Network Biology of Genes Involved in Cancer: Basal Cell Carcinoma, Squamous Cell Carcinoma, Throat Cancer, Thyroid Cancer and Oral Cancer

Ashish Kumar Thakur[1] Jaisri Jagannathan[2]
[2]Assistanr Professor
[1,2]School of Life Sciences, SIILAS Campus, Jaipur National University, Jaipur-203017

Abstract— Network biology and data mining processes are helpful enough to provide detailed information about the genes and their networks which paves the way for drug designing. Here, in this study the genes implicated in cancer such as Basal cell carcinoma (BCC), Squamous cell carcinoma (SCC), Throat cancer, Thyroid cancer and Oral cancer are retrieved and the network comprising co-expression, co-localization, genetic interactions, physical interactions, pathways, predicted, shared-protein domain is generated. Network properties and hubs for the predicted network were identified. The hubs predicted for example, smoothened, fizzled class receptor (SMO) in Basal cell carcinoma can be considered as potential drug targets for drug designing.

Keywords: Network Biology, Hubs, Genemania, Cancers

I. INTRODUCTION

Several diseases are caused by a combination of genetic mutations. Thus, in case of a complex disease we cannot assume that a single genetic mutation can be a cause. Many diseases including cancer fall in this category of complex diseases. The difficulty faced in studying genetic causes of cancer is that different cancer cases might be caused by different genetic mutations. If it is caused by combinatorial effects of many mutations, the individual effect of each mutation might be small and thus hard to discover.

So, how can we approach the study of cancers? A useful clue is provided by the fact that different genes and small molecules interact with each other to form a complex interaction network. Thus a mutation in one gene can be propagated through interactions and the effect of other genes in the complex interaction network.

Network biology provides powerful tools for the study of complex diseases and these approaches leverages the idea that complex diseases can be better understood from the perspective of gene networks than at the individual gene level. [2, 3]

A. Basal Cell and Squamous Cell Carcinoma

The two most common kinds of skin cancer are basal cell carcinoma and squamous cell carcinoma also called nonmelanoma skin cancer. These cancers are carcinomas that begin in the cells that cover or line an organ. Basal cell carcinoma accounts for more than 80 percent of all skin cancers and is most common of all cancers. Squamous cell carcinoma also rarely spreads, but does so more often than basal cell carcinoma. Both basal and squamous cell cancers are found mainly on areas of the skin that are exposed to the sun-the head, face, neck, hands and arms but skin cancer can occur anywhere.

B. Oral Cancer

Oral cancer refers to cancer that develops in any of the parts that makes up the mouth. It's a type of head and neck cancer.

C. Throat or Larynx Cancer

Throat cancer refers to cancerous tumors that develop particularly in your throat, voice box or tonsils.

D. Thyroid Cancer

Thyroid cancer is a rare type of cancer that affects the thyroid gland, a small gland at the base of the neck.

II. MATERIAL AND METHOD

A. Retrieval of Disease Genes

The genes associated with the following cancers in human are retrieved from the databases such as Malacards, Kyoto Encyclopedia of Genes and Genomes (KEGG), gene-disease association database (DisGeNET) and literature mining is used to retrieve the gene from PubMed abstracts. [3]

B. Logical Relations between the Gene Sets from Different Databases

A Venn diagram of the gene sets is created that shows all possible logical relations between the gene sets retrieved from different resources.

C. Approving Unique Symbols and Names for Human Protein Coding Genes

The retrieved gene lists of each disease are then compiled with an online tool (Multi-symbol checker) associated with the HGNC database (HUGO Gene Nomenclature Committee) and all approved genes are then considered as the final gene lists.

D. Disease Network and Their Analysis

The disease networks are constructed based on the gene sets using GeneMania. It helps us to predict the functions and also finds other genes that are related to a set of input genes, using a very large set of functional association data set. Association data include protein and genetic interactions, pathways, co-expression, co-localization and protein domain similarity. As for large gene sets GeneMania is also accessible via cytoscape plugin. The CytoHubba plugin is used to explore the important nodes in the biological network. Cytoscape is an open source bioinformatics software platform for visualizing molecular interaction networks and features are present in the form of plugins. [6, 1, 5]

III. RESULTS

A. Resources for Disease-Gene Retrieval

As a case study, we focus on web-based computational approach that uses integrated data resources including gene literature and data mining for the retrieval of genes associated with diseases. Malacards (analyzing disease-associated gene-sets in gene Analytics to yield affiliated pathways, phenotypes, compounds and genes), DisGeNET is a discovery platform on gene-disease associations from several public sources and the literature, Kyoto Encyclopedia of Genes and Genomes (KEGG) is a collection of databases dealing with genomes, pathways, diseases and drugs. Apart from these databases literature mining is used to extract genes from PubMed abstracts. [3]

Type of cancer	Malacards	DisGeNET	KEGG	Literature mining
Basal cell carcinoma	19	2	4	4
Squamous cell carcinoma	14	108	4	0
Throat cancer	15	10	6	2
Oral cancer	21	45	8	124
Thyroid cancer	50	9	12	0

Table 1: shows the frequency of genes in 5 diseases from public resources.

B. Comparison of Gene across Databases

Genes associated with diseases were retrieved from several databases and their occurrence was compared. For example TP53 is reported to be common in BCC, SCC, Oral and Throat cancer.

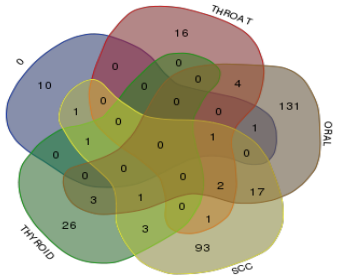

Fig. 1: represents the Venn diagram to demonstrate common genes in BCC, SCC, Oral, Thyroid and Throat cancer obtained from sources such as Malacards, DisGeNET, KEGG and literature mining.

C. Creation of Disease Network

Here, the functions of proteins encoded by these genes are computed using GeneMania. The networks are predicted on the basis of co-expression, co-localization, physical interactions, predicted interactions, pathways, genetic interactions and their shared protein domains. [6]

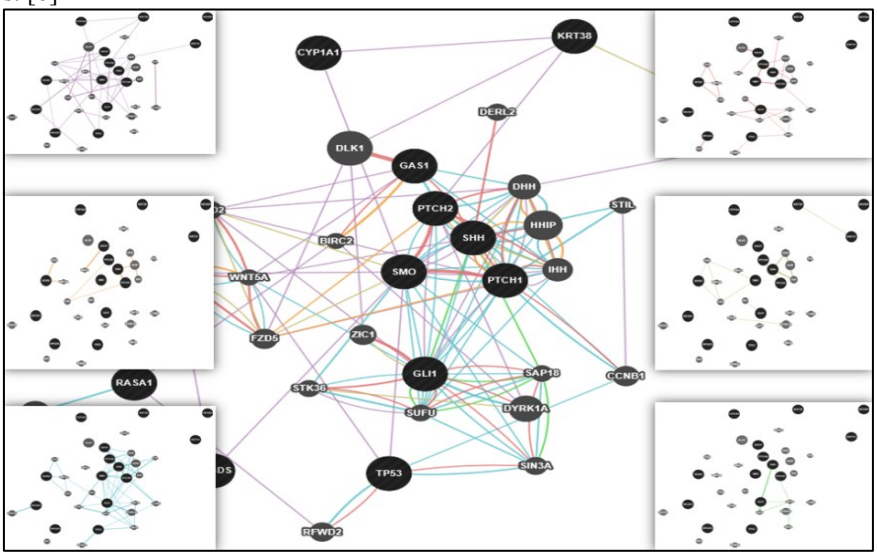

Fig. 2: shows the BCC Disease Network

Parameters	BCC	SCC	THROAT	ORAL	THYROID
Co-expression	7.67%	41.33%	41.33%	46.68%	43.91%
Shared protein domain	5.23%	6.93%	6.93%	2.10%	0.28%
Physical interactions	27.29%	21.70%	21.70%	29.10%	7.66%
Predicted	4.03%	8.84%	8.84%	6.35%	27.38%
Pathway	49.46%	4.51%	4.51%	3.85%	13.57%
Genetic Interactions	6.32%	9.08%	9.08%	6.86%	1.01%
C0-localization	0	7.62%	7.62%	5.06%	6.19%

Table 2: shows fraction of gene categorized into different experimental resources.

D. Property Analysis of the Networks

To visualize the molecular networks and its properties we use Cytoscape (open source software for the visualization and manipulation of networks and its properties). [5]

Properties	BCC	Throat cancer	Oral cancer	SCC	Thyroid cancer
Clustering coefficient	0.167	0.199	0.143	0.136	0.176
Connected components	1	1	1	1	1
Network diameter	4	4	6	5	4
Network radius	1	1	1	1	1
Shortest paths	244(23%)	537(29%)	13162(46%)	6968(40%)	860(30%)
Characteristic path length	1.803	2.05	1.971	1.994	1.862
Avg. number of neighbors	5.758	7.163	30.452	22.833	9.778
Number of nodes	33	43	177	132	54
Number of edges	95	154	2695	1507	264
Network density	0	0	0	0	0
Isolated nodes	0	0	0	0	0
Number of self-loops	0	0	0	0	0
Multi-edge node pairs	0	0	0	0	0

Table 3: shows simple properties of network analyzed using Cytoscape.

Types of cancer	In-degree				Out-degree			
	A	B	Correlation	R-squared	A	B	Correlation	R-squared
Basal Cell Carcinoma	6.746	-0.792	0.738	0.645	8.654	-0.858	0.793	0.553
Throat Cancer	7.102	-0.541	0.58	0.323	0.409	-0.89	0.732	0.657
Oral Cancer	8.313	-0.31	0.106	0.113	3.932	-0.565	0.672	0.457
Squamous Cell Carcinoma	12.287	-0.513	0.482	0.339	0.691	-0.493	0.712	0.474
Thyroid Cancer	4.118	-0.34	0.364	0.117	7.599	-0.623	0.512	0.495

Table 4: shows the In-degree and Out-degree distribution of networks using fit power law.

E. Identification of Hub

With the help of Cytohubba plugin in cytoscape we identify the hubs (a highly connected node of a network that may act as a target for a ligand or drug). [1]

BCC	SCC	Throat cancer	Oral cancer	Thyroid cancer
SMO	JUN	GSTP1	STAT3	CTNNB1
GLI1	CDH13	TP53	CDKN1A	BRAF
SHH	MYC	EGFR	MYC	TG
PTCH1	EGFR	CDK4	CSNK2A1	RET
DHH	STAT3	CCND1	EGFR	NRAS
IHH	TP53	MYC	CXCR4	PIK3CA
PTCH2	VEGFA	AURKB	PTK2	TPO
SUFU	CDH1	ACVRL1	CCND2	PIK3CD
FZD5	COL17A1	NME1	SRC	HRAS
FZD2	CEBPA	SERPINB5	STAT5A	CDKN1A

Table 5: shows top 10 highly connected nodes in each network.

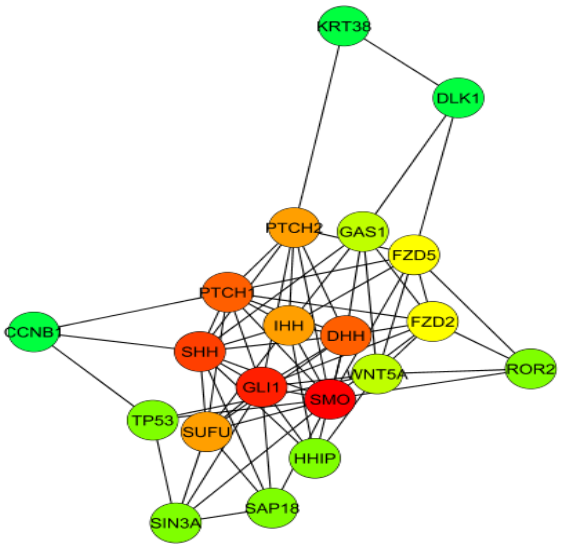

Fig. 3: shows top 20 highly connected nodes in BCC network ranked by Degree method.

IV. Discussion and Conclusions

We analyzed the interaction networks of genes and its relationship to disease and found that the interactions helped us to determine whether a genetic mutation is associated to be disease-related or not. This method could provide a useful filter for experimentalists searching for new candidate protein targets for drug designing and other R&D processes.

References

[1] Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. "CytoHubba: identifying hub objects and sub-networks from complex interactome" 2014; 8 Suppl 4:S11. Doi: 10.1186/1752-0509-8-S4-S11. Epub 2014 Dec 8.

[2] Hongxiang Feng, Xiaowei Wang, et al. "Identification of Genetic Mutations in Human Lung Cancer by Targeted Sequencing", Cancer Informatics, 2015:14.

[3] Jaisri Jagannadham, Hitesh Kumar Jaiswal and Kamal Rawal "Deciphering relationships in disease networks using computational approaches: Fatty Liver, PCOD, Osteoarthritis, cholelithiasis & hyperlipdemia" CODEN (USA): IJPRIF, ISSN: 0974-4304 Vol.8, No.1, pp 127-134, 2015

[4] Matthew B Carson, Hui Lu "Network-based prediction and knowledge mining of disease genes" Carson and Lu BMC Medical Genomics 2015, 8(Suppl 2):S9

[5] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. "Cytoscape: a software environment for integrated models of biomolecular interaction networks" 2003 Nov;13(11):2498-504

[6] Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q."The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function" 2010 Jul; 38(Web Server issue):W214-20. Doi: 10.1093/nar/gkq537.