OPINION

Investigating the shallow end and calculating the information content of studies on transcriptomic

Nikolay Gogol

Gogol N. Investigating the shallow end and calculating the information content of studies on transcriptomic J. Environ. Microbiol.2022;4(3):1-2.

ABSTRACT

An important tool for studying organismal biology is transcriptomics. With the introduction of new parallel sequencing technology, a new field of transcriptomics has emerged, allowing for the identification and quantification of each and every transcript present in a sample through ever-deeper sequencing. For all transcriptomics experiments, this might not be the optimal way to use parallel sequencing technology. I investigated shallow RNAseq's capacity to capture the majority of this information and used the Shannon entropy technique to estimate the amount of information present in a transcriptomics experiment. This investigation demonstrated that a subset of the most abundant 5,000 transcripts or less within any given sample can effectively capture nearly all of the network or genomic information provided in a variety of transcriptomics experiments. Thus, it seems that using parallel sequencing technology, large-scale factorial analysis with a high level of replication should be doable and economical.

Key Words: *Transcriptomics; Shannon entropy; Polymorphisms; Genotypes; Next-Generation*

INTRODUCTION

Transcriptomics, the process of directly measuring thousands of transcripts at once, has significantly contributed to the progress of biological research as well as the emergence of new fields like genomics and systems biology. Transcriptomics is now used in almost all areas of biological research, including genetics, biochemistry, ecology, and evolution. This has improved our understanding of how regulatory and evolutionary pressures shape an organism's transcriptome and, more importantly, made it possible to determine the purpose of countless additional genes. Because RNAseq is now feasible for a wider range of organisms thanks to recent technological improvements, the transition from microarray-based to RNAseq-based transcriptomics has happened quickly.

Although a single duplicate of the full genetic matrix would require 25,000,000 genotypes, these experiments are restricted to the ability to robotically control the organism and quantify a single phenotype inside a 5000 gene matrix of paired epistatic pairings. Due to the factorial nature of the experiment, comprehensive transcriptomics on 25,000,000 lines is not thought to be technically or economically viable.

Utilizing crossings between natural genotypes to allow segregation to shuffle 100 to 1000 polymorphisms and then measuring the transcriptome in the offspring have been various approaches to the same goal of systems genetics. However, because they typically have fewer than 500 individuals for a population that may have at least 1000 distinct causal polymorphisms with the potential to influence the transcriptome, the population sizes merely scrape the surface of the possible allele combinations. In order to properly examine the factorial structure of the natural variation network in this example, a population of 1,000,000 people would need to sample the 1000x1000 matrix containing all conceivable pairwise combinations between the causative polymorphisms. The 500 initial participants in this scenario would only sample 0.05% of the potential.

Consequently, methods must be developed to enable genomics of considerably bigger genotype collections in order to completely comprehend how networks may differ in nature.

Another area where the demand for factorial experiments using transcriptomics is increasing is systems regulation. Research on the transcriptional circadian clock is one of the best examples of this, since it demonstrates how regulatory networks are essential to an organism's functionality by integrating various inputs to effectively govern the clock's output. Therefore, a multifactorial examination of the environment and how variation in all of the environmental cues interacts to shape the phenotype of the organism would be necessary for a complete transcriptome knowledge of the clock in an organism.

Editorial Office, Journal of Environmental Microbiology, UK

Correspondence: Nikolay Gogol, Journal of Environmental Microbiology, UK , Email: nikolaygogol@gmail.com

Received:- 01-June-2022, Manuscript No. PULJEM-23-6000; Editor assigned:- 03-June-2022, Pre-QC No. PULJEM-23-6000 (PQ); Reviewed:- 07-June-2022, QC No. PULJEM-23-6000 (Q); Revised:- 10-June-2022; Manuscript No. PULJEM-23-6000 (R); Published:- 15-June-2022, DOI: 10.37532/puljem.22.4 (3).1-2

The enormous volume of samples that must be examined for transcriptomics is the most challenging aspect of the aforementioned factorial experiments. This need for a high sample size requires the development of techniques and strategies for quickly and affordably carrying out these factorial analyses. The adoption of next-generation sequencing technologies, which have been demonstrated to be capable of high-throughput parallel DNA sequencing for quick, large-scale mapping studies, may be one solution. The application of next-generation sequencing for RNA seq, however, has mostly concentrated on the identification and measurement of additional transcripts, i.e., deep sequencing, to capture the expression.

Transcriptomes exhibit substantial co-expression, which is largely influenced by the structure of the underlying regulatory network, which presents a challenge for transcriptomic optimization. The goal of finding a particular subset of transcripts that measure key nodes in this network has frequently been motivated by the co-expression structure of the transcriptome, and theoretically, the entire state of the transcriptome could be described by tracking the expression of a small number of carefully chosen genes. Due to the fact that the essential nodes frequently alter depending on the biological issue, locating this set has proven to be challenging. An alternative would be to select a set of genes at random.

I hypothesised that it could be possible to use shallow RNAseq analysis for factorial transcriptomic research by determining where the information lies in microarray transcriptome studies, given the similarity in transcriptomics results between the platforms. This ought to aid in improving the method for factorial analysis in transcriptomics.

Utilizing the parallel sequencing capability of next-generation sequencing technologies to sequence transcriptomes at a shallow depth for the factorial trials is one possible solution to this problem. This might then be utilised to examine the transcriptome data as if it were physiological measures using a network design.