Indian Res. J. Genet. & Biotech. 5 (1) : 1-10 (2013)

Identification of Significant Pathways Responsible For Autism through Molecular Network Analysis

YRSN Tejaswini[1]*, Y Anurupa Devi[2], Yog Raj Ahuja[3] and Q Hasan[4]

[1]Vasavi Medical and Research Centre, Khairatabad, Hyderabad, [2]Jawaharlal Nehru Technological University
[3]School of Health sciences, Karunya University, [4]Kamineni Hospitals, LB Nagar, Hyderabad

(Received : October 2012; Revised : December 2012; Accepted : January 2013)

Abstract :

Autism is a spectrum of developmental disorders characterized by impairments in social interaction, communication, often accompanied by stereotypical or repetitive behaviours. There are numerous hypotheses holding to the etiology and pathology of Autism but actual mechanism of the ailment is still unknown. Although a number of rare mutations and dosage abnormalities are specific to autism, these explain no more than 10% of all cases making the problem more complex. In this regard, shift from a narrow focus on individual candidate genes towards a broader view of affected protein networks and associated biological pathways have achieved the significant role. we have used network biology approach to identify important molecules and pathways which play significant role in autism through molecular interaction map (MIP) containing 248 nodes linked with 892 edges. Our studies elucidate the relationship between topological properties of MIP and the role played by molecules in biological systems. Further by applying the graph theory hub proteins were obtained and the pathways in which they are involved were analyzed. Our results showed link between many signalling pathways forming 22% in combination with other pathways like adherens junction. These insights provide useful clues in understanding how and to what extent each pathway is contributing in pathophysiology of this heterogenous disorder.

Introduction

Autism is a developmental disorder characterized by impairments in social interaction and communication, often accompanied by stereotypical or repetitive behaviors (Liu, et al., 2010). The condition manifests within the first 3 years and has a prevalence rate of 60-70 per 10,000 children in broader diagnostic criteria as per the most recent estimates (Fombonne, et al., 2009). Although a number of rare mutations and dosage abnormalities are specific to autism, these explain no more than 10% of all cases (Turner, et al., 2011). There are numerous hypotheses holding to the etiology and pathology of Autism but actual mechanism of the ailment is still unknown. Since, Pharmacologic research that targeted interfering symptom domains associated with autism, has showed low or medium efficacy, The increasing rates of this disorder every year (Kao, et al., 2010) demand to unearth the alternate ways in mining the clues that would finally provide definite molecular targets.

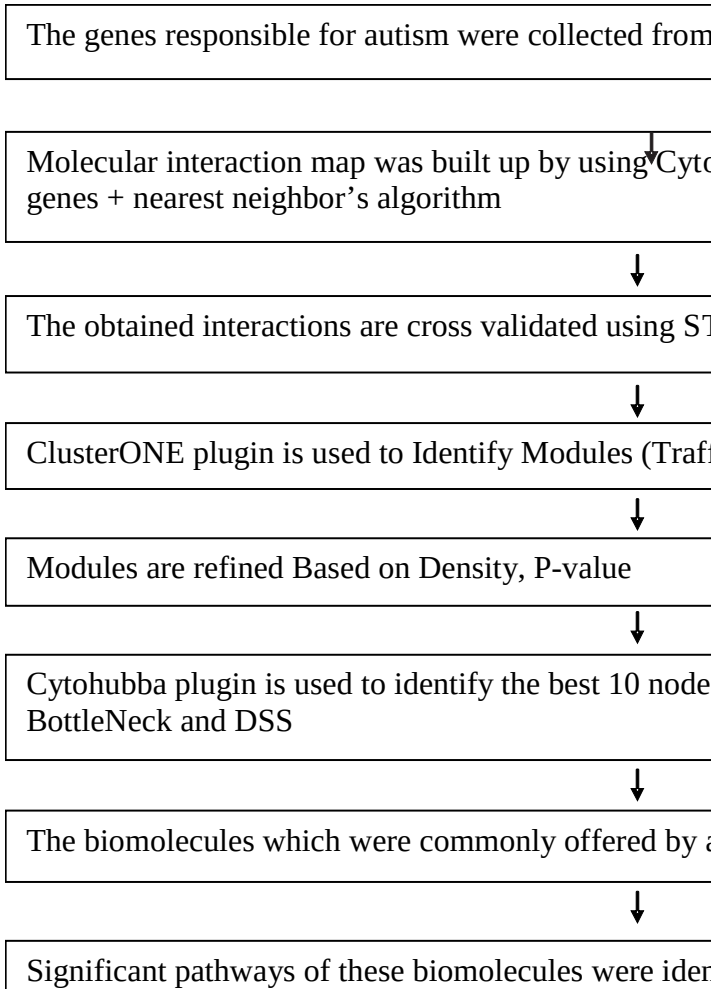
While the involvement of single gene mutations in individual autism cases cannot be excluded, the concept of a complex genetic model with multiple genes contributing to disease susceptibility remains highly plausible (Zhao, et al., 2007). In this regard the shift from a narrow focus on individual candidate genes towards a broader view of affected gene networks and associated biological pathways has achieved the significant role. The Biological networks are being used as a means to decipher the important key controllers inside the complex networks. These essential nodes/hubs may serve as candidates of drug targets for developing novel therapy for various diseases. In this

* Corresponding author's e-mail : yrsntejaswnini@gmail.com
Published by Indian Society of Genetics, Biotechnology Research & Development- Biotech Bhawan, 114 IInd Floor, Pushpanjali Comercial Complex, Shastripuram Road, Agra
Online www.isgbrd.co.in

approach, the diseases can be seen as emergent from a complex network of underlying molecular activity influenced by genes and environment. Indeed, complex networks are a natural way of representing any data with complicated dependency relationships.

Methodology

The genes responsible for autism were collected from GeneCards database (www.genecards.org) which provides a comprehensive, authoritative compendium of annotative information on responsible genes for a particular disease based on text mining algorithm (Safran, et al., 2010). A specific set of 60 genes which come above the score limit (score>0.3) belonging to various Categories like protein-coding, RNA gene etc were selected. Molecular interaction map was built up by using Cytoscape (Shannon, et al., 2003) Mimi plugin (Jayapandian, et al., 2007) based on Query genes + nearest neighbor algorithm. The obtained interactions are cross validated using string Database which correctly uncovers and annotates all functional interactions (Szklarczyk, et al., 2011).ClusterONE clustering with Overlapping Neighborhood Expansion), the graph clustering algorithm is used for detecting protein complexes in protein-protein interaction networks with associated confidence values. The generated modules are then refined based on density (> = 0.7) and P-value (< = 0.00). As a further process of refinement cytohubba (Lin, et al., 2008) is used to obtain best 10 biomolecules based on the topology-based scoring methods which include Degree (Jeong, et al., 2001), BottleNeck (Yu, et al., 2007), Edge percolated component (Chin, et al., 2003) MNC, DMNC and The double screening scheme (Lin, et al., 2008). The biomolecules which were commonly offered by all algorithms are extracted to identify their pathways to generate a relationship of various pathways in the disease mechanism.



Results and discussion

The core areas affected in autism involve rapid and coherent integration of information from multiple, higher-level association areas (Geschwind and Levitt, 2007). Accordingly, the predominant genetic model supposes the presence of multigenic inheritance of common polymorphisms contributing to autism risk in multiplex families (Abrahams and Geschwind, 2008) leading to disruption of normal function. Such functions could be easily perturbed by minor, but relatively widespread disruptions in a set of pathways. To mine these details we have constructed a Molecular Interaction Map for the selected 60 nodes (refined by score) by taking into account of its neighboring interaction molecules through MiMI plug-in of Cytoscape. Totally 248 molecules as nodes and 892 edges as interactions were obtained. The molecular interaction map can be explained as a mathematical graph, permitting analysis with graph theoretical algorithms. Molecules like genes, proteins, transcriptional factors are denoted as nodes in the graph and interactions between them are called as edges. This MIP is a scale free network which obeys power law distribution of connectivity.

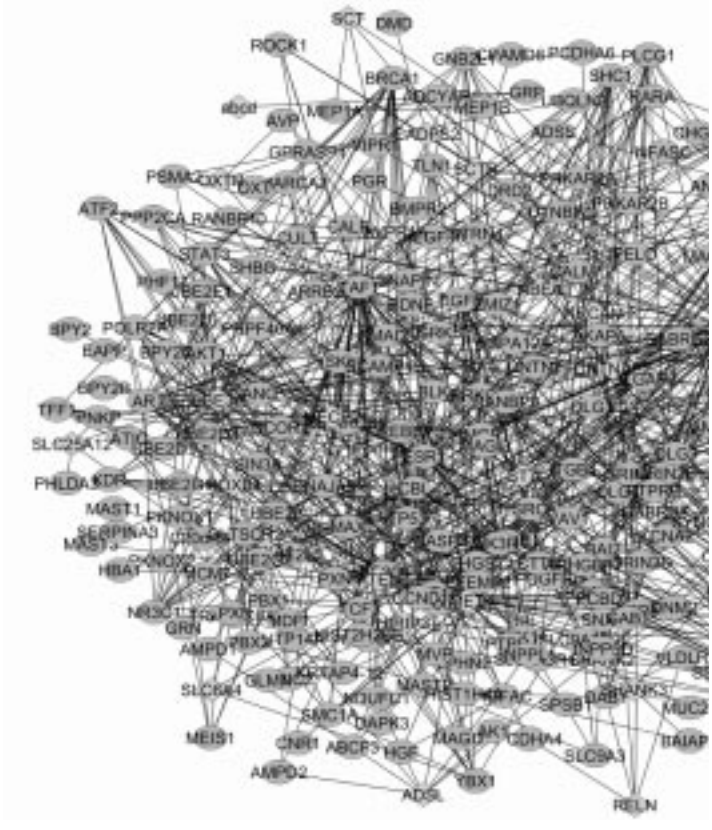


Figure 1: Molecular interaction map containing query genes and their interactions

Network Analysis

We have represented Molecular Interaction graph as an undirected graph M (N, E), consisting of set of nodes as N and set of edges as E. The size of the graph is given by the number of its nodes. The degree of its nodes indicates the number of interactions to a single node with the other nodes. The network obtained by MiMI was analyzed to know the wide range of pathways which directly or indirectly play role in autism. The analysis of all the genes obtained through MIMi gave almost 57 varied pathways out of which some of the genes were continuously repeating for Neurodegenerative Diseases, protein kinase cascade etc but majority of molecules in the interaction map were showing a high frequency of focal adhesion, signaling, Melanogenesis and cell cycle pathways. Figure 2 shows the important pathways present in the molecular interaction map on x- axis and on Y-axis is the frequency of these pathways occurring for the respective Biomolecules.

Figure 2: Major pathways present in 248 nodes of the molecular interaction map

The NetworkAnalyzer plug-in (Assenov, et al., 2008) is used to calculate the topological properties of each module individually that were tabulated in Table 1.

Figure 3 and Figure 4 represent the two highly connected modules that were obtained after refined by merging the cohesive subgroups from the molecular interaction map based on the Quality of the cluster, measured by the in-weight divided by the sum of the in-weight and the out-weight. The rationale behind this measure is that a good cluster contains many heavyweight edges within the cluster itself, and it is connected to the rest of the network only by a few lightweight edges. P- Value showing the validity of the cluster.

Figure 3 and Figure 4 represent the two modules formed from molecular interaction network using clusterONE plug-in of cytoscape.

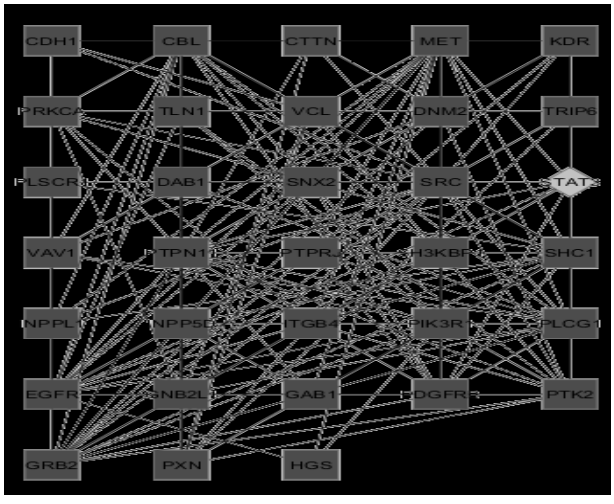


Figure 3: Module 1(34 nodes) obtained from molecular interaction network using cluster ONE plug-in of cytoscape.

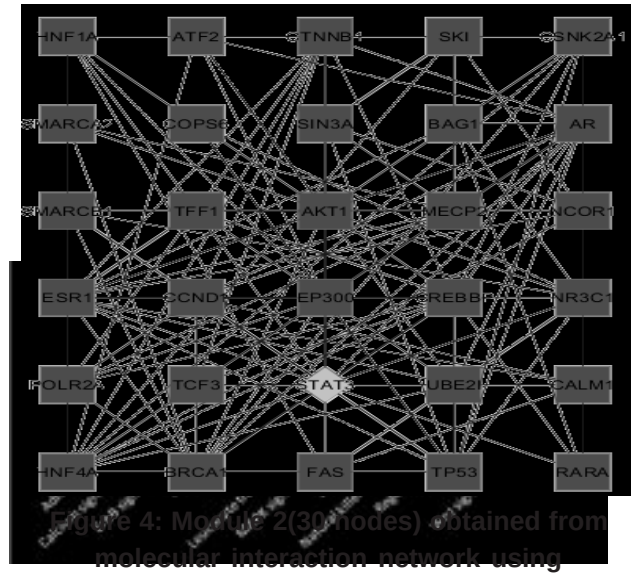


Figure 4: Module 2(30 nodes) obtained from molecular interaction network using clusterONE plug-in of cytoscape.

Table 1: The Topological properties of each module obtained through NetworkAnalyzer plug-in Once the two

Topological Parameters	Module1	Module2
Cluster Coefficient	0.656	0.469
Network Diameter	3	3
Network Density	0.303	0.30
Network Heterogenity	0.625	0.463
Network Radius	2	2
Network Centralization	0.443	0.303
Charecterisic Path Length	1.748	1.795
Avg No: Of Neighbours	9.697	8.8
No: Of Nodes	34	30

modules are selected the topological parameters of each node in the modules are calculated to identify relative importance of nodes based on graph theory by considering each module as undirected network. Table 2 and Table 3 represent the topolocial parameters of nodes present in Module

Table 2 : The Topological parameters of individual nodes in Module 1

NAME	TOPOLOGICAL COEFICIENT	CLOSENESS	NEIGHBOU-RHOOD CONNECTIVITY	CLUSTER COEFICIENT	DEGREE	RADIALITY	STERSS
CBL	0.42	0.67	13.38	0.53	16	0.83	176.00
CDH1	0.57	0.53	18.25	0.67	4	0.71	30.00
CTTN	0.48	0.54	15.40	0.60	5	0.72	16.00
DAB1	0.68	0.49	19.00	1.00	3	0.66	0.00
DNM2	0.43	0.54	12.86	0.57	7	0.72	28.00
EGFR	0.34	0.76	10.82	0.37	22	0.90	492.00
GAB1	0.54	0.56	16.33	0.86	9	0.74	10.00
GNB2L1	0.40	0.52	11.50	0.40	6	0.70	26.00
GRB2	0.40	0.70	12.67	0.52	18	0.85	222.00
HGS	0.72	0.48	19.33	1.00	3	0.65	0.00
INPP5D	0.54	0.53	16.33	1.00	6	0.71	0.00
INPPL1	0.61	0.54	19.60	1.00	5	0.72	0.00
ITGB4	0.50	0.53	15.40	0.60	5	0.71	8.00
KDR	0.55	0.54	16.43	0.81	7	0.72	10.00
MET	0.33	0.73	10.45	0.34	20	0.88	398.00
PDGFRB	0.42	0.64	13.43	0.55	14	0.81	186.00
PIK3R1	0.41	0.67	13.19	0.53	16	0.83	222.00
PLCG1	0.40	0.67	12.75	0.49	16	0.83	238.00
PLSCR1	0.65	0.52	20.67	1.00	3	0.70	0.00
PRKCA	0.45	0.57	14.38	0.50	8	0.75	64.00
PTK2	0.41	0.64	13.21	0.49	14	0.81	226.00
PTPN11	0.40	0.70	12.89	0.50	18	0.85	244.00
PTPRJ	0.64	0.50	18.00	0.83	4	0.67	2.00
PXN	0.41	0.56	12.44	0.56	9	0.74	64.00
SH3KBP1	0.63	0.52	17.67	0.93	6	0.69	2.00
SHC1	0.40	0.68	12.94	0.53	17	0.84	166.00
SNX2	0.75	0.47	18.67	0.67	3	0.63	2.00
SRC	0.33	0.78	10.70	0.34	23	0.91	574.00
STAT3	0.45	0.56	12.13	0.61	8	0.74	38.00
TLN1	0.43	0.57	13.63	0.57	8	0.75	68.00
TRIP6	0.54	0.48	13.50	1.00	4	0.64	0.00
VAV1	0.55	0.54	16.43	0.76	7	0.72	20.00
VCL	0.38	0.51	9.50	0.46	8	0.68	30.00

Table 3: The Topological parameters of individual nodes in Module 2

NAME	TOPOLOGICAL COEFICIENT	CLOSENESS	NEIGHBOU-RHOOD CONNECTIVITY	CLUSTER COEFICIENT	DEGREE	RADIALITY	STERSS
AKT1	0.47	0.54	12.14	0.62	7.00	0.71	32.00
AR	0.37	0.69	10.75	0.43	16.00	0.85	286.00
ATF2	0.46	0.52	11.14	0.52	7.00	0.69	34.00
BAG1	0.56	0.48	12.25	0.50	4.00	0.63	12.00
BRCA1	0.39	0.64	11.23	0.50	13.00	0.82	174.00
CALM1	0.42	0.55	10.43	0.19	7.00	0.72	92.00
CCND1	0.42	0.60	12.20	0.56	10.00	0.78	96.00
COPS6	0.53	0.49	13.33	0.67	3.00	0.66	2.00
CREBBP	0.36	0.69	10.56	0.42	16.00	0.85	290.00
CSNK2A1	0.42	0.56	11.00	0.39	8.00	0.74	64.00
CTNNB1	0.41	0.59	11.00	0.45	11.00	0.77	130.00
EP300	0.35	0.67	9.75	0.39	16.00	0.84	268.00
ESR1	0.34	0.71	9.94	0.34	17.00	0.86	430.00
FAS	0.48	0.48	9.80	0.40	5.00	0.63	14.00
HNF1A	0.46	0.52	11.14	0.52	7.00	0.69	36.00
HNF4A	0.35	0.67	10.27	0.37	15.00	0.84	366.00
MECP2	0.34	0.53	7.78	0.33	9.00	0.70	94.00
NCOR1	0.36	0.57	10.13	0.39	8.00	0.75	122.00
NR3C1	0.32	0.63	9.17	0.27	12.00	0.80	242.00
POLR2A	0.38	0.58	10.78	0.47	9.00	0.76	110.00
RARA	0.45	0.50	11.25	0.50	4.00	0.67	20.00
SIN3A	0.42	0.46	8.25	0.50	4.00	0.61	14.00
SKI	0.41	0.48	9.00	0.67	4.00	0.63	20.00
SMARCA2	0.44	0.52	11.00	0.53	6.00	0.69	26.00
SMARCB1	0.44	0.50	10.17	0.53	6.00	0.67	24.00
STAT3	38.00	0.56	3.00	0.61	8.00	12.13	0.74
TCF3	0.55	0.48	11.80	0.40	5.00	0.64	12.00
TFF1	0.45	0.56	13.00	0.67	6.00	0.74	42.00
TP53	0.40	0.63	11.25	0.44	12.00	0.80	222.00
UBE2I	0.40	0.59	11.67	0.47	9.00	0.77	116.00

In order to visually depict the relationship between the topological parameters Regression graph has been generated. In Figure 5 first graph represents the linear relationship between closeness (X-axis) and Degree (Y- axis) and second graph represents the linear relationship between Cluster Coefficient (X-axis) and Topological coefficient (Y- axis) for the Module1.

Likewise, In Figure 6 first graph represents the linear relationship between closeness (X-axis) and Degree (Y- axis) and second graph represents the linear relationship between Cluster Closeness (X-axis) and Betweeness (Y- axis) for the Module2. The R^2 value obtained confirms the uniformity in the Result.

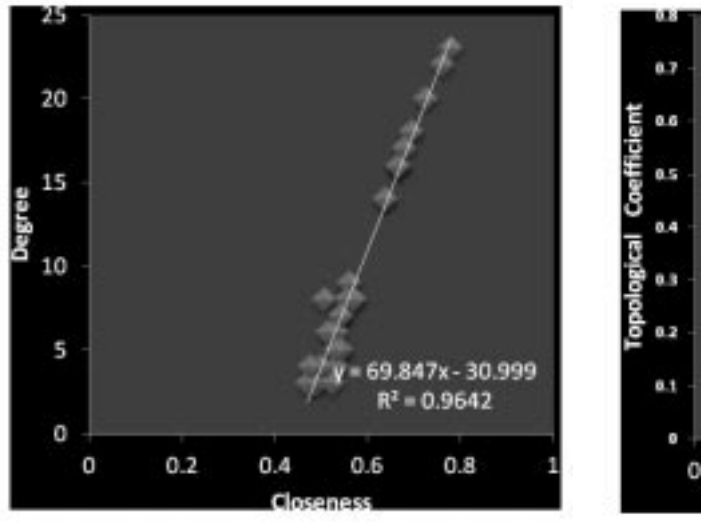


Figure 5: Linear relationship between the Topological parameters for Module1

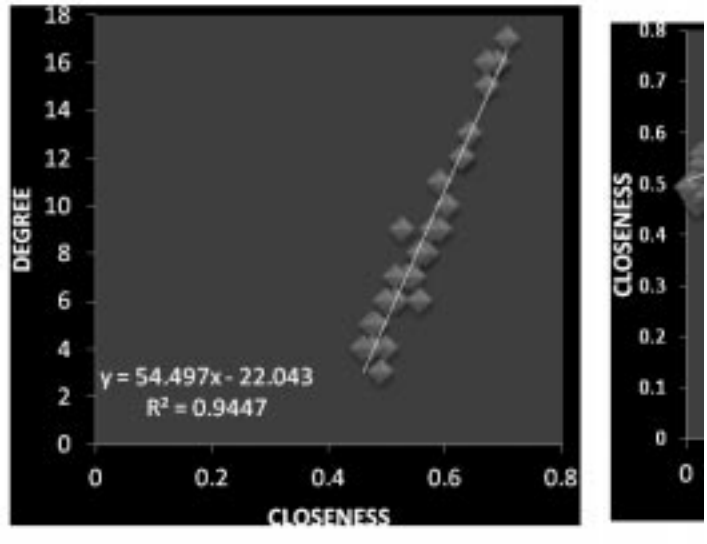


Figure 6: Linear relationship between the Topological parameters for Module2

Further, the two was analysed using the cytoHubba plugin (http:// hub.iis.sinica.edu.tw/cytoHubba/) to explore the key regulatory nodes in the network. The top 10 hubs (i.e. highly connected nodes) were identified by using the all the algorithms (Degree, EPC, DMNC, BottleNeck and DSS), displayed in Table 3 and Table 4. The proteins that were identified commonly by at least 5 algorithms are selected to identify the major pathways in which they are involved.

Table 3: The top 10 hub proteins identified by using the all the algorithms of cytohubba plug-in for module 1

Algorithm	Top 10 Hub Proteins
DSS	CBL, GAB1, GRB2, PDGFRB, PIK3R1, PLCG1, PTK2, PTPN11, PXN, SHC1
Degree	CBL, EGFR, GRB2, MET, PIK3R1, PLCG1, PTK2, PTPN11, SHC1, SRC
BottleNeck	EGFR, GRB2, MET, PIK3R1, PLCG1, PTK2, PTPN11, SRC, TLN1, VCL
EPC	CBL, EGFR, GRB2, MET, PDGFRB, PIK3R1, PLCG1, PTPN11, SHC1, SRC
MNC	CBL, EGFR, GRB2, MET, PIK3R1, PLCG1, PTK2, PTPN11, SHC1, SRC
DMNC	CBL, GAB1, GRB2, INPP5D, INPPL1, KDR, SH3KBP1, SHC1, TRIP6, VAV1

Table 4: The top 10 hub proteins identified by using the all the algorithms of cytohubba plug-in for module 2

Algorithm	Top 10 Hub Proteins
DSS	AR, BRCA1, CCND1, CREBBP, CTNNB1, EP300, HNF4A, POLR2A, TP53, UBE2I
Degree	AR, BRCA1, CCND1, CREBBP, CTNNB1, EP300, ESR1, HNF4A, NR3C1, TP53
BottleNeck	AR, BRCA1, CALM1, CREBBP, EP300, ESR1, HNF4A, NCOR1, NR3C1, TP53
EPC	AR, BRCA1, CCND1, CREBBP, CTNNB1, EP300, ESR1, HNF4A, NR3C1, TP53
MNC	AR, BRCA1, CCND1, CREBBP, CTNNB1, EP300, ESR1, HNF4A, NR3C1, TP53
DMNC	AKT1, AR, BRCA1, CCND1, CREBBP, CTNNB1, SIN3A, STAT3, TFF1, TP53

Finally the hub proteins that were selected commonly by at least 5 algorithms in both the modules are extracted and their pathways are identified (Table 5). Most of the hub proteins obtained are involved in signaling, while the rest of the proteins are involved in Focal adhesion, Glioma, and Regulation of transcription.

Table 5: Pathways for hub proteins identified by at least 5 algorithms of cytohubba plug-in

Hub proteins	chromosome	Pathways
CBL	11	ErbB signaling pathway ; Jak-STAT signaling pathway
SHC1	1	ErbB signaling pathway ; Focal adhesion
GRB2	17	Focal adhesion; Gap junction; Jak-STAT signaling pathway; Glioma
PLCG1	20	ErbB signaling pathway ; Glioma
PTPN11	12	Jak-STAT signaling pathway
PIK3R1	5	ErbB signaling pathway; Regulation of actin cytoskeleton ; Jak-STAT signaling pathway; Glioma; Melanogenisis; Focal adhesion
AR	X	cell-cell signaling; regulation of transcription

BRCA1	17	DNA repair
CCND1	11	cell cycle
CREBBP	16	response to hypoxia; signal transduction ; regulation of transcription
EP300	22	cell cycle ; regulation of transcription; signal transduction; regulation of transcription
TP53	17	apoptosis
CTNNB1	3	Wnt receptor signaling pathway
HNF4A	20	blood coagulation; transcription

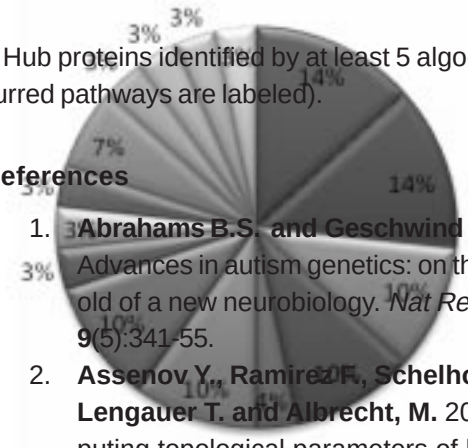


Figure 6: percentage of Occurrence of Pathways in Hub proteins identified by at least 5 algorithms of cytohubba plug-in (highly occurred pathways are labeled).

Conclusion

In an effort to identify the important pathways involved in the Autism mechanism, the set of biomolecules which are highly connected (Modules) are identified from Molecular interaction map and the topological parameters for each individual node within the modules (Module 1 and Module 2) are calculated. Further, the best 10 nodes in each module are obtained by using all the algorithms of cytohubba and the commonly occurring hub proteins in at least 5 algorithms are extracted to identify the pathways in which they are involved. Our Analysis has revealed that ErbB signaling pathway and Jak-STAT signaling pathways are the major pathways in autism. In addition, Focal adhesion, Glioma, and Regulation of transcription are also involved in the pathology of the disease.

References

1. **Abrahams B.S. and Geschwind DH.** 2008 Advances in autism genetics: on the threshold of a new neurobiology. *Nat Rev Genet.* **9**(5):341-55.
2. **Assenov Y., Ramirez F., Schelhorn, S.E., Lengauer T. and Albrecht, M.** 2008 Computing topological parameters of biological networks. *Bioinformatics*, **24**(2):282-284,
3. **Chin C.S. and Manoj P.S.** 2003 Global snapshot of a protein interaction network—a percolation based approach. *Bioinformatics*, **19**, 2413–2419.
4. **Fombonne E., Quirke S. and Hagen A.** 2009 Prevalence and interpretation of recent trends in rates of pervasive developmental disorders. Mcgill J Med 12: 73.

5. **Geschwind DH. and Levitt P.** 2007 Autism spectrum disorders: developmental disconnection syndromes. Curr Opin Neurobiology. Feb;17(1):103-11.
6. **Jeong H., Mason S.P., Baraba, S.I. A.L. and Oltvai Z.N.** (2001) Lethality and centrality in protein networks. *Nature*, 411, 41–42. 24.
7. **Kao H.T., Buka S.L., Kelsey K.T., Gruber D..F and Porton B.** 2010 The Correlation between Rates of Cancer and Autism: An Exploratory Ecological Investigation. PLoS ONE 5(2): e9372.
8. **Ka-Yuet Liu., Marissa King, and Peter S.** 2010 Bearman Social Influence and the Autism Epidemic AJS. March ; 115(5): 1387–1434.
9. **Jayapandian Magesh, Chapman Adriane, Glenn V., Tarcea Cong Yu, Aaron Elkiss, Angela Ianni Bin Liu, Arnab Nandi, Carlos Santos, Philip Andrews, Brian Athey, David States and Jagadish Michigan H. V.** 2007 Molecular Interactions (MiMI): putting the jigsaw puzzle together Nucleic Acids Research, 35, Database issue
10. **Shannon Paul , Markiel Andrew, Owen Ozier, Nitin S. Baliga,Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker Cytoscape**: 2003 A Software Environment for Integrated Models of Biomolecular Interaction Networks *Genome Res.* 13: 2498-2504
11. **Safran M., Dalah I., Alexander J., Rosen N., Iny Stein T., Shmoish M., Nativ N., Bahir I., Doniger T., Krug H., Sirota-Madi A., Olender T., Golan Y., Stelzer G., Harel A. and Lancet D.** *GeneCards Version 3: the human gene integrator* Database 2010;
12. **Szklarczyk D. Franceschini A., Kuhn M. and von Mering C.**, 2011 The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. January; 39(Database issue): D561–D568.
13. **Turner T., Pihur V. and Chakravarti A.** 2011 Quantifying and Modeling Birth Order Effects in Autism. PLoS ONE 6(10): e26418.
14. **Yu H., Kim P.M., Sprecher E., Trifonov V. and Gerstein M.** 2007The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS Comput. Biol., 3, e59.
15. **Zhao X., Leotta A., Kustanovich V., LaJonchere C., Geschwind D.H.** 2007 A unified genetic theory for sporadic and inherited autism. Proceedings of the National Academy of Sciences of the United States of America 104: 12831–12836.

****