Studies in Computational Intelligence

Volume 932

The series "Studies in Computational Intelligence" (SCI) publishes new developments and advances in the various areas of computational intelligence—quickly and with a high quality. The intent is to cover the theory, applications, and design methods of computational intelligence, as embedded in the fields of engineering, computer science, physics and life sciences, as well as the methodologies behind them. The series contains monographs, lecture notes and edited volumes in computational intelligence spanning the areas of neural networks, connectionist systems, genetic algorithms, evolutionary computation, artificial intelligence, cellular automata, self-organizing systems, soft computing, fuzzy systems, and hybrid intelligent systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

Indexed by SCOPUS, DBLP, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at http://www.springer.com/series/7092

Ripon Patgiri · Anupam Biswas ·
Pinki Roy
Editors

Health Informatics: A Computational Perspective in Healthcare

Springer

Editors
Ripon Patgiri
Department of Computer Science
and Engineering
National Institute of Technology Silchar
Silchar, Assam, India

Anupam Biswas
Department of Computer Science
and Engineering
National Institute of Technology Silchar
Silchar, Assam, India

Pinki Roy
Department of Computer Science
and Engineering
National Institute of Technology Silchar
Silchar, Assam, India

ISSN 1860-949X ISSN 1860-9503 (electronic)
Studies in Computational Intelligence
ISBN 978-981-15-9734-3 ISBN 978-981-15-9735-0 (eBook)
https://doi.org/10.1007/978-981-15-9735-0

Preface

Computing technique is one of the key technologies that is being currently used to perform medical diagnostics in the healthcare domain, thanks to the abundance of medical data being generated and collected. Nowadays, medical data is available in many different forms like MRI images, CT scan images, EHR data, test reports, histopathological data, doctor–patient conversation data, etc. This opens up huge opportunities for the application of computing techniques, to derive data-driven models that can be of very high utility, in terms of providing effective treatment to patients. Moreover, machine learning algorithms can uncover hidden patterns and relationships present in medical datasets, which are too complex to uncover, if a data-driven approach is not taken. With the help of computing systems, today, it is possible for researchers to predict an accurate medical diagnosis for new patients, using models built from previous patient data. Apart from automatic diagnostic tasks, computing techniques have also been applied in the process of drug discovery, by which a lot of time and money can be saved. Utilization of genomic data using various computing techniques is other emerging areas, which may in fact be the key to fulfilling the dream of personalized medications. Medical prognostics is another area in which machine learning has shown great promise recently, where automatic prognostic models are being built that can predict the progress of the disease as well as can suggest the potential treatment paths to get ahead of the disease progression. Our book on *Health Informatics: A Computational Perspective in Healthcare* presents at attracting research works, to demonstrate the potential and the advancements of computing approaches to utilize healthcare centric and medical datasets.

Silchar, India

Dr. Ripon Patgiri
Dr. Anupam Biswas
Dr. Pinki Roy

Contents

About the Editors

Dr. Ripon Patgiri is currently working as an Assistant Professor at the Department of Computer Science & Engineering, National Institute of Technology Silchar. He has received his B.Tech., M.Tech. and Ph.D. degree from the Institutions of Electronics and Telecommunication Engineers, Indian Institute of Technology Guwahati and National Institute of Technology Silchar, respectively. His research interests are big data, bioinformatics and distributed systems. He has published several papers in reputed journals, conferences and books. Also, he was General Chair of 6th International Conference on Advanced Computing, Networking and Informatics. Currently, he is General Chair of International Conference on Big Data, Machine Learning and Applications to be held during 16–19 December 2019 at National Institute of Technology Silchar. Moreover, he is an organizing chair of 25th International Symposium Frontiers of Research in Speech and Music (FRSM 2020), to be held during 08–09 October 2020. He is also an organizing chair of International Conference on Modeling, Simulations and Optimizations (CoMSO 2020), to be held during 3–5 August 2020. Furthermore, he is a Guest Editor of "Big Data: Exascale computation and beyond" in EAI Transaction on Scalable Information Systems and Guest Editor of "Internet of Things: Challenges and Solutions" in "EAI Transactions on Internet of Things". He reviewed many research articles from KSII Transactions on Internet and Information Systems, Electronics Letters, EAI Endorsed Transactions on Energy Web, EAI Endorsed Transactions on Scalable Information Systems, ACM Transactions on Knowledge and Data Engineering, IET Software, International Journal of Computational Vision and Robotics, Journal of Computer Science, International Journal of Advanced Computer Science and Applications and IEEE Access. Also, he served as TPC member in many conferences. He is a senior member of IEEE, member of ACM, EAI and ACCS and associate member of IETE.

Dr. Anupam Biswas is currently working as an Assistant Professor at the Department of Computer Science & Engineering, National Institute of Technology Silchar. He has received his B.Tech., M.Tech. and Ph.D. degree from Dibrugarh University, Motilal Nehru National Institute of Technology Allahabad and Indian

Institute of Technology (BHU) Varanasi respectively. His research interests are social networking, review mining, sentiment analysis, machine learning and soft computing. He has received the Best Paper Award for the paper titled "Community Detection in Multiple Featured Social Network using Swarm Intelligence" in International Conference on Communication and Computing (ICC-2014), Bangalore. Also, he has received Reviewer Award from Applied Soft Computing Journal (IF 3.541), Elsevier, 2015 and 2017, and Physica A: Statistical Mechanics and its Applications (IF 2.243), Elsevier, 2016. He has published several papers in reputed journals, conferences and books. He is a reviewer of IEEE Transactions on Fuzzy Systems (TFS), IEEE Transactions on Evolutionary Computation (IEEETEVC), IEEE Systems Journal (IEEE-SJ), IEEE Transactions on Systems, Man and Cybernatics: System (IEEE TSMC), Applied Soft Computing (ASOC), ACM Transactions on Knowledge Discovery from Data (TKDD), ACM Transactions on Intelligent Systems and Technology (TIST) and Physica A: Statistical Mechanics and its Applications and Information Sciences.

Dr. Pinki Roy is currently working as an Assistant Professor at the Department of Computer Science & Engineering, National Institute of Technology Silchar. Dr. Pinki Roy received her B.Tech. degree in Computer Science & Engineering from Dr. Babasaheb Ambedkar Technological University, Lonere, Maharashtra (2002, First class with distinction) and M.Tech. degree (2004, First class with distinction) from Dr. Babasaheb Ambedkar Technological University, Lonere, Maharashtra. She has received her Ph.D. degree in the year 2014 in the field of Language Identification from National Institute of Technology, Silchar, Assam-788010, India. She was working as a Lecturer in Naval Institute of Technology, Colaba, Mumbai, India. (from February 2004 to August 2004). Her research interests include language identification, speech processing, machine intelligence and cloud computing. She has published several papers in reputed journals, conferences and books. She has received several awards which are listed below—1. "Distinguished Alumnus Award", Dr. Babasaheb Ambedkar Technological University, Lonere, Maharashtra, 2014. 2. "Young Scientist Award", Venus International Foundation, Chennai, 2015. Awarded for major contribution in research during Ph.D. 3. "Rastriya Gaurav Award", India International Friendship Society, New Delhi, 2015. 4. "Bharat Excellence Award", Friendship Forum, New Delhi, 2016. 5. "Best Golden personalities Award", Friendship Forum, New Delhi, 2016. 6. "Global Award for Education", Friendship Forum, New Delhi, 2016. 7. Honoured as one of the "Most Distinguished Lady Alumni" by Computer Engineering Department of Dr. Babasaheb Ambedkar Technological University, Lonere, Maharashtra, India.

PROTEIN INTERACTION AND DISEASE GENE PRIORITIZATION IN NEUROPSYCHIATRIC DISORDERS.

Brijendra Gupta

Abstract Protein prioritization through genes is a novel approach to derive disease and traits over gene encodings. We have sets of genes over the chromosomes which are combined, crossover and mutated to different sets. The very essential observation for disease genes comes under the analysis of linkage region where we find the genetic elements of diseases. Disease susceptibility through genes and its variants are part of genome wide Association studies. For a list of candidate genes, gene location in the linkage region maps the particular disease phenotype. Disease associated candidate genes and its translated protein formation are considered as a diagnostic biomarker. We have a common representative database online mendelian inheritance in man (OMIM) that links genes, genomic region and phenotypes. To get effective computation out of biological resources, gene prioritization gives us the best experimental validation. The subsequent targeted genes are mapped on reference genes and biological cellular and metabolic paths are represented to analyse phenotypic variation.

Characteristics of genetic Pathways and focus on findings of more precise as biomarkers are elementary process involved in disease detection and progression. Research efforts over target proteins, their activities and transcription factors is linking metabolism, cellular activity and biological signalling pathway. These components also trace cell division, its differentiation and disease susceptibility. Gene-Gene interaction and its Association channelize protein-protein interaction through biological Pathways. These network are investigated through different database sources. We find a robust as well as integrated analysis of biological process over genotypes and phenotypes. In this chapter, topological properties of protein protein interaction and its focus of phenotype through gene-gene interactions are paid attention through which structure, function and dynamic analysis of network are mapped.

Brijendra Gupta
Dept of cse, Siddhant College of Engineering,Pune
e-mail: Gupbrij@rediffmail.com

1 PRIORITIZATION USING INTERACTION NETWORKS

Many methods using interactions in prioritizing candidate genes are fundamentally counting and comparing "qualified" interactions under the following underlying assumptions

1. Causative genes for a disease reside in the same biological module;
2. Genes in the module have many interactions with other members within the module and relatively few interactions with nonmembers outside the module.

The observation that disease genes share interactions at a much higher rate than random gene pairs was made when protein–protein interaction (PPI) networks were assembled [1].

2 PATHWAYS

Research efforts are therefore currently focusing on the discovery of more precise markers, which are needed so that appropriate treatment decisions can be made for individual patients, and on the characterization of genetic pathways involved in the development and progression of the disease. However, simply identifying the list of genes involved in a disease is only a first step in the process. There is the need for identifying more refined information (e.g., the interaction between the various steps in the pathway). The target proteins include enzymes (e.g., glycogen synthase and other kinases); transcription factors (e.g., c-Jun) and non enzymatic proteins (e.g., histones) that are involved in distinct signaling pathways linked to metabolism, gene expression, cell motility, cell division, cell differentiation, and apoptosis.

A set of genes has a high chance of participating in a common biological pathway if they form a module in both the gene–gene and protein–protein interaction networks (PPINs)[2]. Sometimes, these networks are constructed by combining multiple sources of information by way of investigating associations between them. Thus, we can obtain a robust and integrated view of the underlying biology. This chapter focuses exclusively on the topological properties of interaction networks of proteins and their significance in the systems level. Instead of pursuing a piecemeal study of the single components, we pay attention to the more global analyzes of the structure, function, and dynamics of the networks in which macromolecules work.

3 TARGET, MAPPED AND REFERENCE GENES FOR ADHD, DEMENTIA, MOOD DISORDER, OCD AND SCHIZOPHRENIA

We have taken data from NCBI which are the allelic variants out of many other available genes for those given Diseases. So, these genes are taken as target genes

whose interactions we are taking into consideration with other many genes who are experimentally tested and verified from STRING DATABASE and ENRICHNET for the functional, molecular and cellular bindings with other genes[3][4]. So, the available datasets in these databases are taken as reference genes. The pathway under which we observed the connections of genes in ENRICHNET tool are mapped genes which are highly interacted[5] and with which we have to target other interactions under the available database and the targeted ones to get interaction scores and summed together to determine the priorities out of those targeted genes on behalf on distance minimization and interaction score maximization[6].

3.1 GENE-GENE INTERACTION FOR ADHD

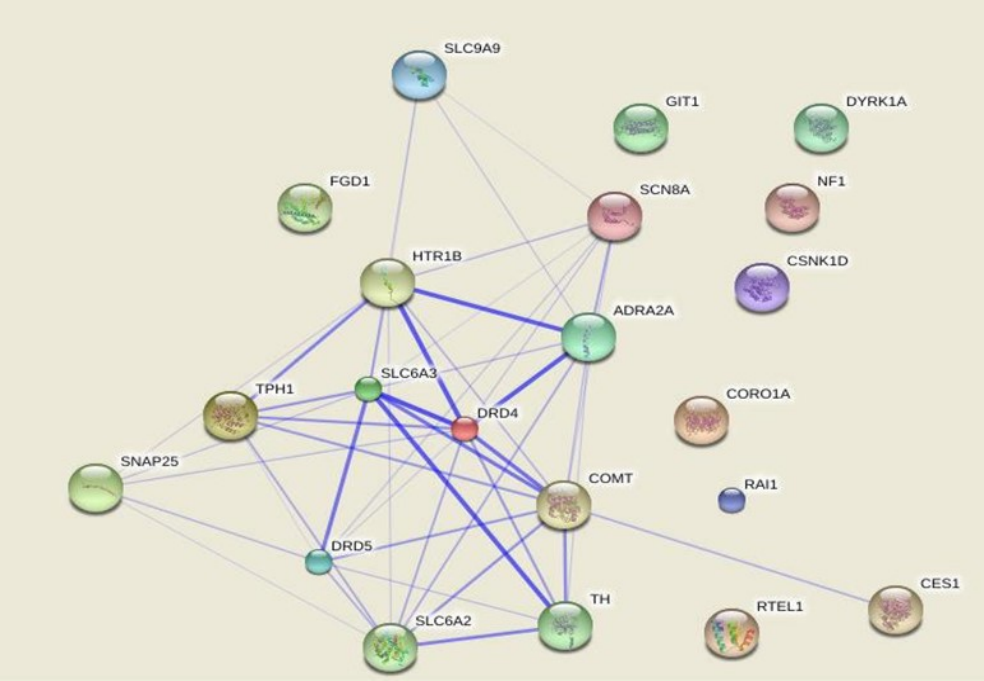

Fig 1.1: Thicker lines represent strong interaction between the genes.

HTR1B: 5-hydroxytryptamine (serotonin) receptor 1B; This is one of the several different receptors for 5- hydroxytryptamine (serotonin), a biogenic hormone that functions as a neurotransmitter, a hormone, and a mitogen. The activity of this receptor is mediated by G proteins that inhibit adenylate cyclase activity .

ADRA2A: Adrenergic,alpha-2A-, receptor;Alpha-2 adrenergic receptors mediate the catecholamine- induced inhibition of adenylate cyclase through the action of G proteins. The rank order of potency for agonists of this receptor is oxymetazoline > clonidine > epinephrine > norepinephrine > phenylephrine > dopamine > p-synephrine >p-tyramine > serotonin =p-octopamine.For 4 antagonists,the rank order is yohimbine > phentolamine = mianserine > chlorpromazine = spiperone = prazosin > propanolol > alprenolol = pindolol.

TPH1: Tryptophan hydroxylase 1 .

SLC6A3: Solute carrier family 6(neurotransmitter transporter, dopamine), member 3; Amine transporter. Terminates the action of dopamine by its high affinity sodium-dependent reuptake into presynaptic terminals .

DRD4: dopamine receptor D4 .

COMT: Catechol-O-methyltransferase; Catalyzes the O-methylation, and thereby the inactivation, of catecholamine neurotransmitters and catechol hormones.Also shortens the biological half-lives of certain neuroactive drugs, like L-DOPA, alpha-methyl DOPA and isoproterenol.

DRD5: Dopamine receptor D5; This is one of the five types (D1 to D5) of receptors for dopamine. The activity of this receptor is mediated by G proteins which activate adenylyl cyclase .

SLC6A2: Solute carrierfamily 6(neurotransmitter transporter, noradrenalin), member 2; Amine transporter.Terminates the action of noradrenaline by its high affinity sodium-dependent reuptake into presynaptic terminals .

TH: Tyrosine hydroxylase; Plays an important role in the physiology of adrenergic neurons .

SNAP25: Synaptosomal-associated protein, 25kDa;t-SNARE involved in the molecular regulation of neurotransmitter release. May play an important role in the synaptic function of specific neuronal systems. Associates with proteins involved in vesicle docking and membrane fusion. Regulates plasma membrane recycling through its interaction with CENPF (206 aa)

CES1: Carboxylesterase 1 (monocyte/macrophage serine esterase 1);Involved in the

detoxification of xenobiotics and in the activation of ester and amide prodrugs. Hydrolyzes aromatic and aliphatic esters, but has no catalytic activity toward amides or a fatty acyl CoA ester .

SCN8A: Sodium channel, voltage gated, type VIII, alpha subunit; Mediates the voltage-dependent sodium ion permeability of excitable membranes. Assuming opened or closed conformations in response to the voltage difference across the membrane, the protein forms a sodium-selective channel through which Na(+) ions may pass in accordance with their electrochemical gradient.

SLC9A9: Solute carrier family 9 (sodium/hydrogen exchanger), member 9; May act in electroneutral exchange of protons for Na(+) across membranes. Involved in the effusion of Golgi luminal H(+) in exchange for cytosolic cations.Involved in organelle ion homeostasis by contributing to the maintainance of the unique acidic pH values of the Golgi and post-Golgi compartments in the cell.

3.2 GENE-GENE INTERACTION FOR DEMENTIA

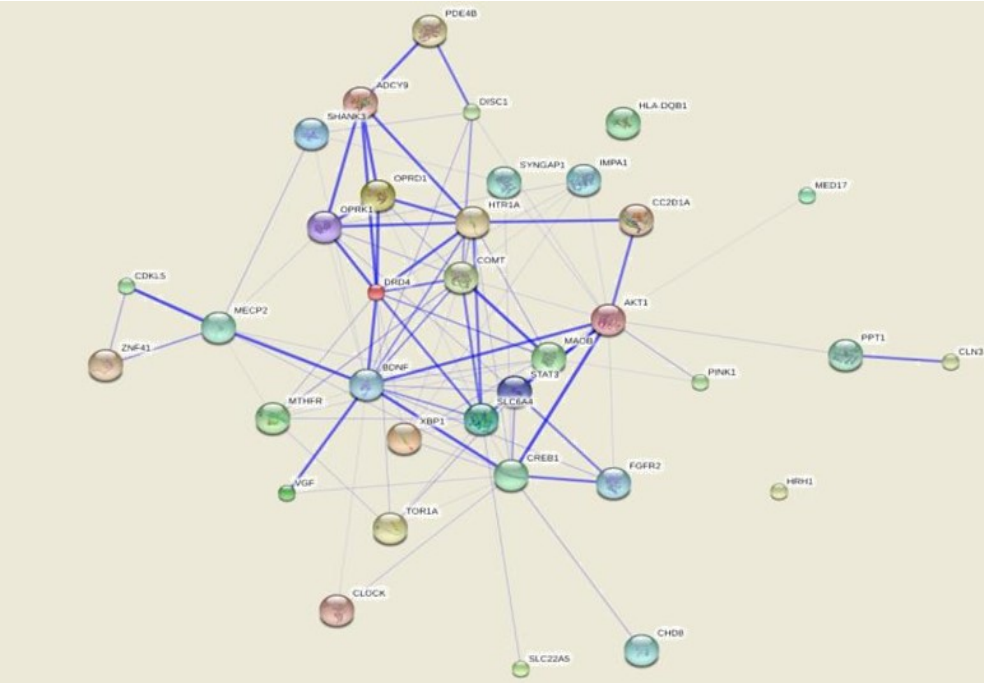

Fig 1.2:Thicker lines represent strong interaction between the genes.

. **PDGFRB:** Platelet-derived growth factor receptor, beta polypeptide; Receptor that binds specifically to PDGFB and PDGFD and has a tyrosine-protein kinase activity. Phosphorylates Tyr residues at the C-terminus of PTPN11 creating a binding site for the SH2 domain of GRB2

HTRA1: HtrA serine peptidase 1; Protease that regulate the availability of IGFs by cleaving IGF-binding proteins

CLN3: ceroid-lipofuscinosis, neuronal 3

3.3 GENE-GENE INTERACTION FOR MOOD DISORDER

The Protein-Protein interaction sub- networks for target and reference set is shown which guides us to analyse the interacted genes for MOOD DISORDER. Disease whose target genes is tested.

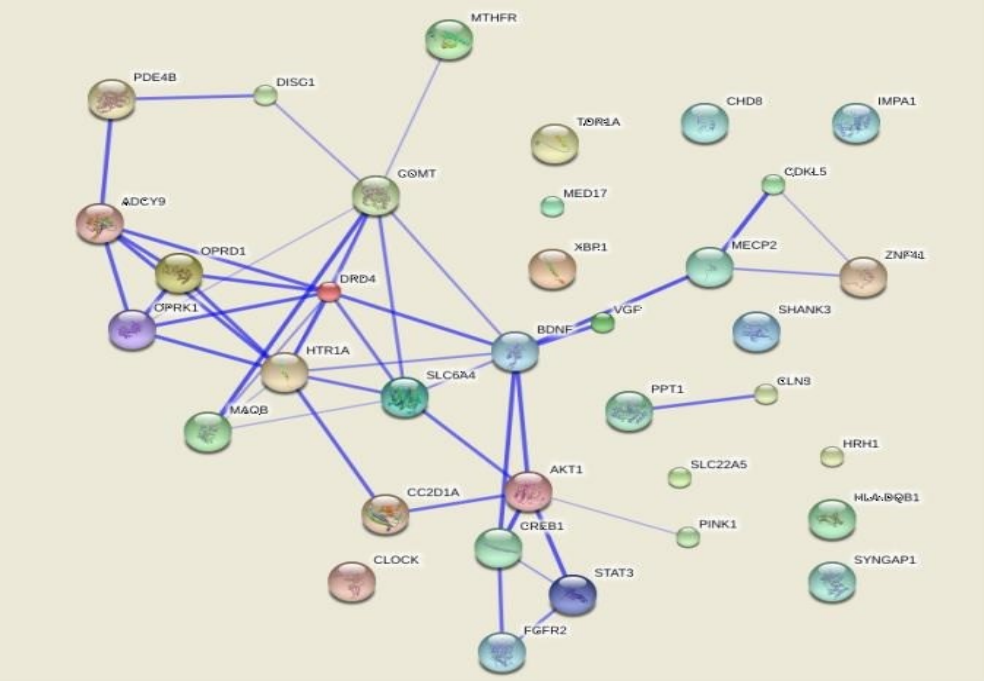

Fig 1.3 :Thicker lines represent strong interaction between the genes.

. **PDE4B:** Phosphodiesterase 4B, cAMP-specific (phosphodiesterase E4 dunce homolog, Drosophila); May be involved in mediating central nervous system effects of therapeutic agents ranging from antidepressants to antiasthmatic and anti-inflammatory agents

DISC1: Disrupted in schizophrenia 1

ADCY9: Adenylate cyclase 9; May play a fundamental role in situations where fine interplay between intracellular calcium and cAMP determines the cellular function. May be a physiologically relevant docking site for calcineurin (By similarity)

COMT: catechol-O-methyltransferase; Catalyzes the O-methylation, and thereby the inactivation, of catecholamine neurotransmitters and catechol hormones. Also shortens the biological half-lives of certain neuroactive drugs, like L-DOPA, alpha-methyl DOPA and isoproterenol

MTHFR: 5,10-methylenetetrahydrofolate reductase (NADPH); Catalyzes the conversion of 5,10- methylenetetrahydrofolate to 5-methyltetrahydrofolate, a co- substrate for homocysteine remethylation to methionine

OPRK1: Opioid receptor, kappa 1; Inhibits neurotransmitter release by reducing calcium ion currents and increasing potassium ion conductance. Receptor for dynorphins.May play a role in arousal and regulation of autonomic and neuroendocrine functions

OPRD1: Opioid receptor, delta 1; Inhibits neurotransmitter release by reducing calcium ion currents and increasing potassium ion conductance.Highly stereoselective.receptor for enkephalins

MAOB: Monoamine oxidase B; Catalyzes the oxidative deamination of biogenic and xenobiotic amines and has important functions in the metabolism of neuroactive and vasoactive amines in the central nervous system and peripheral tissues.MAOB preferentially degrades benzylamine and phenylethylamine

HTR1A: 5-hydroxytryptamine (serotonin) receptor 1A; This is one of the several different receptors for 5- hydroxytryptamine (serotonin), a biogenic hormone that functions as a neurotransmitter, a hormone, and a mitogen. The activity of this receptor is mediated by G proteins that inhibit adenylate cyclase activity

DRD4: Dopamine receptor D4

CC2D1A: coiled-coil and C2 domain containing 1A; Transcription factor that binds specifically to the FRE (five repressor element) and represses HTR1A gene transcription in neuronal cells. The combination of calcium and ATP specifically inactivates the binding with FRE. May play a role in the altered regulation of HTR1A associated with anxiety and major depression. Mediates HDAC- independent repression of HTR1A promoter in neuronal cell

SLC6A4: solute carrier family 6 (neurotransmitter transporter, serotonin), member

4;Terminates the action of serotonin by its high affinity sodium-dependent reuptake into presynaptic terminals

BDNF: brain-derived neurotrophic factor; During development, promotes the survival and differentiation of selected neuronal populations of the peripheral and central 9 nervous systems. Participates in axonal growth, pathfinding and in the modulation of dendritic growth and morphology.Major regulator of synaptic transmission and plasticity at adult synapses in many regions of the CNS. The versatility of BDNF is emphasized by its contribution to a range of adaptive neuronal responses including long-term potentiation (LTP), long-term depression(LTD), certain forms of short-term synaptic plastic.

VGF: VGF nerve growth factor inducible; May be involved in the regulation of cell-cell interactions or in synatogenesis during the maturation of the nervous system

MECP2: methyl CpG binding protein 2 (Rett syndrome); Chromosomal protein that binds to methylated DNA. It can bind specifically to a single methyl-CpG pair. It is not influenced by sequences flanking the methyl-CpGs. Mediates transcriptional repression through interaction with histone deacetylase and the corepressor SIN3A

ZNF41: zinc finger protein 41; May be involved in transcriptional regulation

AKT1: v-akt murine thymoma viral oncogene homolog 1; General protein kinase capable of phosphorylating several known proteins. Phosphorylates TBC1D1. Signals downstream of phosphatidylinositol 3-kinase (PI(3)K) to mediate the effects of various growth factors such as platelet-derived growth factor (PDGF),epidermal growth factor (EGF), insulin and insulin-like growth factor I (IGF-I).Plays a role in glucose transport by mediating insulin-induced translocation of the GLUT4 glucose transporter to the cell surface. Mediates the antiapoptotic effects of IGF-I. Mediates insulin-stimulated protein.

CREB1: cAMP responsive element binding protein 1; This protein binds the cAMP response element (CRE), a sequence present in many viral and cellular promoters. CREB stimulates transcription on binding to the CRE. Transcription activation is enhanced by the TORC coactivators which act independently of Ser-133 phosphorylation. Implicated in synchronization of circadian rhythmicity

STAT3: signal transducer and activator of transcription 3 (acute-phase response factor);Transcription factor that binds to the interleukin-6 (IL-6)-responsive elements identified in the promoters of various acute-phase protein genes. Activated by 10 IL31 through IL31RA FGFR2fibroblast growth factor receptor 2; Receptor for acidic and basic fibroblast growth factors

PPT1 palmitoyl-protein thioesterase 1; Removes thioester-linked fatty acyl groups such as palmitate from modified cysteine residues in proteins or peptides during lysosomal degradation. Prefers acyl chain lengths of 14 to 18 carbons

CLN3: ceroid-lipofuscinosis, neuronal 3

3.4 GENE-GENE INTERACTION FOR OCD

The Protein-Protein interaction sub- networks for target and reference set is shown which guides us to analyse the interacted genes for OCD Disease whose target genes is tested.

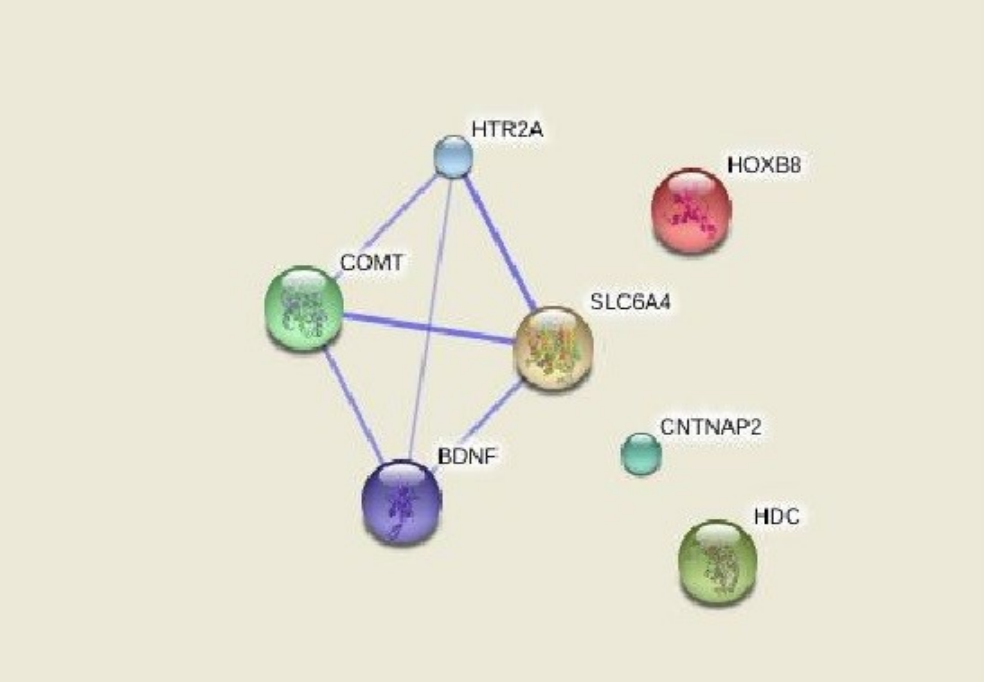

Fig 1.4 :Thicker lines represent strong interaction between the genes.

. **HTR2A:** 5-hydroxytryptamine (serotonin) receptor 2A; This is one of the several different receptors for 5- hydroxytryptamine (serotonin), a biogenic hormone that functions as a neurotransmitter, a hormone, and a mitogen. This receptor mediates its action by association with G proteins that activate a phosphatidylinositol-calcium second messenger system. This receptor is involved in tracheal smooth muscle contraction, bronchoconstriction, and control of aldosterone production

COMT: catechol-O-methyltransferase; Catalyzes the O-methylation, and thereby the inactivation, of catecholamine neurotransmitters and catechol hormones. Also shortens the biological half-lives of certain neuroactive drugs, like L-DOPA, alpha-methyl DOPA and isoproterenol

SLC6A4: solute carrier family 6 (neurotransmitter transporter, serotonin), member 4;Terminates the action of serotonin by its high affinity sodium-dependent reuptake into presynaptic terminals

BDNF: brain-derivedneurotrophic factor; During development, promotes the survival and differentiation of selected neuronal populations of the peripheral and central nervous systems. Participates in axonal growth, pathfinding and in the modulation of dendritic growth and morphology.Major regulator of synaptic transmission and plasticity at adult synapses in many regions of the CNS. The versatility of BDNF is emphasized by its contribution to a range of adaptive neuronal responses including long-term potentiation (LTP), long-term depression (LTD), certain forms of short-term synaptic plastic

3.5 GENE-GENE INTERACTION FOR SCHIZOPHRENIA

The Protein-Protein interaction sub- networks for target and reference set is shown which guides us to analyse the interacted genes for SCHIZOPHRENIA. Disease whose target genes is tested.

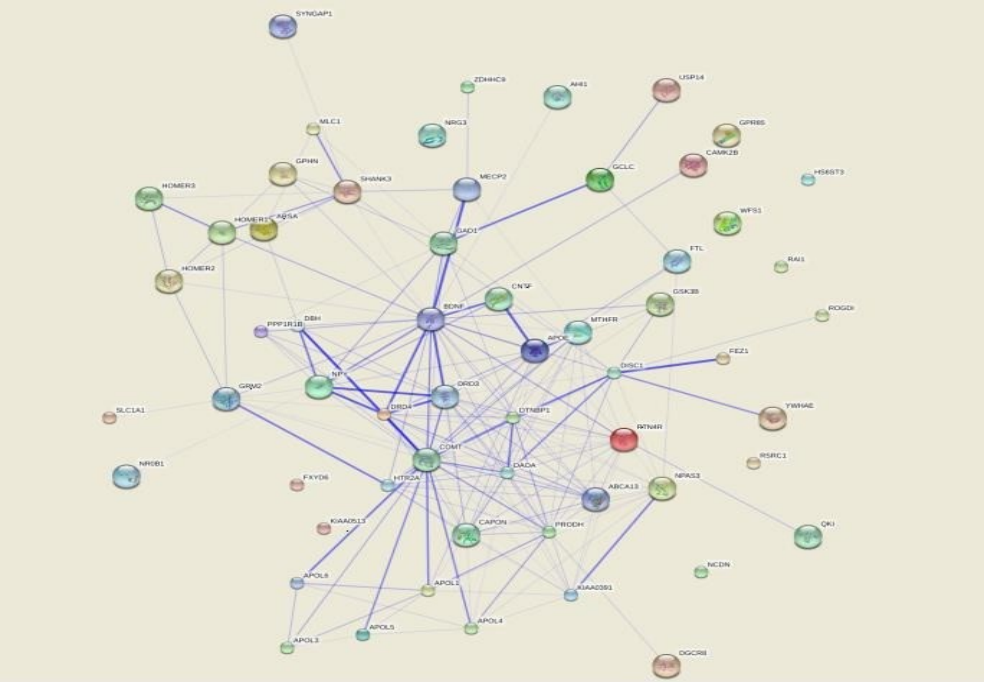

Fig 1.5:Thicker lines represent strong interaction between the genes.

. **MECP2:** methyl CpG binding protein 2 (Rett syndrome); Chromosomal protein that binds to methylated DNA. It can bind specifically to a single methyl-CpG pair. It is not influenced by sequences flanking the methyl-CpGs. Mediates transcriptional repression through interaction with histone deacetylase and the corepressor SIN3A

GOLG: A-7 complex is a palmitoyltransferase specific for HRAS and NRAS

GAD1: glutamate decarboxylase 1 (brain, 67kDa); Catalyzes the production of GABA

APOE: apolipoprotein E; Mediates the binding, internalization, and catabolism of lipoprotein particles. It can serve as a ligand for the LDL (apo B/E) receptor and for the specific apo-E receptor (chylomicron remnant) of hepatic tissues

MTHFR: 5,10-methylenetetrahydrofolate reductase (NADPH); Catalyzes the conversion of 5,10- methylenetetrahydrofolate to 5-methyltetrahydrofolate, a co- substrate for homocysteine remethylation to methionine

DRD4: dopamine receptor D4

FEZ1: fasciculation and elongation protein zeta 1 (zygin I); May be involved in axonal outgrowth as component of the network of molecules that regulate cellular morphology and axon guidance machinery. Able to restore partial locomotion and axonal fasciculation to C.elegans unc-76 mutants in germline transformation experiments

DISC1: disrupted in schizophrenia 1

APOL1: apolipoprotein L, 1; May play a role in lipid exchange and transport throughout the body. May participate in reverse cholesterol transport from peripheral cells to the liver

APOL4: apolipoprotein L, 4; May play a role in lipid exchange and transport throughout the body. May participate in reverse cholesterol transport from peripheral cells to the liver

APOL5: apolipoprotein L, 5; May affect the movement of lipids in the cytoplasm or allow the binding of lipids to organelles

APOL6: apolipoprotein L, 6; May affect the movement of lipids in the cytoplasm or allow the binding of lipids to organelles

NPY: neuropeptide Y; NPY is implicated in the control of feeding and in secretion of gonadotrophin-release hormone

DTNBP1: dystrobrevin binding protein 1; Plays a role in the biogenesis of lysosome-related organelles such as platelet dense granule and melanosomes

HTR2A: 5-hydroxytryptamine (serotonin) receptor 2A; This is one of the several different receptors for 5- hydroxytryptamine (serotonin), a biogenic hormone that functions as a neurotransmitter, a hormone, and a mitogen. This receptor mediates its action by association with G proteins that activate a phosphatidylinositol-calcium second messenger system. This receptor is involved in tracheal smooth muscle contraction, bronchoconstriction, and control of aldosterone production

YWHAE: tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein,epsilon polypeptide; Adapter protein implicated in the regulation of a large spectrum of both general and specialized signaling pathway. Binds to a large number of partners, usually by recognition of a phosphoserine or phosphothreonine motif. Binding generally results in the modulation of the activity of the binding partner

4 DISEASE GENE PRIORITIZATION USING RANDOM WALK WITH RESTART

RANDOM WALK WITH RESTART FOR DISEASE GENE PRIORTIZATION

- Method is used to prioritize genes based on cumulative interaction score corresponding to mapped genes of target and reference in a pathway regarding disease. The network is traversed starting from a gene to find its cumulative score with high restart probability to find its cumulative score in interaction of pathway.

RANDOM WALK WITH RESTART STEPS

- INPUT-List of target genes/proteins L, list of reference datasets P, molecular interacion adjacency matrix A for graph G=V,E, restart probability p(0.9)
- OUTPUT-vector of cummulative distance scores for each reference dataset in P
- STEP1-map pathway sets P and the gene/protein list L onto graph G
- STEP2-v:=vector of length IVI with enies for mapped elements of L set o 1, otherwise 0;
- STEP3- u:=v;
- STEP4- u(old):= vector of length IVI with all entries set to 0;
- STEP5- A:= normalize(A), so that each column sums to 1;
- STEP6-while (sum|u-u(old)|)>=1E-06,do
  – U(old):=u;
  – u:=(1-p)Au(old) + pv;
- STEP7- distance scores:= vector of length IPI;
- STEP8-for i<= 1 to IPI do
  – Distance$_s cores[i] := 1 - u[P[i]], converttodistancescores$

4.1 RANDOM WALK WITH RESTART-ADHD

INPUT-L= 21 GENE-GIT1 SLC9A9 RAI1 CORO1A CSNK1D DYRK1A SCN8A SNAP25 FGD1 TH TPH1 HTR1B SLC6A2 NF1 SLC6A3 DRD5 DRD4 COMT CES1 ADRA2A

- P=124 gene- parkinson's data is similar so need to be tested all its genes .
- includes 3 target genes- (SLC6A3 ,TH , COMT)
- Mapped number of genes in desired pathways=3 i.e. (SLC6A3 ,TH , COMT)
- Now v=145 in which 3 genes are target and is used for random walk over the v nodes. STEP1: map pathway sets P and L onto graph G
  STEP2:V=[10100010000100001111111]
  STEP3:U=[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0], (U initiazes to V) STEP 4: Uold=[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0], STEP5:The normalization table A is given under:

ADHD G	ADRA2A	CES1	COMT	CORO1A	CSNK1D	DRD4	DRD5	DYRK1A	FGD1	GIT1	HTR1B	NF1	RAI1	RTEL1	SCN8A	SLC6A2	SLC6A3	SLC9A9	SNAP25	TH	TPH1
ADRA2A	0	0	0.0911	0	0	0	0	0	0	0	0.22	0	0	0	0	0.16	0	0.51	0	0.17	0
CES1	0	0	0.113	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COMT	0	0	0	0	0	0	0	0	0	0	0.09	0	0	0	0	0.21	0	0	0	0.17	0
CORO1A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CSNK1D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DRD4	0.707584	0	0.218	0	0	0	0.16	0	0	0	0.22	0	0	0	0	0.16	0.52	0	1	0.14	1
DRD5	0	0	0.15	0	0	0	0	0	0	0	0	0	0	0	0	0.15	0	0	0	0.09	0
DYRK1A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FGD1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GIT1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HTR1B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.09	0	0	0	0	0
NF1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RAI1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RTEL1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SCN8A	0	0	0.07	0	0	0	0	0	0	0	0.09	0	0	0	0	0	0	0	0	0	0
SLC6A2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.19	0
SLC6A3	0.292416	0	0.21	0	0	0	0.48	0	0	0	0.1	0	0	0	0	0	0	0.43	0	0.25	0
SLC9A9	0	0	0	0	0	0	0	0	0	0	0.05	0	0	0	0	0	0	0	0	0	0
SNAP25	0	0	0	0	0	0	0.24	0	0	0	0.06	0	0	0	0	0.09	0.17	0	0	0	0
TH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TPH1	0	0	0.15	0	0	0	0.13	0	0	0	0.16	0	0	0	0	0.16	0.3	0	0	0	0

4.1.1 FILTERS FOR SCORING INTERACTION

In table 1.1, we have taken the set of candidate genes and they are scored with the reference gene lists in which the filter helps us give us the slected number of genes which are interacted with the filtered candidate genes.

This filter scoring is taken from STRING database and it is then used to update the scores with the target genes to form higher interactive network by the formula:

u:=(1-p)Au(old)+pv.

STEP 7: distance score minimization and interaction score maximization

STEP 8: distance score for all nodes is taken and interaction score is listed by STRING.

Using gene-gene interaction score , we first find significant disease pathway and overall gene in the pathway network and then mapping our candidate gene, we find the highest correlated genes that are called the ‘target genes’.

#node	node2	combined_sco
HTR1B	ADRA2A	0.906
COMT	SCN8A	0.247
COMT	TPH1	0.532
SLC6A2	SNAP25	0.235
TH	DRD4	0.514
TH	DRD5	0.318
SLC6A2	HTR1B	0.262
HTR1B	COMT	0.374
COMT	DRD5	0.54
HTR1B	TPH1	0.675
SLC6A3	DRD4	0.965
SCN8A	DRD5	0.286
SCN8A	SLC6A3	0.229
SCN8A	SLC9A9	0.21
COMT	CES1	0.407
DRD5	SNAP25	0.371
COMT	DRD4	0.784
SLC6A2	ADRA2A	0.428
SNAP25	DRD4	0.34
SLC9A9	SLC6A3	0.269
TH	SLC6A2	0.693
SCN8A	DRD4	0.296
SCN8A	ADRA2A	0.462
DRD5	DRD4	0.253

Table 1.1 : score filters for candidate genes.

.

4.1.2 INTERACTION SCORE FOR ADHD

In table 1.2, the interaction score is calculated for each genes through RWR and each nodes(genes) are taken with equal probability. Distance score is minimized and interaction score is maximized for ADHD.

ADHD GEI	ADRA2	CES1	COMT	CORO1A	CSNK	DRD4	DRD5	DYRK1A	FGD1	GIT1	HTR1B	NF1	RAI1	RTEL1	SCN8A	SLC6A2	SLC6A3	SLC9A9	SNAP25	TH	TPH1
ADRA2A	0	0	0.327	0	0	0	0	0	0	0	0.906	0	0	0	0	0.428	0	0.281	0	0.613	0
CES1	0	0	0.407	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COMT	0	0	0	0	0	0	0	0	0	0	0.374	0	0	0	0	0.567	0	0	0	0.613	0
CORO1A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CSNK1D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DRD4	0.905	0	0.784	0	0	0	0.253	0	0	0	0.911	0	0	0	0	0.432	0.965	0	0.34	0.514	0.581
DRD5	0	0	0.54	0	0	0	0	0	0	0	0	0	0	0	0	0.399	0	0	0	0.318	0
DYRK1A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FGD1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GIT1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HTR1B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.262	0	0	0	0	0
NF1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RAI1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RTEL1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SCN8A	0	0	0.247	0	0	0	0	0	0	0	0.371	0	0	0	0	0	0	0	0	0	0
SLC6A2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.693	0
SLC6A3	0.374	0	0.752	0	0	0	0.75	0	0	0	0.43	0	0	0	0	0	0	0.269	0	0.898	0
SLC9A9	0	0	0	0	0	0	0	0	0	0	0.228	0	0	0	0	0	0	0	0	0	0
SNAP25	0	0	0	0	0	0	0.371	0	0	0	0.263	0	0	0	0	0.235	0.319	0	0	0	0
TH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TPH1	0	0	0.532	0	0	0	0.202	0	0	0	0.675	0	0	0	0	0.428	0.563	0	0	0	0

Table 1.2: The interaction score for ADHD

4.1.3 UPDATED INTERACTION SCORE FOR ADHD

In Table 1.3 : The interaction score is updated and summed for each genes through RWR having interactions with the target genes for ADHD.

ADHD GENE	TH	SLC6A3	COMT	INTERACTION SCORE SUM
ADRA2A	0.613	0	0.327	0.94
CES1	0	0	0.407	0.407
COMT	0.613	0	0	0.613
CORO1A	0	0	0	0
CSNK1D	0	0	0	0
DRD4	0.514	0.965	0.784	2.263
DRD5	0.318	0	0.54	0.858
DYRK1A	0	0	0	0
FGD1	0	0	0	0
GIT1	0	0	0	0
HTR1B	0	0	0	0
NF1	0	0	0	0
RAI1	0	0	0	0
RTEL1	0	0	0	0
SCN8A	0	0	0.247	0.247
SLC6A2	0.693	0	0	0.693
SLC6A3	0.898	0	0.752	1.65
SLC9A9	0	0	0	0
SNAP25	0	0.319	0	0.319
TH	0	0	0	0
TPH1	0	0.563	0.532	1.095

Table 1.3: The interaction score updation for ADHD.

Here, the target gene TH score is taken for the candidate genes ADRA2A, COMT, DRD4, DRD5, SLC6A2 and SLC6A3. The target gene SLC6A3 score is taken for the candidate gene DRD1. The target gene COMT score is taken for the candidate genes ADRA2A, CES1, COMT, DRD4, DRD5, SCN8A,SLC6A2 and SLC6A3.

4.1.4 PRIORITIZED GENES FOR ADHD

In Table 1.4, genes expanded Sorting for ADHD is done and the highly interacted genes are sorted with the gene names and prioritized in network generation resulting for ADHD.

The genes are prioritized according to their overall interaction score during RWR from a candidate gene to all target genes. Adding scores corresponding to walk of the target genes is presented in Table 1.2.

S.NO.	ADHD GENE	SUM
1	DRD4	2.263
2	SLC6A3	1.65
3	TPH1	1.095
4	ADRA2A	0.94
5	DRD5	0.858
6	SLC6A2	0.693
7	COMT	0.613
8	CES1	0.407
9	SNAP25	0.319
10	SCN8A	0.247
11	CORO1A	0
12	CSNK1D	0
13	DYRK1A	0
14	FGD1	0
15	GIT1	0
16	HTR1B	0
17	NF1	0
18	RAI1	0
19	RTEL1	0
20	SLC9A9	0
21	TH	0

Table 1.4 : Genes expanded Sorting for ADHD

4.1.5 RESULTS AND DISCUSSIONS

Firstly, Section 1.1.1.1 gives the interaction scores of the 21 genes. Secondly, in Section 1.1.1.2, the target genes (3 genes i.e. TH, SLC6A3 and COMT) are selected from the pathway interaction use from ENRICHNET tools and the interaction score is calculated with RWR method. Finally, in Section 1.1.1.3, the scores are updated according to the targeted. Here, it is total 10 genes to interact and fame the network for Disease prioritization.

4.2 RANDOM WALK WITH RESTART-DEMENTIA

- L= 71 GENE-MAPT VCP PRNP TARDBP GRN PSEN1 SNCA TNF FUS APP CETP SNCB CLN6 C10ORF2 DCTN1 PSEN2 LAMA2 DNMT1 CP SERPINA3 DARC LRRK2 PINK1 NPC1 ARSA EPM2A CLN3 EIF2B5 HSD11B1 PLP1 ARX POLG PDGFRB FTL ANG ROGDI C9ORF72 OTUD4 DNAJC5 NPC1 SAITOHIN NPEPPS CYP27A1 PARK7 NCKAP1 CXCL11 APBA3 TYROBP SNCAIP CTSF SNCG APBB1 KLK6 SORT1 NEDD4 HTRA1 NOVA2 SIGMAR1 ISCHEMIC HAP1 DYRK1A CXCL12 VSNL1 MEOX2 RELN NR1H2 NOTCH3 HNRNPA2B1 PARK2 KSS FMR1
- P=124 GENE- Parkinson's data is similar so need to be tested all its genes which includes 3 target genes- (PARK7 ,LRRK2 , SNCA)
- Mapped number of genes in desired pathways=3 i.e. (PARK7 ,LRRK2 , SNCA) now v=195 in which 3 genes are target and is used for random walk over the v nodes.
  Similarly, the algorithm is executed for Dementia, the result comes with interaction score, then updation and then prioritization of gene network for Disease.
  Using gene-gene interaction score , we first find significant disease pathway and overall gene in the pathway network and then mapping our candidate gene, we find the highest correlated genes that are called the 'target genes'.

4.2.1 INTERACTION SCORE FOR DEMENTIA

In the interaction score is calculated for each genes through RWR and each nodes(genes) are taken with equal probability. Distance score is minimized and interaction score is maximized for DEMENTIA.

4.2.2 UPDATED INTERACTION SCORE FOR DEMENTIA

In the interaction score is updated and summed for each genes through RWR having interactions with the target genes for DEMENTIA

Here, the target gene PARK7 score is taken for the candidate genes PSEN1, SNCA and LRRK2. The target gene LRRK2 score is taken for the candidate genes TARDBP, GRN and APP. The target gene PARK2 score is taken for the candidate gene SNCB. The target gene LRRK2 score is taken for the candidate genes TARDBP, GRN and APP. The target gene PARK2 score is taken for the candidate gene SNCB. The target gene PINK1 score is taken for the candidate genes PSEN1, SNCA and LRRK2. The target gene SNCA score is taken for the candidate gene TARDBP, GRN, PSEN1, APP, SNCB and LRRK2. The target gene SNCAIP score is not taken as it is not interacting with other candidate genes. The target gene APBB1 score is taken for the candidate gene APP. The target gene APP score is taken for the candidate genes GRN , CP and SERPINA3. The target gene PSEN1 score is taken for the candidate genes GRN, APP, SNCB,LRRK2 and HSD11B1. The target gene PSEN2 score is

taken for the candidate genes TARDBP, GRN, SNCA, TNF, APP, SNCB, LRRK2, CLN3 and HSD11B1.

4.2.3 PRIORITIZED GENES FOR DEMENTIA

In genes expanded Sorting for Dementia is done and the highly interacted genes are sorted with the gene names and prioritized in network generation resulting for Dementia.

S.NO.	Dementia	Interaction score sum
1	APP	1.154
2	LRRK2	2.815
3	SNCAIP	2.445
4	NOTCH3	2.156
5	SNCA	1.948
6	GRN	1.547
7	PSEN1	1.518
8	SNCB	1.512
9	TARDBP	1.243
10	KLK6	0.969
11	PARK2	0.946
12	APBB1	0.92
13	SORT1	0.907
14	MAPT	0.835
15	PARK7	0.778
16	HSD11B1	0.695
17	APBA3	0.636
18	SERPINA3	0.629
19	NEDD4	0.563
20	NR1H2	0.427

S.NO.	Dementia	Interaction score sum
21	CP	0.272
22	CLN3	0.266
23	POLG	0.24
24	VCP	0
25	PRNP	0
26	TNF	0
27	FUS	0
28	CETP	0
29	CLN6	0
30	C10ORF2	0
31	DCTN1	0
32	PSEN2	0
33	LAMA2	0
34	DNMT1	0
35	DARC	0
36	PINK1	0
37	NPC1	0
38	ARSA	0
39	EPM2A	0
40	EIF2B5	0
41	PLP1	0
42	ARX	0
43	PDGFRB	0
44	FTL	0
45	ANG	0
46	ROGDI	0
47	C9ORF72	0
48	OTUD4	0
49	DNAJC5	0

S.NO.	Dementia	Interaction score sum
50	NPC1	0
51	SAITOHIN	0
52	NPEPPS	0
53	CYP27A1	0
54	NCKAP1	0
55	CXCL11	0
56	TYROBP	0
57	CTSF	0
58	SNCG	0
59	HTRA1	0
60	NOVA2	0
61	SIGMAR1	0
62	ISCHEMIC	0
63	HAP1	0
64	DYRK1A	0
65	CXCL12	0
66	VSNL1	0
67	MEOX2	0
68	RELN	0
69	HNRNPA2B1	0
70	KSS	0
71	FMR1	0

Table 1.5: Genes expanded Sorting for Dementia

4.2.4 RESULTS AND DISCUSSIONS

Firstly, Section 1.1.2.1 gives the interaction scores of the 71 genes. Secondly, in Section 1.1.2.2, the target genes (10 genes i.e. PARK7, LRRK2, PARK2, PINK1, SNCA, SNCAIP, APBB1, APP, PSEN1, andPSEN2) are selected from the pathway interaction use from ENRICHNET tools and the interaction score is calculated with RWR method. Finally, in Section 1.1.2.3, the scores are updated according to the targeted. Here, it is total 23 genes to interact and fame the network for Disease prioritization.

4.3 RANDOM WALK WITH RESTART-MOOD DISORDER

- L= 46 GENE-CHD8 PINK1 MTHFR SHANK3 DISC1 TOR1A HLA-DQB1 SYNGAP1 SLC22A5 IMPA1 CLOCK ZNF41 CDKL5 MECP2 CLN3 CHAC XBP1 SLC6A4 PD AKT1 DRD4 CREB1 COMT BDNF HTR1A STAT3 ANXIETY DISC1 ADCY9 VGF IMPA1 PPT1 HRH1 PDE4B ZNF41 MAOB CDKL5 CHAC XBP1 FGFR2 OPRK1 OPRD1 PINK1 CLN3 DRD4 COMT
- P=124 GENE- Parkinson's data is similar so need to be tested all its genes which includes 3 target genes- (MAOB, AKT1, BDNF)
- Mapped number of genes in desired pathways=3 i.e. (MAOB, AKT1, BDNF)
- Now v=170 in which 3 genes are target and is used for random walk over the v nodes.
  Similarly, the algorithm is executed for Mood Disorder, the result comes with interaction score, then updation and then prioritization of gene network for disease Using gene-gene interaction score , we first find significant disease pathway and overall gene in the pathway network and then mapping our candidate gene, we find the highest correlated genes that are called the 'target genes'.

4.3.1 INTERACTION SCORE FOR MOOD DISORDER

The interaction score is calculated for each genes through RWR and each nodes(genes) are taken with equal probability. Distance score is minimized and interaction score is maximized for MOOD DISORDER.

4.3.2 UPDATED INTERACTION SCORE FOR MOOD DISORDER

In the interaction score is updated and summed for each genes through RWR having interactions with the target genes for MOOD DISORDER.

The target gene MAOB score is taken for the candidate genes MTHFR, TOR1A, SLC6A4, DRD4, COMT, HTR1A, and PINK1. The target gene AKT1 score is taken

for the candidate gene XBP1 and SLC6A1. Similarly, it is done for all candidate genes in respect of the target genes.

4.3.3 PRIORITIZED GENES FOR MOOD DISORDER

In genes expanded Sorting for MOOD DISORDER is done and the highly interacted genes are sorted with the gene names and prioritized in network generation resulting for MOOD DISORDER.

S.NO.	mood	INTERACTION SCORE SUM
1	AKT1	2.34
2	SLC6A4	2.115
3	COMT	1.582
4	DRD4	1.453
5	DRD4	1.453
6	HTR1A	1.327
7	VGF	1.172
8	XBP1	0.954
9	CREB1	0.947
10	MECP2	0.946
11	STAT3	0.911
12	XBP1	0.621
13	COMT	0.609
14	TOR1A	0.593
15	CLOCK	0.518
16	MTHFR	0.462
17	DISC1	0.379
18	DISC1	0.379
19	FGFR2	0.368
20	MAOB	0.34
21	CDKL5	0.243
22	CDKL5	0.243
23	PINK1	0.219
24	OPRK1	0.21
25	SHANK3	0.2

S.NO.	mood	INTERACTION SCORE SUM
26	CHD8	0
27	PINK1	0
28	HLA-DQB1	0
29	SYNGAP1	0
30	SLC22A5	0
31	IMPA1	0
32	ZNF41	0
33	CLN3	0
34	CHAC	0
35	PD	0
36	BDNF	0
37	ANXIETY	0
38	ADCY9	0
39	IMPA1	0
40	PPT1	0
41	HRH1	0
42	PDE4B	0
43	ZNF41	0
44	CHAC	0
45	OPRD1	0
46	CLN3	0

Table 1.6: Genes expanded Sorting for MOOD DISORDER

4.3.4 RESULTS AND DISCUSSIONS

Firstly, Section 1.1.3.1 gives the interaction scores of the 46 genes. Secondly, in Section 1.1.3.2, the target genes (6 genes i.e. MAOB, AKT1, STAT3, BDNF, CREB1, and PINK1) are selected from the pathway interaction use from ENRICHNET tools and the interaction score is calculated with RWR method. Finally, in Section 1.1.3.3,

the scores are updated according to the targeted. Here, it is total 25 genes to interact and fame the network for Disease prioritization.

4.4 RANDOM WALK WITH RESTART-OCD

- L= 8 GENE-CNTNAP2 SLC6A4 HTR2A HOXB8 HDC COMT BDNF SLC6A4
- P=124 gene- Parkinson's data is similar so need to be tested all its genes which includes 3 target genes- (HDC, COMT, BDNF)
- Mapped number of genes in desired pathways=3 i.e. (HDC, COMT, BDNF)
- now v=132 in which 3 genes are target and is used for random walk over the v nodes.
  Similarly, the algorithm is executed for OCD, the result comes with interaction score, then updation and then prioritization of gene network for disease.
  Using gene-gene interaction score , we first find significant disease pathway and overall gene in the pathway network and then mapping our candidate gene, we find the highest correlated genes that are called the 'target genes'.

4.4.1 INTERACTION SCORE FOR OCD

In the interaction score is calculated for each genes through RWR and each nodes(genes) are taken with equal probability. Distance score is minimized and interaction score is maximized for OCD.

4.4.2 UPDATED INTERACTION SCORE FOR OCD

In The interaction score is updated and summed for each genes through RWR having interactions with the target genes for OCD.

The target gene HDC score is not taken as it is not interacting with other candidate genes. The target gene COMT score is taken for the candidate gene SLC6A1. The target gene BDNF score is taken for the candidate genes SLC6A4 , HTR2A and COMT.

4.4.3 PRIORITIZED GENES FOR OCD

In Genes expanded Sorting for OCD is done and the highly interacted genes are sorted with the gene names and prioritized in network generation resulting for OCD.

S.NO.	OCD	INTERACTION SCORE SUM
1	SLC6A4	1.367
2	SLC6A4	1.367
3	COMT	0.609
4	HTR2A	0.411
5	CNTNAP2	0
6	HOXB8	0
7	HDC	0
8	BDNF	0

Table 1.7: Genes expanded Sorting for OCD

4.4.4 RESULTS AND DISCUSSIONS

Firstly, Section 1.1.1.1 gives the interaction scores of the 8 genes. Secondly, in Section 1.1.1.2, the target genes (3 genes i.e. HDC, COMT and BDNF) are selected from the pathway interaction use from ENRICHNET tools and the interaction score is calculated with RWR method. Finally, in Section 1.1.1.3, the scores are updated according to the targeted. Here, it is total 4 genes to interact and fame the network for Disease prioritization.

4.5 RANDOM WALK WITH RESTART-SCHIZOPHRENIA

- L= 61 GENE-MTHFR PRODH HTR2A DRD3 RTN4R DISC1 SHANK3 BDNF DRD4 WFS1 ZDHHC9 NR0B1 MECP2 NPY FTL CNTF COMT SLC1A1 APOE ROGDI RSRC1 KIAA0513 KIAA0391 QKI NPAS3 HS6ST3 DBH DGCR8 AHI1 NCDN ABCA13 CAMK2B DAOA USP14 APOL6 APOL5 APOL4 APOL3 DTNBP1 GCLC FXYD6 MLC1 NOS1AP NRG3 GAD1 GPR85 YWHAE GSK3B FEZ1 HOMER3 HOMER2 HOMER1 PPP1R1B GRM2 GPHN APOL1 RAI1 ARSA MTHFR SHANK3 SYNGAP1
- P=124 gene- Parkinson's data is similar so need to be tested all its genes which includes 3 target genes- (GAD1, GSK3B,BDNF)
- Mapped number of genes in desired pathways=3 i.e. (GAD1, GSK3B,BDNF)
- Now v=185 in which 3 genes are target and is used for random walk over the v nodes.

  Similarly, the algorithm is executed for SCHIZOPHRENIA, the result comes with interaction score, then updation and then prioritization of gene network for Disease.

Using gene-gene interaction score , we first find significant disease pathway and overall gene in the pathway network and then mapping our candidate gene, we find the highest correlated genes that are called the ‘target genes’.

4.5.1 INTERACTION SCORE FOR SCHIZOPHRENIA

In the interaction score is calculated for each genes through RWR and each nodes (genes) are taken with equal probability. Distance score is minimized and interaction score is maximized for SCHIZOPHRENIA.

4.5.2 UPDATED INTERACTION SCORE FOR SCHIZOPHRENIA

The interaction score is updated and summed for each genes through RWR having interactions with the target genes for SCHIZOPHRENIA.

The target gene GAD1 score is taken for the candidate gene NPY, DTNBP1, GCLC, MLC1, PPP1R1B and GPHN. The target gene GSK3B score is taken for the candidate gene APOE. The target gene YWHAE score is not taken as it is not interacting with other candidate genes. Similarly, it is done for all candidate genes in respect of the target genes.

4.5.3 PRIORITIZED GENES FOR SCHIZOPHRENIA

In Genes expanded Sorting for SCHIZOPHRENIA is done and the highly interacted genes are sorted with the gene names and prioritized in network generation resulting for SCHIZOPHRENIA.

Table 1.8: Genes expanded Sorting for SCHIZOPHRENIA

S.NO.	SCHIZOPHRENIA	INTERACTION SCORE SUM
1	DRD4	1.668
2	NPY	1.434
3	DTNBP1	1.234
4	APOE	1.214
5	PPP1R1B	0.961
6	MECP2	0.946
7	DRD3	0.896
8	GCLC	0.8
9	GAD1	0.793
10	PRODH	0.764
11	RTN4R	0.718
12	APOL1	0.714
13	APOL4	0.679
14	GSK3B	0.658
15	APOL5	0.657
16	GPHN	0.657
17	COMT	0.609
18	APOL3	0.462
19	HOMER1	0.462
20	CAMK2B	0.43
21	HTR2A	0.411

22	GRM2	0.379
23	DBH	0.347
24	DAOA	0.34
25	ABCA13	0.297
26	HOMER2	0.241
27	AHI1	0.229
28	MLC1	0.203
29	MTHFR	0.202
30	MTHFR	0.202
31	SHANK3	0.2
32	SHANK3	0.2
33	DISC1	0
34	BDNF	0
35	WFS1	0
36	ZDHHC9	0
37	NR0B1	0
38	FTL	0
39	CNTF	0
40	SLC1A1	0
41	ROGDI	0
42	RSRC1	0
43	KIAA0513	0
44	KIAA0391	0
45	QKI	0
46	NPAS3	0
47	HS6ST3	0
48	DGCR8	0
49	NCDN	0
50	USP14	0
51	APOL6	0
52	FXYD6	0
53	NOS1AP	0

54	NRG3	0
55	GPR85	0
56	YWHAE	0
57	FEZ1	0
58	HOMER3	0
59	RAI1	0
60	ARSA	0
61	SYNGAP1	0

4.5.4 RESULTS AND DISCUSSIONS

Firstly, Section 1.1.9.1 gives the interaction scores of the 61 genes. Secondly, in Section 1.1.9.2, the target genes (6 genes i.e. GAD1, GSK3B, BDNF, YWHAE, CAMK2B and COMT) are selected from the pathway interaction use from ENRICHNET tools and the interaction score is calculated with RWR method. Finally, in Section 1.1.9.3, the scores are updated according to the targeted. Here, it is total 32 genes to interact and fame the network for Disease prioritization.

Taking 5 topmost genes that cause diseases from this chapter, the data are as below:

S.NO.	ADHD GENE	Dementia	mood disorder	OCD	SCHIZOPHRENIA
1	DRD4	APP	AKT1	SLC6A4	DRD4
2	SLC6A3	LRRK2	SLC6A4	COMT	NPY
3	TPH1	SNCAIP	COMT	HTR2A	DTNBP1
4	ADRA2A	NOTCH3	DRD4	CNTNAP2	APOE
5	DRD5	SNCA	HTR1A	BDNF	PPP1R1B

Table 1.9: 5 topmost genes that cause diseases

5 CONCLUSION

In the above sections, we have found the RWR method useful in scoring the interactions with the help of both STRING database and ENICHNET tool which covers the total allelic variant genes to be tested throughout the steps of initialization and getting scores for all the target genes in determining distance scores minimization and interaction score maximization from the reference genes availability. So, in this respect RWR method helps us to find genes and to reach to the conclusion of the highly interacted genes network formation for the above Disease prioritization. The prioritized genes can be used in any computational model for classification purpose.

References

[1] Josin, G.M. and P.F. Liddle, Neural network analysis of the pattern of functional connectivity between cerebral areas in schizophrenia. Biol Cybern, 2001. 84(2): p. 117-22.
[2] Ideker, T., Sharan, R.: Protein networks in disease. Genome research 18(4) (2008) 644– 652
[3] Karni, S., Soreq, H., Sharan, R.: A network-based method for predicting disease-causing genes. Journal of Computational Biology 16(2) (2009) 181–189.
[4] Vanunu, O., Magger, O., Ruppin, E., Shlomi, T., Sharan, R.: Associating genes and protein complexes with disease via network propagation. PLoS Comp. Bio. 6(1) (January 2010)
[5] Frith, C.D., et al., Regional brain activity in chronic schizophrenic patients during the performance of a verbal fluency task.Br J Psychiatry, 1995. 167(3): p. 343-9.
[6] Liddle, P.F., et al., Functional connectivity during word generation in schizophrenia. Scizophr Res, 1997. 23: p. 168.