Gene co-expression network

Gene co-expression networks (GCNs) are powerful tools used in genomics and bioinformatics to analyze the relationships between genes based on their expression patterns across different conditions, tissues, or developmental stages. Here's an overview of key concepts related to gene co-expression networks:

1. Concept of Co-Expression Co-Expression: Refers to the phenomenon where genes show similar expression patterns across multiple samples. This can suggest that the genes may be involved in the same biological processes or pathways. 2. Construction of Gene Co-Expression Networks Data Acquisition: Gene expression data is usually obtained from high-throughput technologies like RNA-Seq or microarrays. Correlation Analysis: Pairwise correlation (e.g., Pearson or Spearman correlation) is calculated between gene expression profiles. Genes that exhibit high correlation coefficients are considered co-expressed. Network Construction: A network is constructed where nodes represent genes and edges (connections) represent significant correlations between them. Various thresholds can be set to determine which correlations are included in the network. 3. Types of Networks Undirected Networks: Edges do not have a direction and represent a symmetric relationship between genes. Directed Networks: Edges indicate a directional influence or regulatory relationship, often based on additional data (like transcription factor-target interactions). 4. Analysis of Gene Co-Expression Networks Modules and Clusters: GCNs can be analyzed to identify modules (clusters of highly interconnected genes) that may represent functional gene groups. Centrality Measures: Metrics such as degree, betweenness, and closeness centrality can help identify key regulatory genes within the network. Pathway Enrichment Analysis: Modules can be subjected to enrichment analysis to determine if they are significantly associated with specific biological pathways. 5. Applications of Gene Co-Expression Networks Gene Function Prediction: Co-expressed genes may share similar functions, so GCNs can help infer the roles of uncharacterized genes. Disease Studies: GCNs can reveal gene interactions relevant to diseases, helping identify potential biomarkers or therapeutic targets. Comparative Genomics: By comparing GCNs across different species, researchers can identify conserved pathways and evolutionary relationships. 6. Tools and Software Several software tools and platforms are available for constructing and analyzing gene co-expression networks, including:

WGCNA (Weighted Gene Co-expression Network Analysis): A widely used R package for constructing and analyzing weighted networks. Cytoscape: A powerful open-source software platform for visualizing complex networks and integrating them with any type of attribute data. Gephi: A tool for network visualization and exploration. 7. Challenges and Limitations Noise in Data: High variability and noise in gene expression data can lead to false positives in co-expression relationships. Dynamic Nature of Gene Expression: Co-expression can vary under different conditions, so context is crucial. Biological Interpretation: Interpreting the biological significance of co-expression relationships can be complex and requires additional validation.

A gene co-expression network (GCN) is an undirected graph, where each node corresponds to a gene, and a pair of nodes is connected with an edge if there is a significant co-expression relationship between them.

Having gene expression profiles of a number of genes for several samples or experimental conditions, a gene co-expression network can be constructed by looking for pairs of genes that show a similar expression pattern across samples, since the transcript levels of two co-expressed genes rise and fall together across samples. Gene co-expression networks are of biological interest since co-expressed genes are controlled by the same transcriptional regulatory program, functionally related, or members of the same pathway or protein complex

In a study, dataset GSE50161 was used to construct a co-expression network for weighted gene co-expression network analysis. Two modules (dubbed brown and turquoise) were found to have the strongest correlation with glioblastoma (Glioblastoma). Functional enrichment analysis indicated that the brown module was involved in the cell cycle, DNA replication, and pyrimidine metabolism. The turquoise module was primarily related to circadian rhythm entrainment, glutamatergic synapses, and axon guidance. Hub genes were screened by survival analysis using The Cancer Genome Atlas and Human Protein Atlas databases and further tested using the GSE4290 and Gene Expression Profiling Interactive Analysis databases. The eight hub genes (NUSAP1, SHCBP1, KNL1, SULT4A1, SLC12A5, NUF2, NAPB, and GARNL3) were verified at both the transcriptional and translational levels, and these gene expression levels were significant based on the World Health Organization classification system. These hub genes may be potential biomarkers and therapeutic targets for the accurate diagnosis and management of Glioblastoma [1].

[1]
Li C, Pu B, Gu L, Zhang M, Shen H, Yuan Y, Liao L. Identification of key modules and hub genes in glioblastoma multiforme based on co-expression network analysis. FEBS Open Bio. 2021 Jan 10. doi: 10.1002/2211-5463.13078. Epub ahead of print. PMID: 33423377.