Metabolome Analysis: Disease Biomarker Discovery

Chapter · January 2018

DOI: 10.1016/B978-0-12-809633-8.20137-4

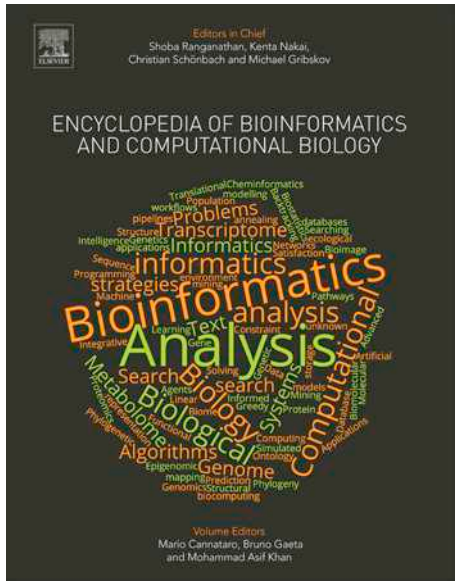
5 authors, including:

Tiratharaj Singh
Jaypee University of Information Technology
120 PUBLICATIONS **446** CITATIONS
SEE PROFILE

Ankita Shukla
CSIR - Institute of Himalayan Bioresource Technology
14 PUBLICATIONS **21** CITATIONS
SEE PROFILE

Taoufik Bensellak
Ecole de Sciences Appliqués de Tanger
8 PUBLICATIONS **1** CITATION
SEE PROFILE

Ahmed Moussa
National School of Applied Scienes, Tangier, Morocco
72 PUBLICATIONS **249** CITATIONS
SEE PROFILE



Tiratha R. Singh, Ankita Shukla, Bensellak Taoufik, Ahmed Moussa and Brigitte Vannier (2019) Disease Biomarker Discovery. In: Guenther, R. and Steel, D. (eds.), Encyclopedia of Bioinformatics and Computational Biology, vol. 3, pp. 476–488. Oxford: Elsevier.

Disease Biomarker Discovery

Tiratha R Singh and Ankita Shukla, Jaypee University of Information Technology, Solan, India
Bensellak Taoufik and Ahmed Moussa, École Nationale Des Sciences Appliquées de Tanger, Tangier, Morocco
Brigitte Vannier, University of Poitiers, Poitiers France

Metabolic Networks: A Background for Biomarkers

Network biology is a branch of science that deals with the interactions among biomolecules that include genes, transcripts, proteins, metabolites, etc. With the advent of system biology, networks are being used widely across many branches of biology (proteomics, genomics, transcriptomics, and metabolomics) as a convenient representation of the interaction between specific biological elements. These graphical representations denote the molecular-level blueprint of interactions and mechanisms of regulation inside a cell. These biological networks include gene regulatory network, transcriptomic network, protein–protein interaction network, and metabolic network. The network biology approach helped to cover the overall aspects of the necessary facets that need to be considered while finding the probable therapeutic intervention for the particular disease type. The interaction data come through the high-throughput methods that are gathered from individual studies and large-scale screens that finally get assembled into a topological form (i.e., network format) that holds significant biological properties. In a recent scenario, more attention has been given to the gene and protein networks to study a complex form of diseases (Shukla and Singh, 2017). Although it has been found that the metabolic networks seem to play a significant role in the complex disease regulation like in case of cancer's Warburg effect (Vander Heiden *et al.*, 2009), which signifies uncontrolled cell division even in anaerobic conditions that involve numerous metabolites and the reaction mechanisms. The metabolic network comprises of metabolites and enzymes that take the role of nodes and the reactions describing their transformations and is represented as directed edges in **Fig. 1** (Bourqui *et al.*, 2007).

Biochemical reactions happening inside a metabolic network allow an organism to grow, reproduce, and respond to the environment and maintain its structure (Xu *et al.*, 2016). In a biochemical pathway, the metabolic network centralizes its attention towards mass flow that generates essential components like amino acids, sugars, and lipids, and the energy required by the biochemical reactions (Zhu *et al.*, 2007). In a metabolic network, it's not only metabolites that perform the overall metabolism but there are genes and proteins too that commence their task in regulatory mechanisms; this is what makes the metabolic networks more efficient from the disease perspective and their immediate applications too for therapeutic interventions (Berkhout *et al.*, 2013). This shows that metabolic networks typically show the representation of not only metabolites but also for genes and proteins and therefore provide wide perspective in disease studies. In a cell, metabolism holds chemical processes by which cells break down food and nutrients into usable building blocks and then reassemble those building blocks to form the biological molecules known as metabolites (DeBerardinis and Thompson, 2012). The metabolites consumed are called the substrates of the reaction; however those produced are called the products. Most metabolic reactions do not occur spontaneously, or we can say that they occur at a very low rate; therefore enzymes are used to enhance the pace of the reaction to get it completed (Cooper, 2000). This breakdown and reassembly in a pathway entails a set of successive chemical reactions that convert initial inputs into useful end products via a series of steps and this complete set of reactions in the pathway forms the metabolic network (Sridharan *et al.*, 2015). To understand the interacting mechanism in a network it's necessary to understand the architecture of the network topology. In a metabolic network, nodes represent the chemicals produced and consumed by the reactions that include small molecules (i.e. carbohydrates, lipids, amino acids, and nucleotides), and the edges denote the metabolic flow or the regulatory effects of a specific reaction (Lee *et al.*, 2008). Understanding the complex network often requires a bottom-up approach that carries its path towards systems biology perspective (Shahzad and Loor, 2012). Thus there is need to examine a system, not only in terms of individual components but as a whole, which can be done by considering the elementary constituents individually as well as when they are connected. Numerous components of a system and their interactions are best characterized as networks and they are mainly represented as graphs where thousands of nodes are connected with thousands of vertices (Cho *et al.*, 2012).

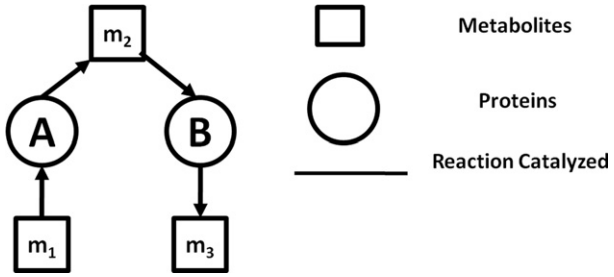


Fig. 1 Basic metabolic network.

Network analysis has suggested that biological networks have two imperative structural properties (Mahadevan and Palsson, 2005). First, it has been shown that several of these networks, including metabolic networks, are scale-free and possess a "small world" property (Barabasi and Oltvai, 2004). Second, scale-free networks are suggested to have high error tolerance and low attack tolerance (Crucitti *et al.*, 2004). In general, interactions in a network between biological entities (genes, transcripts, proteins, metabolites, etc.) can be classified on the basis of the nature of the interaction into two broad categories, i.e., influence networks and the flow networks. For influence networks, the nature of interactions are "influence-based" such as protein–protein interaction or signaling networks, i.e., connections mark the presence or absence of the reaction (Mahadevan and Palsson, 2005). This class might be extended with the case where the type of interaction is important in addition to presence or absence of the interaction, e.g., gene regulatory networks where transcription factors can either activate or repress gene expression. While in flow networks, where a specific variable like mass or energy flow may be conserved at each node, such as metabolic networks. However, it should be noted that the fundamental properties of biological networks in these two classes can be significantly different (Barabasi and Oltvai, 2004). It has been seen widely that along with the complete network graph, the subgraphs are also seen to play an essential functional role like network motifs (that represent most frequently occurring subgraph) and studies have shown the structural organization of the feedback loops (**Fig. 2**). Similarly, single-input and multiple-input motifs (**Fig. 2**) (Sehgal *et al.*, 2015) can influence the dynamics and the regulation of metabolic pathways (Beber *et al.*, 2012). For effective analysis it's important to model the network graphs correctly for which there are a wide variety of approaches depending on the features of interest through which network dynamics can be modeled, like for small networks with explicit kinetics, which can be modeled with differential equations, while for larger networks dynamics can be accessed by flux balance analysis or stochastic kinetic modeling (Boccaletti *et al.*, 2006). Also, for the case where only stoichiometric information is available, more basic approaches like network expansion or Petri nets can be utilized (Peleg *et al.*, 2005).

Varieties of graph representations are available in network biology but studies have shown that bipartite graph (**Fig. 3**) (two nodes represent metabolites with edges joining each metabolite to the reaction) is the most correct representation of the metabolic network (Veeramani and Bader, 2010). The edges in the representative graphs are directed because some metabolites (the substrates) go into the reaction and some (the products) come out of it. Metabolic networks are represented through nodes as metabolites and the links as reactions that are catalyzed by specific gene products; this representation is different from protein–protein interaction networks, where the nodes are the gene products and the links correspond to interactions. The analysis of protein–protein interaction networks has suggested that the deletion of the most highly connected proteins correlates well with a lethal phenotype (Mahadevan and Palsson, 2005). In contrast, a node in metabolic networks cannot be deleted by genetic techniques, but links can (Jeong *et al.*, 2001).

Now the question comes as to how the resultant computational network are formed and how their global representation is possible. The answer lies in the computer readable file formats for the biological networks, i.e., Systems Biology Markup Language (SBML), a global format which could be utilized for the reusability of network models. It is a XML-like machine-readable language that is proficient to represent models to be analyzed by a computer. SBML can represent metabolic networks, cell signaling pathways, regulatory networks, and many other kinds of systems (Hucka *et al.*, 2003). An increasing number of diseases are now

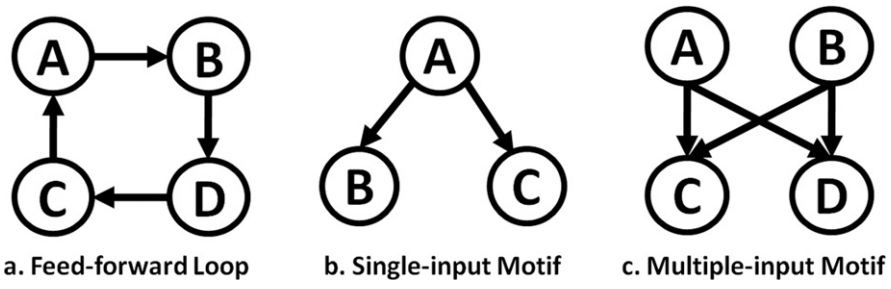


Fig. 2 Most frequent regulatory motif.

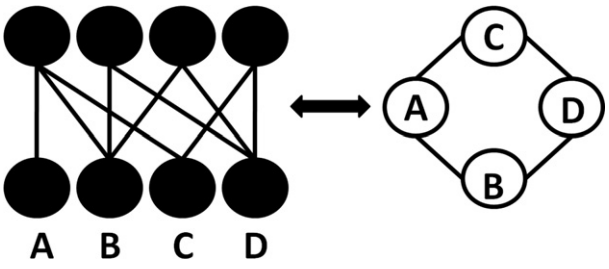


Fig. 3 Bipartite graph with corresponding simple representation.

seen to be a result of drastic perturbations of cellular functions that involve large sets of genes, where connections are complicated to understand. Diseases in particular like cancer, CVD, or diabetes causes huge perturbations in cell metabolism. Therefore, the study of metabolic networks primary perturbation in interactions and fluxes can aid better understanding of the physiopathology of such diseases. It notably permits an understanding of how and why alteration in activity or expression levels of a few enzymes can transmit substantial perturbations into entire cellular functions. Therefore these metabolites could have been proposed as putative biomarkers for the human diseases.

Metabolome Analysis: Computational Protocols and Methods

The metabolome comprises of the biochemical composition of the small-molecule metabolites present in a cell that are involved in the metabolic reaction mechanism and required for the maintenance, growth, and normal functioning of the cell (Mohney and Milburn, 2015). The metabolome was first described by Oliver and colleagues in 1998 (Oliver *et al.*, 1998), during their pioneering work on yeast metabolism that therefore confers the discipline of metabolomics; that follows the analysis of the metabolome. Metabolomics is sometimes also called metabonomics or metabolic profiling (Lindon *et al.*, 2011). Various experimental, statistical, and computational approaches have been discovered so far to perform the metabolome analysis. The most commonly used experimental techniques in metabolomics are mass spectrometry (MS), i.e., gas chromatography (GC–MS) or liquid chromatography (LC-MS), and nuclear magnetic resonance spectroscopy (NMR) spectra (Zhang *et al.*, 2012). Along with this statistical approaches have been devised to perform the analysis and it comprises of various standard analysis methods such as t-test and ANOVA, as well as more sophisticated methodologies such as univariate and multivariate analysis methods (Bartel *et al.*, 2013).

Also, bioinformatic approaches fall into the scenario due to the boom in high-throughput data and this includes databases for the data retrieval as well as web servers for carrying out the analysis (Maudsley *et al.*, 2011). Tremendous efforts have been implemented to archive all types of biological data, i.e., genes, proteins, gene products, metabolites, etc. With the advent of the genomic era, the amount of biochemical knowledge has exploded in the last two decades, which has necessitated its storage in large databases. Variety of top-down (gene to protein to metabolite) and bottom-up (chemical entity to biological function) approaches have been implied resulting in a rich expanse of metabolic knowledge from the biochemical network (Cakir and Khatibipour, 2014). Bioinformatics is a pioneer in preserving the data obtained through the experimental or natural resources and also provided the numerous tools for the analysis purpose. The databases provide the contextual biochemical basis for metabolomics data interpretation. By providing information regarding metabolites like defining which enzymatic reactions consume or produce them, and which pathways they're involved in, researchers can use this data to interpret their experiments towards a higher level of annotation.

METLIN, was the first metabolomics database, was established in 2004 (Smith *et al.*, 2005). Thereby in 2005, the Human Metabolome Project was launched to find and catalog all of the metabolites in human tissue and biofluids. This metabolite information is kept in the Human Metabolome Database, which produced its first draft in 2007 (Wishart *et al.*, 2007). There are many databases containing metabolomics data, and each has different information, ranging from NMR and MS spectra to metabolic pathways. Additionally, and more specifically to the development of metabolomics, mass spectral databases like the Golm Metabolome Database (Hummel *et al.*, 2007), which links mass spectrum and chromatographic retention time to specific compounds, have been developed. Some tools designed for higher level metabolomic analysis can take GC–MS spectra as input and use them for the identification stages of metabolomics. The purpose of metabolic databases is to organize the metabolites in a way that helps researchers in an easy identification and analysis. The information found in metabolite databases has continuously been updated to provide state-of-the-art data to the scientific community. Metabolomics is a new field and therefore new approaches are still being discovered and the existing ones are still improving. These databases are embraced with various types of information including concentration, anatomical location, and related disorders.

Other databases are MassBank (Horai *et al.*, 2008), lipid metabolites and pathways strategy (LIPID-MAPS) (LIPID), Madison metabolomics consortium database, and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000). The major database till now is the KEGG, which is divided into several subdatabases with LIGAND, REACTION PAIR and PATHWAY being the most relevant to metabolomics (Booth *et al.*, 2013). In KEGG however there is a dauntingly large sum, i.e., 10,664 reactions and 18,107 metabolites (Kanehisa and Goto, 2000). These databases have been undergoing continuous updation and annotation for and so contain a great deal of valuable information. KEGG and MetaCyc (Karp *et al.*, 2002) are currently the largest (most number of organisms) and most in-depth comprehensive databases available. There are other databases too as Reactome (Joshi-Tope *et al.*, 2005), Model SEED (Devoid *et al.*, 2013), and BiGG (Schellenberger *et al.*, 2010), that can be more useful than the large databases if a specific organism is desired. The KEGG and MetaCyc databases each contain a generalized conserved set of pathways based on metabolic pathways. For KEGG, organism-specific annotations are available to query while for MetaCyc, individual "Cyc" databases have been generated for a number of organisms, some are just computationally derived while others are extensively manually curated such as AraCyc for *Arabidopsis* (Mueller *et al.*, 2003) and EcoCyc for *E.coli* strain *K-12 MGI655* (Karp *et al.*, 2014). A more recent development is the cheminformatic databases like PubChem (Wang *et al.*, 2009) and ChEBI (Degtyarenko *et al.*, 2007), which provide a chemically ontological approach to catalog small molecules that are active in biological systems. These databases, therefore, can provide fruitful information regarding the metabolic datasets. Finally, it is important to note that these databases can be cross-referenced and linked to each other as well as against more widely known databases such as the well-known Chemical Abstract Service (CAS) among many others. Along with the above mentioned

Table 1 Selected databases and tools w.r.t metabolomics

Databases		
BioCyc	Collection of 10992 Pathway/Genome Databases (PGDBs)	http://biocyc.org
BiGG Models	Knowledgebase of genome-scale metabolic network reconstructions.	http://bigg.ucsd.edu/
BMRB	Repository for data from nuclear magnetic resonance spectroscopy (NMR) spectroscopy on proteins, peptides, nucleic acids, and other biomolecules	http://www.bmrb.wisc.edu/
BRENDA	Comprehensive enzyme repository, provide metabolic pathway details	http://www.brenda-enzymes.info/index.php
ChEBI	Freely available dictionary of molecular entities focused on "small" chemical compounds	https://www.ebi.ac.uk/chebi/
DIMEdb	Database of biologically relevant metabolite structures and annotations	http://dimedb.ibers.aber.ac.uk/
DRUGBANK	Unique bioinformatics and cheminformatics resource that combines detailed drug (i.e., chemical, pharmacological, and pharmaceutical) data with comprehensive drug target (i.e., sequence, structure, and pathway) information	https://www.drugbank.ca/
ECMDB	Contains extensive metabolomic data and metabolic pathway diagrams about *Escherichia coli (strain K12, MG1655)*	http://ecmdb.ca/
GMD	Facilitates the search for and dissemination of reference mass spectra from biologically active metabolites quantified using gas chromatography coupled to mass spectrometry (MS)	http://gmd.mpimp-golm.mpg.de/
HMDB	Archive information about small-molecule metabolites found in the human body	http://www.hmdb.ca/
KEGG	Resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information	http://www.genome.jp/kegg/pathway.html
MANET	Maps evolutionary relationships of molecule (metabolic, protein) architectures directly onto biological networks	http://manet.illinois.edu/
MassBank	Contains high resolution mass spectral data	http://massbank.eu/MassBank/
MetaboLights	Database for metabolomics experiments and derived information	http://www.ebi.ac.uk/metabolights/
MetaNetX	Automated model construction and genome annotation for large-scale metabolic networks	http://www.metanetx.org/
Reactome	Navigable map of human biological pathways, ranging from metabolic processes to hormonal signaling	http://www.reactome.org
SMPDB	Support pathway elucidation and pathway discovery in metabolomics, transcriptomics, proteomics and systems biology	http://smpdb.ca/
YMDB	Manually curated database of small-molecule metabolites found in or produced by *Saccharomyces cerevisiae*	http://www.ymdb.ca/
Computational analysis tools		
Arcadia	Visualization tool for metabolic pathways	http://arcadiapathways.sourceforge.net/
CellNetAnalyzer	MATLAB toolbox providing a various (partially unique) computational methods and algorithms for exploring structural and functional properties of metabolic, signaling, and regulatory networks	http://www.mpi-magdeburg.mpg.de/projects/cna/cna.html
GLAMM	Unified web interface for visualizing metabolic networks, reconstructing metabolic networks from annotated genome data, visualizing experimental data in the context of metabolic networks, and investigating the construction of novel, transgenic pathways	http://glamm.lbl.gov/
IMPaLA	Perform pathway overrepresentation and enrichment analysis with expression and/or metabolite data	http://impala.molgen.mpg.de
iPath	Web-based tool for the visualization, analysis and customization of the various pathways maps	http://pathways.embl.de
JDesigner	Graphical modeling environment for biochemical reaction networks	http://jdesigner.sourceforge.net/Site/JDesigner.html
KaPPA-View	Web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway	http://kpv.kazusa.or.jp/en/
LIPID MAPS	Online tools for lipid research	http://www.lipidmaps.org/tools/index.html
MapMan	A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes	http://mapman.gabipd.org/web/guest/mapman
MetaMapp	Mapping and visualizing metabolomic data by integrating information from biochemical pathways and chemical and mass spectral similarity	http://uranus.fiehnlab.ucdavis.edu:8080/MetaMapp/homePage
MetPA	A web-based metabolomics tool for pathway analysis and visualization	http://metpa.metabolomics.ca
MassTRIX	Annotate metabolites in high precision MS data	http://masstrix3.helmholtz-muenchen.de/masstrix3/
MetaboAnalyst	Web server designed to permit comprehensive metabolomic data analysis, visualization and interpretation	http://www.metaboanalyst.ca/MetaboAnalyst/
MetaPath Online	For the analysis of metabolic networks	https://scopes.biologie.hu-berlin.de/

(*Continued*)

Table 1 Continued

Databases		
Meta P-server	A web based, easy-to-use analysis tool for the statistical analysis of metabolomics data	http://metabolomics.helmholtz-muenchen.de/metap2/
MetExplore	Find information about metabolite relationships in metabolic networks	http://metexplore.toulouse.inra.fr/metexplore/
Metscape	A plug-in for Cytoscape, used to visualize and interpret metabolomic data in the context of human metabolic networks	http://metscape.ncibi.org
MGV	A versatile generic graph viewer for multiomics data also it offers a comprehensive set of tools for analysis and visualization of graphs.	http://www.microarray-analysis.org/mayday
MPEA	Metabolite pathway enrichment analysis	http://ekhidna.biocenter.helsinki.fi/poxo/mpea/
MSEA	A web-based tool to identify biologically meaningful patterns in quantitative metabolomic data	http://www.msea.ca
Omix	Network drawing tool along with programmable visualization framework	http://www.omix-visualization.com/?from=http://www.13cflux.net#sthash.vLyVKteK.dpbs
Paintomics	A web based tool for the joint visualization of transcriptomics and metabolomics data	http://www.paintomics.org
TICL	A web tool for network-based interpretation of compound lists inferred by high-throughput metabolomics	http://mips.helmholtz-muenchen.de/proj/cmp/home.html
Vanted	Network visualization and analysis tool for creating and editing the network and mapping experimental data onto networks	https://immersive-analytics.infotech.monash.edu/vanted/

Table 2 PAM50 genes list

PAM50 genes				
ACTR3B	CDCA1 (NUF2)	FOXA1	MDM2	PGR
ANLN	CDH3	FOXC1	MELK	PHGDH
BAG1	CENPF	GPR160	MIA	PTTG1
BCL2	CEP55	GRB7	MKI67	RRM2
BIRC5	CXXC5	KIF2C	MLPH	SFRP1
BLVRA	EGFR	KNTC2(NDC80)	MMP11	SLC39A6
CCNB1	ERBB2	KRT14	MYBL2	TMEM45B
CCNE1	ESR1	KRT17	MYC	TYMS
CDC20	EXO1	KRT5	NAT1	UBE2C
CDC6	FGFR4	MAPT	ORC6L(ORC6)	UBE2T

databases there are different types of databases built so far only the selected ones based upon accuracy and applications are mentioned in alphabetical order in **Table 1**.

Similarly, there are many computational tools available for the metabolic data handling like for building, editing, enrichment, and interpreting metabolic network models, including Arcadia (Villeger *et al.*, 2010), GLAMM (Bates *et al.*, 2011), CellNetAnalyzer (Klamt and von Kamp, 2011), MetaMapp (Barupal *et al.*, 2012), MetPA (Xia and Wishart, 2010a), Vanted (Rohn *et al.*, 2012) for network visualization. Likewise there are tools for drawing a network, Omix (Droste *et al.*, 2013) which have been widely used for creating a network model. TICL (Antonov *et al.*, 2009) is used for the interpretation of compound lists that are inferred by high-throughput metabolomics. Enrichment analysis could be done with the help of MSEA (Xia and Wishart, 2010b), MPEA (Kankainen *et al.*, 2011). MassTRIX (Suhre and Schmitt-Kopplin, 2008) performs the annotating metabolites in high precision MS data. There are other tools also, which have been mentioned in **Table 2** along with their descriptions.

Biomarker Discovery: A Challenge and Plausible Solutions Through Bioinformatics

Biomarkers are measured indicators of biological and pathogenic conditions or pharmacological responses to a therapeutic intervention (Strimbu and Tavel, 2010). Based on pathophysiological, epidemiological, therapeutic, or other scientific evidence they are intended to be useful in terms of clinical significance (i.e., to know whether they will benefit or harm) (Baumgartner *et al.*, 2011; Downing, 2001). Biomarkers have a generous impact on the care of patients; for those who are suspected to have the disease or those who have or have no visible disease symptoms (Baumgartner *et al.*, 2011). For a long time biomarkers have served as a plausible diagnostic key to unraveling disease conditions, especially in case of cancers. Depending on the condition type they can be categorized as diagnostic, prognostic, and screening biomarkers (Madu and Lu, 2010). Currently, screening biomarkers are of high interest due to their ability to predict future events, but there are only a few accepted biomarkers for disease screening available today (Melander *et al.*, 2009). Therefore it is necessary to have considerable search, verification, biological and

biochemical interpretation, and independent validation of disease biomarkers, which requires advancement in high-throughput technologies. To achieve this goal there is necessity to have interdisciplinary expertise, which requires the teamwork of clinicians, biologists, biochemists, and bioinformaticians to carry out biomarker cohort studies with professional planning, implementation, and control. Bioinformatics plays a key role in the biomarker discovery process via bridging the gap between initial discovery phases such as experimental design, clinical study execution, and bioanalytics, including sample preparation, separation, high-throughput profiling, and independent validation of identified candidate biomarkers (Baumgartner *et al.*, 2011) (**Fig. 4**).

It is well known that a disease or a phenotype is rarely a consequence of an abnormality of a single gene or its expression but instead reflects the interactions of various processes in a complex network; such network could combine multiple genes (proteins) and metabolites. For example, plants can produce numerous metabolites to handle different environmental conditions however the biosynthetic pathways for most of these compounds have not yet been revealed (Schlapfer *et al.*, 2017). From these facts, the need for a disease signature, a set of compounds generally presented as a network, becomes evident. Such disease-specific signature could be helpful in understanding all its mechanisms and evolution and in an earlier diagnosis. Multiple works and frameworks aimed to either use genes expression or metabolic data to extract a set of disease-relevant compounds; it's safe to say biomarkers (Strimbu and Tavel, 2010). Identifying such crucial biomarkers responsible for disease characteristics and revealing its mechanisms can be used to infer its evolution and development and offers better targets for drug development, treatment individualization, and dose regimen. These biomarkers are selected by analytic methods or pathway and network-centric methods (Wang *et al.*, 2015). A typical case would be gene expression data of two pairs of samples in both disease and normal states help in discovering genes and metabolites which can be potential biomarkers (Li *et al.*, 2013; Shlomi, 2010; Li *et al.*, 2012). Cancer is a heterogeneous disease, for instance, breast cancer. Biomarkers at the DNA, RNA, and protein levels were developed to better understand the biology of breast cancer, leading to the possibility to classify the disease into subtypes and subgroups, which may lead to new therapeutic opportunities (Le Du *et al.*, 2013). Many tests are available for the diagnosis and each one is based on a set of genes; we can list some of most known ones such as PAM50. PAM50 stands for Prediction Analysis of Microarray 50 (Sweeney *et al.*, 2014), fifty genes were probed like *ACTR3B*, *ANLN*, *BAG1*, *BCL2*, *BIRC5*, etc. Elaborating a set of biomarkers allowed the development of many tools online (cbioportal (Gao *et al.*, 2013)) and offline, in libraries and packages. These tools offer the possibility of analyzing targeted datasets and understanding the disease stage. Machine learning tools are also a very effective way of dealing with such complex diseases, predicting treatment effectiveness and new treated disease trajectory. Since these, new models have been well designed, which fully explains the relationship between each biomarker.

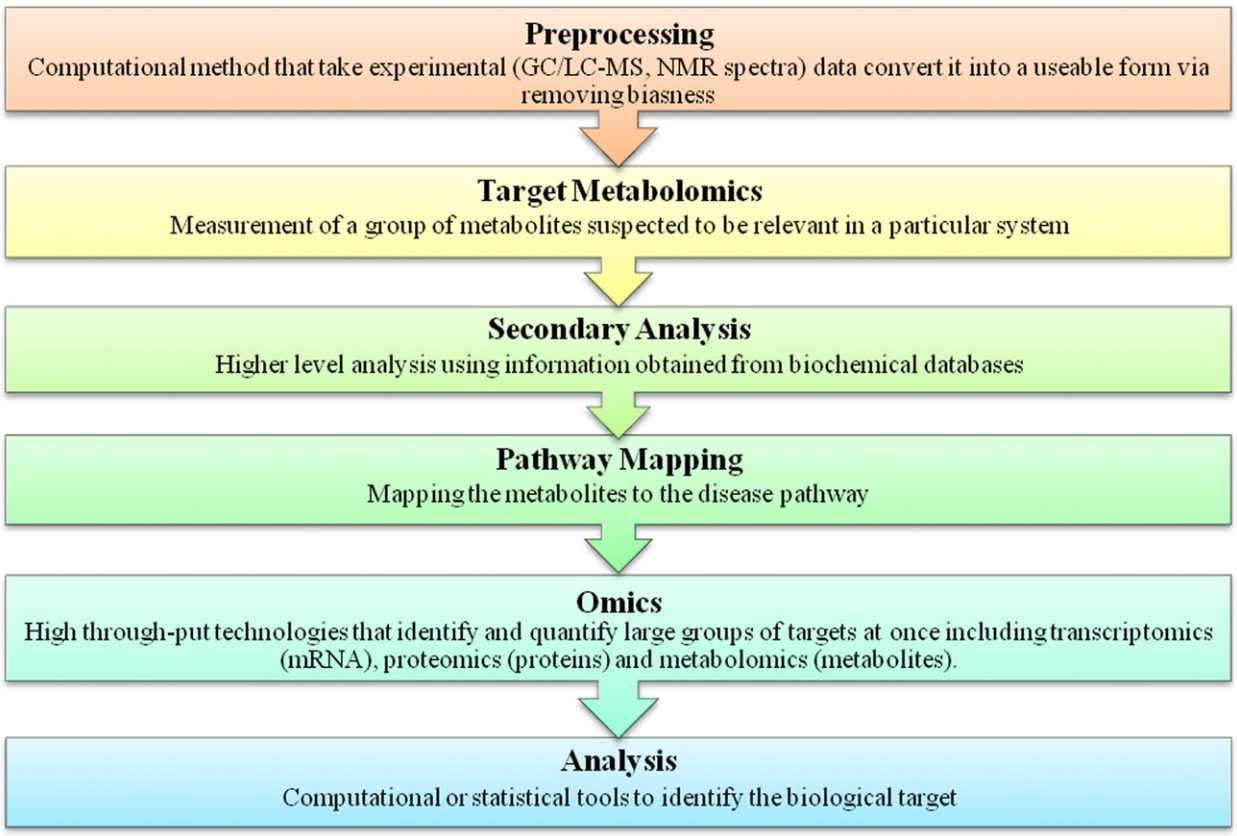


Fig. 4 Computational pipeline for metabolome analysis.

Gene Expression and Metabolic Network: Applications in Biomarker Discovery for Complex Diseases

DNA microarray technologies permit systematic approaches to the biological discovery that have a profound impact on biological research, pharmacology, and medicine (Yousef *et al.*, 2014). The ability to obtain quantitative information from the complete transcription profile of cells provides a powerful means to explore basic biology, disease diagnosis, drug development, mold therapeutics to specific pathologies, and generate databases (Young, 2000). Gene expression studies bridge the gap between DNA information and its trait information by dissecting the biochemical pathways into intermediate components, that is, between genotype and phenotype (Xiong *et al.*, 2001). The gene expression studies, therefore, open new avenues for identifying complex disease genes and biomarkers for disease diagnosis and also for assessing drug efficacy and toxicity. One particularly powerful application of gene expression analyses is in biomarker discovery, which can be used for disease risk assessment, early detection, prognosis, predicting response to therapy, and preventive measures.

For years, scientists studied one gene at a time and genes were indeed studied in isolation from the larger context of other involved genes. Nowadays, genomics via high-throughput techniques helps to study the genome of organisms as a whole thus allowing a wide picture of gene characteristics. One of the most popular high-throughput techniques are arrays, which are an orderly arrangement of a large number of samples allowing large-scale studies (Yousef *et al.*, 2014). This gave rise to the genomic era, which emerged from the sequencing of genomes from many organisms. The development of the first arrays started many years ago to study a large number of genes at a time (Hergenhahn *et al.*, 2003) and has widely expanded since then. Today the approach is also applicable to RNA probes, proteins, antibodies, and even biological samples allowing new types of research (Yousef *et al.*, 2014). Currently, other types of high-throughput techniques are also developing, for instance, to study the transcripts and metabolites.

Today, genomics has induced two new paradigms in biology; the first paradigm is a new approach that allows the study of the complex network through which genes and proteins communicate. It is attained via an amalgamation of the researcher having expertise in the field of biology, engineering, chemistry, and computer science; this multidisciplinary approach allows the development of systems biology. The second paradigm is a direct consequence of information derived from genomics studies where raw data needs to be analyzed and then to be used in the systemic approach. This led the development of bioinformatics,

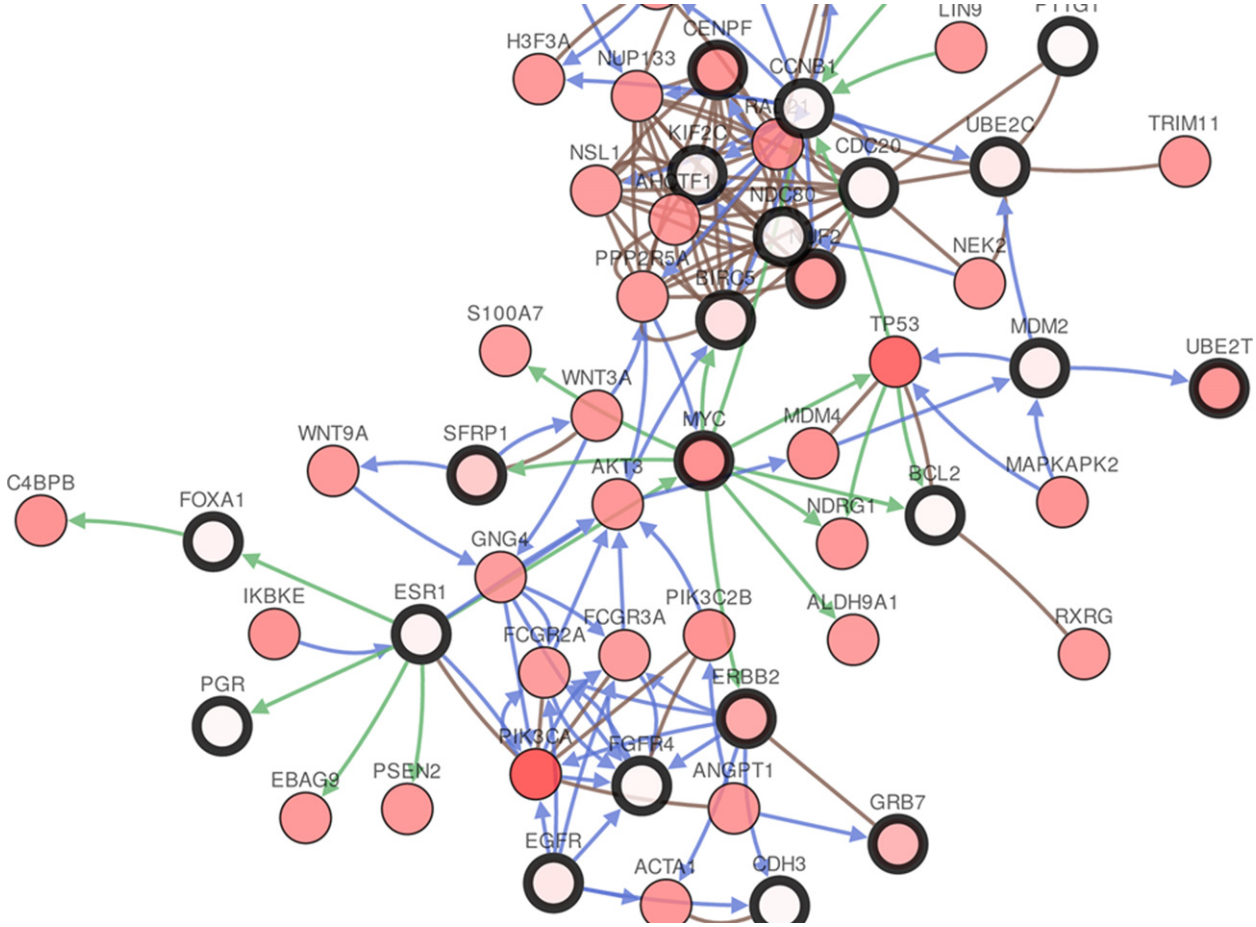


Fig. 5 Network interaction of the PAM50 genes set.

which requires the use of computers to manage biological information. The practical applications of gene expression analyses are numerous and only beginning to be realized. One particularly powerful application of gene expression analyses is in biomarker identification, which can be used for disease risk assessment, early detection, prognosis, prediction response to therapy, and preventative measures.

Therefore approaches to cancer biomarker discovery comprise genomics, epigenomics, transcriptomics, and proteomic analyses. Current efforts in the laboratory focus on the identification of biomarkers in chronic lymphocytic leukemia, lung cancer, and colon cancer (McDermott *et al.*, 2013). Along with the mRNA other small RNAs are also known to be a predictive indicator in disease studies; and it has been found that alterations in gene expression patterns due to dysregulation of miRNAs is a common cause in tumorigenesis (Chen *et al.*, 2012). High concentrations of cell-free miRNAs that originate from the primary tumor have been found in the plasma of cancer patients, and several lines of evidence indicate that plasma miRNAs are associated with specific vesicles called exosomes (Yang *et al.*, 2016). This led to the discovery of new biomarkers that comprises plasma miRNAs and seems to be promising in disease prognosis (Jeffrey, 2008). Recent discovery of quantifiable circulating cancer-associated miRNAs exposes the immense potential of their use as novel minimally invasive biomarkers for breast and other cancers (Heneghan *et al.*, 2009).

Discovery of Biomarkers Through Computational Pipeline: A Cancer Based Study

The classification of samples from gene expression datasets usually involves a small number of samples. The problem of selecting those biomarker genes that are vital for differentiating the different sample classes being compared poses a challenging problem in

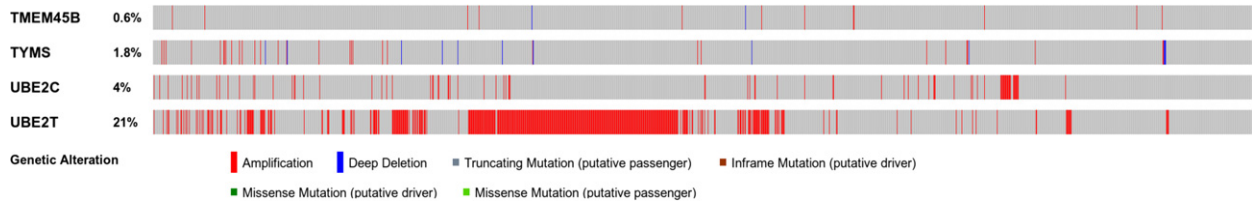


Fig. 6 Oncoprint for only four biomarkers.

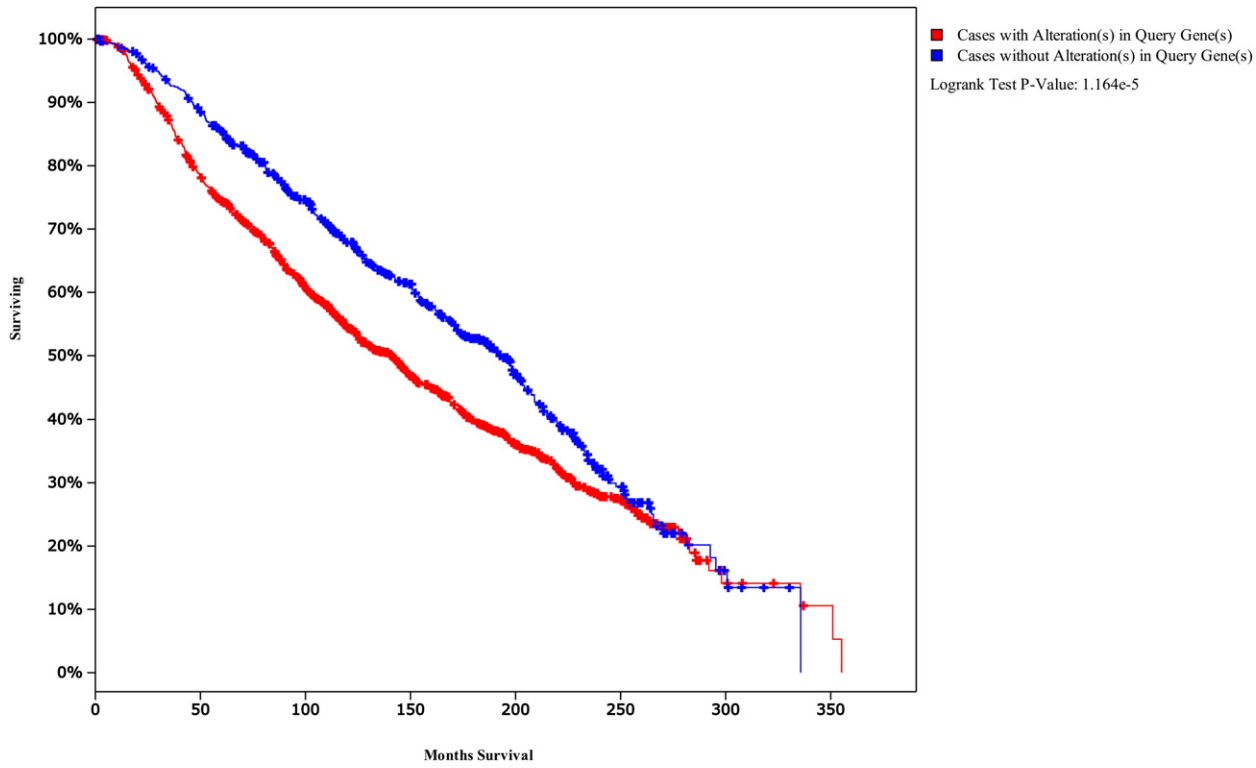


Fig. 7 Survival study.

case of the high dimensional data analysis. A variety of methods to address this problem have been implemented and these methods can be divided into two main categories: (1) Filtering based methods and (2) model-based or wrapper approaches (Wang *et al.*, 2005; Inza *et al.*, 2004). The filter method (Ben-Bassat, 1982) assesses the goodness of the proposed feature subset looking only at the intrinsic characteristics of the data, based on the relation of each single gene with the class label by the calculation of simple statistics computed from the empirical distribution. This approach is extensively used as a feature subset selection method in the microarray (Aris and Recce, 2002). While in the wrapper approach (Kohavi and John, 1997), which is a very powerful machine learning application, search is conducted in the space of genes, evaluating the goodness of each found gene subset by the estimation of the accuracy percentage of the specific classifier to be used.

One of the most important properties that should be considered for the biomarkers is the robustness. It is an important issue for the successful discovery of biomarkers, as it may greatly influence subsequent biological validations. In addition, a more robust set of markers may strengthen the confidence of an expert in the results of a selection method (Abeel *et al.*, 2010). Robustness, a property that allows a system to maintain its functions under certain external and internal perturbations, is a ubiquitously observed feature of biological systems (Kitano, 2004). Studying the relationship between the topology and robustness of

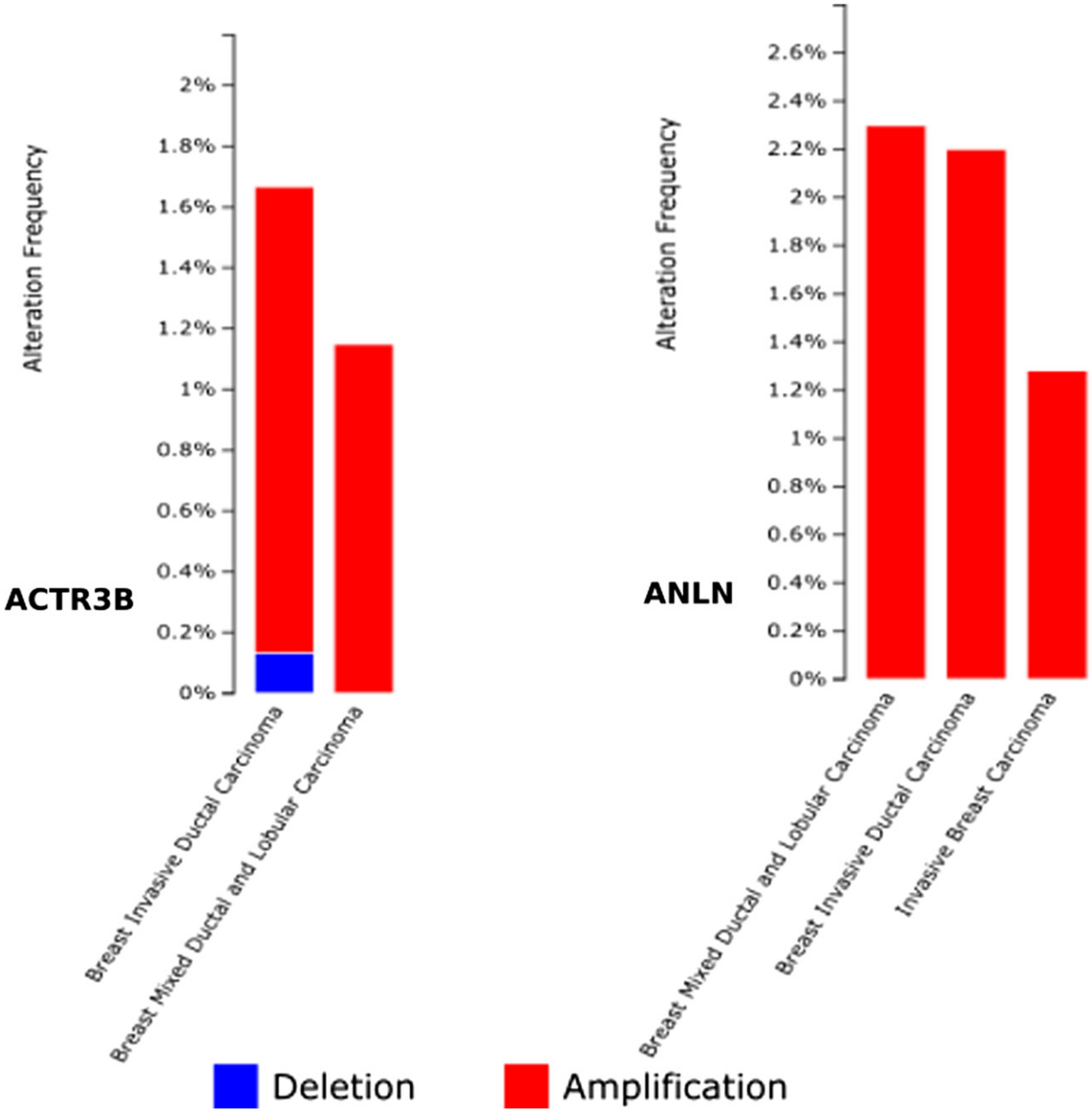


Fig. 8 Breast cancer histogram with subtypes.

metabolic networks may help to understand the functional organization principle of cells and could have important implications for disease studies (Schuster and Holzhütter, 1995) and drug target identifications (Bakker *et al.*, 2000).

The robustness of the metabolic network with respect to specific enzymes can be qualitatively estimated by the preservation or decay of the network after removal of the nodes or edges corresponding to these enzymes. The following topological features of metabolic networks may ensure the robustness of metabolism; these are (1) modularity, which contributes to the robustness of networks by decreasing the cross talk between different functional modules, detaining perturbations and damages to separable parts and preventing deleterious effects from spreading to the whole system (Stelling *et al.*, 2004); (2) bow-tie structure, which aids in forming a robust conserved core because it is the most tightly connected part of the network and there are multiple routes between any pair of nodes. Such connecting patterns provide an advantage in generating a coordinated response to various stimuli and increases the robustness of the whole system (Kitano, 2004); and (3) scale-free topology; the key feature of scale-free networks is the high-degree of error tolerance; that is, the ability of their nodes to communicate is unaffected by the failure of some randomly chosen nodes (Crucitti *et al.*, 2004). Thus the scale-free nature of metabolic networks indicates its high resistance towards random perturbations and thus could explain why some enzyme dysfunction at the metabolic level is without substantial phenotypic effect (Barabasi and Albert, 1999). The studies have shown that the scale-free networks are extremely vulnerable to attacks, i.e., the removal of a few hub nodes that play a crucial role in maintaining a network's connectivity will destroy the whole network (Crucitti *et al.*, 2004). Studies conducted by Mahadevan and Palsson showed that low-degree nodes are almost as likely to be critical to the overall network functions as high-degree nodes (Mahadevan and Palsson, 2005) by calculating the number of lethal reactions among all the reactions connected to every metabolite in the substrate graph of metabolic networks.

To show an example of analysis using biomarkers a dataset from (Pereira *et al.*, 2016) was used. It presents a somatic mutation profiling study of 2433 breast cancers, which approved the classification of the tumors into 10 integrative clusters (IntClusts).

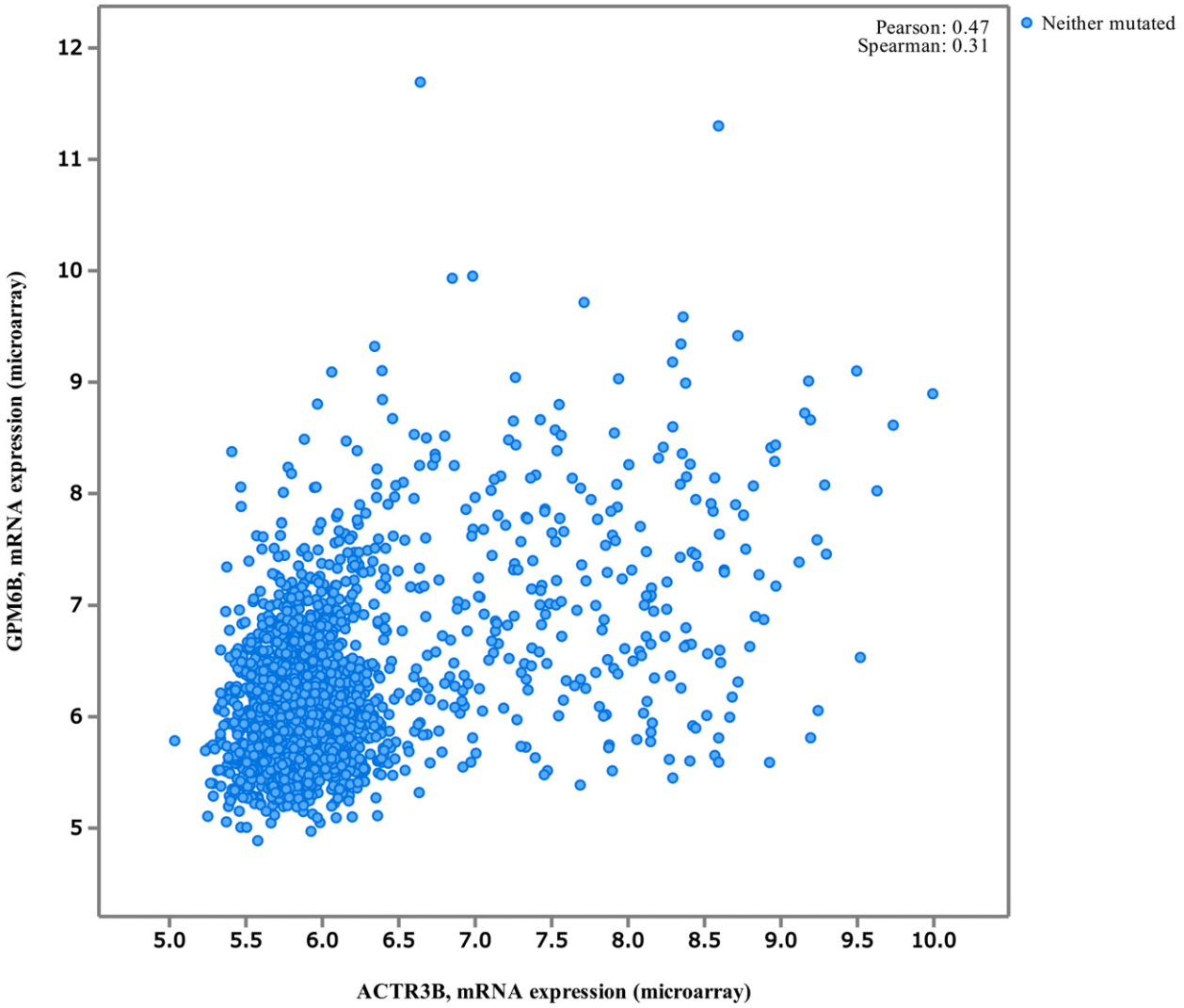


Fig. 9 Coexpression of two biomarkers, ACTR3B & GPM6B.

During this study (Pereira *et al.*, 2016) CNAs (copy number aberrations) were the main driver in an unsupervised clustering approach due to their influence on genes expression. The current analysis used the dataset above against the fifty genes found in PAM50 (**Table 2**). These genes served as a gene set for the cbioportal, so the mutations profiling could be performed. Figures below show different knowledges extracted, cancer types, and mutation happening along genes, genes with similar or exclusive expression, and finally survival estimation. Gene networks present an excellent input for models analysis using algorithmic methods such as Bayesian networks, leading to inferring and predicting the cancer evolution, trajectory, and progression (Caravagna *et al.*, 2016). **Fig. 5** shows the interaction of the PAM50 gene set where nodes are representing genes and edges (arrows) denote the interactions between them.

Thereby an oncoprint of four biomarkers (TMEM45B, TYMS, UBE2C, UBE2T) has been captured (**Fig. 6**) that shows the type of mutation occurring to the gene set.

After this, a survival study has been conducted (**Fig. 7**), which is performed for the p-value of 1.164e-5 and red color denotes the alteration in query gene while blue color shows the query gene without alteration. It has been noticed that the cases with alterations have higher survival rate than the nonaltered ones.

Then alteration frequency of ACTR3B and ANLN were determined through the histogram (**Fig. 8**) that denotes the alteration and the deletion frequency. Finally Pearson and Spearsman correlation coefficient induced shows the coexpression of the two biomarkers, that is, ACTR3B & GPM6B (**Fig. 9**).

Concluding Remarks

The advent of high-throughput technologies, and as a result, the generation of various kinds of omics data, has challenged both the experimental or computational scientific communities. Therefore an "omics cascade" came into the scenario where the genotype to the phenotype can be connected through various intermediate steps for better understanding the disease biology and their overall impact on the phenotypic observations at the physical or organism level. Understanding the genotype to phenotype scenario and its consequences will provide an edge towards better implementation of computational methods for their experimental validations. It will be an added advantage for the scientific community to amalgamate the computational and experimental procedures for the betterment of mankind. It is anticipated that this comprehensive text will provide systematic and ordered information on the metabolic world with its exact connection to the biomarker identification procedure.

See also: Biological Database Searching. Identification and Extraction of Biomarker Information. Integrative Analysis of Multi-Omics Data. Natural Language Processing Approaches in Bioinformatics

References

Abeel, T., Helleputte, T., Van De Peer, Y., Dupont, P., Saeys, Y., 2010. Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. Bioinformatics 26, 392–398.

Antonov, A.V., Dietmann, S., Wong, P., Mewes, H.W., 2009. TICL – A web tool for network-based interpretation of compound lists inferred by high-throughput metabolomics. The FEBS Journal 276, 2084–2094.

Aris, V., Recce, M., 2002. A method to improve detection of disease using selectively expressed genes in microarray data. In: Proceedings of CAMDA'00, pp. 69–81. (S.M. Lin and K.F. Johnson, editors).

Bakker, B.M., Mensonides, F.I., Teusink, B., *et al.*, 2000. Compartmentation protects trypanosomes from the dangerous design of glycolysis. Proceedings of the National Academy of Sciences 97, 2087–2092.

Barabasi, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286, 509–512.

Barabasi, A.L., Oltvai, Z.N., 2004. Network biology: Understanding the cell's functional organization. Nature Review Genetics 5, 101–113.

Bartel, J., Krumsiek, J., Theis, F.J., 2013. Statistical methods for the analysis of high-throughput metabolomics data. Computational and Structural Biotechnology Journal 4, e201301009.

Barupal, D.K., Haldiya, P.K., Wohlgemuth, G., *et al.*, 2012. MetaMapp: Mapping and visualizing metabolomic data by integrating information from biochemical pathways and chemical and mass spectral similarity. BMC Bioinformatics 13, 99.

Bates, J.T., Chivian, D., Arkin, A.P., 2011. GLAMM: Genome-Linked Application for Metabolic Maps. Nucleic Acids Research 39, W400–W405.

Baumgartner, C., Osl, M., Netzer, M., Baumgartner, D., 2011. Bioinformatic-driven search for metabolic biomarkers in disease. Journal of Clinical Bioinformatics 1, 2.

Beber, M.E., Fretter, C., Jain, S., *et al.*, 2012. Artefacts in statistical analyses of network motifs: General framework and application to metabolic networks. Journal of the Royal Society Interface 9, 3426–3435.

Ben-Bassat, M., 1982. Pattern recognition and reduction of dimensionality. Handbook of Statistics 2, 773–910.

Berkhout, J., Teusink, B., Bruggeman, F.J., 2013. Gene network requirements for regulation of metabolic gene expression to a desired state. Scientific Reports 3, 1417.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.-U., 2006. Complex networks: Structure and dynamics. Physics Reports 424, 175–308.

Booth, S.C., Weljie, A.M., Turner, R.J., 2013. Computational tools for the secondary analysis of metabolomics experiments. Computational and Structural Biotechnology Journal 4, e201301003.

Bourqui, R., Cottret, L., Lacroix, V., *et al.*, 2007. Metabolic network visualization eliminating node redundance and preserving metabolic pathways. BMC Systems Biology 1, 29.

Cakir, T., Khatibipour, M.J., 2014. Metabolic network discovery by top-down and bottom-up approaches and paths for reconciliation. Frontiers in Bioengineering and Biotechnology 2, 62.

Caravagna, G., Graudenzi, A., Ramazzotti, D., *et al.*, 2016. Algorithmic methods to infer the evolutionary trajectories in cancer progression. Proceedings of the National Academy of Sciences of the United States of America 113, E4025–E4034.

Chen, P.S., Su, J.L., Hung, M.C., 2012. Dysregulation of microRNAs in cancer. Journal of Biomedical Science 19, 90.

Cho, D.Y., Kim, Y.A., Przytycka, T.M., 2012. Chapter 5: Network biology approach to complex diseases. PLOS Computational Biology 8, e1002820.
Cooper, G.M., 2000. The central role of enzymes as biological catalysts. Sinauer Associates.
Crucitti, P., Latora, V., Marchiori, M., Rapisarda, A., 2004. Error and attack tolerance of complex networks. Physica A: Statistical Mechanics and its Applications 340, 388–394.
DeBerardinis, R.J., Thompson, C.B., 2012. Cellular metabolism and disease: What do metabolic outliers teach us? Cell 148, 1132–1144.
Degtyarenko, K., De Matos, P., Ennis, M., *et al.*, 2007. ChEBI: A database and ontology for chemical entities of biological interest. Nucleic Acids Research 36, D344–D350.
Devoid, S., Overbeek, R., Dejongh, M., *et al.*, 2013. Automated genome annotation and metabolic model reconstruction in the SEED and Model SEED. Systems Metabolic Engineering: Methods and Protocols. 17–45.
Downing, G., 2001. Biomarkers definitions working group. Biomarkers and surrogate endpoints. Clinical Pharmacology & Therapeutics 69, 89–95.
Droste, P., Nöh, K., Wiechert, W., 2013. Omix – A visualization tool for metabolic networks with highest usability and customizability in focus. Chemie Ingenieur Technik 85, 849–862.
Gao, J., Aksoy, B.A., Dogrusoz, U., *et al.*, 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science Signaling 6, pl1.
Heneghan, H., Miller, N., Lowery, A., Sweeney, K., Kerin, M., 2009. MicroRNAs as novel biomarkers for breast cancer. Journal of Oncology 2009.
Hergenhahn, M., Muhlemann, K., Hollstein, M., Kenzelmann, M., 2003. DNA microarrays: Perspectives for hypothesis-driven transcriptome research and for clinical applications. Current Genomics 4, 543–555.
Horai, H., Aranita, M., Nishioka, T., 2008. MassBank: Mass spectral database for metabolome analysis. In: Proceedings of the 56th ASMS Conference on Mass Spectrometry and Allied Topics, Denver, CO.
Hucka, M., Finney, A., Sauro, H.M., *et al.*, 2003. The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. Bioinformatics 19, 524–531.
Hummel, J., Selbig, J., Walther, D., Kopka, J., 2007. The Golm Metabolome Database: A database for GC–MS based metabolite profiling. Metabolomics. 75–95.
Inza, I., Larranaga, P., Blanco, R., Cerrolaza, A.J., 2004. Filter versus wrapper gene selection approaches in DNA microarray domains. Artificial Intelligence in Medicine 31, 91–103.
Jeffrey, S.S., 2008. Cancer biomarker profiling with microRNAs. Nature Biotechnology 26, 400–401.
Jeong, H., Mason, S.P., Barabasi, A.L., Oltvai, Z.N., 2001. Lethality and centrality in protein networks. Nature 411, 41–42.
Joshi-Tope, G., Gillespie, M., Vastrik, I., *et al.*, 2005. Reactome: A knowledgebase of biological pathways. Nucleic Acids Research 33, D428–D432.
Kanehisa, M., Goto, S., 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research 28, 27–30.
Kankainen, M., Gopalacharyulu, P., Holm, L., Orešič, M., 2011. MPEA – Metabolite pathway enrichment analysis. Bioinformatics 27, 1878–1879.
Karp, P.D., Riley, M., Paley, S.M., Pellegrini-Toole, A., 2002. The metacyc database. Nucleic Acids Research 30, 59–61.
Karp, P.D., Weaver, D., Paley, S., *et al.*, 2014. The EcoCyc database. EcoSal Plus 6.
Kitano, H., 2004. Biological robustness. Nature Reviews Genetics 5, 826–837.
Klamt, S., von Kamp, A., 2011. An application programming interface for CellNetAnalyzer. Biosystems 105, 162–168.
Kohavi, R., John, G.H., 1997. Wrappers for feature subset selection. Artificial intelligence 97, 273–324.
Le Du, F., Ueno, N.T., Gonzalez-Angulo, A.M., 2013. Breast Cancer Biomarkers: Utility in clinical practice. Current Breast Cancer Reports 5.
Lee, D.-S., Park, J., Kay, K., *et al.*, 2008. The implications of human metabolic network topology for disease comorbidity. Proceedings of the National Academy of Sciences 105, 9880–9885.
Li, L., Jiang, H., Ching, W.-K., Vassiliadis, V.S., 2012. Metabolite biomarker discovery for metabolic diseases by flux analysis. In: Proceedings of the 2012 IEEE 6th International Conference on Systems Biology (ISB), pp. 1–5. IEEE.
Li, L., Jiang, H., Qiu, Y., Ching, W.K., Vassiliadis, V.S., 2013. Discovery of metabolite biomarkers: Flux analysis and reaction-reaction network approach. BMC Systems Biology 7 (Suppl. 2), S13.
Lindon, J.C., Nicholson, J.K., Holmes, E., 2011. The Handbook of Metabonomics and Metabolomics. Elsevier.
Lipid Maps, LIPID Metabolites And Pathways Strategy, Welcome Trust.
Madu, C.O., Lu, Y., 2010. Novel diagnostic biomarkers for prostate cancer. Journal of Cancer 1, 150–177.
Mahadevan, R., Palsson, B.O., 2005. Properties of metabolic networks: Structure versus function. Biophysical Journal 88, L07–L09.
Maudsley, S., Chadwick, W., Wang, L., *et al.*, 2011. Bioinformatic approaches to metabolic pathways analysis. Methods in Molecular Biology 756, 99–130.
McDermott, J.E., Wang, J., Mitchell, H., *et al.*, 2013. Challenges in biomarker discovery: Combining expert insights with statistical analysis of complex omics data. Expert Opinion on Medical Diagnostics 7, 37–51.
Melander, O., Newton-Cheh, C., Almgren, P., *et al.*, 2009. Novel and conventional biomarkers for prediction of incident cardiovascular events in the community. JAMA 302, 49–57.
Mohney, R., Milburn, M., 2015. Providing insight into complex disease: Metabolomics links genetic loci to phenotype. The FASEB Journal 29, LB292.
Mueller, L.A., Zhang, P., Rhee, S.Y., 2003. AraCyc: A biochemical pathway database for Arabidopsis. Plant Physiology 132, 453–460.
Oliver, S.G., Winson, M.K., Kell, D.B., Baganz, F., 1998. Systematic functional analysis of the yeast genome. Trends in Biotechnology 16, 373–378.
Peleg, M., Rubin, D., Altman, R.B., 2005. Using Petri net tools to study properties and dynamics of biological systems. Journal of the American Medical Informatics Association 12, 181–199.
Pereira, B., Chin, S.-F., Rueda, O.M., *et al.*, 2016. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nature Communications 7.
Rohn, H., Junker, A., Hartmann, A., *et al.*, 2012. VANTED v2: A framework for systems biology applications. BMC Systems Biology 6, 139.
Schellenberger, J., Park, J.O., Conrad, T.M., Palsson, B.O., 2010. BiGG: A biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. BMC Bioinformatics 11, 213.
Schlapfer, P., Zhang, P., Wang, C., *et al.*, 2017. Genome-wide prediction of metabolic enzymes, pathways, and gene clusters in plants. Plant Physiology 173, 2041–2059.
Schuster, R., Holzhütter, H.G., 1995. Use of mathematical models for predicting the metabolic effect of large-scale enzyme activity alterations. The FEBS Journal 229, 403–418.
Sehgal, M., Gupta, R., Moussa, A., Singh, T.R., 2015. An integrative approach for mapping differentially expressed genes and network components using novel parameters to elucidate key regulatory genes in colorectal cancer. PLOS One 10, e0133901.
Shahzad, K., Loor, J.J., 2012. Application of top-down and bottom-up systems approaches in ruminant physiology and metabolism. Current Genomics 13, 379–394.
Shlomi, T., 2010. Metabolic network-based interpretation of gene expression data elucidates human cellular metabolism. Biotechnology & Genetic Engineering Reviews 26, 281–296.
Shukla, A., Singh, T.R., 2017. Computational network approaches and their applications for complex diseases. Translational Bioinformatics and its Application. Springer.
Smith, C.A., O'maille, G., Want, E.J., *et al.*, 2005. METLIN: A metabolite mass spectral database. Therapeutic drug monitoring 27, 747–751.
Sridharan, G.V., Ullah, E., Hassoun, S., Lee, K., 2015. Discovery of substrate cycles in large scale metabolic networks using hierarchical modularity. BMC Systems Biology 9, 5.
Stelling, J., Sauer, U., Szallasi, Z., Doyle, F.J., Doyle, J., 2004. Robustness of cellular functions. Cell 118, 675–685.
Strimbu, K., Tavel, J.A., 2010. What are biomarkers? Current Opinion in HIV and AIDS 5, 463.
Suhre, K., Schmitt-Kopplin, P., 2008. MassTRIX: Mass translator into pathways. Nucleic Acids Research 36, W481–W484.
Sweeney, C., Bernard, P.S., Factor, R.E., *et al.*, 2014. Intrinsic subtypes from PAM50 gene expression assay in a population-based breast cancer cohort: Differences by age, race, and tumor characteristics. Cancer Epidemiology, Biomarkers and Prevention 23, 714–724.

Vander Heiden, M.G., Cantley, L.C., Thompson, C.B., 2009. Understanding the Warburg effect: The metabolic requirements of cell proliferation. Science 324, 1029–1033.
Veeramani, B., Bader, J.S., 2010. Predicting functional associations from metabolism using bi-partite network algorithms. BMC Systems Biology 4, 95.
Villeger, A.C., Pettifer, S.R., Kell, D.B., 2010. Arcadia: A visualization tool for metabolic pathways. Bioinformatics, 26. . pp. 1470–1471.
Wang, J., Zuo, Y., Man, Y.G., *et al*., 2015. Pathway and network approaches for identification of cancer signature markers from omics data. Journal of Cancer 6, 54–65.
Wang, Y., Tetko, I.V., Hall, M.A., *et al*., 2005. Gene selection from microarray data for cancer classification – A machine learning approach. Computational Biology and Chemistry 29, 37–46.
Wang, Y., Xiao, J., Suzek, T.O., *et al*., 2009. PubChem: A public information system for analyzing bioactivities of small molecules. Nucleic Acids Research 37, W623–W633.
Wishart, D.S., Tzur, D., Knox, C., *et al*., 2007. HMDB: The human metabolome database. Nucleic Acids Research 35, D521–D526.
Xia, J., Wishart, D.S., 2010a. MetPA: A web-based metabolomics tool for pathway analysis and visualization. Bioinformatics 26, 2342–2344.
Xia, J., Wishart, D.S., 2010b. MSEA: A web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. Nucleic Acids Research 38, W71–W77.
Xiong, M., Fang, X., Zhao, J., 2001. Biomarker identification by feature wrappers. Genome Research 11, 1878–1887.
Xu, C., Hu, S., Chen, X., 2016. Artificial cells: From basic science to applications. Material Today (Kidlington) 19, 516–532.
Yang, Q., Diamond, M.P., Al-Hendy, A., 2016. The emerging role of extracellular vesicle-derived miRNAs: Implication in cancer progression and stem cell related diseases. Journal of Clinical Epigenetics 2.
Young, R.A., 2000. Biomedical discovery with DNA arrays. Cell 102, 9–15.
Yousef, M., Najami, N., Abedallah, L., Khalifa, W., 2014. Computational approaches for biomarker discovery. Journal of Intelligent Learning Systems and Applications 6, 153.
Zhang, A., Sun, H., Wang, P., Han, Y., Wang, X., 2012. Modern analytical techniques in metabolomics analysis. Analyst 137, 293–300.
Zhu, X., Gerstein, M., Snyder, M., 2007. Getting connected: Analysis and principles of biological networks. Genes & Development 21, 1010–1024.