Transcriptomics: Better Resolutions

Simone Gupta*

McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, USA

In the post Human Genome Project era there are efforts to understand translation of the genomic sequence into the transcriptome. The human transcriptome is represented by >100,000 distinct transcripts presently described for ~20,000 protein-coding genes. Additionally, mRNA isoforms are produced by alternative processing of primary RNA transcripts. This alternative splicing affects >90% of the human genes and has been suggested to the primary driver of phenotypic complexity [1]. Despite this diversity in the coding sequence, the non-protein-coding molecules contribute to >95% of the transcriptome [2].

Numerous technologies and experimental platforms have facilitated the investigation of the complexity in the transcriptome that greater than that of the simple genome sequence.

The recent RNA sequencing or RNA-Seq, involves high-throughput sequencing of short cDNA fragments obtained from the pool of RNA (total or fractionated, such as poly(A)+ or ribosomal RNA depleted) to provide single-base resolution to the transcriptome. The traditional expression analysis primarily addressed the identification of differentially expressed transcripts with respect to measured variables of interest, such as differing environments, treatments, phenotypes, or clinical outcomes. The advent of RNA-Seq has provided a broad spectrum of applications and enabled researchers to address a wider range of biological problems. This technology enables cataloguing all species of transcript, including coding and non-coding mRNAs; to determine the transcriptional structure of genes, splicing patterns and other post-transcriptional modifications.

Despite, the breadth of possibilities RNA-Seq measurements and analysis of expression remains a field of active research. The major concerns and scrutiny is attributed to the numerous technical and analytical limitations. Early concerns regarding library preparation, sequencing error, read mapping, and gene expression quantification have been resolved by a number of studies; however, there is no standardized approach for quality control and data adjustment of RNA-Seq data after the generation of gene expression estimates. As a caution, without an appropriate approach to data analysis, reproducibility of these studies remains limited [3]. There are numerous studies that are providing frameworks and strategies to assess possible sample contamination and assess the biologic validity of each data analysis step to ultimately enable confident downstream analyses [4].

An important consideration in gene expression is still the biological source for RNA profiling. To elaborate for disease relevant questions there is a clear and compelling need to conduct gene expression studies in tissues that are specifically relevant to the disease of interest as opposed to cell lines. It is reassuring that studies have reported that underlying mechanism in some human diseases are accompanied by changes of cell populations in corresponding tissues [10]. There are computational methods of analyzing gene expression in samples of varying composition that can improve analyses of quantitative molecular data in many biological contexts [6,11]. There is also the development of sequencing-based technologies that are increasingly being targeted to individual cells, which will allow many new and longstanding questions to be addressed.

The maturation of single-cell transcriptomics should provide in-depth knowledge of the precise transcript map and the regulatory landscape in individual single cells at different levels of resolution (single cell, cell-type or tissue). This level of resolution should be beneficial towards insightful biomarker discovery and disease diagnostics.

References

*Corresponding author: Simone Gupta, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, USA, Tel: 410.949.0806; E-mail: simonegupta@gmail.com

Received March 12, 2014; Accepted April 04, 2014; Published April 12, 2014


Copyright: © 2014 Gupta S. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.