Int.J.Curr.Microbiol.App.Sci (2023) 12(03): 171-176

EXCELLENT PUBLISHERS

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 Volume 12 Number 3 (2023)
Journal homepage: http://www.ijcmas.com

Review Article https://doi.org/10.20546/ijcmas.2023.1203.021

Trancsriptome Sequencing: Emerging Tool to Reveal Genomic Data and Gene Function

Akhil Gupta* and S. K. Singh

GDC Thatyur Tehri Garwal, Uttrakhand, India

**Corresponding author*

Keywords

Transcriptome, genome, next generation sequencing, molecular markers

Article Info

Received: 05 February 2023
Accepted: 26 February 2023
Available Online: 10 March 2023

ABSTRACT

All the information present in a cell, an organism possesses to survive is known as genetic information. It stores, processes and transmits biological data from generation to generation. The first step in decoding these genetic instructions is to copy part of the base sequence from DNA into RNA.RNA sequencing provide insight into the transcriptome of a cell. RNA-sequencing has multiple applications in agriculture, animal husbandry, and health. For research that involves monitoring the cellular response to stimuli, treatment or environment change, understanding the differential gene expression patterns can provide insights into their physiological responses. Profiling the various RNA molecules provides insights into factors defining specific phenotypes for e.g. the color of fruit, medicinal properties of herbs, seed phenotype, secondary metabolites in plants, etc. Further, Transcriptome analysis is widely used to gain insight into processes such as cellular differentiation, transcription regulation, polymorphism, biomarker discovery & prediction of some useful genes and their functional validation.

Introduction

A transcriptome is the full content of messenger RNA molecules expressed by an organism. It is the set of all RNA transcripts, including coding and non-coding, in an individual or a population of cells. In contrast with the genome, which is characterized by its stability, the transcriptome actively changes. Transcriptome can be seen as precursor of proteome (set of proteins expressed by an organism). In order to study the transcriptome of an organism, complementary DNA (cDNA) is synthesized using mRNA as template. mRNA can be isolated from different tissues of an organism at various time intervals to be able to capture maximum number of genes expressed in an organism in one time. (He *et al.,* 2022; Singh *et al.,* 2020). Microarray technique can be used to examine changes in the transcriptome. Microarrays is used to measure the expression of thousands of genes at the same time, as well as to provide gene expression profiles, which describe changes in the transcriptome in response to a particular condition. Genome-wide analyses and high-throughput screening was domain of

biomedical applications and genetic model organisms. With the rapid development next-generation sequencing and simultaneous development of bioinformatic tools, the situation has thoroughly changed. Now sequencing is being done at least 1000 times faster rate in comparison to Sanger method.

Genome- wide thinking is carving its way into new disciplines like evolutionary biology and conservation biology that were historically confined to small-scale genetic approaches. (Ren *et al.,* 2020; Ellergren *et al.,* 2012). Conservation genomics" encompasses the idea that genome-scale data will improve the capacity of resource managers to protect species. Since the rate of extinction in the last century conservatively estimated to be 22 times faster than the historical rate. Accessibility to genome-level information is transforming these areas, as it attempts to answer long-standing questions like the genetic basis of local adaptation and speciation or the evolution of gene expression profiles that until recently were out of reach (Doben, 2016; Anders 2012)

DNA sequencing system founded by the British chemist Sanger in 1975 and the cDNA PCR method by Karry Mullis in 1984 it became possible to sequence RNA by sequencing the cDNA it synthesized. With the discovery of reverse transcriptase it is now possible to convert the mRNA to stable DNA.

Based on these breakthrough work, microarray (chip) technology emerged. Microarray technology is based on the principle of molecular hybridization to mount hundreds of known partial sequence DNA probes on a solid support slide or nylon membrane using an automated instrument Arrayer. A large number of genes can be quantitatively detected through one hybridization. (Arvind *et al.,* 2020)

Transcriptome of different biological samples can be sequenced at faster rate with the help of cDNA. This makes microarray a systematic way to know the gene function. However, the design and functional application of the technology are mostly based on known genes, and the analysis of unknown genes is still less. In addition, it is difficult for the chip to detect multiple transcripts which are formed by alternative splicing (Puvvala, 2019).

In recent years, another new technology, RNA-Seq, has come emerged. RNA-Seq is based on next generation sequencing (NGS). NGS, or high-throughput sequencing, is represented by Roche's 454 technology, Illumina's Solexa technology, and ABI's SOLiD technology (Martin 2011, Jain 2012). It is characterized by shorter reads and stronger ability in sequencing millions of DNA molecules in parallel with shorter time and lower price in contrast to the first-generation sequencing. (Chitten *et al.,* 2020)

RNA-seq, also called whole-transcriptome shotgun sequencing, refers to the use of high-throughput sequencing technologies for characterizing the RNA content and composition of a given sample. RNA-seq is one of the most important tools for transcriptome study.

It overcomes the drawbacks of microarray and brings deeper insight for transcriptome research. RNA-seq promotes the understanding of gene expression under different conditions, and allows for the discovery of new genes, transcription patterns, and RNAs which helps to understand cell function and metabolic mechanism (Hittalmani, 2017)

RNA-Sequencing can determine the absolute number of each molecule in a cell population. Further, RNA-Sequence information promotes new gene discovery. (Ekblom *et al.,* 2011) The annotations to transcripts in existing databases may not be comprehensive. RNA-Sequence information can lead to genome annotations, which allow the identification of new genes. (Thakur *et al.,* 2020)

In addition to this, RNA-Sequence information also demonstrates its huge potential in identifying sequence differences such as fusion gene, identification and coding sequence polymorphism

studies. Alternative splicing leads to one gene to produce multiple mRNA transcripts, and different mRNAs may be translated into different proteins with different functions. (Nguyen, 2018)

An important aspect of transcriptomics research is the discovery and analysis of ncRNA (non-coding RNA). Studies showed that least 93% of the human genome is transcribed into RNA. With less than 2% as protein encoding region, the remaining 91% of the genome could be transcribed into non-protein-encoded RNA molecule, i.e. ncRNA (non coding RNA)

Transcriptome sequence thus constitutes a resource to develop a large number of molecular markers such as single-nucleotide polymorphisms and microsatellites (Baird *et al.,* 2008; Bao *et al.,* 2011; Jain *et al.,* 2012). Transcriptome sequencing of crop plants demonstrate the efficiency and reliability of transcriptome analysis for rapid identification of desirable genes. Immature seed transcriptome also revealed the identity of many genes that are differentially expressed in cultivated and wild accessions. The differential gene expression analysis through transcriptome sequencing helped in the identification of hundreds of genes that are differentially expressed in cultivated and wild cucumber.

Comparative transcriptome analyses using RNAseq and Gene ontology and Mapman analysis between cultivated Pigeonpea {*Cajanus cajan* (L.)} and wild species {*Cajanus platycarpus* (Benth.) Maesen} showed the transcripts that are significantly expressed in wild relatives pertaining to signaling, transcription factors, and stress-responsive genes.

Transcriptome reconstruction is an important application of RNA-Seq, which makes transcript characterization, genome annotation, novel gene detection and alternative splicing discovery possible. In situations where full sequencing cannot be afforded, but the application requires the use of many markers (e.g. genome scans), the transcriptome provides a useful functionally relevant subset of the genome.

The great advantage of RNA-seq data over other next- generation-sequencing applications is that it allows users to investigate differences in gene expression patterns between populations, for example in the context of speciation (Wolf *et al.,* 2010) or eco-type-specific adaptation. RNA-seq further provides information on RNA splice events; these are not readily detected by standard smicroarrays (Mortazavi *et al.,* 2008).

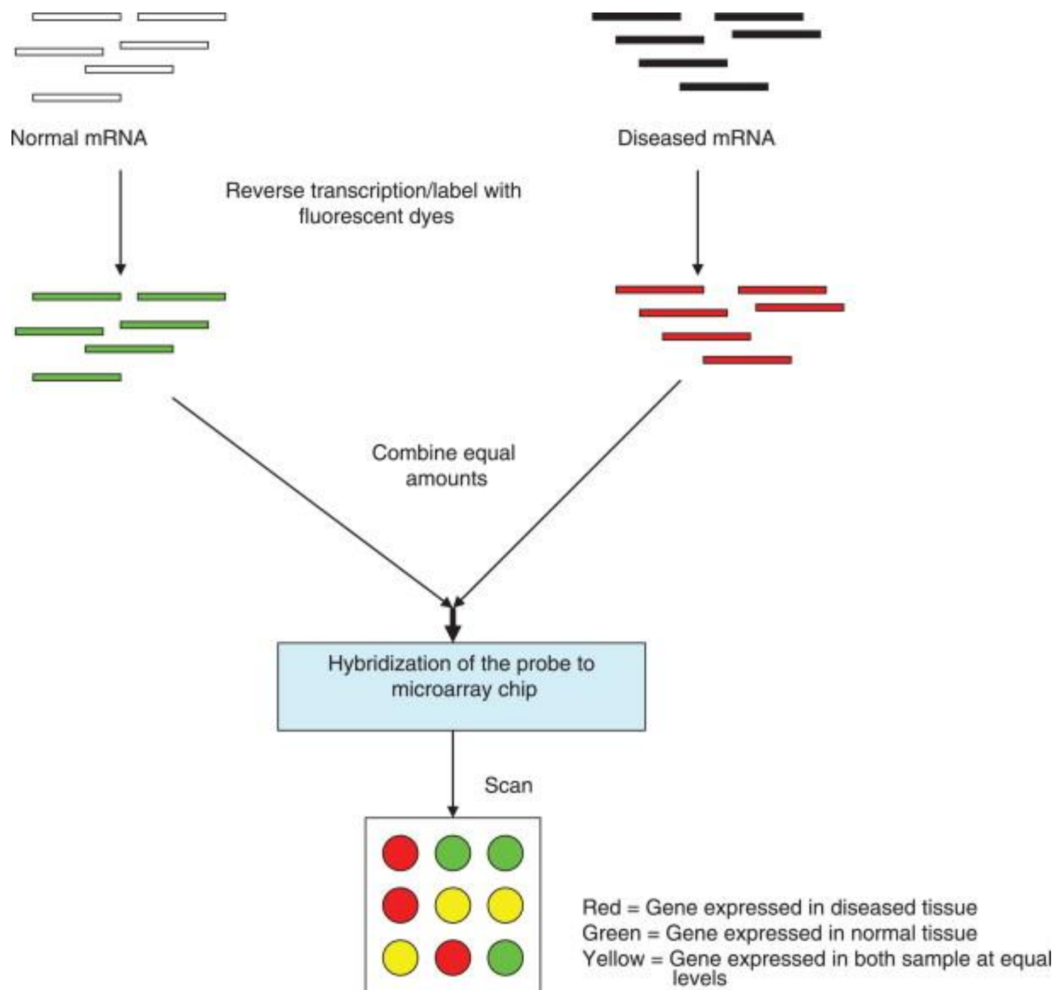
Thus, it is necessary to preserve the RNA direction information. Recently, several RNA-seq sample preparation methods that could retain direction information during library preparation have been reported in the literature, such as QuantSeq and SLAMseq. QuantSeq is a 3' mRNA-Seq library preparation protocol (Gupta *et al.,* 2021).

Gene expression of similar cells can be heterogeneous, which is determined by different derived genomes, cell cycle and micro-environment. However, the heterogeneity is lost with the conventional RNA-seq method, as it obtains the commonality of a large amount of cellular gene.

Transcriptome analysis of individual cells is derived from single-cell cDNA amplification method reported by Brady *et al.,* (1990). Single-cell RNA-Seq technology provides information on the expression of nearly 10,000 genes in a single cell, providing a powerful tool for identifying transcriptome features of various cell types in biological tissues and for revealing the heterogeneity of gene expression between cells.(Gurjar *et al.,* 2019)

Serial analysis of gene expression (SAGE): SAGE is a sequence-based approach which was first introduced in 1995 by Velculescu and coworkers. It allows identification of a large number of transcripts present in tissues and the quantitative comparison of transcriptomes. (Singh *et al.,* 2020)

Fig.1



Massively Parallel Signature Sequencing (MPSS): MPSS is a recently developed high-throughput transcription proûling technology, has the ability to proûle almost every transcript in a sample without requiring prior Knowledge of the sequence of the transcribed genes. (He *et al.,* 2022) MPSS is one of the few technologies that produce data in a digital format.

Amplicon sequencing, Amplicons are DNA products of a polymerase chain reaction (PCR). The term amplicon is often used interchangeably with PCR product.

Next-Generation Sequencing (NGS), can be used to obtain the sequence of a PCR fragment that targets a specific genomic region. Transcriptome sequencing employs high-throughput sequencing technologies to get access to almost all transcripts of specific tissues or cells in a certain state by comprehensive and rapid cDNA sequencing, It has become basis as well as the starting point for the research of gene expression. (Nguyen *et al.,* 2018) The main objectives of transcriptome research include the classification of all transcriptional products, the determination of the transcription structure of genes and the changes in the expression levels.

Whole genome sequencing (WGS), also known as full genome sequencing, complete genome sequencing, or entire genome sequencing, is the process of determining the entirety, or nearly the entirety, of the DNA sequence of an organism's genome at a single time (Muhae-ud-Din *et al.,* 2020)

This entails sequencing all of an organism's chromosomal DNA as well as DNA contained in the mitochondria and, for plants, in the chloroplast.

Transcriptome sequencing allows genome-wide analysis of large, complex plant genomes and the potential to identify biologically significant SNPs. Molecular genetic markers have found wide application in plants. Molecular markers are used increasingly in plant breeding to guide selection and plant identification and play an important role in determining plant identity and or purity in forensic and industrial situations. (Hittalmani *et al.,* 2017) Protection of intellectual property rights such as plant variety rights is strongly supported by the use of molecular markers. Sequencing genomic DNA enriched for genes by hybridization with probes for all or some of the known genes simplifies sequencing and analysis of differences in gene sequences between large numbers of genotypes and genes especially when working with complex genomes. Further, they are used for study of fitness and adaptation in various crops in different agroclimatic conditions (Gupta *et al.,* 2021). Association (GWAS and candidate gene based) mapping for important traits in cereal crops have led to QTL analysis and identification of useful genes.

References

Anders S, Reyes A, Huber W., (2012) Detecting differential usage of exons from RNA-seq data. Genome Research, 22: 2008–2017. https://doi.org/10.1101/gr.133744.111

Arvind, K., Rajesh, M. K., Josephraj kumar, A., Grace, T., (2020). Dataset of de novo assembly and functional annotation of the transcriptome of certain developmental stages of coconut rhinoceros beetle, *Oryctes rhinoceros* L. Data in Brief 28, 105036. https://doi.org/10.1016/j.dib.2019.105036

Brady G, Barbara M, Iscove NN. 1990. Representative in vitro cDNA amplification from individual hemopoietic cells and colonies. Methods Mol. Cell Biol. 2:17–25

Baird N A, Etter P D, Atwood T S, *et al.,* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE, 3: e3376. https://doi.org/10.1371/journal.pone.0003376

Bao S, Jiang R, Kwan W *et al.,* (2011) Evaluation of next-generation sequencing software in mapping and assembly. Journal of Human Genetics, 56: 406–414. https://doi.org/10.1038/jhg.2011.43

Chittem, K.; Yajima, W. R.; Goswami, R. S.; del Río Mendoza, L. E. (2020) Transcriptome analysis of the plant pathogen *Sclerotinia sclerotiorum* interaction with resistant and susceptible canola (*Brassica napus*) lines. PLoS ONE, 15: e0229844. https://doi.org/10.1371/journal.pone.0229844

Dobon, A.; Bunting, D. C. E.; Cabrera-Quio, L. E.; Uauy, C.; Saunders, D. G. O (2016). The host-pathogen interaction between wheat and yellow rust induces temporally coordinated waves of gene expression. BMC Genom., 17: 380. https://doi.org/10.1186/s12864-016-2684-4

Ekblom R, Galindo J., (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity, 107:1 – 15. https://doi.org/10.1038/hdy.2010.152

Ellegren H, Smeds L, Burri R *et al.,*. (2012) The genomic landscape of spe- cies divergence in *Ficedula flycatchers*. Nature, 491: 756–760. https://doi.org/10.1038/nature11584

Gupta, P. K. (2021). GWAS for genetics of complex quantitative traits: Genome to pangenome and SNPs to SVs and k-mers. BioEssays, 43: pp 1-17, e2100109. https://doi.org/10.1002/bies.202100109

Gurjar, M. S.; Aggarwal, R.; Jogawat, A.; Kulshreshtha, D.; Sharma, S.; Solanke, A. U.; Dubey, H.; Jain, R. K. (2019) Genome sequencing and secretome analysis of *T. indica* inciting Karnal bunt of wheat provides pathogenesis-related genes. 3 Biotech, 9: 219. https://doi.org/10.1007/s13205-019-1743-3

He, T. Ren, Z. Muhae-Ud-Din, G. Guo, Q. Liu, T. Chen, W. Gao, L. (2022) Transcriptomics Analysis of Wheat Tassel Response to *Tilletia laevis* Kühn, Which Causes Common Bunt of Wheat. Front. Plant Sci., 13: 823907. https://doi.org/10.3389/fpls.2022.823907

Hittalmani S, Mahesh H B, Shirke M D, Biradar H, Uday G, Aruna Y R, and Mohanrao A., (2017). Genome and tran- scriptome sequence of finger millet (*Eleusine coracana* (L.) Gaertn.) provides insights into drought tolerance and nu- traceutical properties. BMC Genom 18: 465. https://doi.org/10.1186/s12864-017-3850-z

Jain M., (2012) Next-generation sequencing technologies for gene expression profiling in plants. Briefings in Functional Genomics, 11: 63–70. https://doi.org/10.1093/bfgp/elr038

Martin J A, Wang Z., (2011) Next-generation transcriptome assembly. Nature Reviews Genetics, 12, 671–682. https://doi.org/10.1038/nrg3068

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods. 2008;5:621–628.

Muhae-Ud-Din, G. Chen, D. Liu, T. Chen, W. Gao, L. (2020) Characterization of the wheat cultivars against *Tilletia controversa* Kühn, causal agent of wheat dwarf bunt. Sci. Rep., 10: 9029. https://doi.org/10.1038/s41598-020-65748-w

Nguyen, T. T., Suryamohan, K., Kuriakose, B., Janakiraman, V., Reichelt, M., Chaudhuri, S., Guillory, J., Divakaran, N., Rabins, P. E., Goel, R., Deka, B., Sarkar, S., Ekka, P., Tsai, Y.-C., Vargas, D., Santhosh, S., Mohan, S., Chin, C.-S., Korlach, J., Thomas, G., Babu, A., Seshagiri, S., 2018. Comprehensive analysis of single molecule sequencing derived complete genome and whole transcriptome of *Hyposidra talaca* nuclear polyhedrosis virus. Sci Rep 8: 1–11. https://doi.org/10.1038/s41598-018-27084-y

Puvvala, S. S., Muddanuru, T., Thangella, P. A. V., Kumar, O. A., Chakravartty, N., Vettath, V. K., Katta, A. V. S. K. M., Lekkala, S. P., Kuriakose, B., Gupta, S., Mulpuri, S., Lachagari, V. B. R., (2019). Deciphering the transcriptomic insight during organogenesis in Castor (*Ricinus communis* L.), Jatropha (*Jatropha curcas* L.) and Sunflower (*Helianthus annuus* L.). 3 Biotech 9: 434. https://dx.doi.org/10.1007/s13205-019-1960-9

Ren, Z. Liu, J. Din, G. M. U. Zhang, H. Du, Z. Chen, W. Liu, T. Zhang, J. Zhao, S. Gao, L. (2020) Transcriptome analysis of wheat spikes in response to *Tilletia controversa* Kühn which cause wheat dwarf bunt. Sci. Rep., 10: 21567. https://doi.org/10.1038/s41598-020-78628-0

Singh, J. Aggarwal, R. Gurjar, M. S. Sharma, S. Jain, S. Saharan, M. S. (2020) Identification and expression analysis of pathogenicity related genes in *Tilletia indica* inciting Karnal bunt of wheat. Australas. Plant Pathol., 49: 393–402. https://doi.org/10.1007/s13313-020-00711-x

Thakur, V., Bains, S., Pathania, S., Sharma, S., Kaur, R., Singh, K., (2020). Comparative transcriptomics reveals candidate transcription factors involved in costunolide biosynthesis in medicinal plant-Saussurealappa. International Journal of Biological Macromolecules 150: 52–67. https://doi.org/10.1016/j.ijbiomac.2020.01.312

How to cite this article:

Akhil Gupta and Singh, S. K. 2023. Trancsriptome Sequencing: Emerging Tool to Reveal Genomic Data and Gene Function. *Int.J.Curr.Microbiol.App.Sci.* 12(03): 171-176.
doi: https://doi.org/10.20546/ijcmas.2023.1203.021