International Conference on Machine Learning and Computational Intelligence-2017
International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 7 | ISSN : 2456-3307

Research Issues in Biological Data Mining: A Review

Kanica Sachdev, Manoj Kumar Gupta
Computer Science and Engineering Department SMVDU, J&K, India

ABSTRACT

Biological data is evolving at a very fast rate in the recent years. Large datasets of biological data are now available for analysis and inference. Biological data mining techniques help in the understanding of this data to help biologists to study and visualize the relation between this data under different conditions. This paper presents the biological data mining research areas and the corresponding tools that have been developed in these areas. It studies the various techniques of biological data mining data to provide an idea of the current state of research and introduces future directions for researchers to work in these fields.

Keywords: Biological Data Mining, Visual Data Mining, Biclustering, Pathway Analysis

I. INTRODUCTION

Data mining refers to the analysis of the existing data present in databases in order to generate certain new information [1]. It involves the identification of various associations and patterns from the complex and heterogeneous data by applying various techniques of statistics and machine leaning. It is basically categorized into two sections: descriptive and predictive [2]. Descriptive data mining refers to the characterization and depiction of the existing data. Predictive data mining on the other hand refers to interpretation of the existing information for the purposes of prediction. There have been tremendous advancements in the field of data mining in the past decade. Numerous methods and tools have been developed for the clustering of extensive amount of data, analyzing spatial/ temporal data, sequential and structured pattern analysis, outlier analysis from existing databases etc [3].

Biological data too has seen immense growth in the last decade. Data relating to medical issues and diseases has been increasing rapidly. With the development of new technologies related to medical research and applications, the biological data has been expanding in volume as well as diversity [4]. The scale of current biological data has already gone beyond petabytes and exabytes of storage. Figure 1 shows the accelerated growth in biological data around the year 2011. The research in the field of biology has caused the generation of explosive amounts of medical and clinical data that includes DNA microarrays, to biomedical images, to patient and health records [5]. This has created the need to develop more efficient algorithms and techniques to handle the complex and heterogeneous data, to integrate the varied data from different sources and to develop principles for the manipulation of this data [6]. New methods and tools need to be established that can help analyze this huge volume of biological information so as to develop a better understanding of biological processes. Data mining of biological information and databases hence becomes the greatest challenge for the researchers [7].

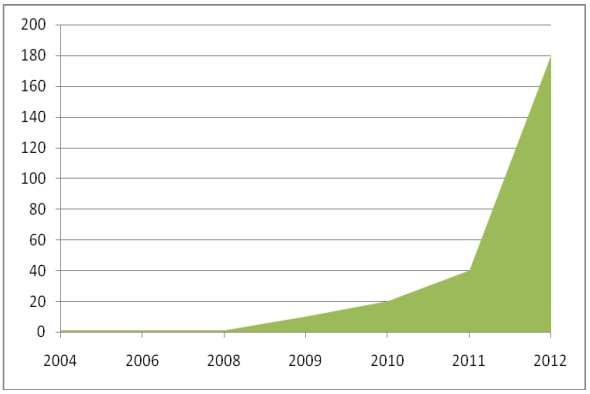

Figure 1. Growth of biological data (in terabases) over recent years [8]

This paper aims to highlight the important research issues in the data mining of biological data. Numerous issues like preprocessing/ cleaning of data, visual analysis of data, pathway analysis of biological data, biclustering etc exist. The goal of this paper is to introduce certain prominent research areas and the noteworthy contributions in those areas. It can serve as a starting point for researchers to understand the fields of research related to biological data mining. Section I gives a basic introduction of biological data mining and its need. Section II highlights the research areas and the prominent works of researchers in those areas. Section III concludes and summarizes the paper.

II. RESEARCH ISSUES IN BIOLOGICAL DATA MINING

Due to the recent growth in medical processes and biological data, many researchers have aimed to develop new techniques that can help in the mining and understanding of this data. Various research issues exist in this field of biological data mining. This paper aims to highlight certain pressing fields of research in this area i.e. visual data mining, biclustering and biological pathway analysis. Visual data mining refers to the representation of data in a form that is easy to analyze and comprehend. Biclustering refers to the grouping of data based on various expression values in order to identify patterns and structures. Biological pathway analysis is also somewhat related to visual data mining. It helps to determine the sequence of evolution of various molecules and genes in order to discover certain changes in them. The following sub sections explain each of these issues and highlight some of the major works done in these areas. Future directions and scope of these areas is also proposed for researchers to give directions for further research.

Visual Data Mining

Visual data mining is a combination of information visualization and computer graphics in the field of life sciences. It refers to the representation of various forms of information like macromolecular structures, genes, sequences, magnetic resonance imaging records etc. It helps to perceive and communicate data, to develop new ideas, as well as to apprehend the biological processes [9]. Data visualization is an important research area in the field of data mining of biological data. The need of data visualization of biological data arises due to various reasons. Firstly, biological data is huge in terms of volume as well as type. The human genome, for example, consists of 3 billion base pairs [10]. Secondly, various biological technologies producing data in the form of DNA microarrays, serial analyses of gene expression (SAGE) etc are developing expeditiously. Hence it becomes difficult to analyze this large quantity of data. Also, visualization tools need to be developed to integrate the heterogeneous sources of data and to model various biological systems. They are required in order to visualize the raw data present in the form of textual annotations tables, images etc and the distributed information stored in diverse spatially and temporally differing data sets. Thus, visual data mining aids the knowledge discovery and comprehension of biological data.

Numerous techniques and tools have been proposed for biological data visualization. The earliest contributions in this field include ACeDB system [11]

and the Entrez web browser [12]. These systems integrated the data, the database management system and the user interface as a complete tool provided to the user. The former is a software package coded in C language which is used to handle the physical as well as genetical data and DNA sequences of animals, plants and prokaryotes. It contains a web interface that is flexible to adapt to any database schema. Entrez web browser from NCBI (National Centre for Biotechnology Information) has also integrated the searching and retrieval modules in which the web interface gives access to all datasets concurrently by entering a single query. It is capable of retrieving related structures and sequences. It visualizes various chromosome maps and gene sequences. These integrated systems helped the biologists to easily mine data visually. However, these systems used data that contained only a snapshot of information occurring at the distribution time which later becomes obsolete. The data may be regularly updated over the internet but it is a tedious job and may be prone to errors. Also, these software packages need to be locally installed on a machine and hence cannot be accessed from any location.

Consequently, the BioViews browser applet was introduced which was written in Java [13]. It represents the biological features on physical maps and DNA sequences. The API is connected to various datasets and can retrieve diverse features and present the hyperlinked data on the features that are selected. The browser was built on top of extensible graphic components that can be reused by other programmers without knowing the internal coding details. The widgets in this browser also provide the feature of semantic zooming so as to view the data at differ detailed levels for a better understanding. Zomit was another architecturally independent applet tool that was developed [14]. The earlier systems used in the visualization of biological data generated new pages when a link from a particular page was followed. They provided no relationship between the two linked pages. Zomit overcame this drawback and helped to keep track of the various views for helping the biologists to maintain semantic link between different views.

Other tools that were developed for biological data visualization include Apollo [15] and GBrowse [16]. Apollo is a Java application that allowed the biologists to view genome annotations as well as edit them. It is an interactive tool for biologists that helps them to evaluate the data related to each annotation. GBrowse included many features like scrolling and zooming different regions of a genome, enabling/ disabling tracks, altering the relative order and appearance of tracks etc. It contained open source components and had a simple installation and integration process. Subsequently, GDVTK (Genome Data Visualization Toolkit) was developed [17]. Unlike Apollo that was a standalone application and needs to be locally installed, GDVTK was developed as a library. In comparison to GBrowse, it required less CPU time and memory as it handled the web requests using Java Servlet. GBrowse on the other hand used CGI (Common Gateway Interface) to handle requests that creates a corresponding page on the server to serve each request. However, an important limitation of GDVTK was that it required certain level of technical expertise due to the complexities of J2EE.

Recent tools include IGV (Integrative Genomics Viewer) [18] and BioCircos.js [19]. IGV can handle diverse datasets efficiently and provides an effortless user experience. The main emphasis of IGV is to support array based as well as next generation sequencing data. IGV may be used to visualize the genomic data from public database, however its main focus is to help biologists to visualize and understand their individual data or the data from their contemporaries. Thus IGV provides efficient data visualization on standalone desktop systems. BioCircos.js is a lightweight script used for interactive

visualization of biological data. It helps in visualizing bimolecular interactions, gene variations and genomic features. It contains Background module to show axis circles and Text module for the annotations. It supports numerous platforms and can be used on all major web browsers. A comparison of various data visualization tools on the basis of their architecture and main distinguishing feature is given in Table 1.

Although many data visualization tools have been developed for biological data, certain open issues still need to be addressed. The data produced by various experiments contain huge amount of noise. This inserts uncertainty in the visualization representations. This uncertainty needs to be addressed so that the biologists can clearly understand and apprehend data. Also, the quality of various graphical visualization models need to be measured for a better comparison and understanding. Although some measures for comparing the effectiveness of these models have been discussed [20], more comparison measures need to be formulated for an efficient comparison. Additionally, it is difficult to represent the evolving changes in data. Optimizing the display space and the visual considerations in order to represent this high dimensional dynamic information is still a challenge.

Table 1. Summary of data visualization tools

Tool	Architectural details	Main Feature
ACeDB	Database management system with user interface coded in C	Integrated data, database and user interface
Entrez	Web Browser	Can be accessed from any location
BioViews	Java Applet	Built on top of graphic components and contains reusable modules
Zomit	Java Applet	Keeps track of different views
Apollo	Java application	Allows user to view/edit genome annotations
GBrowse	Interactive web pages coded using Java Servlets	Includes features like scrolling, zooming etc.
GDVTK	Java based application framework	Requires less CPU time and memory
IGV	Java application	Mainly for biologists to analyze their individual data
BioCircos.js	JavaScript	Lightweight script with diverse features

Biclustering

Once the data has been visualized using biological data visualization tools, it is still hard to comprehend the results. The process of extracting useful information from the visual data is still a challenging task. An important step in analysis of this data is the

grouping of genes that display similar properties or patterns. This grouping/ clustering was shown to be beneficial for the purpose of identification, classification and annotation. However the process of clustering has certain drawbacks. Firstly, it is based on the presumption that similar genes exhibit same properties over all set of conditions. Although, this presumption may hold true when the data is gathered from a single experiment, but it is not correct when the data is accumulated from various experiments and diverse conditions. Secondly, clustering process divides the data into disjoint groups which assumes that each gene belongs to only one biological process or function. This may not always be the case [21].

To overcome the limitations of clustering, subset of genes with similar properties across a subset of conditions are identified. This is achieved by the process of biclustering. The data is organized in the form of matrix with the rows and columns representing these subsets. In the biclustering of biological data, each row represents ne gene and each column represents one condition. Each cell of this matrix shows the expression level of a gene under a specific condition [22]. Biclustering helps to achieve the major objectives of analysis of gene expression data i.e. identifying genes with similar expressions under numerous conditions, identifying conditions with similar gene expressions and classifying new genes based on the expression of other genes [23].

Many biclustering algorithms have been proposed for the classification of genetical data to identify local patterns, where similar properties are being shown by a subset of genes. Cheng and Church suggested an algorithm for biclustering based on MSR (Mean Squared Residue) [24]. The method begins by excluding the rows and colmns where the value of MSR is very high. When the value of MSR becomes greater than or equal to a threshold value, the rows and columns whose residue has a value lesser than the bicluster value are included back. If more than one bicluster s to be selected, the selected biclusters are masked and the process repeats iteratively.

OPSM (Order Preserving Submatrix) is a deterministic greedy algorithm to detect biclusters [25]. Since all biclusters are order preserving, this method represents a bicluster as an order preserving submatrix. The other algorithms aimed to classify all set of genes across all set of experiments, but this algorithm aims to identify a subset of genes in a subset of experiments. It creates biclusters by developing each bicluster iteratively using a probabilistic score that every bicluster will develop to a certain size. The best biclusters at each step of iteration are retained.

MOTIFS is another non deterministic greedy algorithm that constructs biclusters from a dataset with conserved rows [26]. Firstly it constructs intervals by comparing the significance of the interval to uniform distribution. Then, a seed column is selected randomly. For each of the seed columns, the algorithm identifies rows having the same state. Thus, this algorithm detects biclusters with constant row values. This representation has many applications. If the various clusters correspond to various diseases, we can find out the genes that are conserved in many classes but have different values in different classes. These genes can serve as drug targets. Also, the information about the highly expressed genes can be used in pathway identification.

Prelić et al. proposed a divide and conquer algorithm BiMax that searches binary matrix for rectangles consisting of 1s [27]. The entire data matrix is converted into binary form by any thresholding/binarization method. It starts with the entire matrix of data and iteratively divides it into checker board format. Another important proposed algorithm was QUBIC (QUalitative BIClustering) [28]. This algorithm can detct all significant

biclusters. Also, it is a fast and efficient algorithm that can construct biclusters from thousands of genes under thousands of condition in only a few minutes. It represents the data in the form of bipartite graphs and identifies heavy subgraphs. The data is first converted to a discrete form and then the biclusters are produced recursively from a seed edge of the graph.

Bayesian biclustering and spectral biclustering was also proposed for constructing biclusters [30, 31]. Bayesian

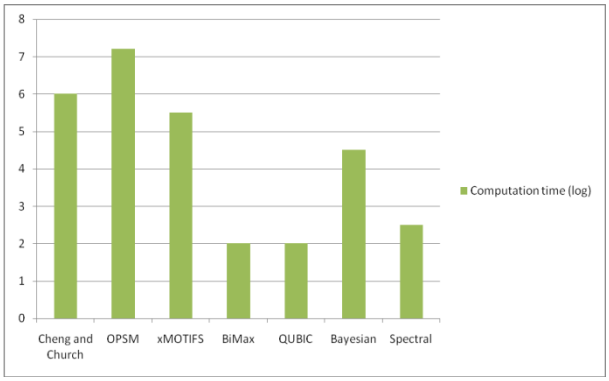

Figure 2. Comparison of biclustering algorithms in terms of running time [29]

biclustering used Gibbs sampling to detect biclusters. It handled the problem of missing data using Monte Carlo imputation. Spectral biclustering constructed checkerboard structures using eigen vectors for each gene expression. These eigen vectors can be identified using SVD (singular value decomposition) or linear algebra techniques.

Figure 2 shows a comparison of the biclustering algorithms in terms of computation time for a dataset comprising of 4000 rows. As it can be seen BiMax and QUBIC method work efficiently and have a lesser computation time compared to other methods.

Many other algorithms and techniques have been proposed and used for the biclustering of gene data and many other combinations of techniques have been employed for the same. Only a few notable methods have been mentioned in this paper. Biclustering of biological data has several applications in data analysis, data mining and filtering. Various other potential applications can also be uncovered by applying these algorithms in other research areas. Future work in this area can include the comparative study of these biclustering methods with already known biological data for the purpose of validation. The algorithms can also be modified to develop more accurate and efficient techniques. The significance of the extracted biclusters also needs to be studied as the selection of a large number of biclusters may be difficult to analyze in real life applications [32].

Biological Pathway Analysis

Biological pathways are defined as a sequence of interactions in the molecules that result in any transition or modification in the cell. These pathways can lead to the collection of molecules like fats/ proteins. The pathways are associated in certain processes like transmission of signals and gene expression regulation. Biological pathways depict how one or more molecules can be used by the organisms to form new products that are essential for sustenance [33]. These pathways are rarely linear in structure. Mostly they consist of several steps which may contain many intermediates. This makes them complicated to visualize and apprehend.

The visualizations of these pathways are static and can be constructed manually. However, this is a time consuming process. Also manual construction does not give additional information about the pathway and its members. This requires an automation for the construction of biological pathways. Many tools exist for the visualization and generation of these pathways.

IPA (Ingenuity Pathway Analysis) is used to analyze the biological pathways that encompasses four algorithms: Upstream Regulator Analysis (URA) to

find out the possible upstream regulators that are directly/ indirectly related to genes, Mechanistic Networks (MN) associates the regulators that belong to the same mechanism in hypothesis network, Causal Network Analysis (CNA) relates upstream regulators to molecules but focuses on the path having more than one link and hence gives a better analysis of possible causes of observed changes and Downstream Effects Analysis (DEA) analyzes the effect on biological functions that are connected to genes whose expression has changed [34]. IPA helps in various analysis applications. It helps to find the most relevant biological functions/ diseases for a particular set of genes. It also predicts the downstream effects of genes on biological functions/ diseases as well as the activation of upstream regulators. It can also be used to compare the affected pathways across the various experiments/ conditions.
GeneSpring is another tool developed by Silicon Gentics for pathway analysis [35]. It had various interactive features like an interface for an organized file management system, a collection of various clustering tools, has numerous data display methods, can handle input data from different formats, had automated annotation feature. It can be used for performing many functions like measuring similarity, hierarchal and k-means clustering, pathway analysis, constructing self organizing maps etc.

The pathway visualization tools generally do not integrate the molecular interactions with the state measurements so that they can be viewed on a common platform and can be studied over various parameters and biological attributes. Cytoscape was designed keeping this need in mind [36]. It presents an environment to combine bimolecular networks and states on a common environment. It can be used to integrate the diverse data, transfer annotations to desired node/ edge, provides automated graphical layout methods and supports graph selection and filtering.

VisANT is another tool that provides an online interactive interface for the interaction of biological data [37]. It provides tools for mining data and visualizing it in terms of pathway, related annotations, sequence and structure. The analyzed and inter related data can be combined and manipulated using numerous built in features of this software.

The proposal of different tools for biological pathway analysis is progressive and open ended due to the growth and development of their targeted applications. Future research can focus on developing more flexible softwares with easy download [procedures, consistent pathway names and description of pathway overlap mechanism. Also, development of a universal scheme for annotation can help to increase the interoperability between different tools. Future pathway analysis tools can also be developed that have increased computational efficiency. The datasets can also include diverse data like genomic databases, pathways, parameters etc to increase the flexibility of their operation.

III. CONCLUSION

This paper presents an overview of various open ended problems in the field of data mining of biological data. Three important issues i.e. visual data mining, pathway analysis and biclustering are mentioned. Although many tools and techniques have been proposed for these issues, still researchers can help to achieve better efficiency and accuracy and design tools with various features that can help scientists and biologists to analyze biological datasets and infer results from the same.

IV. REFERENCES

[1]. Romero, Cristóbal, and Sebastián Ventura. "Educational data mining: a review of the state of the art." IEEE Transactions on Systems, Man, and

Cybernetics, Part C (Applications and Reviews) 40, no. 6 (2010): 601-618.

[2]. Chen, Ming-Syan, Jiawei Han, and Philip S. Yu. "Data mining: an overview from a database perspective." IEEE Transactions on Knowledge and data Engineering 8, no. 6 (1996): 866-883.

[3]. Han, Jiawei. "How can data mining help bio-data analysis?." In Proceedings of the 2nd International Conference on Data Mining in Bioinformatics, pp. 1-2. Springer-Verlag, 2002.

[4]. Cook, Charles E., Mary Todd Bergman, Robert D. Finn, Guy Cochrane, Ewan Birney, and Rolf Apweiler. "The European Bioinformatics Institute in 2016: data growth and integration." Nucleic acids research 44, no. D1 (2015): D20-D26.

[5]. Li, Xiaoli, See-Kiong Ng, and Jason TL Wang, eds. Biological data mining and its applications in healthcare. Vol. 8. World Scientific, 2013.

[6]. Li, Yixue, and Luonan Chen. "Big biological data: challenges and opportunities." Genomics, proteomics & bioinformatics 12, no. 5 (2014): 187-189.

[7]. Yang, Qiang, and Xindong Wu. "10 challenging problems in data mining research." International Journal of Information Technology & Decision Making 5, no. 04 (2006): 597-604.

[8]. Marx, Vivien. "Biology: The big challenges of big data." Nature 498, no. 7453 (2013): 255-260.

[9]. O'Donoghue, Seán I., Anne-Claude Gavin, Nils Gehlenborg, David S. Goodsell, Jean-Karim Hériché, Cydney B. Nielsen, Chris North et al. "Visualizing biological data—now and in the future." Nature methods 7 (2010): S2-S4.

[10]. Venter, J. Craig, Mark D. Adams, Eugene W. Myers, Peter W. Li, Richard J. Mural, Granger G. Sutton, Hamilton O. Smith et al. "The sequence of the human genome." science 291, no. 5507 (2001): 1304-1351.

[11]. Dunham, I., R. Durbin, J. Thierry-Mieg, and D. R. Bentley. "Physical mapping projects and ACEDB." Guide to human genome computing (1994): 111-158.

[12]. Epstein, Jonathan A., Jonathan A. Kans, and Gregory D. Schuler. "WWW Entrez: A Hypertext Retrieval Tool for Molecular Biology." (1994).

[13]. Epstein, Jonathan A., Jonathan A. Kans, and Gregory D. Schuler. "WWW Entrez: A Hypertext Retrieval Tool for Molecular Biology." (1994).

[14]. Pook, Stuart, Guy Vaysseix, and Emmanuel Barillot. "Zomit: biological data visualization and browsing." Bioinformatics (Oxford, England) 14, no. 9 (1998): 807-814.

[15]. Lewis, Suzanna E., S. M. J. Searle, N. Harris, M. Gibson, V. Iyer, J. Richter, C. Wiel et al. "Apollo: a sequence annotation editor." Genome biology 3, no. 12 (2002): research0082-1.

[16]. Stein, Lincoln D., Christopher Mungall, ShengQiang Shu, Michael Caudy, Marco Mangone, Allen Day, Elizabeth Nickerson et al. "The generic genome browser: a building block for a model organism system database." Genome research 12, no. 10 (2002): 1599-1610.

[17]. Sun, Hao, and Ramana V. Davuluri. "Java-based application framework for visualization of gene regulatory region annotations." Bioinformatics 20, no. 5 (2004): 727-734.

[18]. Thorvaldsdóttir, Helga, James T. Robinson, and Jill P. Mesirov. "Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration." Briefings in bioinformatics 14, no. 2 (2013): 178-192.

[19]. Cui, Ya, Xiaowei Chen, Huaxia Luo, Zhen Fan, Jianjun Luo, Shunmin He, Haiyan Yue, Peng Zhang, and Runsheng Chen. "BioCircos. js: an interactive Circos JavaScript library for biological data visualization on web applications." Bioinformatics 32, no. 11 (2016): 1740-1742.

[20]. Bertini, Enrico, Andrada Tatu, and Daniel Keim. "Quality metrics in high-dimensional data visualization: An overview and systematization." IEEE Transactions on Visualization and Computer Graphics 17, no. 12 (2011): 2203-2212.

[21]. Tanay, Amos, Roded Sharan, and Ron Shamir. "Discovering statistically significant biclusters in

gene expression data." Bioinformatics 18, no. suppl_1 (2002): S136-S144.

[22]. Tanay, Amos, Roded Sharan, and Ron Shamir. "Discovering statistically significant biclusters in gene expression data." Bioinformatics 18, no. suppl_1 (2002): S136-S144.

[23]. Madeira, Sara C., and Arlindo L. Oliveira. "Biclustering algorithms for biological data analysis: a survey." IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 1, no. 1 (2004): 24-45.

[24]. Cheng, Yizong, and George M. Church. "Biclustering of expression data." In Ismb, vol. 8, no. 2000, pp. 93-103. 2000.

[25]. Ben-Dor, Amir, Benny Chor, Richard Karp, and Zohar Yakhini. "Discovering local structure in gene expression data: the order-preserving submatrix problem." Journal of computational biology 10, no. 3-4 (2003): 373-384.

[26]. Murali, T. M., and Simon Kasif. "Extracting conserved gene expression motifs from gene expression data." In Pacific symposium on biocomputing, vol. 8, pp. 77-88. 2003.

[27]. Prelić, Amela, Stefan Bleuler, Philip Zimmermann, Anja Wille, Peter Bühlmann, Wilhelm Gruissem, Lars Hennig, Lothar Thiele, and Eckart Zitzler. "A systematic comparison and evaluation of biclustering methods for gene expression data." Bioinformatics 22, no. 9 (2006): 1122-1129.

[28]. Li, Guojun, Qin Ma, Haibao Tang, Andrew H. Paterson, and Ying Xu. "QUBIC: a qualitative biclustering algorithm for analyses of gene expression data." Nucleic acids research 37, no. 15 (2009): e101-e101.

[29]. Eren, Kemal, Mehmet Deveci, Onur Küçüktunç, and Ümit V. Çatalyürek. "A comparative analysis of biclustering algorithms for gene expression data." Briefings in bioinformatics 14, no. 3 (2012): 279-292.

[30]. Gu, Jiajun, and Jun S. Liu. "Bayesian biclustering of gene expression data." BMC genomics 9, no. 1 (2008): S4.

[31]. Kluger, Yuval, Ronen Basri, Joseph T. Chang, and Mark Gerstein. "Spectral biclustering of microarray data: coclustering genes and conditions." Genome research 13, no. 4 (2003): 703-716.

[32]. Madeira, Sara C., and Arlindo L. Oliveira. "Biclustering algorithms for biological data analysis: a survey." IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 1, no. 1 (2004): 24-45.

[33]. Wang, Kai, Mingyao Li, and Hakon Hakonarson. "Analysing biological pathways in genome-wide association studies." Nature reviews. Genetics 11, no. 12 (2010): 843.

[34]. Kramer, Andreas, Jeff Green, Jack Pollard Jr, and Stuart Tugendreich. "Causal analysis approaches in ingenuity pathway analysis." Bioinformatics 30, no. 4 (2013): 523-530.

[35]. Chu, Lillian, Eric Scharf, and Takashi Kondo. "GeneSpringTM: tools for analyzing microarray expression data." Genome Informatics 12 (2001): 227-229.

[36]. Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. "Cytoscape: a software environment for integrated models of biomolecular interaction networks." Genome research 13, no. 11 (2003): 2498-2504.

[37]. Hu, Zhenjun, Joseph Mellor, Jie Wu, and Charles DeLisi. "VisANT: an online visualization and analysis tool for biological interaction data." BMC bioinformatics 5, no. 1 (2004): 17.