

Bioinformatics Methods for Biochemical Pathways and System Biology Analysis

Mahin Ghorbani^{1*}, Hamed Karimi²

¹Department of Biotechnology, Fergusson College, F.C. Road, Pune, Maharashtra, India

²Department of Information Technology, Payam noor university of Farokh-shahr, Farokh-shahr, Chaharmahl va bakhtiari, Iran

ABSTRACT

The analysis of biochemical pathways and system biology has gained significance because of their role in understanding disease and drug discovery. Bioinformatics analysis of the pathways has emerged to understand the genotype-phenotype relationship on a large scale. The techniques integrate the molecular information from the databases with simulation of metabolic networks. These methods also help in representation of genes, proteins and metabolic pathways in combination with dynamic simulated environment. In this paper we reviewed some applicable bioinformatics tools for analytical study of three types of pathways such as metabolic, genetics and signalling pathways along with the information about their principle, work system and their direct access link to the databases and programs. This study helps scientists in fast, economic, high accuracy and large scale based outputs of pathways analysis of their appropriate research involving the biochemical pathways. Bioinformatics tools listed in this paper are named as RegulonDB, WIT, UM-BBD, EcoCyc, MetaCyc, Enzyme and metabolic pathways database, KEGG, KEGG BRITE, Gene network database, Gene path SPAD, STCDB, TRANSPATH Netpath, Signal link2.0.

Keywords: Bioinformatics Tools, Analytical Study, Biochemical Pathways, System Biology

I. INTRODUCTION

A pathway can be defined as a modular section of networking molecules to perform a cellular function. The pathways's products are different based on the need of the cell; they can be structural or functional responses. The pathways are essential for growth and survival of the organism. Inability to fulfil a cellular function due to failure in functioning of a pathway lead to emergence of the diseases[1]. So the pathways are important key targets for understanding the diseases and drug discovery [2]. A 2-D diagram illustrating interconnection of the protein and non-protein substances usually represent a pathway. As an example a simple pathway like $\bullet \rightarrow \rightarrow \bullet$ is a graph whose nodes (small circles) correspond to different biomolecules and modes (arrows) represent the nature of molecule – molecule interaction. A pathway network is a network of interconnected pathways deal with a network of cellular function and regulation. [3,2]. There are several types of pathways such as metabolic, genetics and cell signaling

pathways and different types of codes for their presentation.[1]. Metabolic pathways are agent for metabolism of an organism and involved in generation of essential compounds such as amino acids, lipids, sugar and the energy for their synthesis and their use for building proteins and cellular structures. In metabolic pathway, the metabolite as an initial chemical underwent modification by a chain of chemical reactions which are catalyzed by enzymes in order to fulfill a function. [1,3]. Genetic pathways are also central emphasis in biological systems. Genetic pathways concerned with the regulation of genes. In a simple representation of a genetic pathway, nodes represent genes and phenotypes and arcs stand for direct impact of genes on other genes and phenotypes. In computational analysis of genetic pathway and disease studies related to this pathway, mutations are main tools for definition and classification of genes that participate in a biological mechanism. Determination of the relationship between the genes are then fulfilled using combination of mutations in two or more genes. Genetic networks drawn by generation of

biological mechanism with molecular details which is performed by integration of the relationships between gene pairs. This method of ordering gene function is very fast and economic as compared to the time and cost required to obtain the experimental data.[1]. Cell signaling pathway can be defined as transmission of signals from exterior to interior of the cell using secretion of chemical factors. Inactivation and deficiency in this pathway lead to emerging several disease and cancers, hence, analysis of signaling pathway is a central importance in study of cancers and drug discovery. [3]. There are four types of cell signalling: Hormones secreted in endocrine glands are transported into the bloodstream and can be traveled widely throughout the body. Paracrine signals: are secreted by cells into extracellular medium in nearby cells. Neuronal signals are conducted along axon to distant target cells. Contact-dependent signals involves direct membrane-to-membrane contact of cells with each other. The cells also produce internal signals, e.g. generation of signals for stopping cell division during chromosome duplication and also for starting repair processes or apoptosis. [1,3]. The signaling pathway also called regulatory pathway as the signals and machinery for their processing form a regulatory network. The goal of the network is to control the life process and maintain the equilibrium level for organism to grow and survive. The most important classes of regulatory networks in analysis of signaling pathways are feedback systems (positive and negative feedbacks) and cyclic processes. These classes are essential for proliferation and differentiation processes in multicellular organism. Bioinformatics analysis and computational interpretation of signaling networks help in understanding biological system and related diseases. [3,4].

II. METHODS AND MATERIAL

A. Bioinformatics sources for metabolic pathways analysis:

RegulonDB

(http://www.cifn.unam.mx/Computational_Biology/regulondb) is a database for signal transduction pathways and transcription regulatory system in Escherichia coli. This database provide information about organization of operon and their disintegration into transcription units, promoters, regulatory proteins and their binding sites, ribosomal binding sites terminators, et. RegulonDB

contains both documentation and prediction objects. In addition it is linked with Swiss-prot, with microarray databases for analysis and visualization of microarray experiments.[5]

WIT

The WIT (What Is There) (<http://wit.mcs.anl.gov/WIT2/>) is a comparable computational system for analysis of sequenced genomes and generation of metabolic reconstructions derived from chromosomal sequences and metabolic modules. . [6]

UM-BBD

University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD, <http://umbbd.ahc.umn.edu/>). It includes information about microbial reaction and biodegradation pathways for chemical compounds. It contains database for pathways, reactions, compounds, enzymes, and microorganism entries. [7]

EcoCyc

(<http://www.ecocyc.org/>) is the database for Escherichia coli K-12 MG1655 including its genetic transporters, transcriptional regulation and metabolic pathways. [1].

MetaCyc database: (<http://metacyc.org>) MetaCyc is a database of experimentally clarified metabolic pathways from variety of different organisms. MetaCyc gathered information from over 2600 different organisms including 2260 pathways. These gathered pathways are of both primary and secondary metabolisms. The database also contains information about metabolites, reactions, enzymes and genes.. The purpose of MetaCyc is to list and file the bank of metabolism by stocking a descriptive sample of each experimentally clarified pathway. MetaCyc applications provide online curated database of metabolism, prediction of metabolic pathways from sequenced genomes, reinforcement of metabolic engineering by enzyme database and supporting of metabolomics research via metabolite database. [8].

Enzyme and metabolic pathways database

(<http://emp.mcs.anl.gov>) is a only encyclopedic biochemical data source. The configuration allows various types of tables and stoichiometric matrices to encode metabolic pathways, mechanisms of reactions, rate law and scale of numeric data. [9]

KEGG

(<http://www.genome.jp/kegg/>) stands for Kyoto Encyclopedia of Genes and Genomes is an important pathway database source for study and analysis of high – level functions and efficacies of the biological system, such as the organism, the cell and the ecosystem, from

the information at molecular level. It contains different section such as pathways, reference pathways and ortholog tables [10].

B. Bioinformatics tools for analysis of genetic pathways:

KEGG: as discussed in metabolic pathway part, contain graphical presentation (genome maps) and gene catalogs. [10].

KEGG BRITE

(http://www.genome.jp/kegg/brite_ja.html) stands for KEGG bimolecular reaction pathways for information transfer and expression .KEGG BRITE is a database of binary relations which is not limited only to molecular interactions but also computes and compares graphs involve genes and proteins. In this collection hierachial text(htext)files are created manually .The database consists several sets of pair association including the generalized protein interactions that underlie the KEGG pathway maps, systematic experimental information on protein- protein interactions .[11]

Gene network database

(<http://www.genenetwork.org/webqtl/main.py>) is a bioinformatics tool for genetic pathways especially for analysis of gene regulatory networks. This database provides information about linkage of DNA sequence variants (polymorphism) to related differences in gene and protein expressions. Data sets of Gene network are made up of large sets of genotypes and phenotypes which are obtained from closely related organisms .This data base is helpful for understanding of mechanism of embryogenesis as regulatory gene networks deal with encoding peptide products, transcription factors and enzymes used in differentiation and morphogenesis.[1]

Gene path (<http://magix.fri.uni-lj.si/genepath>): is an automated program used for analysis genetic data and detection of genetic networks. This program creates the genetic network either from such constraints or by combination of constraints with a qualitative logic – based algorithm for identification of networks .The body of the program contains following entities: Genetic data, Expert defined pattern, abductive inference engine and network synthesis methods. .[12]

C. Bioinformatics tools for analysis of signalling pathways:

SPAD or Signalling pathway data base (<http://www.grt.kyushu-u.ac.jp/spad>). This database provides information for signal transduction systems. It is a collective database which assembles information from protein –protein interactions, DNA and DNA and protein sequences. The database consists of four parts based on extracellular signal molecules such as Hormones, Cytokine, Growth factors and stress which are initials of intercellular signal transduction systems. [13].

STCDB

Signal transduction classification database: (<http://www.techfak.uni-bielefeld.de/~mchen/STCDB>). This data base contains information about classification of signal transduction in eukaryotic cells. [14]

TRANSPATH

(<http://www.biobase.de/pages/products/databases.html>) an integrated database on signal transduction and a tool for array analysis. It is a database deal with interpretation of gene regulatory networks that corporate comprehensive information on signal pathway with tools for visual information and analysis. This data base is linked with TRANSFAC (data base about transcription factors and their DNA binding sites) for the purpose of obtaining signaling transduction from binding of ligand to their targets and their products.[15]

Netpath <http://www.netpath> is another cell signaling pathway tools for study of signal transduction systems in human. This database contains diagrams of a number of immune signaling pathways and including 1600 reactions. [16].

Signal link 2.0 (<http://signalink.org/>) is an integrated database system for analysis of signaling pathways. The database includes pathway-specific transcription factors and post translational modifying enzymes, miRNA and scaffolds. [17]

III. RESULT AND DISCUSSION

The role of biochemical pathways and system biology has been highlighted in the study of diseases and drug discovery as they have gained importance in understanding the reasons of disorders and target detection .Bioinformatics tools emerged as alternative methods of study, analysis and drug discovery, for utilization of their benefits such as low cost, rapid process, high accuracy and large scale based studies. [18] For analysis of different multicellular organism's pathways such as metabolic, genetic and cell signalling pathways, bioinformatics tools also have taken a place in

analytical study of the pathways and biology system for the appropriate researcher to take their benefits. Using bioinformatics tools in the pathways analysis will provide the researchers with information on large scale, broad area and subjects involving the pathways. In metabolic pathways analysis, using bioinformatics tools such as RegulonDB, WIT.UM-BBD, EcoCyc, MetaCyc, Enzyme and metabolic pathways database, KEGG by the researchers will provide them with the encyclopaedic information about biochemical products, substrates, catalysing enzymes, amino acids, carbohydrates, lipids and toxic compounds etc. and their metabolic pathways specific diseases related to the failure in their functions. Bioinformatics tools like KEGG, KEGG BRITE, Gene network database, Gene path help the researchers in analysis of genetic pathways and regulatory networks in such a way that give information about the genes, transcriptional factors, miRNA, genes encode enzymes involved in genetic related diseases. Computational analysis of cell signalling pathways using following tools such as SPAD, STCDB, TRANSPATH Netpath, Signal link 2.0 help in understanding signal transduction and signalling pathways specific transcriptional factors also to obtain information from binding of signal pathways specific ligands to their targets and their products which all help in understanding the diseases and drug discovery deal with the appropriate pathways. So study of the pathways using bioinformatics tools has gained significance in biology field to reach the goal of discovery of novel approaches for better treatment of the diseases. Recently researchers discovered many biomarkers, drug targets such as, CDKs, GPCRs, aquaporins and ion channels related to different types of diseases like cancers but detailed and comprehensive information about their pathways and system biology is required, these bioinformatics tools can be helpful for them to utilize their benefits such as economic process, rapid computation, short time interpretation, high accuracy results and more detailed information due to availability of large scale study. So introduction of available tools to the researchers involved in such studies will help them to use them in their research analysis and take the benefits of such tools. But application of such tools in analysis study still requires sufficient knowledge of bioinformatics and computational skills and of course integration of these tools with other bioinformatics tools such as SNP detection (SNP related to specific diseases), gene finding tools, protein prediction tools, and other lab

based tools like gene mapping, microarrays, Q-PCRs, and so on will enhance results of such bioinformatics tools. These tools not only are applicable for human's pathways analysis and also for other animal model specially mouse and zebrafish which recently become a valuable animal model for human diseases and drug discovery. [19-26].

IV. CONCLUSION

Bioinformatics analysis of biochemical pathways and system biology has gained importance to utilize their benefits such as economic process, rapid computation, short time interpretation, high accuracy results and more detailed information due to availability of large scale study. So introduction of available tools to the researchers involved in such studies will help them to use them in their research analysis and take the benefits of such tools. But application of such tools in analysis studies still requires sufficient knowledge of bioinformatics and computational skills. Although these tools are useful and rapid for researchers but still they are not lack of disadvantages such as comprehensive information about all necessary pathways, interpretation tools and other programming related to analysis and detection of other genetic and metabolic parts, so integration of these tools with lab-based techniques will enhance their power and cover disadvantages of the tools.

V. REFERENCES

- [1] Rastogi .S. C., Rastogi P and Mendiratta N.,2008. Bioinformatics Methods and Applications: Genomics Proteomics And Drug Discovery PHI Learning Pvt. Ltd.
- [2] Cascante M, Boros LG, Comin-Anduix B, de Atauri P, Centelles JJ, Lee PW., 2002. Metabolic control analysis in drug discovery and disease. *Nat Biotechnol.*, 20(3):243-9.
- [3] Davidov E, Holland J, Marple E, Naylor S.,2003 . Advancing drug discovery through systems biology. . *Drug Discov Today.*, 8(4):175-83.
- [4] Dandekar T, Sauerborn R.,2002. Comparative genome analysis and pathway reconstruction. *Pharmacogenomics.*,3(2):245-56.
- [5] RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in Escherichia coli K-12., *Nucleic Acids Res.*,1:32(Database issue):D303-6.
- [6] Overbeek R, Larsen N, Pusch GD, D'Souza M, Selkov E Jr, Kyrpides N, Fonstein M, Maltsev N, Selkov E.2000. : integrated system for high-throughput genome sequence analysis and metabolic reconstruction. ... *Nucleic Acids Research.* , 28:123-125

- [7] Ellis LB, Roe D, Wackett LP.,2006. The University of Minnesota Biocatalysis/Biodegradation Database: the first decade., *Nucleic Acids Res.* 1:34(Database issue):D517-21
- [8] Caspi et al 2014, "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases," *Nucleic Acids Research* 42:D459-D471.
- [9] Selkov E., Basmanova S., Gaasterland T., Goryanin I., Grechkin Y., Maltsev N., Nenashev V., Overbeek R., Panyushkina E., Pronevich L., Sel'kov E., Jr., Yunus I., 1996.The Metabolic Pathway Collection from EMP: The Enzymes and Metabolic Pathways Database," *Nucleic Acids* 24 (1), 26-28.
- [10] Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*;28(1):27-30.
- [11] Aoki-Kinoshita KF, Kanehisa M., 2007. Gene annotation and pathway mapping in KEGG. *Methods Mol Biol.*, 396:71-91.
- [12] Zupan B, Bratko I, Demsar J, Juvan P, Curk T, Borstnik U, Beck JR, Halter J, Kuspa A., 2003 .GenePath: a system for inference of genetic networks and proposal of genetic experiments. *Artif Intell Med.* ;29(1-2):107-30
- [13] Tateishi N., Shiotari,H., Kuhara,S., Takagi,T. and Kanehisa,M., 1995. An integrated database SPAD (Signaling Pathway Database) for signal transduction and genetic information. *Genome informatics workshop*, December 11-12. Yokohama, Japan
- [14] Chen M, Lin S and Hofstaedt R .,2004. STCDB: Signal transduction Classification Database .*Nucleic Acids Research*, 40(1): D1187-D1193.
- [15] Krull M, Voss N, Choi C, Pistor S, Potapov A, Wingender E ., 2003.TRANSPATH: an integrated database on signal transduction and a tool for array analysis.. *Nucleic Acids Res.* 1;31(1):97-100.
- [16] Kandasamy, K. and Mohan, SS. et al. 2010. NetPath: A public resource of curated signal transduction pathways. *Genome Biology.* 11:R3.
- [17] Fazekas D, Koltai M, Türei D, Módos D, Pálffy M, Dúl Z, Zsákai L, Szalay-Bekö M, Lenti K, Farkas I J, Vellai T, Csermely P, Korcsmáros T .,2013. SignaLink 2.0 - A signaling pathway resource with multi-layered regulatory networks, *BMC Systems Biology.*, 7:7.
- [18] Ghorbani M and Karimi H.,2014.Ten bioinformatics tools for single nucleotide polymorphism detection, *American Journal of Bioinformatics* .,3 (2): 45.48
- [19] Ghorbani M and Karimi H. Cyclin-Dependent Kinases as valid targets for cancer treatment. *Journal of Pharmacy Research* 2015,9(6),377-382
- [20] Ghorbani M, Karimi H, 'Ion Channels Association with Diseases and their Role as Therapeutic Targets in Drug Discovery', *International Journal of Scientific Research in Science and Technology(IJSRST)*, 1(3):65-69,July-August 2015.
- [21] Ghorbani M, Karimi H, 'Role of Aquaporins in Diseases and Drug Discovery', *International Journal of Scientific Research in Science and Technology(IJSRST)*,1(3):60-64, July-August 2015
- [22] Ghorbani M, Karimi H, 'Role of G-Protein Coupled Receptors in Cancer Research and Drug Discovery', *International Journal of Scientific Research in Science and Technology (IJSRST)*,1(3), pp.122-126, July-August 2015.
- [23] Ghorbani, karimi H, 'Role of Biomarkers in Cancer Research and Drug Development', *International Journal of Scientific Research in Science and Technology(IJSRST)*,1(3), pp.127-132, July-August 2015
- [24] Ghorbani M, Karimi H, 'Role of Microarray Technology in Diagnosis and Classification of Malignant Tumours', *International Journal of Scientific Research in Science and Technology(IJSRST)*, 1(3):117-121, July-August 2015
- [25] Ghorbani M. Karimi H, *Bioinformatics Approaches for Gene Finding* ", *International Journal of Scientific Research in Science and Technology(IJSRST)*,1(4), pp.12-15, September-October-2015
- [26] Ghorbani, M. "Iranian traditional medicine for treatment of type II diabetes, anxiety and hypertension with introduction of zebrafish model system for their screening. *International Journal of Herbal Medicine* 2014; 2 (5): 13-19