OPEN ACCESS

International Journal of Genomics and Data Mining

Srivastava N and Srivastava P. Int J Genom Data Min 02: 125.

Review Article

DOI: 10.29011/2577-0616.000125

System Biology and Network-Based Computational Model Approaches in Biomarker Discovery in Reference to Neurological Disorder

Neha Srivastava[1,2], **Prachi Srivastava**[2*]

[1]AMITY Institute of Biotechnology, AMITY University, Uttar Pradesh, Lucknow, India

[2]Dr. A.P.J. Abdul Kalam Technical University, Lucknow, UP, India

[*]**Corresponding author:** Prachi Srivastava, AMITY Institute of Biotechnology, AMITY University, Uttar Pradesh, Lucknow-226028, India. Tel: +918601909444; Email: psrivastava@amity.edu

Citation: Srivastava N, Srivastava P (2018) System Biology and Network-Based Computational Model Approaches in Biomarker Discovery in Reference to Neurological Disorder. Int J Genom Data Min 02: 125. DOI: 10.29011/2577-0616.000125

Received Date: 18 June, 2018; **Accepted Date:** 25 June, 2018; **Published Date:** 02 July, 2018

Abstract

Neurodegenerative diseases are irredeemable and incapacitating conditions that result in progressive degeneration. It is difficult to define the complexity of neuro-system quantitatively or meaningfully from a system standpoint. Thus, inclined towards the progress in developing new and effective therapeutic intervention, it is important to understand the underlying molecular mechanism and significance of neuro system and their complex molecular interaction. A biomarker discovery is an important need for early disease diagnosis, prognosis and monitoring of new therapy for neurological disorders. The emergence of system biology and network-based computational model approaches provides the underlying molecular mechanism and significance of disease and their complex molecular interaction. Thus, it becomes quite easy to understand the specific nature of neuro system as well as it plays a significant role in integrating the omics data at multiple levels that lead to key success in the development of more accurate and efficient biomarker for neurological disorders. The current review focused on significant contributions of system biology and network-based computational model approaches in biomarker discovery with special reference to neurological disorders.

Keywords: Biomarker discovery; Bioinformatics; System biology; Network-based computational model; Neurological disorders

Introduction

Neurodegenerative disease in human includes the wide range of complex disorders including -Alzheimer Disease (AD), Parkinson Disease (PD), Motor Neuron Disease (MND), and Huntington's Disease (HD) etc. Among these, many of them share common symptom & neuro-pathological condition. Thus, it becomes quite challenging to diagnosis the particular disorder [1]. Therefore, it is essential to understand the molecular mechanism and pathological symptom of disease in order to predict the disease. Biomarker discovery provides a powerful and progressive approach to the mechanistic insight of neurological disorders along with the same level of specificity and sensitivity for evaluation and diagnosis of disease and also used to identify and develops potential drug target as well as a novel compound for the treatment of disease [2]. Identification of novel biomarker will play key success in diagnosis and therapeutic intervention for the wide class of neurological disorders including neurodegenerative and neurodevelopmental disease, from which millions of people are affected around the world every year [3].

Biomarkers are specific pharmacological and physiological biochemical measurement in the body that presence is used for measuring the progress and diagnosis of disease and also in monitoring the treatment [4]. Biomarker acts as an indicator for normal biological as well as pathogenic processes. They are used to indicate presence or onset of disease. Biomarkers discovery to aid accurate diagnosis, predict progression and for use in clinical trials has become a major research need in the current scenario [5]. With the recent advances in the system biology approaches, help to computationally assimilate omics data with network and pathway to understand new underlying biological mechanisms at the system

level that leads to biomarker discovery and its clinical validation. The application of system biology in neurological disorder helps collect information regarding structure and function of neuro system in normal vs. diseases state at different stages in the brain by accommodating omics data including genomics, proteomics, metabolomics & transcriptomics at a different system level. This will provide deeper insight mechanism of the complex feature of neuro system that caused by various factors and changes in the biological system. Thus, the current review focused on system biology and network-based computational model approaches in novel biomarker discovery associated with neurological disorder based on data and knowledge-driven approaches.

Approaches Involved in In-Silico Biomarker Discovery

Omics Data Analysis

The comprehensive analysis of omics data aims to identify significant biological process and pathway from the large dataset and also to find out key gene/protein/ metabolites as a target candidate biomarker. It provides a genome-wide molecular basis for diseases and used to identify a disease-specific biomarker for diagnosis and monitoring of diseases [6]. The omics studies produced from a large amount of high-throughput experimental data is often becoming difficult to interpreted result. With the recent advances in the bioinformatics and the system biology, offers opportunities to interpret data from existing knowledgebase approaches in order to understand the whole mechanism of biological process and disease at the system level. Hence the integration of omics data derived from disease-affected cell and tissue provides the molecular basis for identification of network-based novel biomarker and drug target [7]. Various public repository databases such as the Gene Expression Omnibus (GEO) repository (http://www.ncbi.nlm.nih.gov/geo) and the Array Express archive (http://www.ebi.ac.uk/microarray-as/ae) are available for depositing human disease-oriented omics data. These databases include biological significant information regarding disease network and biomarker which can be analyzed through bioinformatics approaches followed by experimental validation [8].

Gene Expression Analysis

Gene expression analysis is an inventive approach that provides the information regarding the role of the differential expressed gene in the normal biological process and disease state. It compares the expression level of genes in two or more sample. With the use of DNA microarray technology, it's become more convenient to monitor the genome-wide expression pattern of the gene in disease-affected tissue and cell [9]. DNA microarray analysis of omics data can be analyzed by the various representing method such as R-statistical computing program. R statistical programming language is an open source developmental tool for analysis of high-throughput genomic data. It contain the utility for pre-processing Affymetrix, identifying Differentially Expressed Genes (DEGs); followed by statistical analysis using the t-test for comparison between two group or Analysis of Variance (ANOVA test) for comparison between more than three group followed by Multiple Comparison Test (MCT), controlling False Discovery Rate (FDR) and hierarchical clustering analysis. Various R-software packages are available such as Bioconductor 3.5 (htttp://www.bioconductor.com) [10], GENESPRING software (http//www.agilent.com) [11], The Comprehensive R Archive Network (CRAN) (http://cran.r-project.org) to carry out gene expression analysis in silico.

Network Analysis

Network analysis provides the key approach to high thought put data interpretation. It analyzes functionally related genes and networks that are common and biological relevance in response to biomarker discovery from large data scale [12]. The emerging role of system biology approaches provides researcher to link diverse data into knowledgebase to understand the insight mechanism of disease (Figure 1).

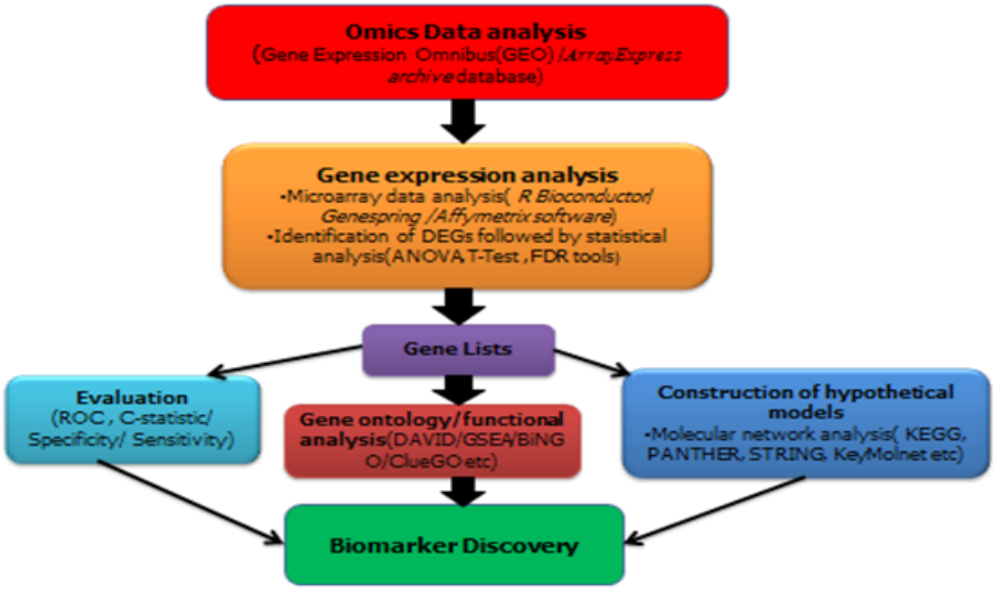

Fig. A road map to *in-silico* biomarker discovery: Form high thought put experimental omics data analysis to molecular network analysis and *in-silico* validation

Figure 1: A road map to *in-silico* biomarker discovery: Form high thought put experimental omics data analysis to molecular network analysis and *in-silico* validation.

Network analysis can help to understand the underlying mechanism of molecular and cellular interaction between genes/ proteins with the surrounding environment. In the network, entities represent nodes i.e. gene, protein & enzyme and edges represent the biologically significant interaction between the two nodes based on experimental, database, text mining, co-expression data etc. A network can be constructed on the basis of i) Gene co-expression data - two genes are similar if their expression level is same throughout the gene expression study. ii) Pathway data - two-gene product are linked if they participated in the same reaction in the pathway. iii) Physical interaction or Protein-Protein Interaction Network (PPI) - the two-gene product is linked if they are found to interact in PPI network [13]. Various pathway analysis tools and databases are available publicly for analysis (Table1).

Databases & Tools	URL	Uses
KEGG (Kyoto Encyclopedia of Genes and Genomes)	HTTP:// www.kegg.jp	KEGG systematically integrates genomic and chemical information to create the whole biological system in silico [14].
PANTHER (Protein Analysis Through Evolutionary Relationships classification system)	http://www.pantherdb.org	It builds the relationship between the evolution of protein sequences to the evolution of protein functions and biological roles, provides a structured representation of protein function in the context of biological reaction networks [15].
STRING (Search Tool for the Retrieval of Interacting Genes Proteins)	http://www.string.embl.de	It is a database of functional associations that derived from a wide range of sources such as high-throughput experimental data, literature and database mining, analyses of co-expressed genes and computational predictions [16].
Gene MANIA	http//www.genemania.org	It is a flexible, user-friendly web interface for generating hypotheses about gene function, analyzing gene lists and prioritizing genes for functional assays.

IPA (Ingenuity Pathways Analysis)	http://www.ingenuity.com	It is a web-based functional analysis tool for comprehensive omics data. Use to Identify the most relevant signaling and metabolic pathways, molecular networks, and biological functions for the list of genes [17].
KeyMolnet (Institute of Medicinal Molecular Design)	http://www.immd.co.jp	KeyMolnet automatically provides network-search algorithm enables us to extract the most relevant molecular network composed of the genes coordinately regulated by putative common upstream transcription factors [18].
Cytoscape	http://www.cytoscape.org	It is an open source software project for integrating biomolecular interaction networks with high-throughput expression data and other molecular states into a unified conceptual framework [19].
Reactome	http://www.reactome.org	Pathway database which provides intuitive bioinformatics tools for the visualization, interpretation, and analysis of pathway knowledge.
Pathway Analyser	http://sourceforge.net/ projects/ pathwayanalyser/	Pathway Analyser (PA) is a tool for systems biologists for analysis of metabolic pathways, particularly by Flux Balance Analysis and ODE simulation.
WikiPathways	http://wikipathways.org	WikiPathways is a community resource for contributing and maintaining content dedicated to biological pathways.

Table 1: Tools & resources of System Biology available for Pathway & Network analysis.

Structural analysis of complex network can be performed based on the various topological parameter such betweenness centrality (shortest paths between all nodes through which a given node pass) and node degree (number of edges connected to the node), the degree distribution (probability of a node having a specific number of edges) the clustering coefficient (the degree to which nodes within a network cluster together), shortest path length (minimal distance, in number of edges, required to connect two nodes), robustness, etc. These biological network parameters provide useful information about the response of the whole system under study [20]. Network analysis also includes identification of clusters in the network (densely connected nodes in the network) and enrichment analysis. Cytoscape (http://www.cytoscape.org) is an open source, software used for biological network visualization, data integration, and interactive network generation. It also includes various plugins that perform network analysis from different data sources based on advance topological parameters. Hence, the construction and analysis of network help to understand the biological mechanism at the system level that leads to playing important role in biomarker discovery.

Functional Enrichment Analysis

Functional analysis is used to identify set of an enriched gene with significant function in entire candidate gene list derived from network analysis. Serval tools & software are available for analysis among them some widely used are Gene Ontology (GO [21]; http://www.geneontology.org), provides core biological knowledge representation for modern biologists, based computationally or experimentally. It represents the gene and its product in term of their biological process, cellular process, and metabolic process. The Database for Annotation, Visualization and Integrated Discovery (DAVID [22]; http:// david.ncifcrf.gov) provides functional enrichment analysis, functional annotation, clustering, bio-Carta and keg pathway mapping, identifying functionally related genes that provide biological significant function derived from the large dataset. Serval Cytoscape plugins such as BiNGO [23], ClueGO [24], Enrichment Analysis and Visualization (ENViz) [25] etc are also available for analysis based on interaction network and topological parameter (Table2).

Tools	URL
Gene Ontology (GO)	http://www.geneontology.org
The Database for Annotation, Visualization and Integrated Discovery (DAVID)	http:// david.ncifcrf.gov
BiNGO (A Biological Network Gene Ontology tool)	http://apps.cytoscape.org/apps/bingo
ClueGO	http://apps.cytoscape.org/apps/cluego
Enrichment analysis and visualization (ENViz)	http://apps.cytoscape.org/apps/enviz

Table2. Publicly available functional enrichment analysis tools

Table 2: Publicly available functional enrichment analysis tools.

Biomarker Evaluation

The efficacy and accuracy of the biomarker is an important

step from the clinical perspective. It provides enlighten as to whether the new biomarker is safe for clinical use to the patient. The major parameter for evaluation includes Area Under the Curve (AUC) of Receiver Operating Characteristics (ROC) curve, sensitivity & specificity. ROC curve analysis is used to analyze that whether the biomarker is capable of selecting between disease onset and healthy individual [26]. It predicts whether an individual will experience even if he/she estimated risk is above the given threshold value c. The c-statistic is the area under the ROC curve [27] that estimates for positive and negative. It includes sensitivity/specificity for all the possible prediction. Default parameter of c-statistic for the diagonal line would be 0.5; perfect discrimination related to c-statistic is 1. Experimental data can be used as a standard parameter for proper evaluation of biomarker. The *In-Silico* data can be cross-validated by the Precision Rate (PPV) - the percentage of Positive Predictive Value (PPV) if disease state is present and Negative Predictive Value (NPV) if disease state is absent (Table 3).

Software packages for Biomarker Discovery	**website**
Simplicity Bio (*In-Silico* Biomarker Discovery)	http://www.simplicitybio.com
FUSION BD (Automated Biomarker Discovery)	http://www.iomics.us/
InSyBio BioNets	https://www.insybio.com/pages/bionets
BRB-Array Tools	https://brb.nci.nih.gov/BRB-ArrayTools/
Affymetrix	https://www.thermofisher.com
Statistical Analysis Software for Biomarker Evaluation	**website**
Commercial Package Stata 8, Stata Corp.	https://www.stata.com/
Decision Curve Analysis, Andrew Vickers	https://www.mskcc.org
Risk Prediction Package, Gary Longton, and Margaret Pepe	https://research.fhcrc.org
GWAS Central	http://www.gwascentral.org

Table 3: List of few software & resources available for evaluation & discovery of Biomarker.

Application of System Biology in Biomarker Discovery in Reference to Neurological Disorder

System biology plays an important role in understanding the complex nature of the disease. Emerging role of system biology and network-based computational model approaches helps to integrate computationally, omics data with pathway and network analysis to find out new biological mechanism that leads to biomarker discovery and their experimental validation. Implementation of system biology approaches in neurological disorder start with integrating omics data from the different system level. This includes genomics, transcriptomics, metabolomics, and proteomics that collect information regarding structure and function of neuro system in normal and diseases state at the different level in the brain. Autism Spectrum Disorder (ASD) is a neurodevelopmental disease that typically appears during early childhood and affect a person's ability to communicate. Etiology of ASD is still not clear. The study investigated the expression profile of serum miRNA from ASD in disease vs. normal state using Taq Man Low-Density Array technology [28]. It reveals upregulated miR-140-3p in ASD vs. normal state. Network functional analysis shows that CD38 and NRIP1 nodes controlled by miR-140-3p, involved in dysregulation in ASD. Further Biomarker analysis proved serum miR-140-3p (ASD vs. normal; Area under ROC curve, AUC: 0.70; sensitivity: 63.33%; specificity: 68%) as a potential biomarker for ASD. In the study, the integrated microarray study and network-based approaches were used to investigates the functional link between Parkinson's Disease (PD) and Type 2 Diabetics Mellitus (T2DM) by comprising 478 genes that are closely associated with PD and T2DM [29].

Their finding reveals seven genes that dysregulate in the blood of PD and T2DM patients. Among them, the gene expression level of APP significantly upregulated in the blood of PD and T2DM patient in comparison to the healthy patient. Thus, the study suggests, the increased level of APP in blood with T2DM patient act as an indicator of neurodegeneration and maybe use as the potential biomarker for PD. A combined study of 2D and Mass Spectrometry (MS) was conducted to screen out protein biomarkers for Traumatic Brain Injury (TBI) in a rat model through proteomic followed by system biology analysis [30]. System biology analysis identified Ubiquitin carboxyl-terminal isozyme 1, tyrosine hydroxylase, and syntaxin-6 as potential biomarker candidate for TBI. Further pathway analysis shows protein take part in neurite outgrowth and cell differentiation. The further result confirmed through semi-quantitative Western blotting analysis compare to control case. In the study [31], develops a network-based model of mutated and differentially expressed disease genes of neurological and psychiatric diseases to validate their association with aging.

Further, the approach was used to identify disease-specific biomarkers for diagnosis and treatment (Table 4).

Disease	Biomarker
Alzheimer disease	Aβ42, Aβ42/Aβ40, BACE1, tau APP, MCP-1, CD14, SOD1, Urine, AD7c, APLP2, NEU2, PCDH11X, SUMF1, TOMM40.
Parkinson Disease	α-synuclein, CALB1, CSF1R, MTCYB,CHAC1, NPFF, DJ-1, 8-OHdG, Urine 8-OHdG.
Huntington disease	Clusterin, AKAP8L, ARFIP2.
Amyotrophic Lateral Sclerosis	ALS2CR8, DERL1, FUS, HOPX, KIF1A, MOBKL2B, NIF3L1, SCN7A, SEMA6A, SLC39A11, STRADB, SUSD1, UNC13A, ZFP64.
Multiple Sclerosis	CSMD1, FCGR1A, FCGR1B, FCGR2A, FCGR2B, FCGR2C, FCGR3A, FCGR3B, IFNB1.
Schizophrenia	AP3B2, HMBS, SETD2, ST6GAL2.
Ataxia Telangiectasia	DDX10, HEPACAM, TCL1A.
Stroke	NINJ2, PROZ.

Table 4: Lists of few specific biomarkers for various neurological disorders discover through OMICS studies.

Limitation

Although system biology plays a key role in finding potential biomarker with good accuracy, it also has a little limitation regarding biomarker prediction. In neurological disorder, the disease is very specific to its pathological condition, thus obtained biomarker should be specific to the particular disorder. Regardless of this, the biomarker obtained from PPI network may have the possibility to be specific to other disease or disorder also. Thus, to overcome such a problem, all the disorder should be included in network and cross-validated to confirm the disease-specific.

Conclusion

The prediction and diagnosis of most of the neurological disorders are still difficult to examine. They only process through serval neurological test or examination. With emerging significance of system biology and network-based computational model approaches, based on interpreting large expression data from omics study and constructing on protein-protein interaction network specific to particular disease or disorder, has provided far most possibility for obtaining a potential novel biomarker through system biology approaches. The role of a biomarker is well established in various other diseases such as cancer, cardiovascular disease etc. Neuro-researcher is now also focusing, their area of interest on biomarker discovery for better prediction, diagnosis & treatment of neurological disorders. Thus, the system biology approaches may play a significant role in understanding the underlying biological and functional mechanism of complex neurobiology of disease at the system level that may lead to the discovery of potential biomarkers for early detection, diagnosis & monitoring of neurological disorders at different stages.

Acknowledgment

The support provided by from all those who directly or indirectly helped us in this works are gratefully acknowledged. The authors have no conflict of interests to disclose.

References

1. Ward & Schofield (2010) Biomarkers for brain disorders. Therapy 4: 321-336.
2. Polvika Jiri, Jiri Polivka, Jr, Kristyna Krakorova, Marek Peterka, Ondrej Topolcan, et al. (2016) Current status of biomarker research in neurology. EPMA J 7: 14.
3. Biomarkers: Potential Uses and Limitations. Richard Mayeux NeuroRx. 2004 Apr; 1(2): 182-188.
4. Gershon, E.S, Goldin, L.R. (1986) Clinical methods in psychiatric genetics. I. Robustness of genetic marker investigative strategies. Acta Psychiatr. Scand 74: 113-118.
5. Lovestone S, Guntert A, Hye A, Lynham S, Thambisetty M, et al. (2007) Proteomics of Alzheimer's disease: understanding mechanisms and seeking biomarkers. Expert Rev Proteomics 4: 227-238.
6. Viswanathan GA, Seto J, Patil S, Nudelman G, Sealfon SC (2008) Getting started in biological pathway construction and analysis. PLoS Comput Biol 4: e16.
7. Albert R, Jeong H, Barabasi AL (2000) Error and attack tolerance of complex networks. Nature 406: 378-382.
8. Satoh J, Tabunoki H, Arima K (2009) Molecular network analysis suggests aberrant CREB-mediated gene regulation in the Alzheimer disease hippocampus. Dis Markers 27: 239-252.
9. MAQC Consortium, Shi L, Reid LH, Jones WD, Shippy R, et al. (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol 24: 1151-1161.
10. Huber, Carey J, Gentleman R, Anders S, Carlson M, et al. (2015) orchestrating high-throughput genomic analysis with Bioconductor. Nature Methods 12: 115-121.
11. Reddy BR, Bhat AR, James BL, Govindan SV, Rohit Mathew, et al. (2016) Meta-Analyses of Microarray Datasets Identifies ANO1 and FADD as Prognostic Markers of Head and Neck Cancer. PLoS One 11: e0147409.
12. Chen BS, Yang SK, Lan CY, Chuang YJ (2008) A systems biology approach to construct the gene regulatory network of systemic inflammation via microarray and databases mining. BMC Med Genomics 1: 46.

13. Wheelock, C.E, Wheelock, A.M, Kawashima, S.; Diez, D, Kanehisa, M, van Erk, M, Kleemann, R.; et al. (2009) Systems Biology approaches and pathway tools for investigating cardiovascular disease. Mol BioSyst 5: 588-602.

14. Kanehisa M, Goto S (2000). "KEGG: Kyoto Encyclopedia of Genes and Genomes". Nucleic Acids Res 28: 27-30.

15. Mi H, Huang X, Muruganujan A, Tang H, Mills C, et al. (2016) "PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements." Nucleic Acids Research 45: D183-D189.

16. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, et al. (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 45: D362-68.

17. Krämer A, Green J, Pollard J Jr, Tugendreich S (2014) Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics 30: 523-530.

18. Sato H, Ishida S, Toda K, Matsuda R, Hayashi Y, et al. (2005) New approaches to mechanism analysis for drug discovery using DNA microarray data combined with KeyMolnet. Curr Drug Discov Technol 2: 89-98.

19. Mason O, Verwoerd M (2007) Graph theory, and networks in Biology. IET Syst Biol 1: 89-119. A comprehensive technical description of graph theory and network applications in the biology field.

20. Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, et al (2012): A travel guide to Cytoscape plugins. Nat Methods 9: 1069-1076.

21. The Gene Ontology Consortium (January 2008). "The Gene Ontology project in 2008". Nucleic Acids Research. 36 (Database issue): D440-444.

22. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44-57.

23. Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. Bioinformatics 21: 3448-3449.

24. Lotia S, Montojo J, Dong Y, Bader GD, Pico AR (2013) Cytoscape app store. Bioinformatics 29: 1350-1351.

25. Steinfeld I, Navon R, Creech ML, Yakhini Z, Tsalenko A (2015) ENViz: a Cytoscape App for integrated statistical analysis and visualization of sample-matched data with multiple data types. Bioinformatics 31: 1683-1685.

26. Matilde, Barbagallo C, Gulisano M, Domini CN, Barone R, et al. (2017) Expression and Regulatory Network Analysis of miR-140-3p, a New Potential Serum Biomarker for Autism Spectrum Disorder. Front Mol Neurosci 10: 250.

27. Santiago JA, Potashkin JA (2013) Integrative Network Analysis Unveils Convergent Molecular Pathways in Parkinson's disease and Diabetes. PLoSONE 8: e83940.

28. Boutté AM, Yao C, Kobeissy F, May Lu XC, Zhang Z, et al. (2012) Proteomic analysis and brain-specific systems biology in a rodent model of penetrating ballistic-like brain injury. Article in Electrophoresis 33: 3693-3704.

29. Kamarudin, Cox T, Kolamunnage-Dona R (2017) Time-dependent ROC curve analysis in medical research: current methods and applications. BMC Medical Research Methodology BMC series - open, inclusive and trusted 2017.

30. Birgit Grund, Sabin C (2010) Caroline Sabin Analysis of Biomarker Data: logs, odds ratios, and ROC curves. Curr Opin HIV AIDS 5: 473-479.

31. Ahmed SS, Ahameethunisa AR, Santosh W, Chakravarthy S, Kumar S, et al. (2011) Systems biological approach on neurological disorders: a novel molecular connectivity to aging and psychiatric diseases. BMC Systems Biology 5: 6.