

# 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

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Introductory Chapter: Transcriptome Analysis

*Miroslav Blumenberg*

The central dogma of molecular biology describes the flow of genetic information from genes to functions of the cells and organisms. This comprises a two-step process: first, DNA, the permanent, heritable, genetic information repository, is transcribed by the RNA polymerase enzymes into RNA, a short-lasting information carrier; second, a subset of RNA, the messenger RNAs, mRNAs, are translated into protein. The **transcriptome**, then, is the complete set of all RNA molecules in a cell, a population of cells or in an organism.

Importantly, not all RNAs are translated into proteins, some serve a structural function, for example, rRNAs in the assembly of ribosomes, others are transporters, e.g., tRNAs, yet others serve regulatory functions, for example, the siRNAs, short interfering RNA, or lncRNAs, long non-coding RNAs; these are not translated into proteins [1]. However, these non-coding RNAs can and often do play roles in human diseases such as cancer, cardiovascular, and neurological disorders. While transcriptomics is most commonly applied to the mRNAs, the coding transcripts, transcriptomics also provides important data regarding content of the cell noncoding RNAs, including rRNA, tRNA, lncRNA, siRNA, and others. Specific approaches address the analysis of splice variant of the same gene in different tissues.

## 1. Transcriptome analysis

Transcriptome Analysis is the study of the transcriptome, of the complete set of RNA transcripts that are produced by the genome, under specific circumstances or in a specific cell, using high-throughput methods. Transcription profiling, which follows changes in behavior of a cell *in toto*, not of a single gene or just a few genes, is used throughout diverse areas of biomedical research, including disease diagnosis, biomarker discovery, risk assessment of new drugs or environmental chemicals etc. Transcription profiling can be applied to loss- and gain-of-function mutants to identify the changes associated with the mutant phenotype. The transcriptomic techniques have been particularly useful in identifying the functions of genes. Transcriptomics also allows identification of pathways that respond to or ameliorate environmental stresses. RNA-Seq can also identify disease-associated gene fusions, single nucleotide polymorphisms and even allele-specific expression.

## 2. Uses of transcriptome analysis

Transcriptome Analysis is most commonly used to compare specific pairs of samples. The differences may be due to different external environmental conditions, e.g., hormonal effects or toxins. More commonly, healthy and disease states

are compared. For example, in cancer, transcriptomics analyses address classification, the mechanisms of pathogenesis and even outcome prediction. Transcriptome studies can classify cancer beyond anatomical location and histopathology. Outcome predictions can establish gene-based benchmarks to predict tumor prognosis and therapy response. These approaches are already in use for personalized medicine, individualized cancer patient therapies.

Organisms and tissues at various stages of development can be molecularly characterized. The transcriptomes of stem cells help to understand the processes of cellular differentiation or embryonic development. Because of its very broad approach transcriptome analysis is a great source for identifying targets for treatment.

## 2.1 Methodologies

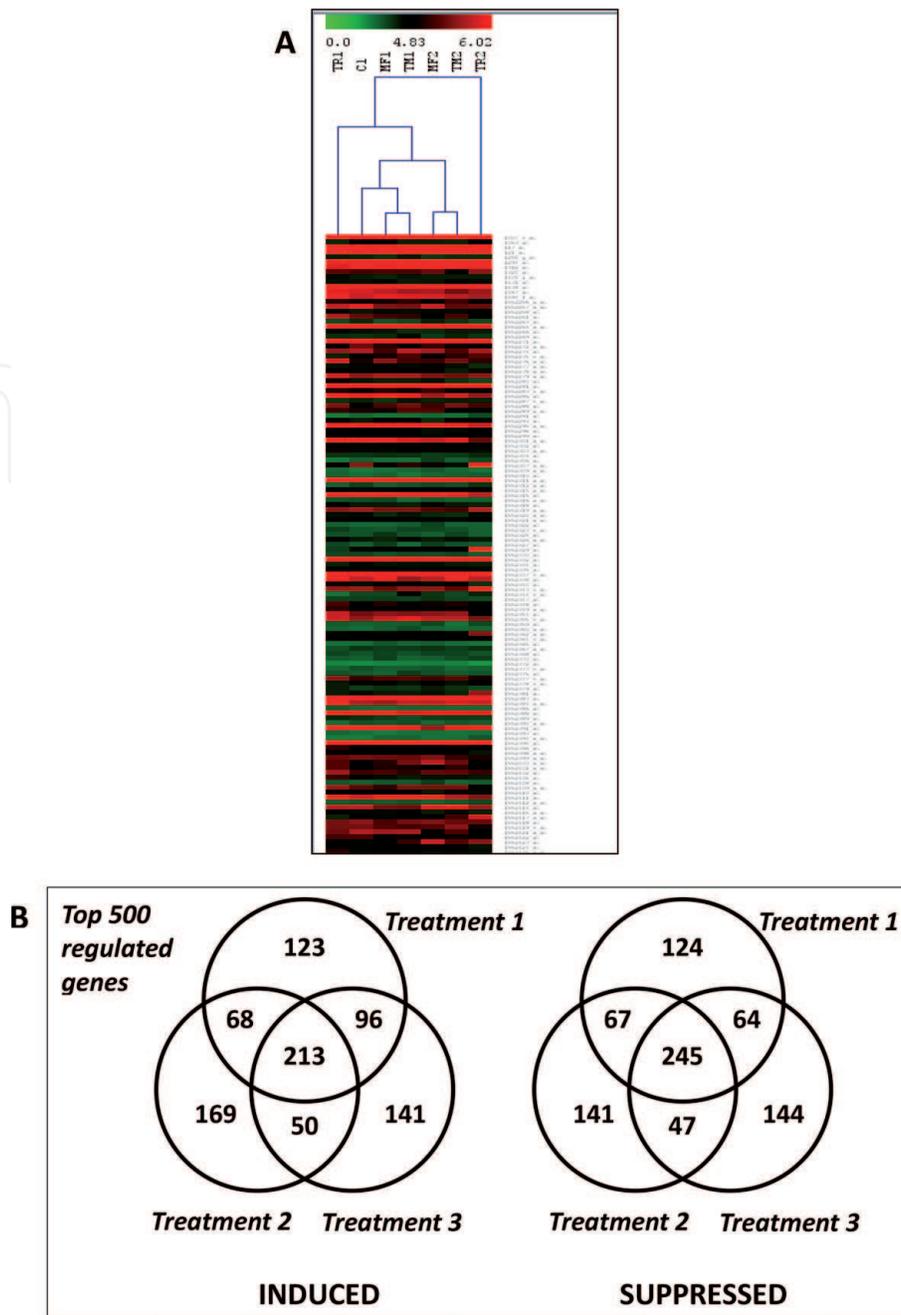
The early approach to study whole transcriptomes used microarrays, a set of defined sequences arranged on a solid substrate [2]. Microarrays almost exclusively represented mRNAs, that is, genes that are translated into proteins.

Nowadays the microarray approach is supplanted by high-throughput RNA sequencing, RNA-Seq, which detects all transcripts in a sample, including the regulatory siRNA and lncRNA transcripts [3]. In this methodology, the bulk RNA is extracted from the sample and copied into stable double-stranded copy DNA, ds-cDNA, which is then sequenced using various sequencing methods [4]. The sequences obtained are aligned to reference genome sequences, available in data banks, to identify which genes are transcribed. Quantitatively, the results provide the expression levels for the transcribed genes. Compared to microarrays, RNA-Seq can measure both the low-abundance and high-abundance RNAs over a five orders of magnitude range and, importantly, RNA-Seq requires much less starting material (nanograms vs. micrograms and even as little as 50 pg) [5]. This made possible analyses of transcriptomes in a single cell, a great advance over bulk tissue RNA analyses [6]. RNA-seq can be used to identify alternative splicing, novel transcripts, and fusion genes (**Table 1**).

In principle, the assembly of RNA-Seq reads is not dependent on reference genomes and can be used for gene expression studies of poorly characterized species with limited genomic resources. It can also be used to identify novel protein coding regions in sequenced genomes. RNA-seq can be performed using many next-generation sequencing platforms, however, each platform has its own requirements of sample preparation and the instrument design.

Method	Read length	Accuracy	Reads per run	Time per run	Cost per 1 million bp	Advantages	Disadvantages
Single-molecule real-time sequencing	>100,000 bases	87%	500,000 per Sequel SMRT cell	30 minutes to 20 hours	\$0.05-\$0.08	Fast.	Expensive.
Pacific Biosciences			10 to 20 gigabases				
Ion semiconductor	Up to 600 bp	99.60%	up to 80 million	2 hours	1	Less expensive equipment. Fast.	Homopolymer errors.
Ion Torrent sequencing							
Pyrosequencing	Up to 700 bp	99.90%	1 million	24 hours	\$10	Long read size. Fast.	Runs are expensive.
454 Life Sciences							Homopolymer errors.
Sequencing by synthesis	50-600 bp	99.90%	1 Million to 2.5 billion	1 to 11 days	\$0.05 to \$0.15	High sequence yield.	Expensive equipment.
Illumina							Requires high DNA concentrations.
Combinatorial probe anchor synthesis	35-300 bp	99.90%	50 to 1300M per flow cell	1 to 9 days	\$0.035- \$0.12		
cPAS- BGI/MGI							
Sequencing by ligation	50+50 bp	99.90%	1.2 to 1.4 billion	1 to 2 weeks	\$0.13	Low cost per base.	Slower.
SOLID sequencing							Issues with palindromic sequences.
Nanopore Sequencing	Up to 500 kb	92-97%	Up to 500 kb	1 min to 48 hrs	Depends	Portable.	Lower throughput. Lower accuracy.
Chain termination Sanger sequencing	400 to 900 bp	99.90%	N/A	20 minutes to 3 hours	\$2,400	Useful for many applications.	Time consuming lab bench steps

**Table 1.**  
*Comparison of RNA-seq methodologies.*



**Figure 1.** Graphic representations of transcriptome analysis data. (A) Heat map with clustering tree. (B) Venn diagrams of regulated genes.

## 2.2 Data analysis, repositories and presentation

Improved sequencing technologies necessitated improved data analysis methods to deal with the increased volume of data produced by each transcriptome experiment. Importantly, the results are deposited into transcriptome databases, essential tools for transcriptome analysis. For example Gene Expression Omnibus, [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov), contains millions of transcription profiling experiments. Such data have potential applications beyond the original aims of an experiment. Typical outputs include quantitative tables of the transcript levels. This requires specific analysis algorithms, often specific to the methodology used. There are software packages to bridge data from disparate methodologies, to identify groups of similar expressed genes, or differentially expressed functionally significant regulatory or metabolic pathways.

The results of transcriptomic analyses are graphically often presented as heat maps, a system of color-coding that represents different levels of expression of given genes in different samples (**Figure 1A**). Such presentations also frequently display a clustering of samples, this helps to identify samples with similar gene expression. Another common graphical presentation uses Venn diagrams, which count the transcripts which are equivalently regulated in multiple samples (**Figure 1B**).

Transcriptome analyses have become indispensable in basic research, translational, and clinical studies. In general, transcriptome analysis is a very powerful hypothesis-generating tool, more than a theory proving one.

### 3. Specific example: transcriptome analysis applied to human skin

Easily accessible, skin was among the first targets analyzed using ‘omics’ and dermatology embraced the approaches very early [7]. A classic example of coordinated transcriptional regulation was observed in cultured fibroblasts after serum stimulation [2]. Serum addition causes not only rapid recommencement of the cell cycle but, characteristically a wound-healing response, a physiological role of fibroblasts in wound healing [8]. Transcriptional responses of epidermal keratinocytes to UV light, hormones, vitamins, infections, inflammatory and immunomodulating cytokines, toxins and allergens have been characterized, as were the changes associated with epidermal differentiation [9, 10].

The expression signatures that define the various cell types in human skin, were used to define 20 specific gene signatures, including those for keratinocytes, melanocytes, endothelia, adipocytes, immune cells, hair follicles, sebaceous, sweat, and apocrine glands. This resource provided a resource named SkinSig, which was then used to analyze 18 skin conditions, providing in-context interpretation of, for example, influx in immune cells in inflammation or differentiation changes in disorders of cornification [11].

In the future we can anticipate a greatly expanded usage of transcriptome analysis. Translated to the bedside, it can provide better understanding and more specific diagnoses of diseases. This, of course, requires additional advances in the technology, both in the lab-bench components reducing the costs and guaranteeing reproducibility and accuracy, as well as in the computer-based components, algorithms that enable physicians to establish diagnosis quickly and reliably. In a generation, this approach will become routine.

#### Author details

Miroslav Blumenberg  
NYU School of Medicine, USA

\*Address all correspondence to: [miroslav.blumenberg@nyulangone.org](mailto:miroslav.blumenberg@nyulangone.org)

#### IntechOpen

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] Botchkareva NV. The molecular revolution in cutaneous biology: Noncoding RNAs: New molecular players in dermatology and cutaneous biology. *The Journal of Investigative Dermatology*. 2017;**137**(5):e105-e111. DOI: 10.1016/j.jid.2017.02.001. PMID: 28411840
- [2] Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JC, et al. The transcriptional program in the response of human fibroblasts to serum. *Science*. 1999;**283**(5398):83-87. PMID: 9872747
- [3] Bayega A, Fahiminiya S, Oikonomopoulos S, Ragoussis J. Current and future methods for mRNA analysis: A drive toward single molecule sequencing. *Methods in Molecular Biology*. 2018;**1783**:209-241. DOI: 10.1007/978-1-4939-7834-2\_11. PMID: 29767365
- [4] Zhang H, He L, Cai L. Transcriptome sequencing: RNA-Seq. *Methods in Molecular Biology*. 2018;**1754**:15-27. DOI: 10.1007/978-1-4939-7717-8\_2. PMID: 29536435
- [5] Shanker S, Paulson A, Edenberg HJ, Peak A, Perera A, Alekseyev YO, et al. Evaluation of commercially available RNA amplification kits for RNA sequencing using very low input amounts of total RNA. *Journal of Biomolecular Techniques*. 2015 April;**26**(1):4-18. DOI: 10.7171/jbt.15-2601-001. PMC 4310221. PMID 25649271
- [6] Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. *Nature Reviews Genetics*. 2015;**16**(3):133-145. DOI: 10.1038/nrg3833. PMID: 25628217
- [7] Mimoso C, Lee DD, Zavadil J, Tomic-Canic M, Blumenberg M. Analysis and meta-analysis of transcriptional profiling in human epidermis. *Methods in Molecular Biology*. 2014;**1195**:61-97. DOI: 10.1007/7651\_2013\_60
- [8] Eming SA, Martin P, Tomic-Canic M. Wound repair and regeneration: Mechanisms, signaling, and translation. *Science Translational Medicine*. 2014;**6**(265):265sr6. DOI: 10.1126/scitranslmed.3009337. PMID: 25473038
- [9] Blumenberg M. Skinomics: Past, present and future for diagnostic microarray studies in dermatology. *Expert Review of Molecular Diagnostics*. 2013;**13**(8):885-894. DOI: 10.1586/14737159.2013.846827. PMID: 24151852
- [10] Santoro S, Lopez ID, Lombardi R, Zauli A, Osiceanu AM, Sorosina M, et al. Laser capture microdissection for transcriptomic profiles in human skin biopsies. *BMC Molecular Biology*. 2018;**19**(1):7. DOI: 10.1186/s12867-018-0108-5. PMID: 29921228 PMCID: PMC6009967
- [11] Shih BB, Nirmal AJ, Headon DJ, Akbar AN, Mabbott NA, Freeman TC. Derivation of marker gene signatures from human skin and their use in the interpretation of the transcriptional changes associated with dermatological disorders. *The Journal of Pathology*. 2017;**241**(5):600-613. DOI: 10.1002/path.4864. Epub 2017 Feb 24. PMID: 28008606 PMCID: PMC5363360