System Approach to Analyze High-Throughput Data

Najafi M[1]*, Golpoor P[1] and Soltanmohammadi E[2]

[1]*Biochemistry Department, Iran University of Medical Sciences, Iran*
[2]*School of Medicine-International Branch, Iran University of Medical Sciences, Iran*

Systems biology is widely applied in translational biomedicine in order to understand and analyze the gene behavior in cellular models over the known biologic pathways. The field is mainly devoted to coordinate the simulation and biological efforts on challenging with bioinformatics and other related areas (OMICS). The main goals of systems biology lie on the clustering of gene profiles, targeting of ligands, integrating and developing of biologic networks by analyzing high-throughput data. Furthermore, it studies the global behavior of components rather than their individual characteristics [1].

Annotation

The experimental results obtained in each investigation could be enriched with other data. There are many databases that prepare these data on the basis of genome, transcription and proteome studies. The data can also be obtained from the bioinformatics databases that prepare the molecular and cellular characteristics of network components. Some of databases most used to annotate in systems biology are reported in Table 1. BioMart (biomart.org) and 12D (*ophid.utoronto.ca*) are other assembled databases that might be used in the enrichment of results. The efforts on the enrichment quality are progressively managed by plugins on networks.

Clustering

Several clustering models are presented and applied to construct the biologic networks. A network refers to a collection of nodes contacted to edges. Basically, the models are able to build the network blocks, known as hubs. Each hub presents the local contacts of several nodes on the network. Although the hubs obtain the smallest operation units and may indicate the network functional or structural parts, other inter-hub coordinators are the most important linkers to build the network global motifs and modules. The data networking may manually be constructed after annotating the data. However, network analysis depends on the data integration, topology properties and organization characteristics [2].

Method

Topological characteristics including network centrality and connectivity are resulted on the modeling techniques. Commonly, two computational methods, known as top-down and bottom-down

models, are most used in systems biology. The large-scale data obtained from high-throughput techniques including the protein and gene expression results are generally clustered with the top-down model. On the other hand, the bottom-down model is applied to relatively small data. It is obvious that the use of clustering approaches such as K-means and MCL depends on the primary data and the system understanding of study.

Enrichment

The networks can be improved by apps in the viewer tools (Cytoscape and String). The network components including nodes and edges may be enriched manually by bioinformatics databases (such as BioMart) or automatically by the compatible plugins. The network processing by annotating data can clear the understanding of disease mechanism, signaling integration and gene regulatory systems [3].

Figure 1: Gene co-expression network. The nodes and edges show the expression levels and scores predicted from enrichment of data (GEO; GDS3973) [4].

*Corresponding author: Mohammad Najafi, Biochemistry Department, Iran University of Medical Sciences, Iran, Tel: 86703107; E-mail: nbsmmsbn@iums.ac.ir

Category	Database	Web
Annotation Data	GO Ontology	geneontology.org
	PIR	pir.georgetown.edu
	KEGG	genome.jp/kegg
	Ensembl	ensembl.org
Gene Expression	GEO	ncbi.nlm.nih.gov/geo
	Reactome	reactome.org
Protein Expression	ProteomeXchange	proteomexchange.org
	PRIDE	ebi.ac.uk/pride
Regulatory Elements	TransFac	gene-rgulation.com/pub/databases.html
	Jaspar	jaspar.genereg.net
	Chea	amp.pharm.mssm.edu/lib/chea.jsp

Table 1: Some databases used for data annotation in systems biology.

Biochemical Networks

Based on the interest of biochemists and biologists, the systems biology improves understanding of cellular events by the system-wide analyses of proteome, genome and transcriptome. This approach opens wide areas to study the molecular interaction, gene regulatory, cell signaling, disease mechanism and response networks. Figure 1 show a protein-protein interaction (PPI) network obtained from gene co-expression data (GEO; GDS3973, DU145 cell line).

Perspectives

Although systems biology is improving our understanding of disease mechanisms, treatment responses and signaling integrations but its application depends on the advances of tools and bioinformatics. The main challenges can be considered on the scoring approaches of OMICS data, most described as interactome, and computational tools to annotate the network components. Furthermore, the improvement of clustering methods can effectively improve the global network construction [5].

References

1. Kirk PD, Babtie AC, Stumpf MP (2015) SYSTEMS BIOLOGY. Systems biology (un)certainties. Science 350: 386-388.

2. Schreiber F, Bader GD, Golebiewski M, Hucka M, Kormeier B, et al. (2015) Specifications of Standards in Systems and Synthetic Biology. J Integr Bioinform 12: 258.

3. Ayers D, Day PJ (2015) Systems Medicine: The Application of Systems Biology Approaches for Modern Medical Research and Drug Development. Mol Biol Int 2015: 698169.

4. Jangravi Z, Najafi M, Shabani M (2016) Investigation of histone lysine specific demethylase 5D (KDM5D) isoforms expression in prostate cancer cell lines: a system approach. IBJ (In press).

5. Hernández-Lemus E, Li W, Meyer P (2015) Advances in systems biology - New trends and perspectives. Comput Biol Chem pii: S1476-9271: 00084-00085.