UNRAVELLING THE MOLECULAR LINKAGE OF CO-MORBID DISEASES: DIABETES MELLITUS, HYPERTENSION AND CORONARY ARTERY DISEASE

Varun Moganti[1], Abhinandita Dash[2]

[1]Department of Biotechnology, University College of Engineering,JNTU-Kakinada,India
[2]Department of Biology, Prof CR.Rao AIMSCS, University of Hyderabad Campus, India

Abstract

ABSTRACT : *The incidence of Diabetes Mellitus (DM), Hypertension (HTN) and Coronary artery disease (CAD) in the country has increased alarmingly. Since decades DM and HTN have been proved to be independent risk factors for CAD. Gene and its regulatory action through a protein are vital for the normal metabolism. Any abnormality in regulation would lead to a disease. Our study used the principles of network biology to understand the comorbidity of diseases at the molecular level. We have collected disease genes of DM, HTN and CAD from various public databases and extracted genes common to all the three diseases. We constructed a biological network by considering the protein interaction data obtained from Human Protein Reference Database (HPRD).The network was validated using power law distribution and the genes were ranked using Centiscape. Finally we identified the crucial genes with literature validation which could play a major role in causing disease co-morbidity.*

Keywords –*Biological Network, Coronary Artery Disease, Diabetes Mellitus, Hypertension and Systems Biology*

---***---

1. INTRODUCTION

Diabetes Mellitus and Hypertension are well established independent risk factors for Coronary artery Disease. The co-existence of diabetes and hypertension would have negative impact on the prognosis and outcome of a patient. The presence of hypertension would accelerate the risk of coronary artery diseases. [Ref1] The underlying key pathophysiological mechanism would be vascular endothelial dysfunction, platelet aggregation and platelet dysfunction. [Ref2] The main goal of a physician is to maintain the blood glucose level and normotensive state in order to reduce the risk of cardiovascular diseases. There is always a chance of an economic burden for the treatment of diabetes mellitus and hypertension considering the existing scenario of our country. So the major focus on treatment modality would be prophylactic to prevent the risk of other disease prevalence.

The central dogma of molecular biology is Deoxyribonucleic Acid (DNA) which upon transcription produce Ribonucleic Acid (RNA) and RNA expresses or regulates a particular function through a protein. Proteins are major molecules that co-ordinate, regulate many functions in our body through various mechanisms. Protein interacts with other protein to perform many metabolic activities. A lot of credit goes in favour with the advent of high throughput experiments leading to lot of protein-protein interaction data.[Ref3]The well established high throughput experiments are the pull down assays, tandem affinity purification(TAP), yeast two hybrid(Y2H), mass spectrophotometry, microarrays, phage displays.[Ref4,12] There are many public repositories available which gives the interacting partners of the genes/protein. Protein-Protein Interaction (PPI) data sets could be retrieved from BIND [Ref5], STRING [Ref6], MINT [Ref7], DIP [Ref8], HPRD [Ref9] and many others.

Graph theory has wide range of applications in fundamental areas of mathematics, statistics, physics, chemistry and biology. As we all understand that human biological phenomenon is highly complex and it could be very well studied in elementary constituents, these elementary constituents interact in their own manner to bring about a normal regulation. For example the energy regulation through generation of ATP molecules are regulated in a chain like phenomenon such as Kreb's cycle which is in fact is a molecular and a metabolic phenomenon. Any disruption to the normal physiology leads to an abnormality which could be a pathophysiologic condition. All these reactions could be represented as a network. This approach of study is called as Systems Biology or Network Biology. This is a rapidly growing domain in bioinformatics which deals with Biological Networks. [Ref10]

There are various types of biological networks such as Gene Regulatory Networks, Cell Signaling Networks, PPI networks, Metabolic Networks and so on. [Ref11]. In the current scenario, PPI networks play a major role in Human Medicine and Molecular Biology. The application of protein interaction networks are: prediction of novel disease genes, prediction of genotype-phenotype associations, studying the genetic and molecular basis of diseases. [Ref12, Ref13]

A task of linking disease genes and associated human disorders is done to develop a human diseasome. This was constructed by using a bipartite graph between disease phenome and disease genome. Few key observations summarized in the study were genes contributing to a common disease have increased tendency to participate in PPI, tendency to express in a specific tissue, tend to display co-expression levels. [Ref14]

Human diseasome exploration would connect the small world, i.e. genes, proteins with global networks i.e. the diseases, mortalityand patient care. The overall benefit of studying the network would give rise to another domain called the “Network Medicine”. This would help the physicians and the biologists in understanding and reframe the existing treatment methods of a patient. This would benefit the patient care and would give birth to a highly specialized domain called the genomic medicine or personalized medicine. [Ref15]

Centrality measures are mostly computed in network biology depending upon the type of networks. The widely used network centralities are Degree, Eccentricity, Betweenness, Closeness, Eigen Vector, Radiality, Centroid value, Stress and so on. The biological significance of the centrality measures is well characterised and analysed as studied by Giovanni. [Ref16]

2. RELATED WORK:

The major challenge faced by biomedical researcher is prediction, identification of novel genes and protein function prediction. Role of biological networks particularly PPI and gene regulatory networks with the application of centrality measures have proved significant in various studies performed earlier. Potential of disease gene prediction and novel candidate gene identification could be possible by exploiting PPI networks. [Ref17]

Degree, Page Rank, Shortest path Betweenness, Closeness, Radiality, Integration, Katz Status Index and Motif based centralities were considered to be best centrality measures for gene regulatory networks. Top two per cent of the genes were considered as key regulatory players in gene regulatory network. [Ref18]For disease gene prioritization considering degree centrality would not favour the loosely connected or disconnected nodes within the network. Several statistical adjustment schemes were conducted to improve the performance of the Degree centrality. [Ref19]

Two hundred and seventy six novel cardiovascular candidate genes were identified by using a combined centrality measures such as Degree, Betweenness, Neighbour count of Disease gene, ratio of disease gene in neighbour, clustering co-efficient and Mean shortest path length of disease. [Ref20]

An attempt was made in ranking candidate genes and prioritizing them based on the microarray datasets and protein interaction network. Katz centrality index was proposed to this study and is applied for 40 diseases in 58 gene expression datasets. [Ref21] A database is developed on Autism and related neurological disease called Autworks. This is web application developed to view autism gene network structure within and related disorders associated with autism. [Ref22] New novel genes were identified in prostate cancer, by using centrality measures such as Degree, Eigen vector, Betweenness and Closeness. The data was extracted by using literature mining by using support vector machines. [Ref23]

A recent advance in applying Network medicine was drug repurposing or drug repositioning. It was conducted in two steps by constructing the protein interaction network with the common genes shared by two diseases. The second step would be identifying the drug target’s (gene/protein) presence within the protein interaction network. The presence of the drug target would make it a potential drug repositioning candidate. This would pave way to a rational approach in treating a multiple diseases with a single. [Ref24]

This paper aims to select three diseases which have a high rate of co-morbidity, identify the common genes, extract their protein interacting partners, constructing a protein interaction network and apply centrality measures to rank the crucial role players of the genes.

3. MATERIALS & METHODS:

3.1 Disease Gene Collection:

Genes known to be associated with the three different diseases (DM, CAD, HTN) were collected from various databases. The databases used are in Table1.

Table1: List of Databases used for Collection of Disease Genes

DATABASE	DM	HTN	CAD
GeneCards [Ref25]	Y	Y	Y
Eugenes [Ref26]	Y	Y	Y
OMIM [Ref27]	Y	Y	Y
Entrez [Ref28]	Y	Y	Y
Ensebml	Y	Y	Y

[Ref29]			
T-HOD [Ref30]	Y	Y	N
Uniprot [Ref31]	Y	Y	Y
KEGG [Ref32]	Y	Y	Y
HuGe Navigator [Ref33]	Y	Y	Y
GeneAtlas [Ref34]	Y	Y	Y
BioGUO [Ref35]	N	N	Y
DDBJ [Ref36]	N	Y	Y
GAD [Ref 37]	Y	Y	Y

3.2 Extraction of common genes:

Common genes for these three diseases were extracted using SQL query.

3.3 Interacting partners:

The interacting partners of the genes/proteins were collected using HPRD [Ref9].

3.4 Network Construction & Visualization:

Cytoscape, [Ref38] is an online visualization tool, used to construct a network. The network is constructed by removing the duplicate edges and self-loops.

3.5 Computation of network parameters:

Network parameters are computed using NetworkAnalyzer [Ref39]. It is a Cytoscape plugin that computes and displays a comprehensive set of topological parameters and centrality measures for undirected and directed networks, which includes the number of nodes, edges, and connected components, the network diameter, radius, density, centralization, heterogeneity, clustering coefficient, and the characteristic path length.

3.6 Computation of the centrality measures:

The centrality measure gives the significance of a particular node in the network. CentiScaPe, [Ref 40] a Cyoscape plugin, allows the user to compute the parameters in the network. In our analysis, we computed Degree, Closeness, Betweenness and Centroid value centralities after understanding the biological significance of the genes with respect to the diseases.

3.7 Ranking the genes:

The top hundred genes of each computed centrality measure were considered and a cumulative score was obtained by summing the independent scores of each centrality measures as considered in our study. The genes with maximum score were ranked higher.

3.8 Validation of the ranked genes:

The genes obtained were validated for literature based evidence through PubMed citation index number (PMID No.). The genes having evidence in all the three diseases were identified as comorbid genes.

RESULTS & DISCUSSION:

The number of genes for Diabetes Mellitus was four thousand three hundred and fourteen genes (4314). Hypertension was two thousand six hundred and seventy five(2675) and for Coronary Artery Disease was one thousand six hundred and ninety two (1692). A total of seven hundred and five four genes (754) were identified to be common among all the three diseases. These genes were extracted using SQL query. The nomenclature and the chromosomal position of these genes were listed using HGNC.[Ref41] The interacting proteins were collected from HPRD.

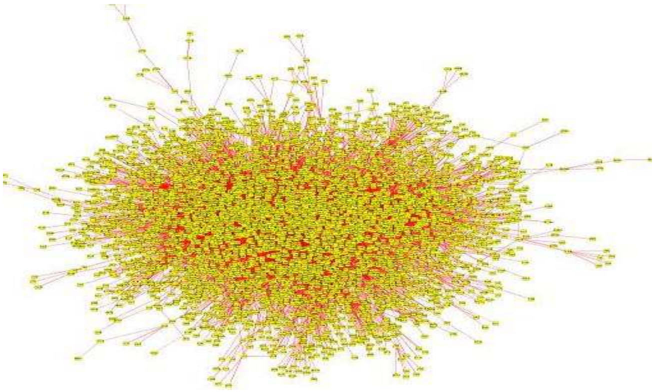

FIGURE1: Biological Network constructed using Cytoscape

The network was constructed and visualized using Cytoscape. The number nodes and edges were four thousand two hundred and eighty seven (4285) and ten thousand four hundred and sixty nine (10169) respectively. The network generated was complex. The network was further analysed using Network Analyzer, a Cytoscape plugin. The simple parameters generated are mentioned in the Table 2. The network was validated using power law distribution. Network Analyzer fits power-law by using least squares method. The parameters of power-law fit are:

$$y=ax^b$$

a=1234.2 b=-1.536

Correlation=0.998 R-squared=0.882

TABLE 2: List of Simple Parameters

S.No.	Parameter	Value
1	No. of connected components	49
2	Clustering co-efficient	0.093
3	Network Diameter	11
4	Network Radius	1
5	Network Density	0.001
6	Characteristic Path Length	4.190
7	Neighbourhood connectivity	4.883

The centrality measures were computed using CentiScape plugin of Cytoscape. Degree, Closeness, Betweenness and Centroid value were selected based upon the understanding of their biological significance particularly in a disease state. We selected only hundred genes falling in each centrality measure. Each gene was assigned a score and a rank was obtained based on the cumulative score of those centrality measures. Initially, we obtained a set of fifty five (55) genes that were ranked by each centrality measure. We validated those genes obtained, with literature evidence by using PubMed citation index. Finally we could identify a total of ten (10) genes showing evidence in all the three diseases.

In this study, we could identify genes which could lead to the co-morbidity among three diseases. These genes are to be studied in detail to understand their mechanisms and pathways associated in each disease. Their physiological role is also to be understood in order to have a clear picture on co-morbidity mechanisms. These genes could be analysed for expression analysis in every disease state to determine the pathophysiologic role of the gene. After expression analysis studies and their results, the following work could be taken one step ahead for therapeutic studies. Drug targets can be identified, prophylactic drugs can be developed, most significantly developing stage of other disease could be predicted and prevention of this disease could be possible.

CONCLUSION:

The dynamics of network biology has lead to many applications in the field of biology and medicine. Drug target repositioning, is now a much explored field as the concept of single drug-multiple diseases gained in lot of importance. This paper is a small attempt to understand the comorbid genes associated with three diseases and trying to identify the top genes with the application of graph theory principles and protein interaction networks. The novelty in our approach is to reduce the impact of complex network and extrapolating the comorbid genes and their role within the protein interaction network. We used the well studied centrality measures such as Degree, Betweenness, Closeness and Centroid value for analysing the networks based on their biological significance we adopted a ranking system by considering the cumulative scores of each centrality measure and assigned rank accordingly.

The efficacy of this work would be determined if people understand the significance of genes that play a role in comorbidity, analyse the expression data and evaluate for potential drug targets.

ACKNOWLEDGEMENTS:

Facilities provided by C.R.Rao Advanced Institute of Mathematics, Statistics and Computer Sciences, Idea and Encouragement by Dr.Ramesh Naidu ,Associate Professor of Medicine, King George Hospital, Andhra Medical College, Prof. Col.Allam Appa Rao Director -AIMSCS and Mr.Ramesh Malothu, Head Of Department , Department of Biotechnology, JawaharLal Nehru Technological University-Kakinada

SUPPLEMENTARY INFORMATION:

Supplementary Information 1: The common genes of all the three diseases, nomenclature & chromosomal location

Supplemenatry Information 2: Top 100 genes ranked by each centrality measure

Supplementary Information 3: Comorbid genes validated through literature evidence (PMID citation number)

REFERENCES

[1] M..Epsteinand J.R Sowers, Diabetes mellitus and hypertension, Hypertension, 19(50),1992, 403-418

[2] J.R.Sowers and M.Epstein, Diabetes Mellitus and Associated Hypertension, Vascular Disease and Nephropathy-An Update, Hypertension, 37, 1991,1053-59

[3] M G. Kann, Protein interactions and disease: computational approaches to uncover the etiology of diseases, Briefings in Bioinformatics,8(5), 2007, 333-346

[4] Trey Ideker andRodedSharan, Protein networks in disease, Genome Research, 18,2008, 644-652

[5] Bader GD, Betel D & Hogue CW, BIND: the Biomolecular Interaction Network, Nucleic Acids Research, Database,31,2003,248-250

[6] Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C,The STRING database in:functional interaction networks of proteins globally integrated and scored, Nucleic Acids Research,(Database issue), 39, 2011,D561-568

[7] Ceol A, ChatrAryamontri A, Licata L, Peluso D, Briganti L, Perfetto L, Castagnoli L &Cesareni G, MINT, the molecular interaction database: update, Nucleic Acids Research (Database issue), 2009, D532-D539.

[8] Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU & Eisenberg D, The Database ofInteracting Proteins: update, Nucleic Acids Research, 32,2004 D449-451

[9] Prasad TSK, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D,Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S,Somanathan DS, Sebastian A, Rani S, Ray S, Harrys Kishore CJ, Kanth S, Ahmed M,Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S,Ranganathan P, Ramabadran S, Chaerkady R &Pandey A, Human ProteinReference Database – Update, Nucleic Acids Research,37, 2009, D767-72.

[10] Georgios A Pavlopoulos, Maria Secrier, Charalampos N Moschopoulos, Theodoros G Soldatos, Sophia Kossida, Jan Aerts, Reinhard Schneider and Pantelis G Bagos, Using graph theory to analyze biological networks, BioDataMining,4(10). 2011, 1-27

[11] Xiaowei Zhu, Mark Gerstein and Michael Snyder, Getting connected: analysis and principles of biological networks, Genes Development, 21,2007, 1010-1024.

[12] M.W. Gonzalez, M. G. Kann, Protein interactions and Disease, PLOS Computational Biology, 8(12), 2012, 1-12

[13] Marc Vidal, Michael E. Cusick, and Albert-La´ szlo´ Baraba´ si, interactome Networks and Human Disease, Cell,144,2011,986-998

[14] Kwang-Il Goh, Michael E. Cusick, David Valle, Barton Childs, Marc Vidal and Albert-La´ szlo´ Baraba´ si ,Human Disease Network, PNAS,104(21).2007 8685-8690.

[15] Frank Emmert-Streib, ShaileshTripathi,Ricardo de Matos Simoes, Ahmed F. Hawwa and Matthias Dehmer, The human disease network, Systems Biomedicine, 1(1),2013,1-8

[16] Giovanni Scardoni and Carlo Laudanna ,Centralities Based Analysis of Complex Networks, Dr.Yagang Zhang (Ed.), New Frontiers in Graph Theory,2012, ISBN: 978-953-51-0115-4, InTech, DOI: 10.5772/35846.

[17] M Oti, B Snel, M A Huynen, H G Brunner, Predicting disease genes using protein-protein interactions, Journal of Medical Genetics, 43,2006,691-698

[18] SinanErten and Mehmet Koyuturk, Role of Centrality in Network-Based Prioritization of Disease Genes, Evolutionary Computation, Machine Learning and Dats Mining I Bioinformatics, Lecture Notes in Computer Science, 6023,2010,13-25

[19] Dirk Koschützki and Falk Schreiber, Centrality Analysis Methods for Biological Networks and Their Application to Gene Regulatory Networks, Gene Regulation and Systems Biology, 2,2008, 193-201

[20] Liangcai Zhang, Xu Li, Jingxie Tai, Wan Li and Lina Chen, Predicting Candidate Genes Based on CombinedNetwork Topological Features: A Case Study in Coronary Artery Disease, PLoS ONE, 7(6),2012,1-12

[21] Jing Zhao, Ting-Hong Yang, Yongxu Huang and PetterHolme, Ranking Candidate Disease Genes from Gene Expression and Protein Interaction: A Katz-Centrality Based Approach, PLoS ONE, 6(9). 2011, 1-9

[22] Tristan H Nelson, Jae-Yoon Jung, Todd F DeLuca1, Byron K Hinebaugh, KristianChe St. Gabriel and Dennis P Wall,Autworks: a cross-disease network biology application for Autism and related disorders, BMC Medical Genomics, 5(56). 2012, 1-4

[23] Arzucan Özgür1, Thuy Vu1, Günes¸ Erkan1 and Dragomir R. Radev, Identifying gene-disease associations using centrality on aliterature mined gene-interaction network, Bioinformatics,24.2008, i277-285

[24] Yutaka Fukuoka, Daiki Takei and Hisamichi Ogawa, A two-step drug repositioning method based on a protein-protein interaction network of genes shared by two diseases and the similarity of drugs, Bioinformation,9(2).2013,89-93

[25] Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D,GeneCards: integrating information about genes, proteins and diseases, Trends Genetics,13 (4), 1997 163-167

[26] D.G Gilbert, euGenes: A eukarytotic genome information system, Nucleic Acids Research, 30. 2002, 145-148

[27] Hamosh, A.; Scott, A. F.; Amberger, J. S.; Bocchini, C. A.; McKusick, V. A, Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders, Nucleic Acids Research (Database issue)33, 2004, D514–D517.

[28] Donna Maglott, Jim Ostell, Kim.D.Pyuitt and TatiuvaTatusova "EntrezGene: gene centred information, Nucleic Acid Research (Database issue), 33, 2005, D54-58

[29] Flicek P, Amode MR, BarrellD,Ensembl 2011, Nucleic Acids Research (Database issue)39,2010, D800–D806

[30] Hong-Jie Dai, Johnny Chi-Yang Wu, Richard Tzong-Han Tsai, Wen-Harn Pan, and Wen-Lian Hsu, T-HOD: A literature-based candidate gene database for hypertension, obesity and diabetes, Database(Oxford), 2013

[31] Universal Consortium: Ongoing and future developments at the Universal Protein Resource, Nucleic Acids Research (Database issue),39, D214-D219

[32] Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M,The KEGG resource for deciphering the genome, Nucleic Acids Research (Database issue),32, D277–80.

[33] W Yu, M Gwinn, M Clyne, A Yesupriya& M J Khoury, A Navigator for Human Genome Epidemiology, Nature Genetics, 40(2), 2008,124-5.

[34] Andrew I.Se, Tim Wiltshire, Serge Batalov, Hilmar Lapp, Keith A.Ching, David Block, Jia Zhang, Richard Soden, Mimi Hayakawa, Gabriel Krieman, Michael.P.Cooke, John.R.Walker and John B. Hogenesch, A GeneAtlasof mouse

and human protein encoding transcriptiomes, PNAS,101(16), 6062-67

[35] Hui Liu, Wei Liu, Yitug Liao, langCheng,Qian Liu, Xiang Ren, Lisag Shi, XinTu,Qiy Kenneth Wang, An-Yuan Guo,CADgene: a comprehensive database for coronary artery disease genes, Nucleic Acids Research (Database issue),39, 2011,991-6

[36] Tateno Y, Imanishi T, Miyazaki S, Fukami-Kobayashi K, Saitou N, Sugawara H,DNA Data Bank of Japan (DDBJ) for genome scale research in life science, Nucleic Acids Research,30 (1),2002, 27–30

[37] KG Becker, K.C Barnes, T.J Bright, S.A Wang, The Genetic Association Database, Nature Genetics, 36,2004, 431 - 432

[38] Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., Hanspers, K., Isserlin, R., Kelley, R.,Killcoyne, S., Lotia, S., Maere, S., Morris, J., Ono, K., Pavlovic, V., Pico, A. R., Vailaya, A., Wang, P.-L. L., Adler, A., Conklin, B. R., Hood, L., Kuiper, M., Sander, C., Schmulevich, I., Schwikowski, B., Warner, G. J., Ideker, T. & Bader, G. D, Integration of biological networks and gene expression data using cytoscape, Nature protocols, 2(10),2007 2366–2382.

[39] Assenov, Y., Ramirez, F., Schelhorn, S.-E., Lengauer, T. & Albrecht, M, Computing topological parameters of biological networks, Bioinformatics 24(2), 2008, 282–284.

[40] Scardoni, G., Petterlini, M. &Laudanna, C,Analyzing biological network parameters with CentiScaPe, Bioinformatics, 25(21), 2009, 2857–2859.

[41] Gray KA, Daughterty LC, Gordon SM, Real RC, Wright MW, Bruford EA, Genenames.org: the HGNC resources, Nucleic Acid Research(Database) 41, 2013, 545:552