Disease Co-Morbidity and the Human Wnt Signaling Pathway: A Network-Wise Study

Losiana Nayak,[1] Harinandan Tunga,[2] and Rajat K. De[1]

Abstract

The human Wnt signaling pathway contains 57 genes communicating among themselves by 70 experimentally established associations, as given in the KEGG/PATHWAY database. It is responsible for a variety of crucial biological functions such as regulation of cell fate determination, proliferation, differentiation, migration, and apoptosis. Abnormal behavior of its members causes numerous types of human cancers, dramatic changes in bone mass density that lead to diseases such as osteoporosis-pseudo-glioma syndrome, Van-Buchem disease, skeletal malformation, autosomal dominant sclerosteosis, and osteoporosis type I syndromes. So far, single genes have been investigated for their disease-causing properties, and single diseases have been traced backwards to discover foul-play of the system pathways. Differential expression of the whole genome has been mapped by microarray. But how all the genes involved in a pathway affect each other in single/multiple disease state(s) and whether the presence of one disease state makes a person prone to another kind of disease(s) (i.e., co-morbidity among diseases associated with a certain important biological pathway) is still unknown. We have developed a human Wnt signaling pathway diseasome and analyzed it for finding answers to such questions. Data used in constructing the diseasome can be downloaded from the publicly accessible webserver http://www.isical.ac.in/-rajat/diseasome/index.php.

Introduction

In order to understand various disease mechanisms of human, a precise disease-gene list is not always enough. A detailed map/network, including cellular components, genes, transcription factors, RNAs, enzymes and metabolites, is much more helpful (Barabasi, 2007; Goh et al., 2007). Diseases may not always be independent of each other. Presence of more than one disease in a patient often complicates the treatment, because they may share some common genes. The more connected a disease is to other diseases, the higher is its prevalence and associated mortality rate (Lee et al., 2008). For example, diseases such as diabetes, obesity, Gaucher disease, and Parkinson disease often co-occur in the same individual. Diseasome-wise studies are needed to understand such situations. A diseasome is the combined set of all known disorders or gene associations in a species. It is created by linking the complete set of genetic disorders (phenome) with the complete list of disease genes (genome) (Goh et al., 2007). In a human disease network, two disorders are linked with each other, if they share at least a common disease gene. On the other hand, in a network of genes associated with diseases, there is an edge between two genes, if both of them are associated with at least a single common disorder. Disease maps are potential knowledge bases that throw light on multiple disease-related complexities. For appropriate disease-specific diagnostic, prognostic, and therapeutic approaches, gene–disease association studies provide valuable information (Tiffin et al., 2009).

Information on interactions among genes can be used to find disease-related genes. The underlying assumption is that if ‘two genes work together, the known association of one with a disease suggests that the other may also be associated with the same disease’ (Tiffin et al., 2009). Methods such as ENDEAVOUR (Tranchevent et al., 2008) and G2D (Perez-Iratxeta et al., 2007) use this assumption. We have used the same assumption but from a different point of view via manual exploration of research articles. A disease network is built from the genes involved in the human Wnt signaling pathway to throw light on the role of Wnt signaling pathway in various diseases and their co-morbidity.

The biomedical literature is too massive a source to be handled in a general sense. Generalized databases including GEPIS (Zhang et al., 2004), KEGG (Kanehisa et al., 2006), OMIM (Hamosh et al., 2005), PubMed (http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed), STRING (Jensen et al.,

[1]Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India.
[2]Department of Computer Science and Engineering, RCC Institute of Information Technology, Kolkata, India.

2009), TiGER (Liu et al., 2008), TRANSFAC (Wingender, 2008), and UniHi (Chaurasia et al., 2009) constitute parts of this vast resource. There are also purpose-specific databases cum tool repositories such as Oncomine 3.0 (Rhodes et al., 2007), DiseaseMeth: a human disease methylation database (Lv et al., 2012), DistiLD Database: diseases and traits in linkage disequilibrium blocks (Palleja et al., 2012), The Disease Ontology (DO) database (Schriml et al., 2012), and web resources such as Network of Cancer Genes [NCG 3.0] (Syed et al., 2010), among others. We have mined some of these sources for our purpose.

Materials and Methods

There is a vast amount of existing literature that supports the role of the human Wnt signaling pathway in various diseases, the most important being cancers. Some of the articles implicate the whole pathway for certain diseases. while the others pinpoint certain gene(s) of the pathway and their role in human pathogenesis. We have scanned literature of both kinds to find gene–disease associations for the human Wnt signaling pathway. Wnt signaling pathway information is taken from the KEGG/Pathway database (Kanehisa et al., 2006). The literature scanning has been done using the links and references provided in KEGG/Pathway database. The pathway entry (entry id hsa:04310) contains multiple references. The initial referrals for finding associations are provided by these links and references. Especially important information in one go, is found in Hatsell et al. (2003). Each gene of the human Wnt signaling pathway in the KEGG/Pathway database is linked to a flat file that contains information about the entry id, gene name, definition, orthology, pathway, class, motif, links to other databases, position in the chromosome, amino acid sequence, and nucleotide sequence. The "links to other databases" section contains links of the Wnt signaling pathway component related files present in other databases, namely, Protein database (http://www.ncbi.nlm.nih.gov/protein/, NCBI), Entrez gene (Maglott et al., 2011), OMIM (Amberger et al., 2011), HGNC (Seal et al., 2011), HPRD (Goel et al., 2011), Ensembl (Flicek et al., 2012), and UniProt (UniProt, 2012).

- The 'NCBI-GI' link retrieves a flat file from the 'Protein' database of NCBI. The Protein database is a collection of sequences obtained from several sources, including translations from annotated coding regions stored in GenBank (Benson et al., 2012), RefSeq (Pruitt et al., 2012) and Third Party Annotation database: http://www.ncbi.nlm.nih.gov/genbank/tpa (TPA) (Benson et al., 2004), as well as records from SwissProt (Boeckmann et al., 2003), http://pir.georgetown.edu/ (PIR), Protein reference database: http://www.prf.or.jp/aboutdb-e.html (PRF), and PDB (Westbrook et al., 2002). The retrieved flat file contains multiple 'Reference' sections dedicated to a query gene. The 'Comment' section gives the summary of experimental or computational findings explained in the articles given in the 'Reference' sections. Both these sections have helped us in scanning literature to establish human Wnt gene-disease associations.
- The 'NCBI-GeneID' link takes into Entrez Gene (http://www.ncbi.nlm.nih.gov/sites/gene) database of NCBI. Entrez Gene is a searchable database of genes, obtained from RefSeq genomes. The file has a 'Summary' section that describes gene function, protein behavior, and differential expression of the query gene in different kinds of tissues. The 'Bibliography' section provides us the required citations for the 'Summary' section. Both these sections have helped in our task of finding gene-disease associations.
- 'OMIM' hyperlink takes a flat file into account in response to a query gene present in OMIM (Online Mendelian Inheritance in Man) database. OMIM is a comprehensive, authoritative, and timely compendium of human genes and genetic phenotypes. The files present in OMIM (http://www.ncbi.nlm.nih.gov/omim) are descriptive in nature with no regular sections. Hence, whole files have to be scanned to detect the gene–disease associations.
- The link 'HGNC' takes a file of the HUGO Gene Nomenclature Committee (http://www.genenames.org/index.html) into account. This file is not helpful in establishing gene–disease associations as it has no summary section.
- 'HPRD' link considers a file of the Human Protein Reference Database (http://www.hprd.org/). All the information in HPRD has been manually extracted from the literature by expert biologists who read, interpret, and analyze the published data. A typical HPRD file has a 'Disease' tab that contains an OMIM id for describing disease relevance of the query gene those we have already manually scanned while considering the OMIM database.
- The 'Ensembl' id takes the user into a file in the Ensembl (http://www.ensembl.org/index.html) database. The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online. This page has many useful sections of information on the human Wnt signaling pathway genes such as gene summary, splice variants, comparative genomics, and genetic variations, but not gene–disease association information.
- 'UniProt' link takes to a structured file in UniProt (http://www.uniprot.org/) database. A standard UniProt file has a 'General annotation (Comments)' section. This section is arranged in multiple subsections. The 'Involvement in disease' subsection has helped us in collecting the gene–disease association of human Wnt signaling pathway components. This subsection contains cited referrals for pathogenicity of a gene, the references being listed at the end of the file.

In summary, the NCBI-GI, NCBI-GeneID, OMIM, and UniProt hyperlinks are useful for our purpose. A list of all the human Wnt signaling pathway genes, as well as their hyperlinks to the other databases, from which we have extracted the gene–disease associations, is furnished in Table 1. We have scanned all these flat and structured files for gene–disease associations of human Wnt signaling pathway in the first phase of our literature scanning. In the second phase, we have independently and individually searched via Google and Google scholar search engines for any kind of association of these genes with human pathogenicity available in research articles. We have tried to incorporate as many recent articles as possible to this piece of work. But this search is not

Table 1. Wnt Signaling Pathway Genes and Associated Links

Sl. no.	*Gene name*	*KEGG entry id*	*NCBI-GI*	*NCBI-Gene ID*	*OMIM ID*	*UniProt ID*
1	LEF1	hsa:51176	195222732	51176	153245	B4DG38, Q659G9, Q9UJU2
2	SMAD4/MADH4	hsa:4089	4885457	4089	600993	Q13485
3	NLK	hsa:51701	149408126	51701	609476	Q9UBE8
4	SOX17	hsa:64321	11967991	64321	610928	Q9H6I2
5	CTBP1	hsa:1487	61743967	1487	602618	Q13363, Q7Z2Q5
6	CREBBP	hsa:1387	119943102	1387	600140	Q4LE28, Q92793
7	RUVBL1	hsa:8607	4506753	8607	603449	Q9Y265
8	MYC	hsa:4609	71774083	4609	190080	P01106
9	JUN	hsa:3725	4758616	3725	165160	P05412
10	FOSL1	hsa:8061	4885243	8061	136515	P15407
11	CCND1	hsa:595	16950655	595	168461	P24385, Q6FI00
12	PPARD	hsa:5467	284807155	5467	600409	Q03181
13	MMP7	hsa:4316	4505219	4316	178990	P09237
14	MAP3K7	hsa:6885	4507361	6885	602614	O43318
15	CTNNB1	hsa:1499	148233338	1499	116806	P35222
16	PSEN1	hsa:5663	4506163	5663	104311	P49768
17	CTNNBIP1	hsa:56998	9889555	56998	607758	Q5T4V2, Q9NSA3
18	**CHD8**	**hsa:57680**	**282165704**	**57680**	**610528**	**Q9HCK8**
19	**PRKACA**	**hsa:5566**	**4506055**	**5566**	**601639**	**P17612**
20	**CSNK1A1L**	**hsa:122011**	**269846834**	**122011**	**-**	**Q8N752**
21	FBXW11	hsa:23291	48928050	23291	605651	Q9UKB1
22	TBL1X	hsa:6907	213021186	6907	300196	O60907
23	DVL1	hsa:1855	32479521	1855	601365	O14640
24	CXXC4	hsa:80319	13376816	80319	611645	Q9H2H0
25	SENP2	hsa:59343	54607091	59343	608261	Q9HC62
26	CSNK2A1	hsa:1457	4503095	1457	115440	P68400
27	FRAT1	hsa:10023	31317236	10023	602503	Q92837
28	APC2	hsa:10297	5031587	10297	612034	O95996
29	**NKD1**	**hsa:85407**	**14916433**	**85407**	**607851**	**Q969G**
30	WNT16	hsa:51384	17402914	51384	606267	Q9UBV49
31	PORCN	hsa:64840	45439329	64840	300651	Q9H237
32	SFRP1	hsa:6422	56117838	6422	604156	Q8N474
33	**CER1**	**hsa:9350**	**4885135**	**9350**	**603777**	**O95813**
34	WIF1	hsa:11197	111125011	11197	605186	Q9Y5W5
35	LRP6	hsa:4040	148727288	4040	603507	O75581
36	DKK1	hsa:22943	7110719	22943	605189	O94907
37	FZD10	hsa:11211	6005762	11211	606147	Q9ULW2
38	RAC1	hsa:5879	9845511	5879	602048	A4D2P1, P63000
39	DAAM1	hsa:23002	21071077	23002	606626	Q 9Y4D1
40	VANGL2	hsa:57216	62955805	57216	600533	Q9ULK5
41	PRICKLE1	hsa:144165	222136680	144165	608500	B3KVG3, Q96MT3
42	WNT9A	hsa:7483	15082261	7483	602863	O14904
43	**MAPK8**	**hsa:5599**	**4506095**	**5599**	**601158**	**P45983**
44	RHOA	hsa:387	10835049	387	165390	P61586, Q9BVT0
45	ROCK1	hsa:6093	4885583	6093	601702	Q13464
46	AXIN1	hsa:8312	27501450	8312	603816	O15169
47	CSNK1E	hsa:1454	4503093	1454	600863	P49674, Q5U045
48	GSK3B	hsa:2932	225903437	2932	605004	P49841, Q6FI27
49	PPP2CA	hsa:5515	4506017	5515	176915	B3KUN1, P67775
50	PLCB1	hsa:23236	12083581	23236	607120	Q9NQ66
51	**CAMK2A**	**hsa:815**	**25952114**	**815**	**114078**	**A8K161, Q8IWE0**
52	**CHP**	**hsa:11261**	**6005731**	**11261**	**606988**	**Q99653**
53	**PRKCA**	**hsa:5578**	**4506067**	**5578**	**176960**	**B5BU22, P17252, Q7Z727**
54	WNT5A	hsa:7474	40806205	7474	164975	P41221
55	NFAT5	hsa:10725	164419746	10725	604708	A2RRB4, Q7LA65
56	TP53	hsa:7157	120407068	7157	191170	P04637, Q3LRW3, Q53GA5
57	SIAH1	hsa:6477	55749557	6477	602212	Q8IUQ4

Bold faced entries depict human Wnt signaling pathway genes (9 in number) not found associated with any disease.

exhaustive, as everyday some new findings about the human Wnt signaling pathway are getting published, and a considerable number of them are not freely accessible to us. In this article, we concentrate on the mined information and its analysis rather than the mining techniques. By no means, it is an exhaustive study; still it gives us important ideas about disease relatedness, interdependence and comorbidity.

Methodology to construct disease (cancerous, noncancerous, and link) networks based on shared genes

From gene–disease associations obtained from the aforementioned sources, disease–disease associations are extracted with a simple thumb rule: "If Disease 1 is associated with Gene A (supported by article X) and Gene A is associated with Disease 2 (supported by article Y), then Disease 1 is associated with Disease 2". A diagrammatic flowchart of the methodology followed for association mining is furnished as Figure 1. These associations are then used to build a Wnt-specific disease network. The disease network is then divided into three disease-specific networks based on disease–disease associations among: (i) cancerous diseases, (ii) noncancerous diseases, and (iii) cancerous and noncancerous diseases. Inherent properties of these networks have then been studied.

Methodology to construct disease (cancerous, noncancerous, and link) networks based on Relative Risk (RR) and ϕ-correlation values

The terms Relative Risk (*RR*) and ϕ-correlation for validating disease associations have been described by Park et al. (2009). Relative Risk (*RR*) for a pair of diseases is a measure by which co-morbid tendency between the two diseases can be quantified. Let I_x and I_y be the number of patients, from a pool of *N* patients having disease *x* and *y*, respectively. Let C_{xy} be the number of patients diagnosed with both disease *x* and *y*. Then RR_{xy} can be defined as $C_{xy}/(I_x.I_y/N)$. ϕ-correlation can be defined as

$$\phi = (RR_{xy} - 1)/\sqrt{I_x I_y/(N - I_x)(N - I_y)} \qquad \text{(Eq. 1)}$$

Each of these two variables has its own advantages and disadvantages. Calculating *RR* is a clear and simple way of understanding the weightage of association between two diseases. But when disease *x* or *y*, or both, are rare in occurrences, $I_x.I_y$ will be low leading to high *RR*, which is apparently not the case. Further processing in terms of a threshold for $I_x.I_y\ /N$ overcomes this problem in some cases. On the other hand, ϕ-correlation has no such bias as it is a range limited parameter. The value of ϕ ranges between -1 and 1. However, the value of ϕ correlation turns out to be very small, in some cases, even if there is evidence of high co-morbidity between two diseases. Moreover, when diseases *x* and *y* are maximum possible co-morbid (i.e., disease *x* always occurs with disease *y*), $\phi \neq 1$ due to Eq. 1. Hence, considering *RR* and ϕ values together ensures the robustness of co-morbidity findings among diseases. $RR > 1$ and $\phi > 0$ indicate that diseases *x* and *y* co-occur more frequently than expected by chance (Park et al., 2009). We have collected the calculated *RR* and ϕ-correlation values for our curated disease associations from an "All patients ICD-9 5 dataset (total 13039018 patients)".

The ICD-9 5 dataset

The ICD-9 5 dataset has been provided by Hidalgo et al. (2009) (http://barabasilab.neu.edu/projects/hudine/resource/data/data.html). In this dataset, the diseases have been categorized by a primary diagnosis and up to 9 secondary

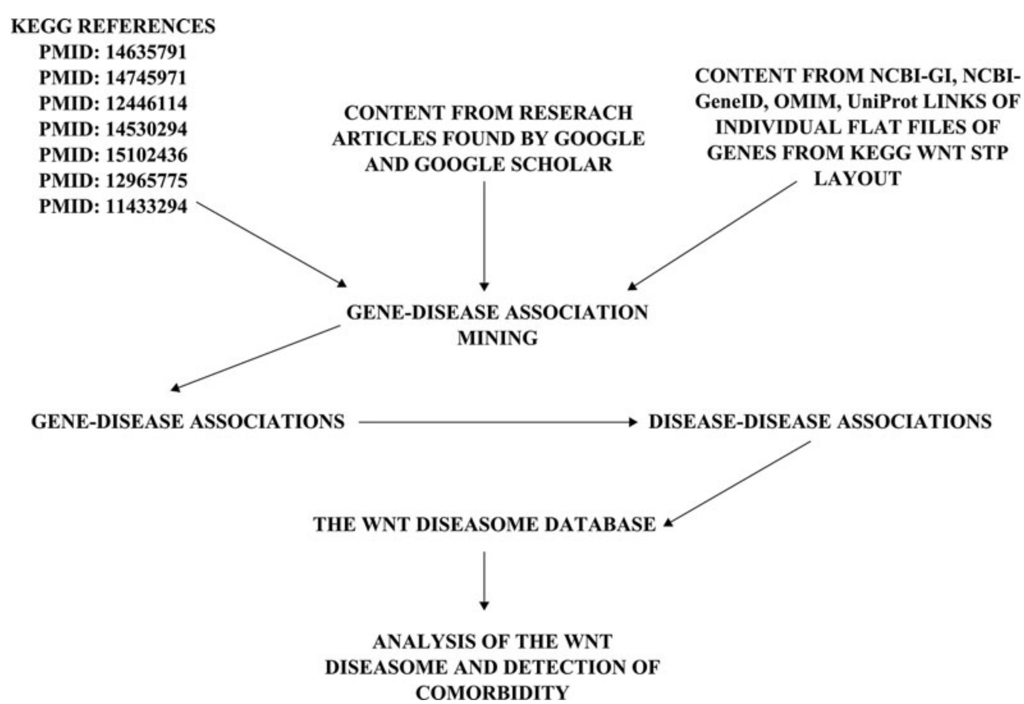

FIG. 1. A diagrammatic flowchart of the methodology followed for association mining.

diagnoses, all being specified by ICD9 codes of up to 5 digits by International Statistical Classification of Diseases (ICD), a detailed classification of diseases. Altogether we have found 112 diseases associated with the genes of the human Wnt signaling pathway. Thus, there is 6216 disease pairs among them. *RR* and ϕ-correlation values of these pairs were searched in the All patients ICD-9 5 dataset, and 1753 disease associations with *RR* and ϕ-correlation values were found. Some of them carried negative ϕ-correlation values. Among them, 1150 valid disease associations have been found with $RR > 1$ and $\phi > 0$. These associations have then been used to create a separate diseasome based on *RR* and ϕ-correlation values. It has been named as the ICD9-5 diseasome. Some of these associations were found to be present in the human Wnt signaling pathway diseasome. We have also constructed a separate set of cancerous, noncancerous, and link networks from the common associations found both in the human Wnt signaling pathway diseasome, as well as in the ICD-9 5 dataset to highlight highly co-morbid disease associations.

Analysis of both the sets of networks has been done by Network Analyzer plug-in release 2.7. (Assenov et al., 2008) of the open-source software package Cytoscape version 2.8.2 (Smoot et al., 2011). Definitions of the network parameters, calculated by Network Analyzer, are provided where they appear first.

Results and Analysis

We have found a total of 112 different kinds of anomaly or disturbances or diseases or symptoms or syndromes. In general, they are termed as diseases throughout this article for better understanding. Out of these 112 diseases, 66 are different kinds of cancer, while 46 are noncancerous and *in vivo* events including eyesight problems, heart and lungs complications, mental illness, nervous disorders, and organ duplications arising due to developmental and environmental issues, and genetic makeup. Diseases related to the Wnt signaling pathway affect multiple tissue types (blood, bone marrow, cartilage, endometrial, epithelial, germ, lymph, mesothelial, muscle, nerve, and skin) of various organs such as artery, bone, brain, breast, ear, elementary canal, esophagus, eye, face, heart, kidney, liver, lungs, mouth, neck, nose, ovary, pancreas, prostate, skeleton, skin, spine, testis, throat, thumb, urinary bladder, uterus, and glands (lymph, thyroid, parathyroid, salivary, parotid, and sebaceous glands). Some diseases are found strictly in children, while the others are found in all age range.

These diseases are associated with 48 out of 57 human Wnt signaling pathway genes. Nine genes (CAMK2A, CER1, CHD8, CHP, CSNK1A1L, MAPK8, NKD1, PRKACA, and PRKCA) have not been found to be associated with any of the aforementioned events. The list of considered genes is given in Table 1. Here the term 'association' means: i) the disease is caused due to mutation of the gene, ii) the gene is experimentally found to have differential expression in diseased tissues, iii) the gene may have role in the disease, iv) the gene is regulated by a third party molecule and possibly that molecule has role in the disease, v) the gene is a therapeutic target to stall the disease progression, and vi) the gene is a therapeutic target to overcome the diseased state. If a gene, by any of the above means, is found to be associated with a disease, we have declared it as a valid gene–disease association. The associations are used to create a gene–disease network (Fig. 2).

The gene–disease network

The bipartite network (Fig. 2) is made of 200 individual associations among 48 genes and 112 diseases. We have calculated various parameters of this network to study its properties. As no direct gene–gene or disease–disease associations are in this network, the clustering co-efficient has been found to be zero. Clustering coefficient of a node n is the ratio N/M, where N is the number of edges among the neighbors of n, and M is the maximum number of edges that could possibly exist between the neighbors of n. The clustering coefficient of a node is always a number between 0 and 1.

Nine different connected components have been noticed in this network, the largest being of 144 nodes. (In an undirected network, two nodes are connected if there is a path of edges between them. Within a network, all the nodes that are pairwise connected form a connected component.) The number of connected components indicates the connectivity of a network. A lower number of connected components suggest a stronger connectivity among those components. On the other hand, a very large connected component ensures stronger connectivity among most of the nodes of a network irrespective of the presence of other very small connected components. Here the rest connected components are of the size of 2 each, except one of size 3. The maximum of shortest path lengths between a pair of nodes of the network, called network diameter (network diameter is the largest distance in terms of edges between two nodes; the network diameter and the shortest path length distribution may indicate small-world properties of the analyzed network), is found to be 11, while the shortest possible path length (network radius) is 1. Network radius is the minimum among the non-zero path-lengths of the nodes in the network. In simple words, maximum 11 edges and minimum 1 edge lie between a pair of disease.

Network centralization score (a score that ranges between 0 and 1 and defines structure of a network. A higher score reveals a lot of nodes are linked with all the rest nodes, to create a structure closer to a star) that defines network topology is 0.168. In star topology, each node is connected to central hub(s) with a point-to-point connection. The hub(s) represent a single point of control as well as a single point of failure. Networks whose topologies resemble a star have a centralization score close to 1, whereas decentralized networks are characterized by having a centralization score close to 0. In that sense, the gene–disease network is decentralized.

The average shortest path length, also known as the characteristic path length, is 4.261. It gives the expected distance between two connected nodes [i.e, on average, 4 edges have to be traversed to go from one node (gene/disease) to another node (disease/gene) in the gene–disease network]. The average number of neighbors (2.484) indicates the average connectivity of a node in the network [i.e., a gene has been involved with at least two diseases]. A normalized version of this parameter is the network density that lies between 0 and 1. It shows how densely the network is populated with edges (ignoring self-loops and duplicated edges). A network which contains no edge with solely isolated nodes has a density of 0. In contrast, the density of a clique (a network where each node is connected with every other node) is 1. A network density of

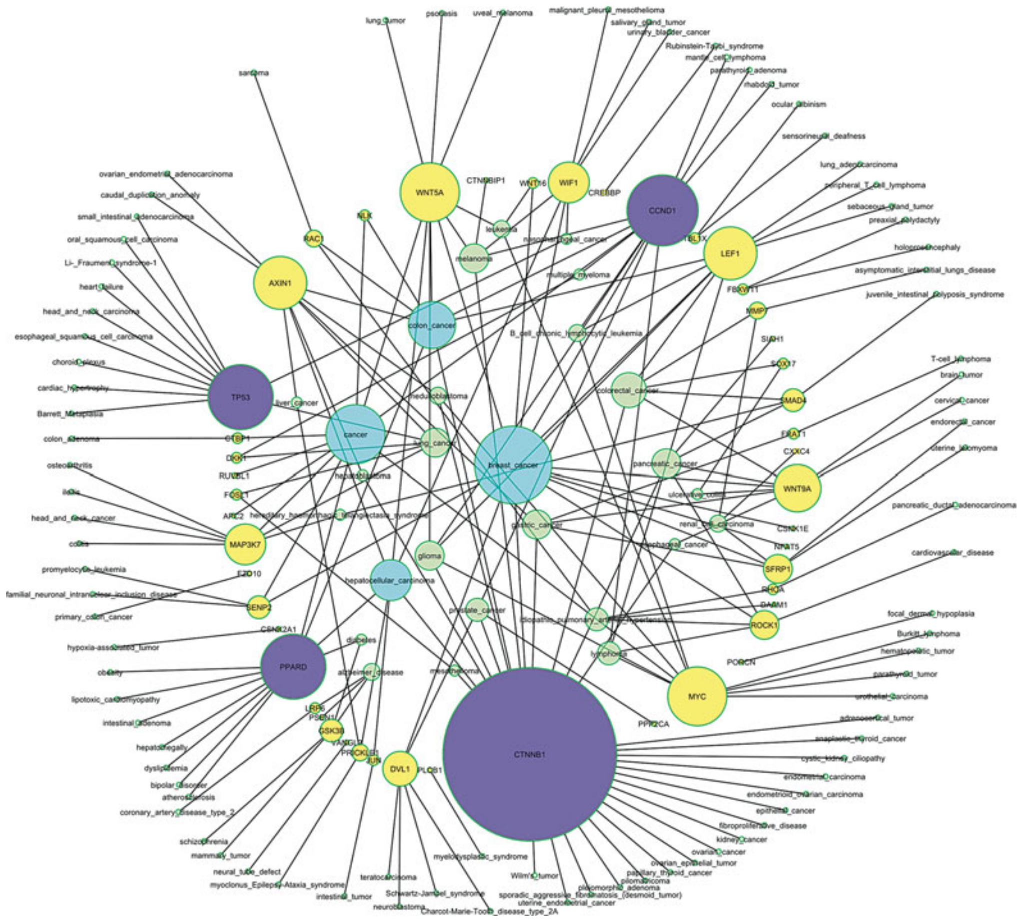

FIG. 2. The bipartite gene–disease network. Number of genes: 48; number of diseases: 112; number of gene–disease associations: 200. Node sizes are indicative of their degree distribution, higher the degree bigger is the circle. Wnt signaling pathway genes are depicted *yellow*, while the four highly connected genes are marked *blue*. Similarly, diseases are marked *green*, while the four highly connected diseases are marked *aqua*.

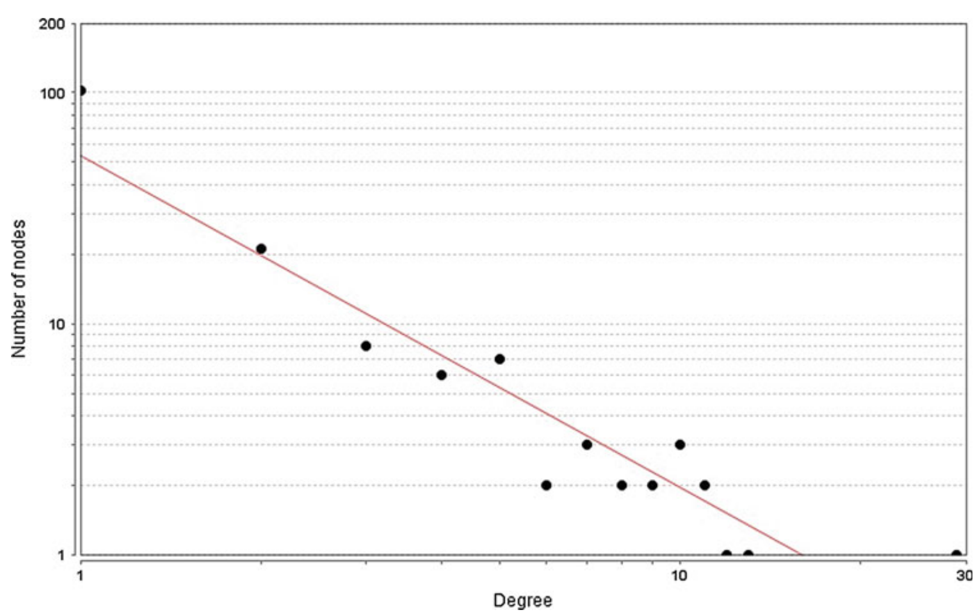

FIG. 3. Logarithmic scatter plot of node–degree distributions of the gene-disease network. A power low ($y = ax^b$) can be fitted to the points of the plot where $y = \log$(number of nodes), $x = \log$(degree), $a = 53.08$ and $b = -1.433$ (correlation value = 0.982 and R-squared value = 0.888). The network is a scale-free network.

0.016 indicates that the gene–disease network is not densely populated. There are many sparse connections in the network. A logarithmic scatter plot of node-degree (node-degree of a node n is the number of edges linked to n) distributions of the gene–disease network is given in Figure 3. A power low ($y = ax^b$) can be fitted to the points of the plot where $y = \log$(number of nodes), $x = \log$(degree), $a = 53.08$, and $b = -1.433$ (correlation value = 0.982 and R-squared value = 0.888). The correlation value represents correlation between the data points and corresponding points on the line. R-squared value is also known as coefficient of determination. It determines the relative confidence of a model for fitting into a

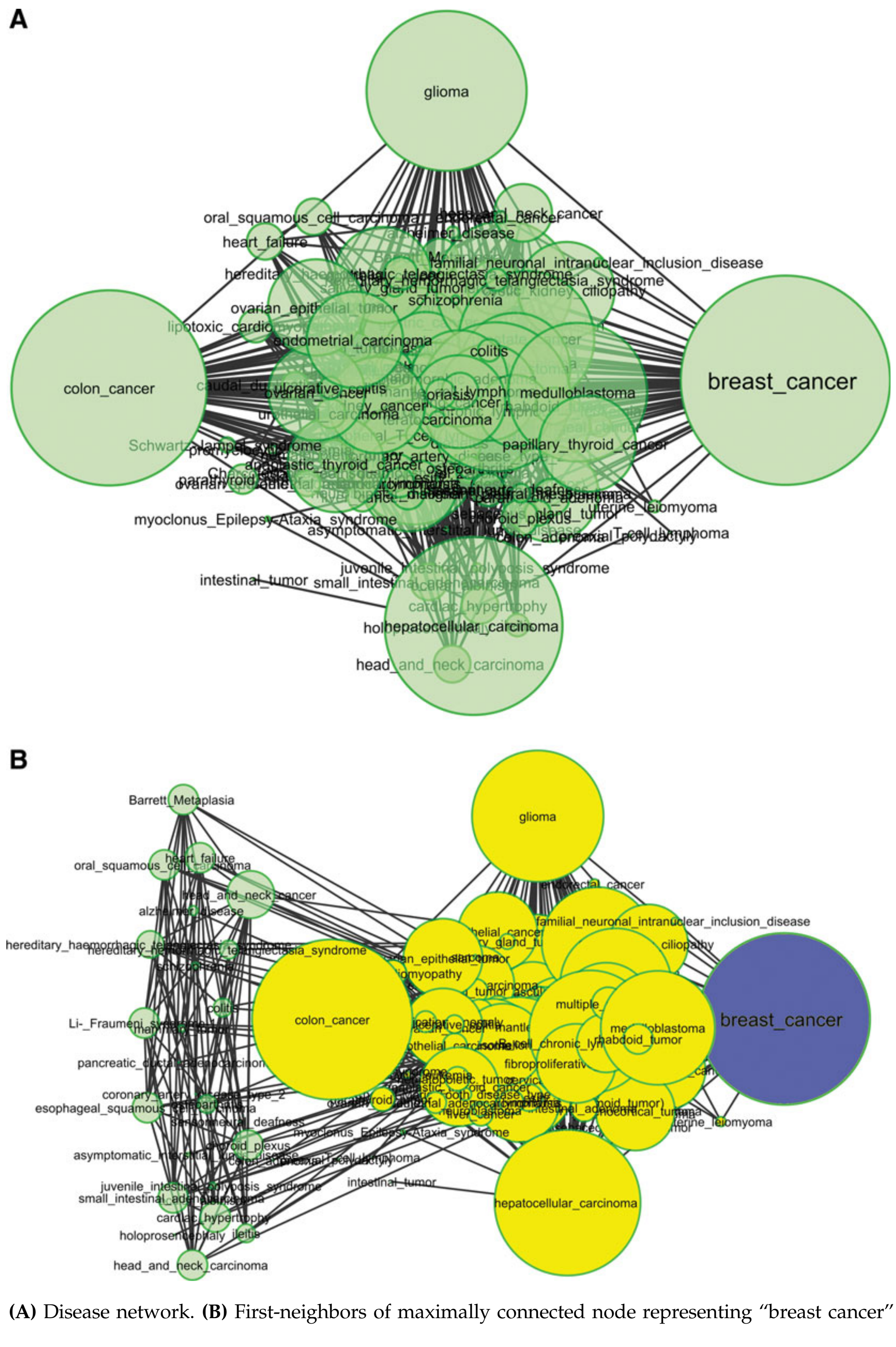

FIG. 4. **(A)** Disease network. **(B)** First-neighbors of maximally connected node representing "breast cancer" (*yellow*).

Table 2. Network Properties of the Disease Network

Parameter	*Value of the parameter*
Clustering co-efficient	0.809
Connected components	3
Network diameter	5
Network radius	1
Network centralization	0.426
Shortest paths	10716 (92%)
Characteristic path length	2.2
Average no. of neighbors	15.241
Network density	0.142

hypothesized property. Here high correlation and R-squared values depict that the gene–disease network is a scale-free network.

There are many highly connected nodes. Colorectal cancer is the only disease common among associated diseases of the highly connected nodes representing genes CTNNB1, CCND1, and PPARD. Altogether they, along with Tp53, are associated with 55 diseases (49.1% of the total number of diseases). Likewise, among the diseases, breast cancer is highly prevalent, which is associated with 13 genes of the Wnt signaling pathway, followed by colon cancer (9 genes) and hepatocellular carcinoma (7 genes). These observations confirm that the one disease–one target gene concept is not applicable to established pathway genes, and diseases are way beyond complex events as more and more number of genes are being found to be associated with them.

The disease network

A network of 107 nodes (diseases) out of total 112 diseases are associated among themselves with 823 edges as seen in Figure 4A. Only five diseases (hypoxia-associated tumor, myelodysplastic syndrome, neural tube defect, primary colon cancer, and Rubinstein-Taybi syndrome) are found to not be associated with any other disease. They turned out to be the 2-node components of the gene–disease network (Fig. 2). Other related parameters of the network are given in Table 2. Here we discuss some of them to reveal the inherent properties of the disease network. A high clustering co-efficient (0.809) indicates presence of separate components (modules) in the network. On an average, each disease is connected to 15 neighbor diseases (13.39% of the total nodes). It indicates the high rate of co-morbidity among the diseases.

Four maximally connected diseases are breast cancer, colon cancer, hepatocellular carcinoma, and glioma in decreasing order. They are depicted as the four corner nodes in Figure 4(A). Nodes are sized proportionately according to their degree. First neighbors of breast cancer included the other three maximally connected nodes along with most of the highly connected nodes of the network, confirming our suspicion that pathway-related diseases are linked among themselves as seen in Figure 4B. A logarithmic scatter plot of node-degree distributions of the disease network is given in Figure 5. A power low ($y = ax^b$) can be fitted to the points of the plot where y = log(number of nodes), x = log(degree), a = 10.151, and b = – 0.548 (correlation value = 0.372 and R-squared value = 0.375). A low correlation and R-squared value depict that this network is not scale free.

Cancerous and noncancerous disease networks

Here we have categorized the aforementioned 112 Wnt signaling pathway associated diseases into two major categories (i.e., 66 cancerous and 46 noncancerous diseases). We have divided the disease network (Fig. 4A) into three separate components, *viz.*, a cancerous disease network (Fig. 6A), a noncancerous disease network (Fig. 6B) and a network linking the cancerous and noncancerous diseases (Fig. 6C).

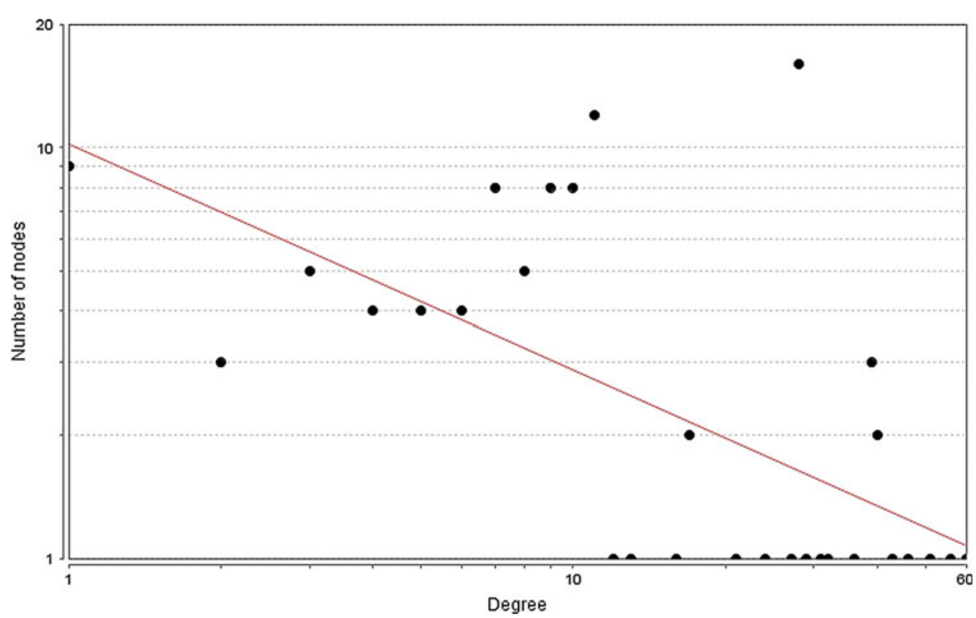

FIG. 5. Logarithmic scatter plot of node–degree distributions of the disease network. A power low ($y = ax^b$) can be fitted to the points of the plot where y = log(number of nodes), x = log(degree), a = 10.151 and b = – 0.548 (correlation value = 0.372 and R-squared value = 0.375). The network is not scale-free.

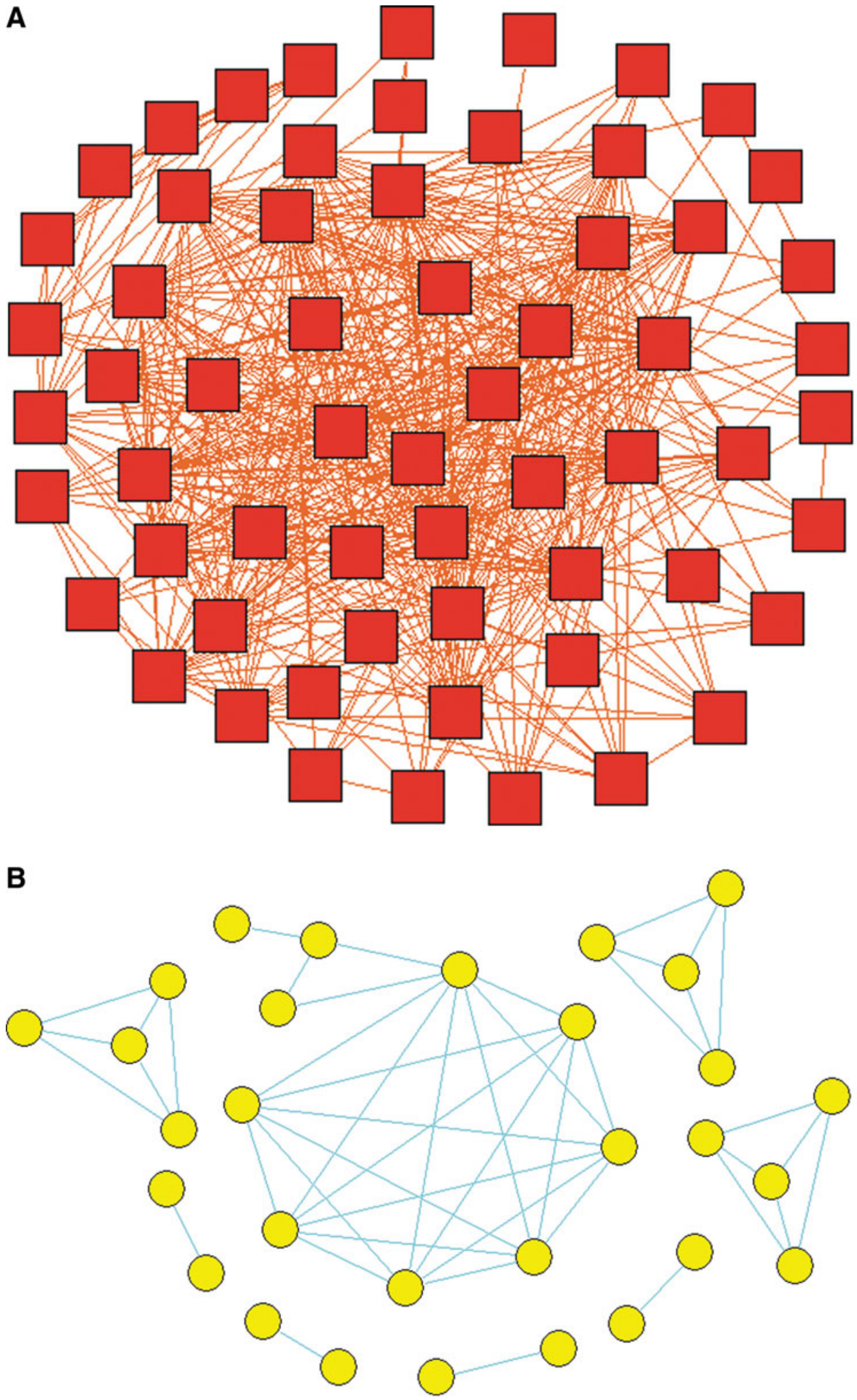

FIG. 6. **(A)** Cancerous disease network. **B)** Noncancerous disease network. **(C)** Links among cancerous and noncancerous diseases. Cancerous diseases are marked with *red squares*, while nodes representing the noncancerous diseases are marked with *yellow circles*.

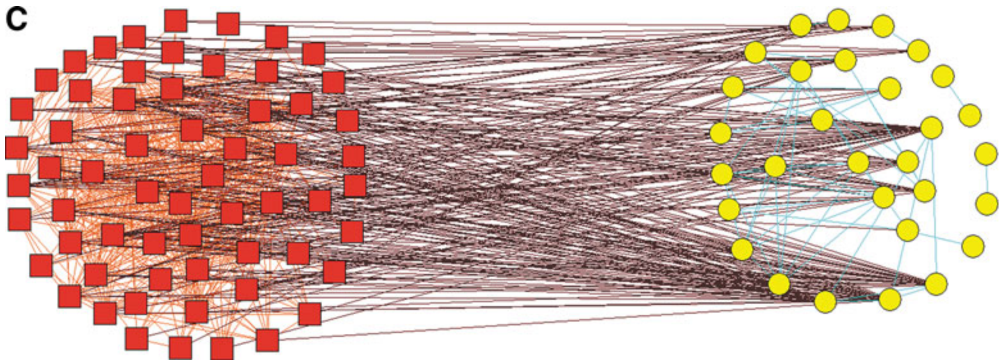

FIG. 6. (Continued).

The cancerous disease network is a close knit cluster of 61 diseases and 471 unique disease–disease interactions as shown in Figure 6A. Five types of cancers are found not to be connected with any other kind of cancers, *viz.*, hypoxia associated tumors, mammary tumor, myelodysplastic syndrome, pancreatic ductal adenocarcinoma, and primary colon cancer. "Breast cancer" is the maximally connected disease with 40 interactions with other types of cancers (Fig. 7). Diseases in the network are densely connected. Clustering coefficient is 0.801 and on an average each disease shared approximately 15 neighbors with each other. The node "breast cancer", its interactions, first neighbors and their interactions cover 95% of the whole cancer disease network, indicating high co-morbidity among Wnt signaling pathway-related cancers (Fig. 8).

The noncancerous disease network (Fig. 6B) is plotted from 30 noncancerous diseases and their 47 unique interactions. Out of 46 noncancerous diseases, 16 have not been found to be present in this network (asymptomatic interstitial lungs disease, caudal duplication anomaly, familial neuronal intranuclear inclusion disease, focal dermal hypoplasia, juvenile intestinal polyposis syndrome, myoclonus-epilepsy-ataxia syndrome, neuronal tube defect, parathyroid tumor, parathyroid adenoma, pleiomorphic adenoma, psoriasis, Rubinstein-Taybi syndrome, salivary gland tumor, sebaceous gland tumor, ulcerative colitis, and uterine leiomyoma). Unlike the cancerous disease network, this network is sparsely connected. Clustering coefficient is less (0.663) and the diseases share very few neighbors with each other (~ 3), indicating less co-morbidity and individualistic behavior of these diseases.

Co-Morbidity

When two disorders or illnesses occur in the same person, simultaneously or one after the other, they are called co-

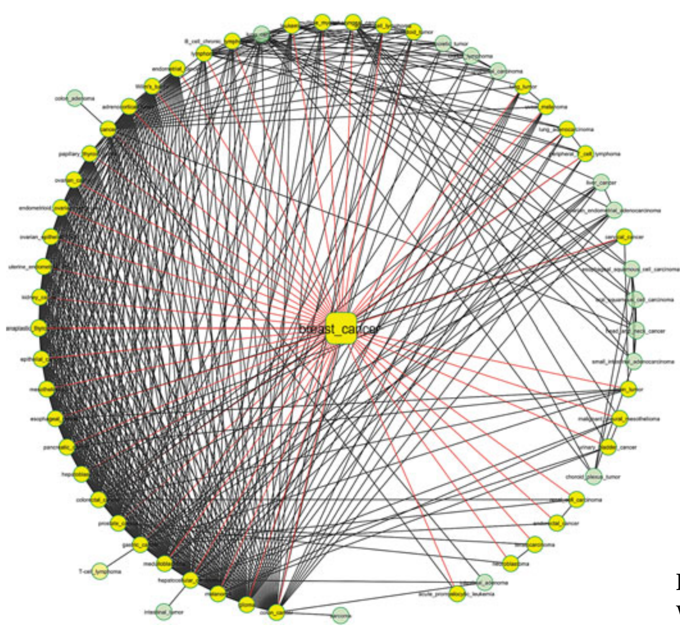

FIG. 7. The maximally connected node "breast cancer" with its adjacent edges (*red*), and first neighbors in the cancerous disease network (*yellow*).

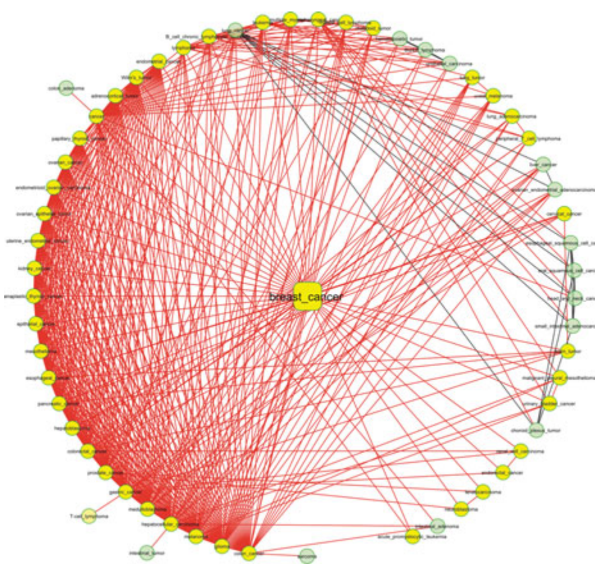

FIG. 8. The maximally connected node "breast cancer" with its adjacent edges (*red*), and first neighbors (*yellow*) and their adjacent edges (*red*) in the cancerous disease network. The noticeable fact is that almost the whole network is attached with the central node "breast cancer" with just two steps traversing adjacent edges of connected nodes.

morbid (Le and Kwon, 2011). These disorders may happen due to a common gene or genes belonging to a common pathway. Interactions between the molecular components (possibly genes belonging to a pathway) may result in potential co-morbidity effects (Park et al., 2009).

In this article, we have established associations among the diseases via common genes in a disease network to get a glimpse of their co-morbid status. Two disease are associated if they share at least one common causative gene. Likewise, there will be multiple diseases and many associations among themselves, if all the genes of a pathway are under consideration. These associations are nothing but indications of existing co-morbidity among these diseases. An overall network view of the gene–disease network (Fig. 2), disease network (Fig. 4A), cancerous disease network (Fig. 6A), noncancerous disease network (Fig. 6B), and the link network among the cancerous and noncancerous diseases (Fig. 9) is provided in Table 3. The denser the network, the higher is its co-morbidity due to the high rate of shared neighbors. As evident from the Table, the cancerous disease network (Fig. 6A) is the densest network (network density: 0.257) with a high degree of interactions among the diseases (average number of neighbors: ~ 15). On the other hand, the noncancerous disease network is quite sparse (network density: 0.108) with less interactions among the diseases (~ 3). These facts conclude that Wnt signaling pathway-related cancerous diseases show high degrees of co-morbidity among themselves, while the noncancerous diseases do not.

Modules in the cancerous disease network

The cancerous disease network (Fig. 6A) is a dense network with a high degree of interactions among the nodes and it could not be resolved into separate components. For a deeper level of inspection, we have divided 61 different types of cancerous diseases according to the tissue or organ they affect. Hence, the cancerous diseases are divided into 15 categories (blood, brain, colon, esophagus, intestine, kidney, liver, lungs, lymph, multiple organs, ovary, thyroid gland, urinary bladder, uterus, and single diseases). When a particular type of cancer is found to affect more than one type of tissue or organ (e.g., colorectal cancer affects colon and rectum), it is listed in the "multiple organ" category. When a single type of tissue/organ specific diseases and associations among them are considered, 10 different modules emerged (Fig. 10). They are named numerically starting from module 1 in an anti-clockwise manner with the original cancerous disease network (Fig. 6A) in the center. Module 1 shows associations among blood-related cancers. Multiple myeloma and B-cell chronic lymphocytic leukemia are associated via gene CCND1. Leukemia and B-cell chronic lymphocytic leukemia, hematopoetic tumor, and B-cell chronic lymphocytic leukemia, and leukemia and hematopoetic tumor are linked via gene MYC. Associated liver cancers are included in module 2. Anaplastic thyroid cancer and papillary thyroid cancer (cancers of the thyroid gland) are associated via gene CTNNB1 in module 3. Brain related cancers (glioma and medulloblastoma) are in module 4.

Module 5 includes kidney-related cancers. Wilm's tumor and kidney cancer, Wilm's tumor and adrenocortical tumor, and kidney cancer and adrenocortical tumor are associated via gene CTNNB1. Module 6 showcases diseases of tissue or organ categories (bone, breast, epithelium, eyes, mammary gland, mesothelium, mouth, nerve, pancreas, pancreatic duct, prostate, rectum, skin, stomach, T-cell, throat, and testis) in which a single disease is listed that we could not consider independently. Instead, we found associations among such

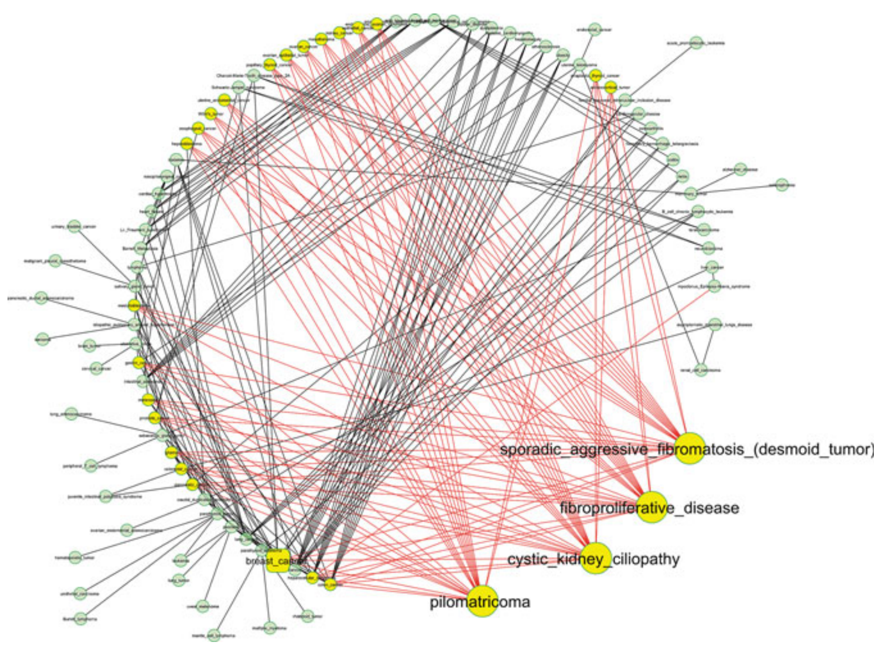

FIG. 9. The maximally connected nodes pilomatricoma, cystic kidney ciliopathy, fibroproliferative disease, and desmoid tumor with their adjacent edges, and first neighbors in the connection network among cancerous and noncancerous diseases.

Table 3. Properties of Various Networks

Type of network	*Clustering coefficient*	*Average no. of neighbors*	*Network density*	*Network heterogeneity*
Gene–disease network	0	2.484	0.016	1.323
Disease network	0.809	15.241	0.142	0.882
Cancerous disease network	0.801	15.443	0.257	0.741
Noncancerous disease network	0.663	3.133	0.108	0.626
Link network among cancerous and noncancerous diseases	0	5.116	0.054	0.968

types of cancers and listed them under a separate module. It turned out to be the dense module of the cancerous disease network. Associations among cancers that affect multiple types of tissue or organ are shown in module 7. Associated cancers of uterus are in module 8. Module 9 lists cancers of ovary. Ovarian epithelial tumor and ovarian cancer, ovarian epithelial tumor and endometrioid ovarian carcinoma, and ovarian cancer and endometrioid ovarian carcinoma are associated via gene CTNNB1. Module 10 contains lymph-related cancers (Burkitt lymphoma, mantle cell lymphoma, lymphoma, peripheral T-cell lymphoma) and their associations. We have not found any association among colon, esophagus, intestine, lungs and urinary bladder-associated diseases.

Modules in the noncancerous disease network

The noncancerous disease network (Fig. 6B) has been divided into eight separate modules, as seen in Figure 11. Module 1 contains eye and ear sensory disorders (ocular albinism and sensorineuronal deafness linked via gene TBL1X). Congenital disorders belong to module 2 (holoprosencephaly and preaxial polydactyly linked via gene FBXW11). Generally, high blood pressure in the arteries of the lungs (idiopathic pulmonary hypertension) makes the heart work harder to force the blood through these vessels. Over time, this leads to cardiovascular diseases. These two disorders belong to module 3. Module 4 contains muscle and nerve-related

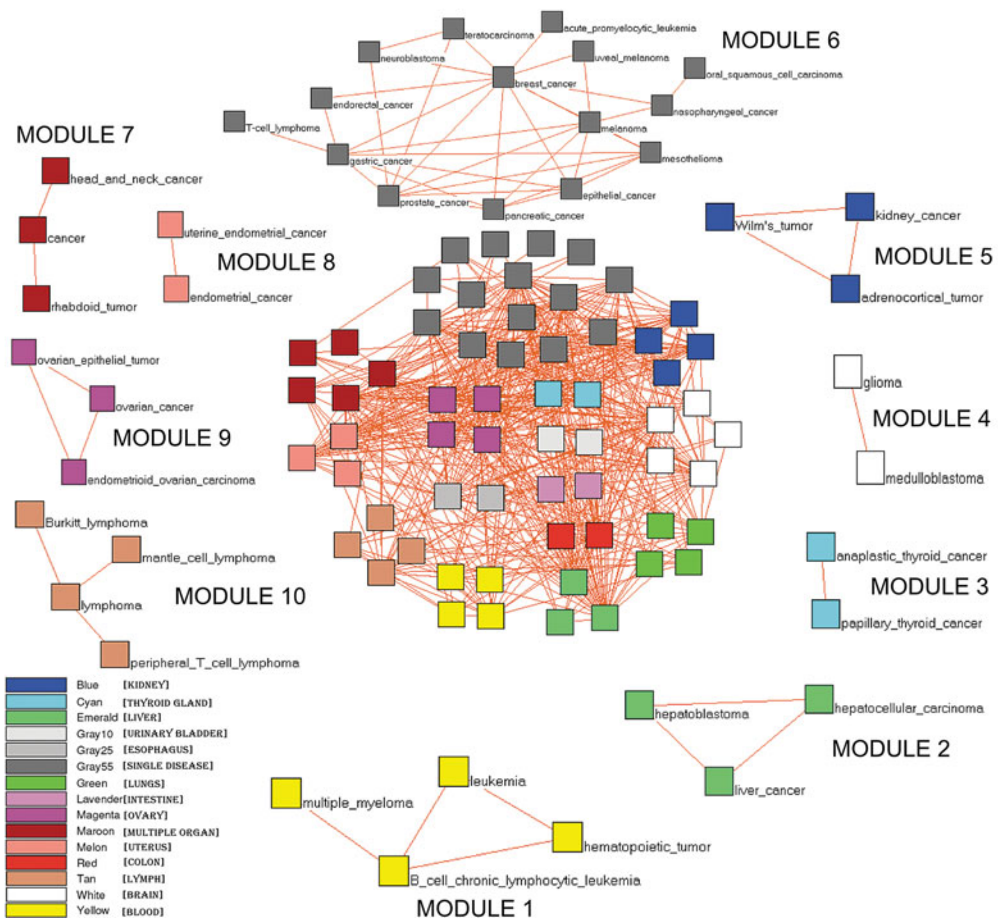

FIG. 10. Separate modules of the cancerous disease network. Modules are named numerically starting with Module 1 (containing multiple myeloma, leukemia, hematopoetic tumor, and B-cell chronic lymphocytic leukemia) from mid-bottom anti-clockwise manner.

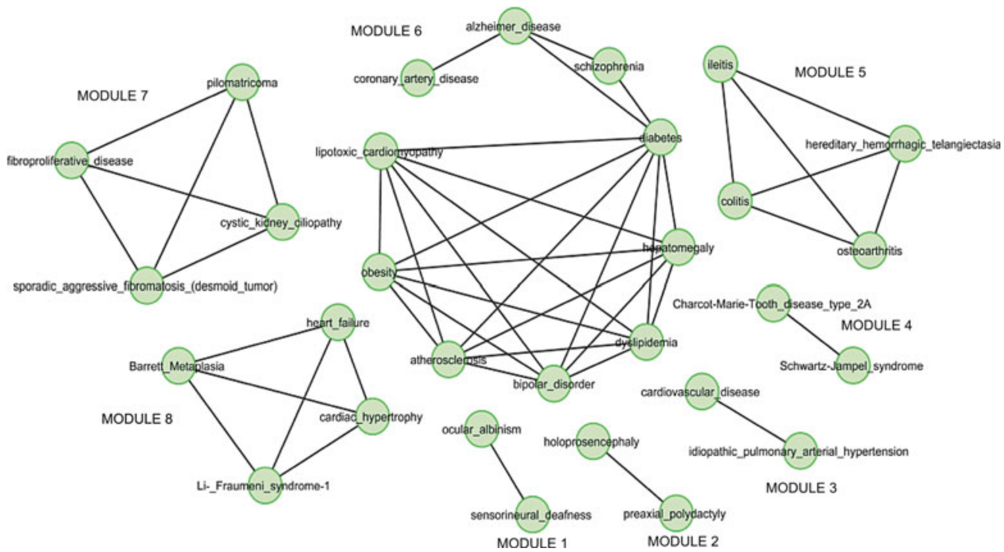

FIG. 11. Separate modules of the noncancerous disease network. Modules are named numerically starting with Module 1 (containing ocular albinism and sensorineuronal deafness from mid-bottom anti-clockwise manner.

disorders affecting human body movement and flexibility (Charcot-Marie-Tooth disease type 2A and Schwartz-Jampel syndrome linked via gene DVL1). Inflammatory bowel diseases came under the same component (ileitis and colitis linked via gene MAP3K7 in module 5).

Module 6 is the largest component of noncancerous disease network that has clustered obesity-related complicacies (diabetes, coronary heart disease, dyslipidemia, hepatomegaly, lipotoxic cardiomyopathy, and obesity) along with brain disorders such as Alzheimer's disease, schizophrenia, and bipolar disorder. It is well known that overweight issues due to bad food habits and overeating in early and adult life increase risk of mental illness and mood disorders. Obesity and diabetes significantly and independently increase risk for Alzheimer's disease (Profenno et al., 2010). It is one of the most common physical health problems among patients with severe and persistent mental illnesses, such as schizophrenia (Citrome and Vreeland, 2008). Studies have reported that up to 60% of individuals with schizophrenia and 68% of those with bipolar disorder are overweight/obese (Kolotkin et al., 2008). On the other hand, patients with bipolar disorder, in particular, are at greater risk for overweight and obesity than individuals in the general population (Keck and McElroy, 2003).

Module 7 represents an aggregate of tumors and cyst-related disorders (pilomatricoma, desmoid tumor, fibroproliferative disease, and cystic kidney ciliopathy). They are the four highly connected noncancerous diseases which have maximum (23) associations with cancerous diseases, as evident from the link

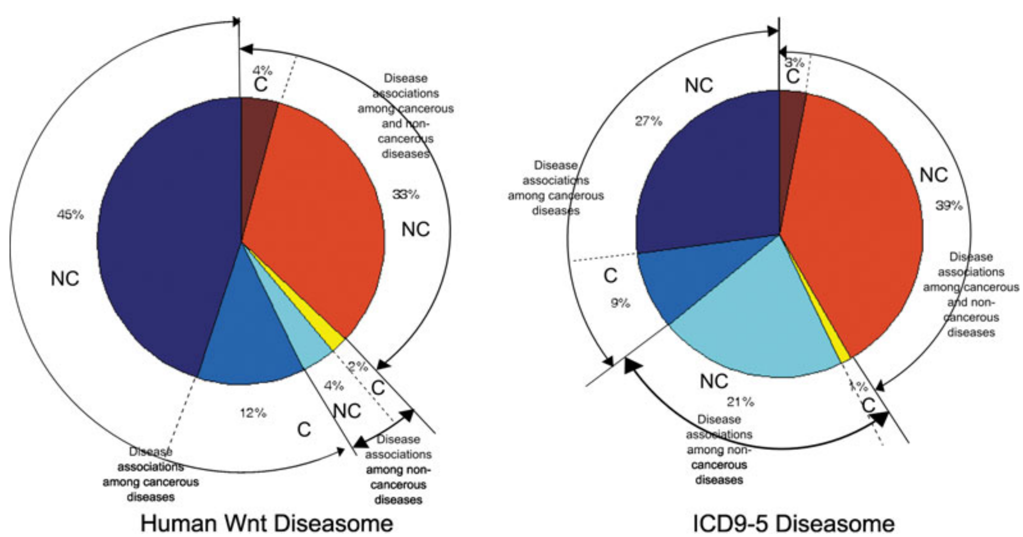

FIG. 12. A diagrammatic representation of common and uncommon disease associations present in the human Wnt diseasome and the ICD9-5 diseasome. C stands for common disease associations found in both human Wnt diseasome and ICD9-5 diseasome; NC, not-common (unique) disease associations found in the respective diseasome.

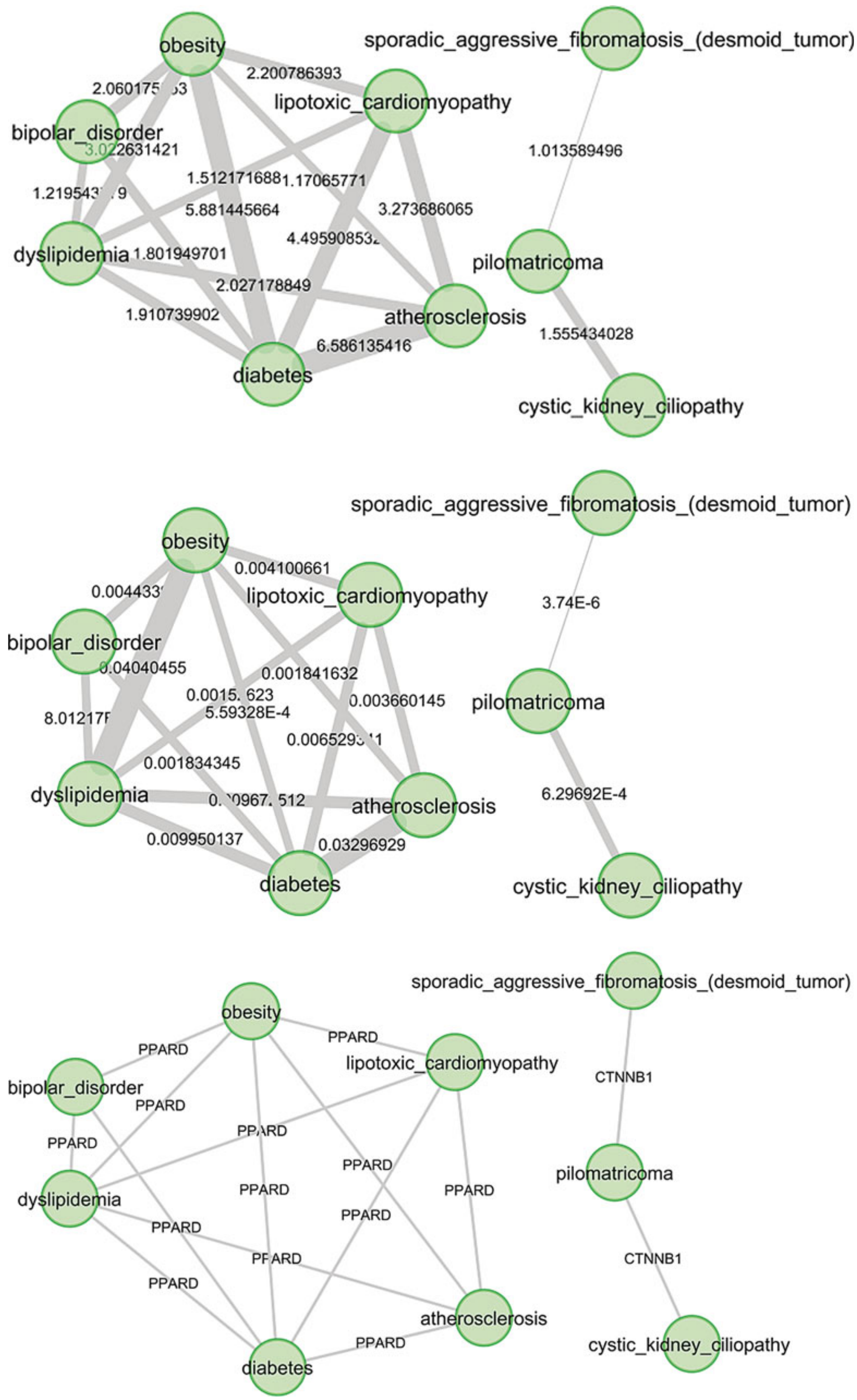

FIG. 13. **(A)** Noncancerous component of human Wnt diseasome found in the ICD9-5 diseasome. Edge width is a map of *RR* value of the disease association. **(B)** Noncancerous component of human Wnt diseasome found in the ICD9-5 diseasome. Edge width is a map of ϕ-correlation value of the disease association. **(C)** Noncancerous component of human Wnt diseasome found in the ICD9-5 diseasome. Edge label indicates the common gene associated with both the diseases, based on which the association has been formed.

network between cancerous and noncancerous diseases (Fig. 9). The network is plotted out of 95 diseases and 243 unique associations among cancerous and noncancerous diseases. Module 8 contains heart-related disorders (cardiac hypertrophy, heart failure) along with Barrett metaplasia, and Li-Fraumeni syndrome-1.

Common Disease–Disease Associations in the Human Wnt Signaling Pathway and the ICD-9 5 Diseasomes

The human Wnt diseasome is made up of 823 disease associations among 107 diseases (Fig. 12). It is to be mentioned here that 5 (112 - 107) diseases were found to be non-co-morbid (neither associated among themselves nor with the other diseases). The cancerous component had 471 disease associations (57%) among 61 cancerous diseases. Out of them, 100 disease associations (12%) are present in the ICD9-5 diseasome, making the rest 45% (371) disease associations unique to the human Wnt diseasome. The noncancerous component had 47 (6%) disease associations among 30 diseases. Out of them, 15 (2%) of them are also present in the ICD9-5 diseasome. The link component had 305 (37%) disease associations among cancerous and noncancerous diseases, out of which 35 (4%) associations are present in the ICD9-5 diseasome (Fig. 12).

Similarly, the ICD9-5 diseasome is made up of 1150 disease associations among the same 112 diseases (Fig. 12). The cancerous component had 413 (36%) disease associations among 66 diseases, the noncancerous component had 258 (22%) disease associations among 46 diseases, while the link component had 479 (42%) disease associations among cancerous and noncancerous diseases. The common associations with the Wnt diseasome constituted 9%, 1%, and 3% corresponding to the cancerous, noncancerous, and link components (Fig. 12). These common disease associations have been put through a separate set of analyses as they are trusted co-morbid disease pairs with involvement of at least one common gene.

Noncancerous component

The noncancerous component has been made up of 15 disease associations among 9 noncancerous co-morbid diseases (Fig. 13A–C). The disease associations constituted two separate modules of 13 and 2 associations, respectively. Figure 13A depicts *RR* values of the disease associations. Let association among disease x and disease y be represented as "x-y". Then some of the disease associations in the large module corresponded to high *RR* value (i.e., diabetes–atherosclerosis (= 6.586135416), diabetes–obesity (= 5.881445664), diabetes–lipotoxic cardiomayopathy (= 4.495908532), atherosclerosis–lipotoxic cardiomayopathy (= 3.273686065), and obesity–dyslipidemia (= 3.022631421). These high *RR* values depict that the disease pairs diabetes–atherosclerosis, diabetes–obesity, diabetes–lipotoxic cardiomayopathy, atherosclerosis–lipotoxic cardiomayopathy, and obesity–dyslipidemia are highly co-morbid (Table 4). In the small module, pilomatricoma showed high co-morbidity with cystic kidney

Table 4. High Co-Morbid Disease Pairs Found Among Common Noncancerous, Link, and Cancerous Components in Both Human Wnt Diseasome and ICD9-5 Diseasome

Disease 1	*Disease 2*	*RR value*	*ϕ-correlation value*	*Common gene*
High co-morbid disease pairs found in noncancerous component				
Diabetes	**Atherosclerosis**	**6.586135416**	**0.03296929**	**PPARD**
Diabetes	Obesity	5.881445664	0.000559328	PPARD
Diabetes	Lipotoxic cardiomayopathy	4.495908532	0.006529341	PPARD
Atherosclerosis	Lipotoxic cardiomayopathy	3.273686065	0.003660145	PPARD
Obesity	**Dyslipidemia**	**3.022631421**	**0.04040455**	**PPARD**
Pilomatricopma	**Cystic kidney ciliopathy**	**1.555434028**	**0.000629692**	**CTNNB1**
High co-morbid disease pairs found in the link component				
Lymphoma	**Idiopathic pulmonary arterial hypertension**	**20.8934514**	**0.00120647**	**ROCK1**
Kidney cancer	**Sporadic aggressive fibromatosis**	**13.83961187**	**0.00135192**	**CTNNB1**
Renal cell carcinoma	**Asymptomatic interstitial lungs disease**	**4.436839703**	**0.011070768**	**MMP7**
Wilm's tumor	**Cystic kidney ciliopathy**	**5.707592579**	**0.006605426**	**CTNNB1**
High co-morbid disease pairs found in the cancerous component				
Hepatoblastoma	Hepatocellular carcinoma	774.6564876	0.007698504	AXIN1
Glioma	Rhabdoid tumor	547.283022	0.006467981	CCND1
Glioma	Wilm's tumor	547.283022	0.006467981	CTNNB1
Glioma	**Medulloblastoma**	**290.5329623**	**0.030853057**	**CTNNB1**
Cancer	Melanoma	215.4461757	0.004055428	CCND1
Nasopharyngeal Cancer	Oral squamous cell Carcinoma	187.6019541	0.012514126	Tp53
Glioma	Papillary thyroid cancer	178.3772196	0.003679096	CTNNB1
Anaplastic thyroid cancer	Glioma	178.3772196	0.003679096	CTNNB1
Cancer	Lung tumor	107.7230879	0.002854249	WNT5A
Hepatoblastoma	**Liver cancer**	**101.6291348**	**0.00679636**	**AXIN1**
Cancer	**Rhabdoid tumor**	**20.47854064**	**0.071501673**	**CCND1**
Kidney cancer	**Wilm's tumor**	**70.84213258**	**0.062356786**	**CTNNB1**

Associations supported by both *RR* and ϕ-correlation values are in **bold-face**.

ciliopathy (pilomatricopma–cystic kidney ciliopathy with *RR* value = 1.555434028).

A plot of the same 15 noncancerous co-morbid disease associations is shown in Figure 13B where each disease association is a map of its corresponding ϕ-correlation value. This plot supported high co-morbidity among pairs diabetes–atherosclerosis, and obesity–dyslipidemia in the large module, as well as pilomatricoma–cystic kidney ciliopathy in the small module. Figure 13C depicts the same 15 noncancerous co-morbid disease associations where edges are based on common gene involvement found in both the diseases of an association. Interestingly, all the diseases of the large module shared PPARD gene as a common gene, while the small module contained diseases with the involvement of the gene CTNNB1.

Link component

The link component is made up of 35 disease associations among co-morbid cancerous and noncancerous diseases (29 in total). The component has been resolved into one major module of 23 disease associations and three minor modules of two disease associations each. *RR* values of these associations are depicted in Figure 14A. High *RR* value has been found among disease pairs lymphoma–idiopathic pulmonary arterial hypertension (= 20.8934514) and kidney cancer–sporadic aggressive fibromatosis (= 13.83961187), making them highly co-morbid (Table 4). Support for both the disease pairs have been found in ϕ-correlation value based plot (Fig. 14B) where edge width is a map of ϕ-correlation value of the disease association. In addition, two more highly co-morbid disease pairs come into picture as shown in Figure 14B. These diseases are renal cell carcinoma–asymptomatic interstitial lungs disease (with ϕ-correlation value = 0.011070768) and Wilm's tumor–cystic kidney ciliopathy (with ϕ-correlation value = 0.006605426). When these disease associations were considered from a common shared gene point of view, 8 common genes (CCND1, CTNNB1, GSK3-β, LEF1, PPARD, ROCK1, and WNT5A) came into the picture (Fig. 14C).

Cancerous component

The cancerous component is made up of 100 disease associations among 34 co-morbid diseases. Module structure cannot be found in this component. *RR* value based plot (Fig. 15A) displays some high co-morbid disease pairs as given in Table 4. A ϕ-correlation value based plot (Fig. 15B) of these disease associations validates two among them, as well as adds two new disease pairs (bold faced pairs in Table 4). Figure 15C depicts common genes of these disease associations in color-coded format. All total 13 genes have been found to be involved with diseases present in the cancerous component. These genes are AXIN1, CCND1, CTNNB1, DVL1, FOSL1, LEF1, MYC, PPARD, Tp53, WIF1, WNT16, WNT5A, and WNT9A.

The Human Wnt Diseasome Database

The gene-disease and disease-disease associations analyzed in this paper can be downloaded from the online human

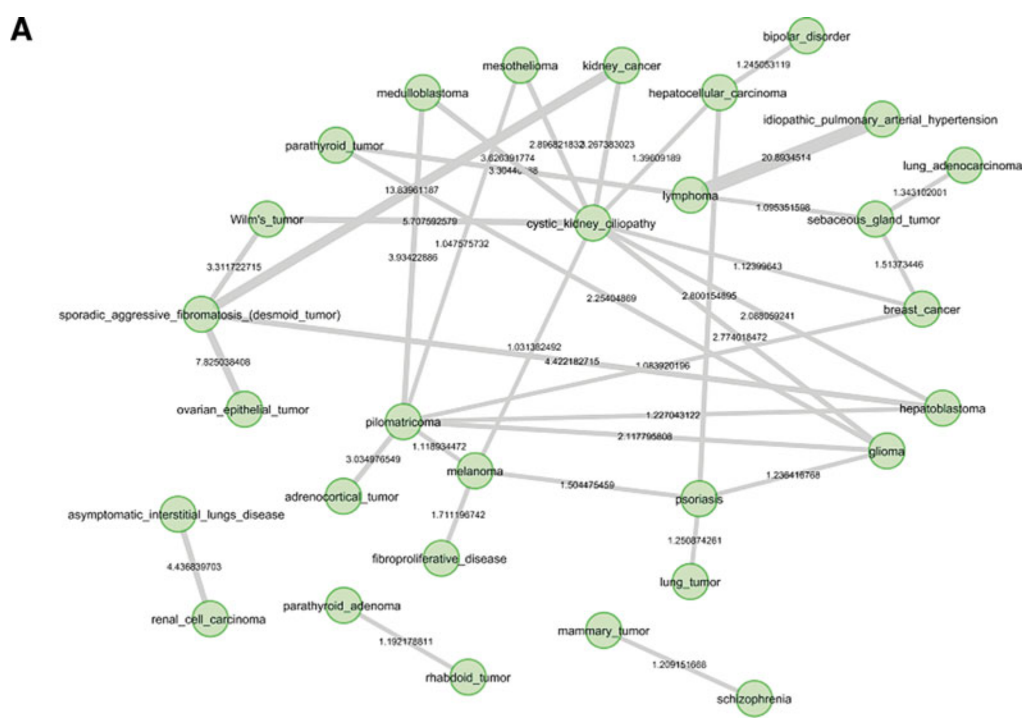

FIG. 14. **(A)** Link component of human Wnt diseasome found in the ICD9-5 diseasome. Edge width is a map of *RR* value of the disease association. (**B)** Link component of human Wnt diseasome found in the ICD9-5 diseasome. Edge width is a map of ϕ-correlation value of the disease association. **(C**) Link component of human Wnt diseasome found in the ICD9-5 diseasome. Edge label indicates the common gene associated with both the diseases, based on which the association has been formed.

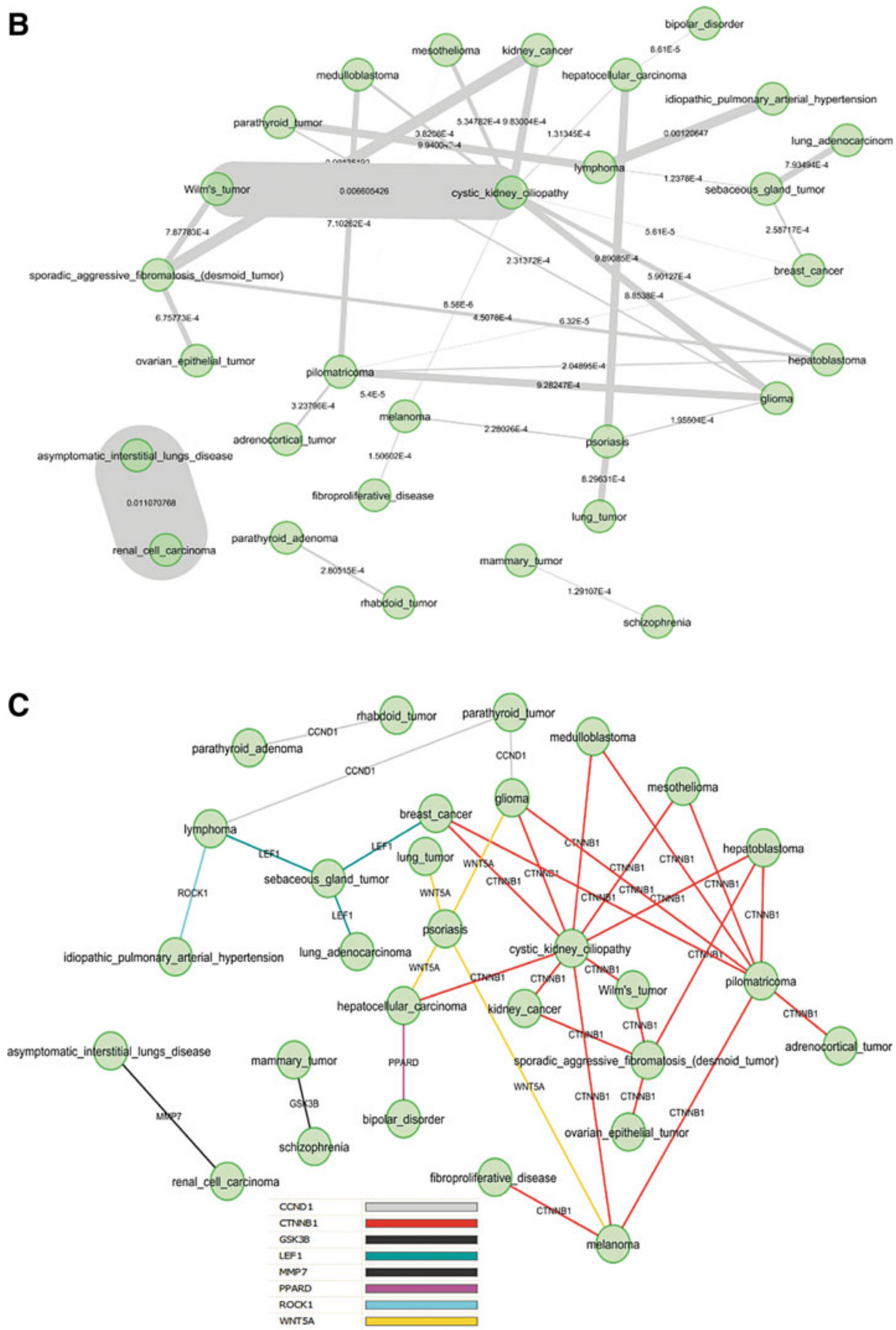

FIG. 14. (Continued).

Wnt diseasome database server in ".csv" format: http://www.isical.ac.in/~rajat/diseasome/index.php. The homepage of the online database provides definition and an introduction of diseasome, along with some useful links for further reference. One can download the whole or partial data from the browse page by selecting all the genes or a few genes of interest in the page. Data can be downloaded at multiple levels (gene–disease associations along with their citations, unique diseases associated with the selected genes and disease–disease associations.

Discussion and Conclusion

In this article we have built and analyzed a manually curated diseasome (disease map) from 57 genes of the Wnt signaling pathway. Disease pathways or diseasomes are potential knowledge bases that throw light on multiple disease-related complexities. Disease pathway architecture studies help us to uncover potential knowledge about their overall function that can contribute to the cure of human diseases. In this article, we have built a gene–disease network from the genes of the human Wnt signaling pathway. A disease network has been inferred from the gene–disease network. It throws light on disease co-morbidity among Wnt signaling pathway associated diseases. When divided into cancerous, noncancerous, and link (among cancerous and noncancerous disease) networks, they have showcased their individual network properties.

However, a data-driven diseasome construction always has its own limiting options, whether disease–disease associations are based on common causative genes (human Wnt diseasome) or common occurrence in patients (ICD9-5 diseasome). Sometimes lack of adequate research work limits knowledge about common genes. Scanning of appropriate knowledge among huge existing literature is crucial. Manual curation of data vouches for surety. But the process is slow and can take ages to complete. Automatic literature scanning is faster. In that case, machine intelligence has its own limitation of false positives and true negatives. Keeping a balance between them is tricky. Information about common occurrence in patients is dependent on hospital datasets. These datasets cannot ensure capturing a complete cross-section of the population. They lack information of the nonhospitalized patients. Thus combining both the approaches takes us one step ahead towards a better analysis. Hence the common disease associations found in both the aforementioned diseasomes have been put through a separate set of analyses.

These analyses gave us 6, 4, and 12 highly co-morbid disease pairs in the noncancerous, link, and cancerous components respectively (Table 4). We found 15 noncancerous, 35 link, and 100 cancerous co-morbid associations among respective components of the constructed human Wnt signaling pathway diseasome. These disease associations have strong evidence of comorbidity with $RR > 1$, $\phi > 0$, and knowledge of common involved gene. Two modules have been found in the noncancerous component. The large module contained diseases sharing PPARD as common gene, while the small module had diseases sharing the gene CTNNB1 (Fig.13C). These modules showcase highly co-morbid diseases sharing a common gene, indicating a pivotal role of these genes in causing diseases. The scenario became complex in the link component. Here, 29 diseases shared 8 common genes. The component resolved into four modules. While the three small modules shared one common gene each, diseases in the large module shared 6 common genes. In the cancerous component, 34 diseases shared 13 common genes. It cannot be resolved into any module indicating high degree of co-morbidity among cancerous diseases (Fig. 15C). In summary, analyzes of these three components showcased an increased rate of co-morbidity among diseases as we switch from noncancerous to cancerous diseases. If the presence of one disease influences occurrence or risk of the other disease(s), then they definitely suggest personalized systems medicine rather than targeting one gene or one disease at a time for therapeutic purposes. The

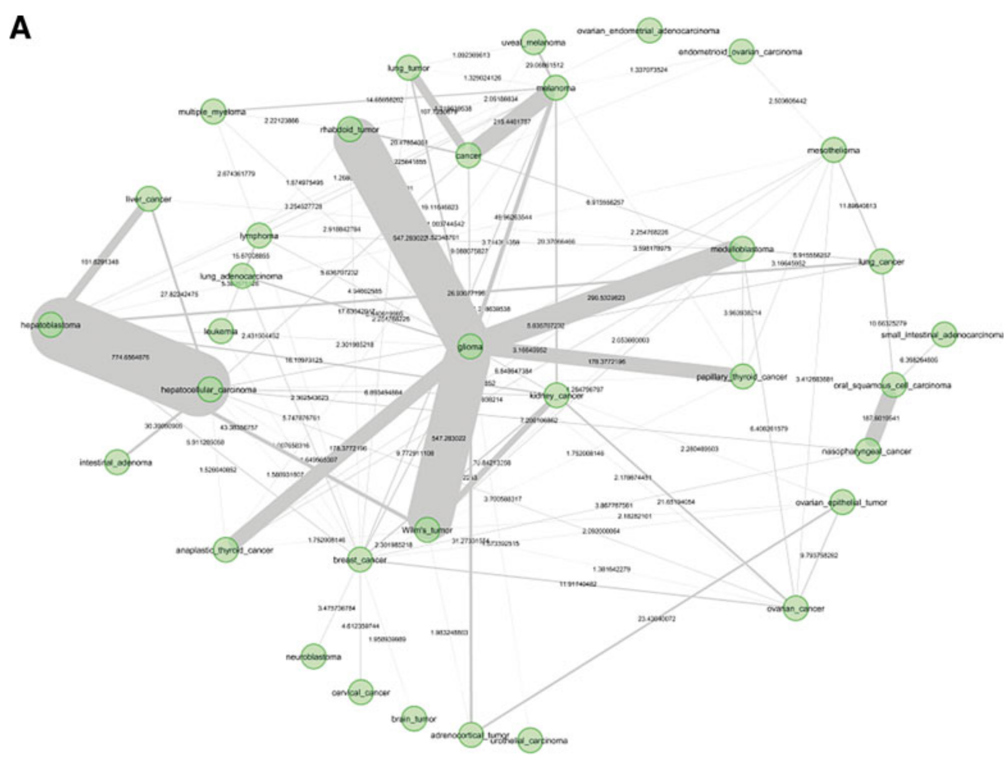

FIG. 15. **(A)** Cancerous component of human Wnt diseasome found in the ICD9-5 diseasome. Edge width is a map of *RR* value of the disease association. **(B)** Cancerous component of human Wnt diseasome found in the ICD9-5 diseasome. Edge width is a map of ϕ-correlation value of the disease association. **(C)** Cancerous component of human Wnt diseasome found in the ICD9-5 diseasome. Edge label indicates the common gene associated with both the diseases, based on which the association has been formed.

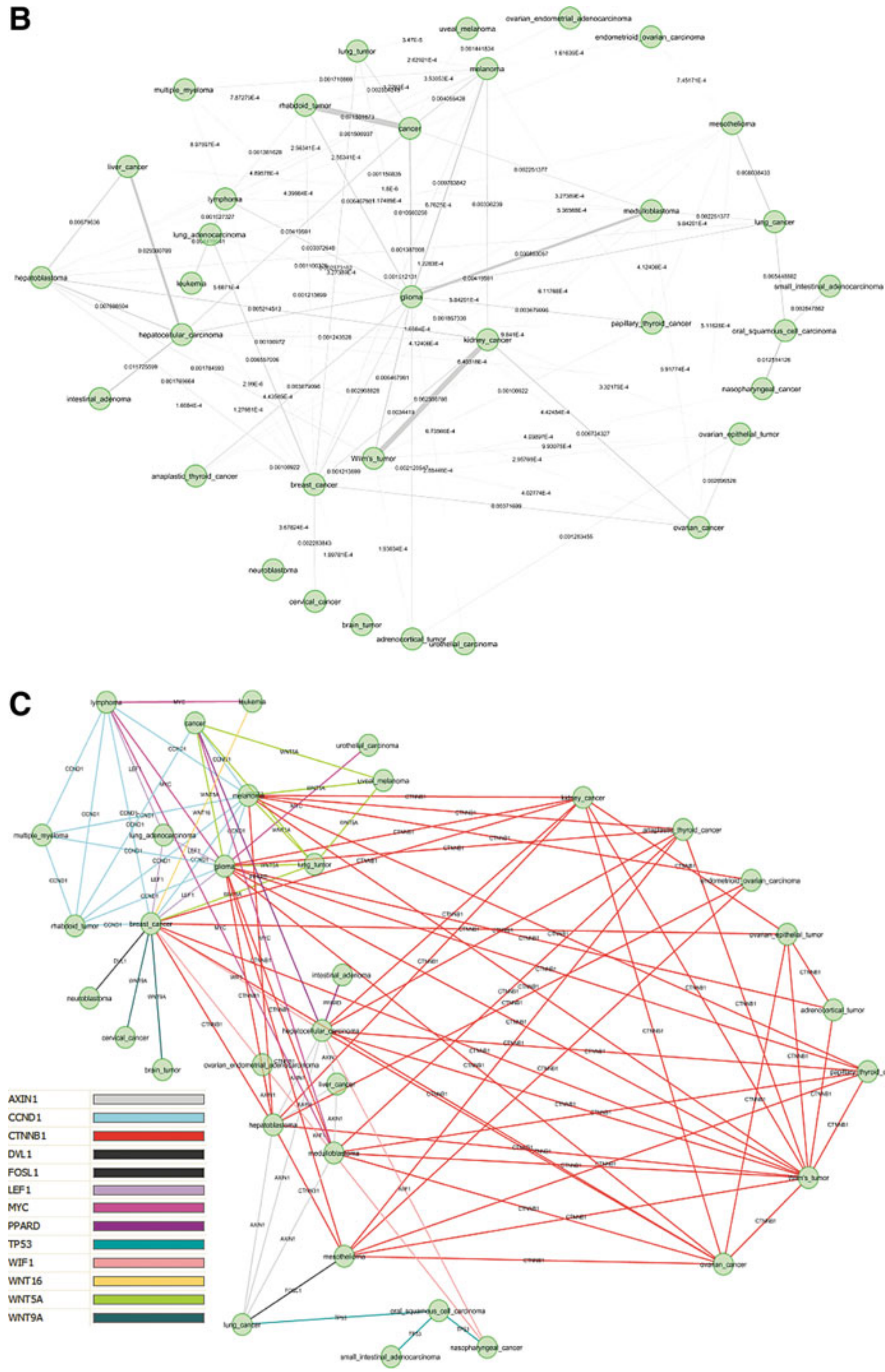

FIG. 15. (Continued).

whole data analyzed in this article is available in a publicly accessible web server.

Author Disclosure Statement

The authors declare that no competing financial interests exist.

References

Amberger J, Bocchini C, and Hamosh A. (2011). A new face and new challenges for Online Mendelian Inheritance in Man (OMIM(R)). Hum Mutat 32, 564–567.

Apweiler R, Martin MJ, O'Donovan C, et al. (2012). Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res 40, D71.

Assenov Y, Ramirez F, Schelhorn SE, Lengauer T, and Albrecht M. (2008). Computing topological parameters of biological networks. Bioinformatics 24, 282–284.

Barabasi AL. (2007). Network medicine—from obesity to the "diseasome". N Engl J Med 357, 404–407.

Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Estell J, and Sayers EW. (2012). GenBank. Nucleic Acids Res 40, D48–D53.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, and Wheeler DL. (2004). GenBank: Update. Nucleic Acids Res 32, D23–D26.

Boeckmann B, Bairoch A, Apweiler R, et al. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res 31, 365–370.

Chaurasia G, Malhotra S, Russ J, et al. (2009). UniHI 4: New tools for query, analysis and visualization of the human protein-protein interactome. Nucleic Acids Res 37, D657–660.

Citrome L, and Vreeland B. (2008). Schizophrenia, obesity, and antipsychotic medications: What can we do? Postgrad Med 120, 18–33.

Flicek P, Amode MR, Barrell D, et al. (2012). Ensembl 2012. Nucleic Acids Res 40, D84–D90.

Goel R, Muthusamy B, Pandey A, and Prasad TS. (2011). Human protein reference database and human proteinpedia as discovery resources for molecular biotechnology. Mol Biotechnol 48, 87–95.

Goh KI, Cusick ME, Valle D, Childs B, Vidal M, and Barabasi AL. (2007). The human disease network. Proc Natl Acad Sci USA 104, 8685–8690.

Hamosh A, Scott AF, Amberger JS, Bocchini CA, and McKusick VA. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res 33, D514–D517.

Hatsell S, Rowlands T, Hiremath M, and Cowin P. (2003). Beta-catenin and Tcfs in mammary development and cancer. J Mammary Gland Biol Neoplasia 8, 145–158.

Hidalgo CA, Blumm N, Barabasi A-L, and Christakis NA. (2009). A dynamic network approach for the study of human phenotypes. PLoS Comput Biol 5:e1000353.

Jensen LJ, Kuhn M, Stark M, et al. (2009). STRING 8—A global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res 37, D412–D416.

Kanehisa M, Goto S, Hattori M, et al. (2006). From genomics to chemical genomics: New developments in KEGG. Nucleic Acids Res 34, D354–D357.

Keck PE, and McElroy SL. (2003). Bipolar disorder, obesity, and pharmacotherapy-associated weight gain. J Clin Psychiatry 64, 1426–1435.

Kolotkin RL, Corey-Lisle PK, Crosby RD, et al. (2008). Impact of obesity on health-related quality of life in schizophrenia and bipolar disorder. Obesity 16, 749–754.

Le D, and Kwon Y. (2011). The effects of feedback loops on disease comorbidity in human signaling networks. Bioinformatics 27, 1113–1120.

Lee DS, Park J, Kay KA, Christakis NA, Oltvai ZN, and Barabasi AL. (2008). The implications of human metabolic network topology for disease comorbidity. Proc Natl Acad Sci USA 105, 9880–9885.

Liu X, Yu X, Zack DJ, Zhu H, and Qian J. (2008). TiGER: A database for tissue-specific gene expression and regulation. BMC Bioinformat 9, 271.

Lv J, Liu H, Su J, et al. (2012). DiseaseMeth: A human disease methylation database. Nucleic Acids Res 40, D1030–D1035.

Maglott D, Ostell J, Pruitt KD, and Tatusova T. (2011). Entrez Gene: Gene-centered information at NCBI. Nucleic Acids Res 39, D52–D57.

Palleja A, Horn H, Eliasson S, and Jensen LJ. (2012). DistiLD Database: Diseases and traits in linkage disequilibrium blocks. Nucleic Acids Res 40, D1036–D1040.

Park J, Le D, Christakis NA, and Barabasi A. (2009). The impact of cellular networks on disease comorbidity. Mol Syst Biol 5, 262.

Perez-Iratxeta C, Bork P, and Andrade-Navarro MA. (2007). Update of the G2D tool for prioritization of gene candidates to inherited diseases. Nucleic Acids Res 35, W212–W216.

Profenno LA, Porsteinsson AP, and Faraone SV. (2010). Meta-analysis of Alzheimer's disease risk with obesity, diabetes, and related disorders. Biol Psychiatry 67, 505–512.

Pruitt KD, Tatusova T, Brown GR, and Maglott DR. (2012). NCBI Reference Sequences (RefSeq): Current status, new features and genome annotation policy. Nucleic Acids Res 40, D130–D135.

Rhodes DR, Kalyana-Sundaram S, Mahavisno V, et al. (2007). Oncomine 3.0: Genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia 9, 166–180.

Schriml LM, Arze C, Nadendla S, et al. (2012). Disease Ontology: A backbone for disease semantic integration. Nucleic Acids Res 40, D940–D946.

Seal RL, Gordon SM, Lush MJ, Wright MW, and Bruford EA. (2011). genenames.org: the HGNC resources in 2011. Nucleic Acids Res 39, D514–D519.

Smoot ME, Ono K, Ruscheinski J, Wang PL, and Ideker T. (2011). Cytoscape 2.8: New features for data integration and network visualization. Bioinformatics 27, 431–432.

Syed AS, D'Antonio M, and Ciccarelli FD. (2010). Network of Cancer Genes: A web resource to analyze duplicability, orthology and network properties of cancer genes. Nucleic Acids Res 38, D670–D675.

Tiffin N, Andrade-Navarro MA, and Perez-Iratxeta C. (2009). Linking genes to diseases: It's all in the data. Genome Med 1, 77.

Tranchevent LC, Barriot R, Yu S, et al. (2008). ENDEAVOUR update: A web resource for gene prioritization in multiple species. Nucleic Acids Res 36, W377–W384.

Westbrook J, Feng Z, Jain S, et al. (2002). The Protein Data Bank: Unifying the archive. Nucleic Acids Res 30, 245–248.

Wingender E. (2008). The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. Brief Bioinform 9, 326–332.

Zhang Y, Eberhard DA, Frantz GD, et al. (2004). GEPIS–quantitative gene expression profiling in normal and cancer tissues. Bioinformatics 20, 2390–2398.

Address correspondence to:
Dr. Rajat K. De
Machine Intelligence Unit
Indian Statistical Institute
203 B. T. Road
Kolkata 700108
India

E-mail: rajat@isical.ac.in