Gene Technology
ISSN: 2329-6682

Gupta, Gene Technology 2014, 4:1
DOI: 10.4172/2329-6682.1000111

Short Commentary Open Access

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 robust gene expression can be obtained using RNA from autopsy-derived tissue 24 hours after autolysis [5]. However, examination of a tissue which is a heterogeneous mix of several distinct cell populations makes it difficult to distinguish whether gene expression variability reflects shifts in cell proportions or variable cell-type specific expression [6].

In addition to cell-type variability, gene expression data is also confounded by various known and unknown sources of variation such as batch effects, environmental influences and sample history. These unknown confounders that plague comparative expression analysis are not easily attributable to any recorded measurement. These unknown covariates can be approximated through various data decomposition methods, like Principal Components Analysis (PCA), Surrogate Variable Analysis (SVA) [7], Independent Surrogate Variable Analysis (ISVA) [8] and Probabilistic Estimation of Expression Residuals (PEER) [9]. This method can be used to correct for biases and provide accurate estimates of global comparison of gene analysis and for detecting genetic associations in expression data (eQTL) [4].

Although, there are suggestions to correct the variability of expression that may be caused by difference in cell-type proportions. 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

1. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, et al. (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456: 470-476.
2. Frith MC, Pheasant M, Mattick JS (2005) The amazing complexity of the human transcriptome. Eur J Hum Genet 13: 894-897.
3. Nekrutenko A, Taylor J (2012) Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. Nat Rev Genet 13: 667-672.
4. Ellis SE, Gupta S, Ashar FN, Bader JS, West AB, et al. (2013) RNA-Seq optimization with eQTL gold standards. BMC Genomics 14: 892.
5. Gupta S, Halushka MK, Hilton GM, Arking DE (2012) Postmortem cardiac tissue maintains gene expression profile even after late harvesting. BMC Genomics 13: 26.
6. Kuhn A, Thu D, Waldvogel HJ, Faull RL, Luthi-Carter R (2011) Population-specific expression analysis (PSEA) reveals molecular changes in diseased brain. Nat Methods 8: 945-947.
7. Leek JT, Storey JD (2007) Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet 3: 1724-1735.
8. Teschendorff AE, Zhuang J, Widschwendter M (2011) Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. Bioinformatics 27: 1496-1505.
9. Stegle O, Parts L, Piipari M, Winn J, Durbin R (2012) Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat Protoc 7: 500-507.
10. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, et al. (2011) Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature 474: 380-384.
11. Gong T, Szustakowski JD (2013) DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. Bioinformatics 29: 1083-1085.

***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

Citation: Gupta S (2014) Transcriptomics: Better Resolutions. Gene Technology 4: 111. doi: 10.4172/2329-6682.1000111