Opinion Article | Open Access

Application of RNA technology in gene expression

M Anna[*]

Department of Biological Engineering, Massachusetts Institute of Technology, Massachusetts, USA

[*]Corresponding author. E-mail: mollyan@gmail.com

DECSRIPTION

DNA encodes all of the information required to determine all of the properties and functions of each individual cell in every living organism. Cells can dynamically access and translate specific instructions from this blueprint *via* 'gene expression,' or by selectively turning on and off a specific set of genes. The information encoded in the chosen genes is transcribed into RNA molecules, which can then be translated into proteins or used directly to control gene expression. Thus, the set of RNAs transcribed under specific conditions and at specific times reflects a cell's current state and can reveal pathological diseases underlying mechanisms. More intriguingly, the study of differential gene expression allows for the comparison of gene expression profiles from different tissues and conditions in order to identify genes that play a significant role in phenotype determination.

For example, comparing healthy versus diseased tissues can reveal new information about the genetic factors involved in pathology. For the study of gene expression, Next Generation Sequencing (NGS)-based RNA profiling methodology is replacing microarrays. The RNA-sequencing framework allows for the investigation of all RNAs present in a sample at high resolution, characterizing their sequencing and quantifying their abundances at the same time. Millions of short strings, known as 'reads,' are sequenced from random positions of the input RNAs in practice. These reads can then be computationally mapped onto a reference genome to reveal a 'transcriptional map,' with the number of reads aligned to each gene indicating the level of expression.

The powerful features of RNA-sequence, such as high resolution and broad dynamic range, have accelerated transcriptomics research to unprecedented levels, yielding an impressive amount of data globally.

Several computational tools have been developed and updated at a rapid pace to support this exponential growth and to deal with the various stages of data analysis. Nonetheless, the above analysis scheme, which appears simple at first glance, is far more complex and involves several processing steps. Because it allows for the investigation and comparison of gene expression levels at unprecedented resolution, RNA-sequence has quickly become the method of choice for the study of differential gene expression.

However, converting massive and complex RNA-sequence data sets into biologically meaningful results is not easy. The interpretation of RNA-sequence data necessitates the development of a computational pipeline that includes several steps such as read mapping, count computation, normalization, and differential gene expression testing. RNA-sequence is a novel NGS-based methodology for investigating differential gene expression at high resolution. Data interpretation, on the other hand, is not simple and necessitates several analysis steps, including read mapping, counts computation, counts normalization, and DE testing.

Depending on the algorithm and heuristics used, read mapping tools provide varying solutions. Even in the presence of high-similarity sequence, special care must be taken when handling reads mapping on multiple genomic locations to estimate correct gene expression levels. Gene expression levels in RNA-sequence studies are measured by counts or the number of reads mapped to each gene. Counts are frequently affected by gene-and sample-specific covariates such as gene length and library size. Before comparing samples to detect differentially expressed genes, differences in library size between samples must be corrected.