txGeneNetwork: transcript-level functional analysis using network biology

Lucio Rezende Queiroz, Gloria Regina Franco, Izabela Mamede Costa Andrade da Conceição

UFMG - Departamento de Bioquímica e Imunologia, UFMG, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

The differential expression analysis outcomes are highly variable due to the vast amount of differentially expressed genes from a single transcriptomics experiment. It is a challenging task to correlating these genes with a concise biological function or molecular pathway. Multiple gene set enrichment (GSE) methods were developed to address that problem. However, most GSE analyses result in hundreds of enriched gene sets, with considerable overlap of each gene set's genes. Thus, defining biologically relevant pathways that correlate with a given study design is a challenging task. Another common issue in most gene expression analyses is that gene-level only quantification is performed. A gene can produce multiple distinct coding and non-coding transcripts through alternative splicing, a process present in all multicellular eukaryotes. Gene-level summarized quantification sometimes leads to incorrect conclusions about a gene's contribution in a biological process, ignoring individual isoform contribution to the analyzed condition. Therefore, integrating transcript-level expression could provide invaluable insights into the significant contribution of a given gene. Current methods and workflows based on biological networks used to visualize gene expression data and integrate with GSE do not explore the relationships between genes and their isoforms, especially non-coding transcripts. An additional consideration is that most network biology workflows consist of several steps in different tools, accessed in variable interfaces, that can be hard to reproduce. We present a comprehensive, freely available, and open-source R/Bioconductor workflow to access and interpret transcripts' functional relevance using network biology. We correlate the differentially expressed transcripts and genes from transcript-level expression data, also performing GSE analysis based on transcripts effect size. Different methods are offered for functional enrichment analysis. We then use a graph-based network approach to interconnect the relations between pathways, related genes, and the transcript isoforms which arise from those genes. We then integrate isoform subtypes, the direction of their expression, properties, and topology of the network with other network science metrics to rank biological processes and transcript relevance. The final network forms an intuitive and functional visualization for GSE analysis. Its topology could bring insights into transcript function, isoform switch, and integration between gene sets, helping unravel transcriptomic signatures for coding and non-coding transcripts.

Funding:

Link to Video: