plant physiology. 2007;144(1):32-42.

[14] Barba M, Czosnek H, Hadidi A. Historical perspective, development and applications of next-generation sequencing in plant virology. Viruses. 2014;6(1):106-36.

[15] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews Genetics. 2009;10(1):57-63.

[16] Bumgarner R. Overview of DNA microarrays: types, applications, and their future. Current protocols in molecular biology. 2013;Chapter 22:Unit-22.1.

[17] Govindarajan R, Duraiyan J, Kaliyappan K, Palanisamy M. Microarray and its applications. Journal of pharmacy & bioallied sciences. 2012;4(Suppl 2):S310-2.

[18] Fu X, Fu N, Guo S, Yan Z, Xu Y, Hu H, et al. Estimating accuracy of RNA-Seq and microarrays with proteomics. BMC genomics. 2009;10:161.

[19] Choudhuri S. Chapter 3 - Genomic Technologies**The opinions expressed in this chapter are the author's own and they do not necessarily reflect the opinions of the FDA, the DHHS, or the Federal Government. In: Choudhuri S, editor. Bioinformatics for Beginners. Oxford: Academic Press; 2014. p. 55-72.

[20] Held GA, Grinstein G, Tu Y. Relationship between gene expression and observed intensities in DNA microarrays--a modeling study. Nucleic acids research. 2006;34(9):e70.

[21] Okoniewski MJ, Miller CJ. Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. BMC bioinformatics. 2006;7:276.

[22] Nagaraj SH, Gasser RB, Ranganathan S. A hitchhiker's guide to expressed sequence tag (EST) analysis. Briefings in bioinformatics. 2007;8(1):6-21.

[23] Parkinson J, Blaxter M. Expressed sequence tags: analysis and annotation. Methods in molecular biology (Clifton, NJ). 2004;270:93-126.

[24] Cai J, Shin S, Wright L, Liu Y, Zhou D, Xue H, et al. Massively parallel signature sequencing profiling of fetal human neural precursor cells. Stem cells and development. 2006;15(2):232-44.

[25] Chu TJ, Peters DG. Serial analysis of the vascular endothelial transcriptome under static and shear stress conditions.

Physiological Genomics. 2008;34(2):185-92.

[26] Reinartz J, Bruyns E, Lin JZ, Burcham T, Brenner S, Bowen B, et al. Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. Briefings in functional genomics & proteomics. 2002;1(1):95-104.

[27] Kodzius R, Kojima M, Nishiyori H, Nakamura M, Fukuda S, Tagami M, et al. CAGE: cap analysis of gene expression. Nature methods. 2006;3(3):211-22.

[28] Fryer RM, Randall J, Yoshida T, Hsiao LL, Blumenstock J, Jensen KE, et al. Global analysis of gene expression: methods, interpretation, and pitfalls. Experimental nephrology. 2002;10(2):64-74.

[29] Zhao X, Valen E, Parker BJ, Sandelin A. Systematic clustering of transcription start site landscapes. PLoS One. 2011;6(8):e23409.

[30] Slatko BE, Gardner AF, Ausubel FM. Overview of Next-Generation Sequencing Technologies. Current protocols in molecular biology. 2018;122(1):e59-e.

[31] Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell. 2008;133(3):523-36.

[32] Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, et al. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature. 2008;453(7199):1239-43.

[33] Morin R, Bainbridge M, Fejes A, Hirst M, Krzywinski M, Pugh T, et al. Profiling the HeLa S3 transcriptome

using randomly primed cDNA and massively parallel short-read sequencing. BioTechniques. 2008;45(1):81-94.

[34] Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature methods. 2008;5(7):621-8.

[35] Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nature methods. 2008;5(7):613-9.

[36] Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K, et al. Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature. 2016;539(7628):309-13.

[37] Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, Iuso A, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. Nature communications. 2017;8:15824.

[38] MacParland SA, Liu JC, Ma XZ, Innes BT, Bartczak AM, Gage BK, et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. Nature communications. 2018;9(1):4383.

[39] James KL, de Silva TI, Brown K, Whittle H, Taylor S, McVean G, et al. Low-Bias RNA Sequencing of the HIV-2 Genome from Blood Plasma. Journal of virology. 2019;93(1).

[40] Bai Y, Wang D, Li W, Huang Y, Ye X, Waite J, et al. Evaluation of the capacities of mouse TCR profiling from short read RNA-seq data. PLoS One. 2018;13(11):e0207020.

[41] Townsley BT, Covington MF, Ichihashi Y, Zumstein K, Sinha NR. BrAD-seq: Breath Adapter Directional sequencing: a streamlined, ultra-simple and fast library preparation protocol for strand specific mRNA library construction. Frontiers in plant science. 2015;6:366.

[42] Langevin SA, Bent ZW, Solberg OD, Curtis DJ, Lane PD, Williams KP, et al. Peregrine: A rapid and unbiased method to produce strand-specific RNA-Seq libraries from small quantities of starting material. RNA biology. 2013;10(4):502-15.

[43] Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC molecular biology. 2006;7:3.

[44] Slomovic S, Laufer D, Geiger D, Schuster G. Polyadenylation of ribosomal RNA in human cells. Nucleic acids research. 2006;34(10):2966-75.

[45] Hrdlickova R, Toloue M, Tian B. RNA-Seq methods for transcriptome analysis. Wiley interdisciplinary reviews RNA. 2017;8(1).

[46] Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nature protocols. 2012;7(8):1534-50.

[47] Archer SK, Shirokikh NE, Preiss T. Selective and flexible depletion of problematic sequences from RNA-seq libraries at the cDNA stage. BMC genomics. 2014;15(1):401.

[48] Armour CD, Castle JC, Chen R, Babak T, Loerch P, Jackson S, et al. Digital transcriptome profiling using selective hexamer priming for cDNA synthesis. Nature methods. 2009;6(9):647-9.

[49] Zhao S, Zhang Y, Gamini R, Zhang B, von Schack D. Evaluation of

two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA+ selection versus rRNA depletion. Scientific reports. 2018;8(1):4781.

[50] Picelli S, Björklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome research. 2014;24(12):2033-40.

[51] Kulski JK. Next-generation sequencing—an overview of the history, tools, and "Omic" applications. 2016:3-60.

[52] Guo J, Xu N, Li Z, Zhang S, Wu J, Kim DH, et al. Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(27):9145-50.

[53] Specifications for the NextSeq 1000 and NextSeq 2000 Systems n.d. [Available from: https://www.illumina. com/systems/sequencing-platforms/ nextseq-1000-2000/specifications.html.

[54] Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. Human molecular genetics. 2010;19(R2):R227-40.

[55] Introducing The Sequel IIe System - Sequencing Evolved n.d. [Available from: https://www.pacb.com/products-and-services/sequel-system/ latest-system-release/.

[56] Kasianowicz JJ, Brandin E, Branton D, Deamer DW. Characterization of individual polynucleotide molecules using a membrane channel. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(24):13770-3.

[57] Reuter JA, Spacek DV, Snyder MP. High-throughput sequencing technologies. Molecular cell. 2015;58(4):586-97.

[58] Oxford Nanopore Technologies n.d. [Available from: https://nanoporetech. com/products.

[59] Metzker ML. Sequencing technologies - the next generation. Nature reviews Genetics. 2010;11(1):31-46.

[60] Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008;26(10):1135-45.

[61] Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475(7356):348-52.

[62] Thompson JF, Steinmann KE. Single molecule sequencing with a HeliScope genetic analysis system. Current protocols in molecular biology. 2010;Chapter 7:Unit7.10.

[63] Snyder M, Du J, Gerstein M. Personal genome sequencing: current approaches and challenges. Genes & development. 2010;24(5):423-31.

[64] Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, et al. Assessing the performance of the Oxford Nanopore Technologies MinION. Biomolecular detection and quantification. 2015;3:1-8.

[65] Weirather JL, de Cesare M, Wang Y, Piazza P, Sebastiano V, Wang XJ, et al. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. F1000Research. 2017;6:100.

[66] Petersen LM, Martin IW, Moschetti WE, Kershaw CM,

Tsongalis GJ. Third-Generation Sequencing in the Clinical Laboratory: Exploring the Advantages and Challenges of Nanopore Sequencing. Journal of clinical microbiology. 2019;58(1).

[67] Faster sequencing and data processing n.d. [Available from: https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology/2-channel-sbs.html.

[68] Ferrarini M, Moretto M, Ward JA, Šurbanovski N, Stevanović V, Giongo L, et al. An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome. BMC genomics. 2013;14:670.

[69] Teng JLL, Yeung ML, Chan E, Jia L, Lin CH, Huang Y, et al. PacBio But Not Illumina Technology Can Achieve Fast, Accurate and Complete Closure of the High GC, Complex Burkholderia pseudomallei Two-Chromosome Genome. Frontiers in microbiology. 2017;8:1448.

[70] Zhang J, Su L, Wang Y, Deng S. Improved High-Throughput Sequencing of the Human Oral Microbiome: From Illumina to PacBio. The Canadian journal of infectious diseases & medical microbiology = Journal canadien des maladies infectieuses et de la microbiologie medicale. 2020;2020:6678872.

[71] Greig DR, Jenkins C, Gharbia S, Dallman TJ. Comparison of single-nucleotide variants identified by Illumina and Oxford Nanopore technologies in the context of a potential outbreak of Shiga toxin-producing *Escherichia coli*. GigaScience. 2019;8(8).

[72] Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC genomics. 2012;13:341.

[73] Lahens NF, Ricciotti E, Smirnova O, Toorens E, Kim EJ, Baruzzo G, et al. A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression. BMC genomics. 2017;18(1):602.

[74] Suzuki S, Ranade S, Osaki K, Ito S, Shigenari A, Ohnuki Y, et al. Reference Grade Characterization of Polymorphisms in Full-Length HLA Class I and II Genes With Short-Read Sequencing on the ION PGM System and Long-Reads Generated by Single Molecule, Real-Time Sequencing on the PacBio Platform. Frontiers in immunology. 2018;9:2294.

[75] Tan MH, Austin CM, Hammer MP, Lee YP, Croft LJ, Gan HM. Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (*Amphiprion ocellaris*) genome assembly. GigaScience. 2018;7(3):1-6.

[76] Guerrero-Sanchez VM, Maldonado-Alconada AM, Amil-Ruiz F, Verardi A, Jorrín-Novo JV, Rey MD. Ion Torrent and lllumina, two complementary RNA-seq platforms for constructing the holm oak (*Quercus ilex*) transcriptome. PLoS One. 2019;14(1):e0210356.

[77] Dhar R, Seethy A, Pethusamy K, Singh S, Rohil V, Purkayastha K, et al. De novo assembly of the Indian blue peacock (*Pavo cristatus*) genome using Oxford Nanopore technology and Illumina sequencing. GigaScience. 2019;8(5).

[78] Li W, Li K, Zhang QJ, Zhu T, Zhang Y, Shi C, et al. Improved hybrid de novo genome assembly and annotation of African wild rice, *Oryza longistaminata*, from Illumina and PacBio sequencing reads. The plant genome. 2020;13(1):e20001.

[79] Huang B, Rong H, Ye Y, Ni Z, Xu M, Zhang W, et al. Transcriptomic analysis of flower color variation in the ornamental crabapple (Malus spp.) half-sib family through Illumina and PacBio Sequel sequencing. Plant physiology and biochemistry : PPB. 2020;149:27-35.

[80] Advantages of paired-end and single-read sequencing n.d. [Available from: https://www.illumina.com/science/technology/next-generation-sequencing/plan-experiments/paired-end-vs-single-read.html.

[81] Corley SM, MacKenzie KL, Beverdam A, Roddam LF, Wilkins MR. Differentially expressed genes from RNA-Seq and functional enrichment results are affected by the choice of single-end versus paired-end reads and stranded versus non-stranded protocols. BMC genomics. 2017;18(1):399.

[82] Tavassoly I, Goldfarb J, Iyengar R. Systems biology primer: the basic methods and approaches. Essays in biochemistry. 2018;62(4):487-500.

[83] Kitano H. Systems biology: a brief overview. Science (New York, NY). 2002;295(5560):1662-4.

[84] Dix A, Vlaic S, Guthke R, Linde J. Use of systems biology to decipher host-pathogen interaction networks and predict biomarkers. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases. 2016;22(7):600-6.

[85] Cesur MF, Durmuş S. Systems Biology Modeling to Study Pathogen-Host Interactions. Methods in molecular biology (Clifton, NJ). 2018;1734:97-112.

[86] Westermann AJ, Gorski SA, Vogel J. Dual RNA-seq of pathogen and host. Nature reviews Microbiology. 2012;10(9):618-30.

[87] Falkow S, Isberg RR, Portnoy DA. The interaction of bacteria with mammalian cells. Annual review of cell biology. 1992;8:333-63.

[88] Saliba A-E, C Santos S, Vogel J. New RNA-seq approaches for the study of bacterial pathogens. Current Opinion in Microbiology. 2017;35:78-87.

[89] Westermann AJ, Förstner KU, Amman F, Barquist L, Chao Y, Schulte LN, et al. Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. Nature. 2016;529(7587):496-501.

[90] Baddal B, Muzzi A, Censini S, Calogero RA, Torricelli G, Guidotti S, et al. Dual RNA-seq of Nontypeable *Haemophilus influenzae* and Host Cell Transcriptomes Reveals Novel Insights into Host-Pathogen Cross Talk. mBio. 2015;6(6):e01765-15.

[91] Aprianto R, Slager J, Holsappel S, Veening JW. Time-resolved dual RNA-seq reveals extensive rewiring of lung epithelial and pneumococcal transcriptomes during early infection. Genome biology. 2016;17(1):198.

[92] Ritchie ND, Evans TJ. Dual RNA-seq in *Streptococcus pneumoniae* Infection Reveals Compartmentalized Neutrophil Responses in Lung and Pleural Space. mSystems. 2019;4(4).

[93] Rienksma RA, Suarez-Diez M, Mollenkopf HJ, Dolganov GM, Dorhoi A, Schoolnik GK, et al. Comprehensive insights into transcriptional adaptation of intracellular mycobacteria by microbe-enriched dual RNA sequencing. BMC genomics. 2015;16(1):34.

[94] Pisu D, Huang L, Grenier JK, Russell DG. Dual RNA-Seq of Mtb-Infected Macrophages In Vivo Reveals Ontologically Distinct Host-Pathogen Interactions. Cell reports. 2020;30(2):335-50.e4.

[95] Montoya DJ, Andrade P, Silva BJA, Teles RMB, Ma F, Bryson B, et al. Dual RNA-Seq of Human Leprosy Lesions Identifies Bacterial Determinants Linked to Host Immune Response. Cell reports. 2019;26(13):3574-85.e3.

[96] Yimthin T, Cliff JM, Phunpang R, Ekchariyawat P, Kaewarpai T, Lee JS, et al. Blood transcriptomics to characterize key biological pathways and identify biomarkers for predicting mortality in melioidosis. Emerging microbes & infections. 2020:1-47.

[97] Whitaker-Dowling P, Youngner JS. VIRUS-HOST CELL INTERACTIONS. Encyclopedia of Virology. 1999:1957-61.

[98] Lisnic VJ, Babic Cac M, Lisnic B, Trsan T, Mefferd A, Das Mukhopadhyay C, et al. Dual analysis of the murine cytomegalovirus and host cell transcriptomes reveal new aspects of the virus-host cell interface. PLoS Pathog. 2013;9(9):e1003611.

[99] Park SJ, Kumar M, Kwon HI, Seong RK, Han K, Song JM, et al. Dynamic changes in host gene expression associated with H5N8 avian influenza virus infection in mice. Scientific reports. 2015;5:16512.

[100] Jones M, Dry IR, Frampton D, Singh M, Kanda RK, Yee MB, et al. RNA-seq analysis of host and viral gene expression highlights interaction between varicella zoster virus and keratinocyte differentiation. PLoS Pathog. 2014;10(1):e1003896.

[101] Kozak RA, Fraser RS, Biondi MJ, Majer A, Medina SJ, Griffin BD, et al. Dual RNA-Seq characterization of host and pathogen gene expression in liver cells infected with Crimean-Congo Hemorrhagic Fever Virus. PLoS neglected tropical diseases. 2020;14(4):e0008105.

[102] Fabozzi G, Oler AJ, Liu P, Chen Y, Mindaye S, Dolan MA, et al.

Strand-Specific Dual RNA Sequencing of Bronchial Epithelial Cells Infected with Influenza A/H3N2 Viruses Reveals Splicing of Gene Segment 6 and Novel Host-Virus Interactions. Journal of virology. 2018;92(17).

[103] Michlmayr D, Kim EY, Rahman AH, Raghunathan R, Kim-Schulze S, Che Y, et al. Comprehensive Immunoprofiling of Pediatric Zika Reveals Key Role for Monocytes in the Acute Phase and No Effect of Prior Dengue Virus Infection. Cell reports. 2020;31(4):107569.

[104] Wesolowska-Andersen A, Everman JL, Davidson R, Rios C, Herrin R, Eng C, et al. Dual RNA-seq reveals viral infections in asthmatic children without respiratory illness which are associated with changes in the airway transcriptome. Genome biology. 2017;18(1):12.

[105] Sridhar S, To KK, Chan JF, Lau SK, Woo PC, Yuen KY. A systematic approach to novel virus discovery in emerging infectious disease outbreaks. The Journal of molecular diagnostics : JMD. 2015;17(3):230-41.

[106] Chen L, Liu W, Zhang Q, Xu K, Ye G, Wu W, et al. RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. Emerging microbes & infections. 2020;9(1):313-9.

[107] Cao M, Zhang S, Li M, Liu Y, Dong P, Li S, et al. Discovery of Four Novel Viruses Associated with Flower Yellowing Disease of Green Sichuan Pepper (Zanthoxylum Armatum) by Virome Analysis. Viruses. 2019;11(8).

[108] Wright AA, Cross AR, Harper SJ. A bushel of viruses: Identification of seventeen novel putative viruses by RNA-seq in six apple trees. PLoS One. 2020;15(1):e0227669.

old parasitological tradition. Journal of helminthology. 2017;91(6):647-56.

[125] Ngara M, Palmkvist M, Sagasser S, Hjelmqvist D, Björklund Å K, Wahlgren M, et al. Exploring parasite heterogeneity using single-cell RNA-seq reveals a gene signature among sexual stage *Plasmodium falciparum* parasites. Experimental cell research. 2018;371(1):130-8.

[126] Greif G, Ponce de Leon M, Lamolle G, Rodriguez M, Piñeyro D, Tavares-Marques LM, et al. Transcriptome analysis of the bloodstream stage from the parasite Trypanosoma vivax. BMC genomics. 2013;14:149.

[127] Choi YJ, Aliota MT, Mayhew GF, Erickson SM, Christensen BM. Dual RNA-seq of parasite and host reveals gene expression dynamics during filarial worm-mosquito interactions. PLoS neglected tropical diseases. 2014;8(5):e2905.

[128] Foth BJ, Tsai IJ, Reid AJ, Bancroft AJ, Nichol S, Tracey A, et al. Whipworm genome and dual-species transcriptome analyses provide molecular insights into an intimate host-parasite interaction. Nature genetics. 2014;46(7):693-700.

[129] Anderson L, Amaral MS, Beckedorff F, Silva LF, Dazzani B, Oliveira KC, et al. *Schistosoma mansoni* Egg, Adult Male and Female Comparative Gene Expression Analysis and Identification of Novel Genes by RNA-Seq. PLoS neglected tropical diseases. 2015;9(12):e0004334.

[130] Pittman KJ, Aliota MT, Knoll LJ. Dual transcriptional profiling of mice and Toxoplasma gondii during acute and chronic infection. BMC genomics. 2014;15(1):806.

[131] Soto C, Satani N. The intricate mechanisms of neurodegeneration in

prion diseases. Trends in molecular medicine. 2011;17(1):14-24.

[132] Bellingham SA, Coleman BM, Hill AF. Small RNA deep sequencing reveals a distinct miRNA signature released in exosomes from prion-infected neuronal cells. Nucleic acids research. 2012;40(21):10937-49.

[133] Carroll JA, Race B, Williams K, Striebel J, Chesebro B. RNA-seq and network analysis reveal unique glial gene expression signatures during prion infection. Molecular brain. 2020;13 (1):71.

[134] Thackray AM, Lam B, Shahira Binti Ab Razak A, Yeo G, Bujdoso R. Transcriptional signature of prion-induced neurotoxicity in a Drosophila model of transmissible mammalian prion disease. The Biochemical journal. 2020;477(4):833-52.

[135] Bakuła Z, Gromadka R, Gawor J, Siedlecki P, Pomorski JJ, Maciszewski K, et al. Sequencing and Analysis of the Complete Organellar Genomes of Prototheca wickerhamii. Frontiers in plant science. 2020;11:1296.

[136] Zeng X, Kudinha T, Kong F, Zhang QQ. Comparative Genome and Transcriptome Study of the Gene Expression Difference Between Pathogenic and Environmental Strains of Prototheca zopfii. Frontiers in microbiology. 2019;10:443.

[137] Vlasova-St. Louis I, Chang CC, Shahid S, French MA, Bohjanen PR. Transcriptomic Predictors of Paradoxical Cryptococcosis-Associated Immune Reconstitution Inflammatory Syndrome. Open Forum Infectious Diseases. 2018;5(7).

[138] Seelbinder B, Wallstabe J, Marischen L, Weiss E, Wurster S, Page L, et al. Triple RNA-Seq Reveals Synergy in a Human Virus-Fungus Co-infection Model. Cell reports. 2020;33(7):108389.

[139] Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. Genome biology. 2016;17:13.

[140] Alkhateeb A, Rueda L. Zseq: An Approach for Preprocessing Next-Generation Sequencing Data. Journal of computational biology : a journal of computational molecular cell biology. 2017;24(8):746-55.

[141] Zhao S, Zhang B, Zhang Y, Gordon W, Du S, Paradis T, et al. Bioinformatics for RNA-seq data analysis. 2016:125-49.

[142] Andrews S. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom; 2010.

[143] Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics (Oxford, England). 2012;28(16):2184-5.

[144] Dai M, Thompson RC, Maher C, Contreras-Galindo R, Kaplan MH, Markovitz DM, et al. NGSQC: cross-platform quality analysis pipeline for deep sequencing data. BMC genomics. 2010;11 Suppl 4(Suppl 4):S7.

[145] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England). 2014;30(15):2114-20.

[146] Martin MJEj. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011;17(1):10-2.

[147] Liu X, Yan Z, Wu C, Yang Y, Li X, Zhang G. FastProNGS: fast preprocessing of next-generation sequencing reads. BMC bioinformatics. 2019;20(1):345.

[148] Martínez-Alcántara A, Ballesteros E, Feng C, Rojas M, Koshinsky H, Fofanov VY, et al. PIQA: pipeline for Illumina G1 genome analyzer data quality assessment. Bioinformatics (Oxford, England). 2009;25(18):2438-9.

[149] Pérez-Rubio P, Lottaz C, Engelmann JC. FastqPuri: high-performance preprocessing of RNA-seq data. BMC bioinformatics. 2019;20(1):226.

[150] Zhou Q, Su X, Jing G, Chen S, Ning K. RNA-QC-chain: comprehensive and fast quality control for RNA-Seq data. BMC genomics. 2018;19(1):144.

[151] Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics (Oxford, England). 2018;34(17):i884-i90.

[152] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England). 2013;29(1):15-21.

[153] Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology. 2013;14(4):R36.

[154] Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, et al. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic acids research. 2010;38(18):e178.

[155] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012;9(4):357-9.

[156] Boratyn GM, Thierry-Mieg J, Thierry-Mieg D, Busby B, Madden TL. Magic-BLAST, an accurate RNA-seq aligner for long and short reads. BMC bioinformatics. 2019;20(1):405.

[157] Baruzzo G, Hayer KE, Kim EJ, Di Camillo B, FitzGerald GA, Grant GR. Simulation-based comprehensive benchmarking of RNA-seq aligners. Nature methods. 2017;14(2):135-9.

[158] Schaarschmidt S, Fischer A, Zuther E, Hincha DK. Evaluation of Seven Different RNA-Seq Alignment Tools Based on Experimental Data from the Model Plant *Arabidopsis thaliana*. International journal of molecular sciences. 2020;21(5).

[159] Engström PG, Steijger T, Sipos B, Grant GR, Kahles A, Rätsch G, et al. Systematic evaluation of spliced alignment programs for RNA-seq data. Nature methods. 2013;10(12):1185-91.

[160] Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics (Oxford, England). 2011;27(17):2325-9.

[161] Li JJ, Jiang C-R, Brown JB, Huang H, Bickel PJJPotNAoS. Sparse linear modeling of next-generation mRNA sequencing (RNA-Seq) data for isoform discovery and abundance estimation. 2011;108(50):19867-72.

[162] Mezlini AM, Smith EJ, Fiume M, Buske O, Savich GL, Shah S, et al. iReckon: simultaneous isoform discovery and abundance estimation from RNA-seq data. Genome research. 2013;23(3):519-29.

[163] Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SLJNb. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. 2015;33(3):290-5.

[164] Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols. 2013;8(8):1494-512.

[165] Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, et al. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics (Oxford, England). 2014;30(12):1660-6.

[166] Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, et al. De novo assembly and analysis of RNA-seq data. Nature methods. 2010;7(11):909-12.

[167] Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics (Oxford, England). 2012;28(8):1086-92.

[168] Anders S, Pyl PT, Huber WJB. HTSeq—a Python framework to work with high-throughput sequencing data. 2015;31(2):166-9.

[169] Liao Y, Smyth GK, Shi WJB. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. 2014;30(7):923-30.

[170] Zhang C, Zhang B, Vincent M, Zhao S. Bioinformatics tools for RNA-seq gene and isoform quantification. 2016;3:140.

[171] Li B, Dewey CNJBb. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. 2011;12(1):323.

[172] Roberts A, Pachter LJNm. Streaming fragment assignment for real-time analysis of sequencing experiments. 2013;10(1):71-3.

[173] Nariai N, Kojima K, Mimori T, Sato Y, Kawai Y, Yamaguchi-Kabata Y, et al. TIGAR2: sensitive and accurate

estimation of transcript isoform expression with longer RNA-Seq reads. 2014;15(S10):S5.

[174] Patro R, Duggal G, Love MI, Irizarry RA, Kingsford CJNm. Salmon provides fast and bias-aware quantification of transcript expression. 2017;14(4):417-9.

[175] Bray NL, Pimentel H, Melsted P, Pachter LJNb. Near-optimal probabilistic RNA-seq quantification. 2016;34(5):525-7.

[176] Patro R, Mount SM, Kingsford CJNb. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. 2014;32(5):462-4.

[177] Li DJEP. Statistical Methods for RNA Sequencing Data Analysis. 2019:85-99.

[178] Hardcastle TJ, Kelly KA. baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC bioinformatics. 2010;11:422.

[179] Anders S, Huber W. Differential expression analysis for sequence count data. Genome biology. 2010;11(10):R106.

[180] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15(12):550.

[181] Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BM, et al. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics (Oxford, England). 2013;29(8):1035-43.

[182] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics (Oxford, England). 2010;26(1):139-40.

[183] Di Y, Schafer D, Cumbie J, Chang J. NBPSeq: Negative Binomial Models for RNA-Sequencing Data R package version 0.3. 0, URL http://CRAN. R-project. org/package= NBPSeq. 2015.

[184] Li J, Witten DM, Johnstone IM, Tibshirani R. Normalization, testing, and false discovery rate estimation for RNA-sequencing data. Biostatistics. 2012;13(3):523-38.

[185] Auer PL, Doerge RWJSaig, biology m. A two-stage Poisson model for testing RNA-seq data. 2011;10(1).

[186] Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511-5.

[187] Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 2013;31(1):46-53.

[188] Tarazona S, García-Alcalde F, Dopazo J, Ferrer A, Conesa A. Differential expression in RNA-seq: a matter of depth. Genome research. 2011;21(12):2213-23.

[189] Li J, Tibshirani R. Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. Stat Methods Med Res. 2013;22(5):519-36.

[190] Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology. 2004;3:Article3.

[191] Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome biology. 2014;15(2):R29.

[192] Wang L, Feng Z, Wang X, Wang X, Zhang X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics (Oxford, England). 2010;26(1):136-8.

[193] van de Wiel MA, Neerincx M, Buffart TE, Sie D, Verheul HMW. ShrinkBayes: a versatile R-package for analysis of count-based sequencing data in complex study designs. BMC bioinformatics. 2014;15(1):116.

[194] Davidson NM, Oshlack A. Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. Genome biology. 2014;15(7):410.

[195] Robles JA, Qureshi SE, Stephen SJ, Wilson SR, Burden CJ, Taylor JM. Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. BMC genomics. 2012;13:484.

[196] Soneson C, Delorenzi M. A comparison of methods for differential expression analysis of RNA-seq data. BMC bioinformatics. 2013;14:91.

[197] Rapaport F, Khanin R, Liang Y, Pirun M, Krek A, Zumbo P, et al. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome biology. 2013;14(9):R95.

[198] Zhang ZH, Jhaveri DJ, Marshall VM, Bauer DC, Edson J, Narayanan RK, et al. A comparative study of techniques for differential expression analysis on RNA-Seq data. PLoS One. 2014;9(8):e103207.

[199] Seyednasrollah F, Laiho A, Elo LL. Comparison of software packages for detecting differential expression in RNA-seq studies. Briefings in bioinformatics. 2015;16(1):59-70.

[200] Rajkumar AP, Qvist P, Lazarus R, Lescai F, Ju J, Nyegaard M, et al. Experimental validation of methods for differential gene expression analysis and sample pooling in RNA-seq. BMC genomics. 2015;16(1):548.

[201] Costa-Silva J, Domingues D, Lopes FM. RNA-Seq differential expression analysis: An extended review and a software tool. PLoS One. 2017;12(12):e0190152.

[202] Mehmood A, Laiho A, Venäläinen MS, McGlinchey AJ, Wang N, Elo LL. Systematic evaluation of differential splicing tools for RNA-seq studies. Briefings in bioinformatics. 2020;21(6):2052-65.

[203] Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. Genome research. 2012;22(10):2008-17.

[204] Hartley SW, Mullikin JC. Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq. Nucleic acids research. 2016;44(15):e127.

[205] Vaquero-Garcia J, Barrera A, Gazzara MR, González-Vallinas J, Lahens NF, Hogenesch JB, et al. A new view of transcriptome complexity and regulation through the lens of local splicing variations. eLife. 2016;5:e11752.

[206] Zhu D, Deng N, Bai C. A generalized dSpliceType framework to detect differential splicing and differential expression events using RNA-Seq. IEEE transactions on nanobioscience. 2015;14(2):192-202.

[207] Trincado JL, Entizne JC, Hysenaj G, Singh B, Skalic M, Elliott DJ, et al. SUPPA2: fast, accurate, and

uncertainty-aware differential splicing analysis across multiple conditions. Genome biology. 2018;19(1):40.

[208] Hu Y, Huang Y, Du Y, Orellana CF, Singh D, Johnson AR, et al. DiffSplice: the genome-wide detection of differential splicing events with RNA-seq. Nucleic acids research. 2013;41(2):e39.

[209] Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome biology. 2010;11(2):R14.

[210] Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC bioinformatics. 2013;14:7.

[211] Wang X, Cairns MJ. SeqGSEA: a Bioconductor package for gene set enrichment analysis of RNA-Seq data integrating differential expression and splicing. Bioinformatics (Oxford, England). 2014;30(12):1777-9.

[212] Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic acids research. 2004;32(Database issue):D277-80.

[213] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics. 2000;25(1):25-9.

[214] Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nature methods. 2015;12(2):115-21.

[215] Choi YJ, Aliota MT, Mayhew GF, Erickson SM, Christensen BM. Dual RNA-seq of Parasite and Host Reveals Gene Expression Dynamics during Filarial Worm–Mosquito Interactions.

PLoS neglected tropical diseases. 2014;8(5):e2905.

[216] Liao ZX, Ni Z, Wei XL, Chen L, Li JY, Yu YH, et al. Dual RNA-seq of Xanthomonas oryzae pv. oryzicola infecting rice reveals novel insights into bacterial-plant interaction. PLOS ONE. 2019;14(4):e0215039.

[217] Sun Y, Zhuang Z, Wang X, Huang H, Fu Q, Yan Q. Dual RNA-seq reveals the effect of the flgM gene of Pseudomonas plecoglossicida on the immune response of *Epinephelus coioides*. Fish & shellfish immunology. 2019;87:515-23.

[218] Haque A, Engel J, Teichmann SA, Lönnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Medicine. 2017;9(1):75.

[219] Penaranda C, Hung DT. Single-Cell RNA Sequencing to Understand Host-Pathogen Interactions. ACS infectious diseases. 2019;5(3):336-44.

[220] Avraham R, Haseley N, Brown D, Penaranda C, Jijon HB, Trombetta JJ, et al. Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses. Cell. 2015;162(6):1309-21.

[221] Avital G, Avraham R, Fan A, Hashimshony T, Hung DT, Yanai I. scDual-Seq: mapping the gene regulatory program of Salmonella infection by host and pathogen single-cell RNA-sequencing. Genome biology. 2017;18(1):200.

[222] Golumbeanu M, Cristinelli S, Rato S, Munoz M, Cavassini M, Beerenwinkel N, et al. Single-Cell RNA-Seq Reveals Transcriptional Heterogeneity in Latent and Reactivated HIV-Infected Cells. Cell reports. 2018;23(4):942-50.

[223] Brazovskaja A, Treutlein B, Camp JG. High-throughput single-cell

transcriptomics on organoids. Current opinion in biotechnology. 2019;55:167-71.

[224] Combes AN, Phipson B, Zappia L, Lawlor KT, Er PX, Oshlack A, et al. High throughput single cell RNA-seq of developing mouse kidney and human kidney organoids reveals a roadmap for recreating the kidney. 2017:235499.

[225] Collin J, Queen R, Zerti D, Dorgau B, Hussain R, Coxhead J, et al. Deconstructing Retinal Organoids: Single Cell RNA-Seq Reveals the Cellular Components of Human Pluripotent Stem Cell-Derived Retina. Stem Cells. 2019;37(5):593-8.

[226] Burgess DJ. Genetic screens: Combining CRISPR perturbations and RNA-seq. Nature reviews Genetics. 2017;18(2):67.

[227] Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. Nature reviews Genetics. 2019;20(11):631-56.

[228] Halpern KB, Shenhav R, Matcovitch-Natan O, Toth B, Lemze D, Golan M, et al. Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. Nature. 2017;542(7641):352-6.

[229] Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, et al. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. Nature protocols. 2015;10(3):442-58.

[230] Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. Spatially resolved, highly multiplexed RNA profiling in single cells. Science (New York, NY). 2015;348(6233):aaa6090.

[231] Jean Beltran PM, Federspiel JD, Sheng X, Cristea IM. Proteomics and integrative omic approaches for understanding host–pathogen interactions and infectious diseases. Molecular Systems Biology. 2017;13 (3):922.

[232] Hsu PY, Calviello L, Wu HL, Li FW, Rothfels CJ, Ohler U, et al. Super-resolution ribosome profiling reveals unannotated translation events in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(45):E7126-e35.

[233] Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science (New York, NY). 2009;324(5924):218-23.

[234] Paulet D, David A, Rivals E. Ribo-seq enlightens codon usage bias. DNA research : an international journal for rapid publication of reports on genes and genomes. 2017;24(3):303-210.

[235] Holmes MJ, Shah P, Wek RC, Sullivan WJ. Simultaneous Ribosome Profiling of Human Host Cells Infected with <span class="named-content genus-species" id="named-content-1">Toxoplasma gondii</span&gt. mSphere. 2019;4(3):e00292-19.

[236] Dai A, Cao S, Dhungel P, Luan Y, Liu Y, Xie Z, et al. Ribosome Profiling Reveals Translational Upregulation of Cellular Oxidative Phosphorylation mRNAs during Vaccinia Virus-Induced Host Shutoff. Journal of virology. 2017;91(5).

[237] Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome biology. 2020;21(1):30.

[238] De Maio N, Shaw LP, Hubbard A, George S, Sanderson ND, Swann J, et al. Comparison of long-read sequencing technologies in the hybrid assembly of

complex bacterial genomes. Microbial genomics. 2019;5(9).

[239] Mahmoud M, Gobet N, Cruz-Dávalos DI, Mounier N, Dessimoz C, Sedlazeck FJ. Structural variant calling: the long and the short of it. Genome biology. 2019;20(1):246.

[240] Linsen SE, de Wit E, Janssens G, Heater S, Chapman L, Parkin RK, et al. Limitations and possibilities of small RNA digital gene expression profiling. Nature methods. 2009;6(7):474-6.

[241] Dabney J, Meyer M. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. BioTechniques. 2012;52(2):87-94.

[242] Raabe CA, Tang TH, Brosius J, Rozhdestvensky TS. Biases in small RNA deep sequencing data. Nucleic acids research. 2014;42(3):1414-26.

[243] Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, et al. Library construction for next-generation sequencing: overviews and challenges. BioTechniques. 2014;56(2):61-4, 6, 8, passim.

[244] Barrett SP, Salzman J. Circular RNAs: analysis, expression and potential functions. Development (Cambridge, England). 2016;143(11):1838-47.

[245] Szabo L, Salzman J. Detecting circular RNAs: bioinformatic and experimental challenges. Nature reviews Genetics. 2016;17(11):679-92.

[246] Fan X, Zhang X, Wu X, Guo H, Hu Y, Tang F, et al. Single-cell RNA-seq transcriptome analysis of linear and circular RNAs in mouse preimplantation embryos. Genome biology. 2015;16(1):148.

[247] Kim JK, Kolodziejczyk AA, Ilicic T, Teichmann SA, Marioni JC.

Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. Nature communications. 2015;6:8687.

[248] Jia C, Hu Y, Kelly D, Kim J, Li M, Zhang NR. Accounting for technical noise in differential expression analysis of single-cell RNA sequencing data. Nucleic acids research. 2017;45(19): 10978-88.

[249] Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. Nature reviews Genetics. 2016;17(5):257-71.

[250] Hardwick SA, Deveson IW, Mercer TR. Reference standards for next-generation sequencing. Nature reviews Genetics. 2017;18(8):473-84.

[251] Munro SA, Lund SP, Pine PS, Binder H, Clevert DA, Conesa A, et al. Assessing technical performance in differential gene expression experiments with external spike-in RNA control ratio mixtures. Nature communications. 2014;5:5125.

[252] Consortium SM-I. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nat Biotechnol. 2014;32(9):903-14.

[253] Li S, Tighe SW, Nicolet CM, Grove D, Levy S, Farmerie W, et al. Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. Nat Biotechnol. 2014;32(9):915-25.

[254] t Hoen PA, Friedländer MR, Almlöf J, Sammeth M, Pulyakhina I, Anvar SY, et al. Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. Nat Biotechnol. 2013;31(11):1015-22.