Construction of Cancer Diseasome through Graph Theoretical Approach

Chetan H, Shujaat Afzal, Tanusree Chaudhuri, B. K. Manjunath
Student, Student, Asst. Professor, Head of the Department
Department of Biotechnology, The Oxford College Of Engineering, Bangalore, India

Abstract

Diseasome is a collection of networks that relates human diseases with the disease causing human genes. A network of disorders and disease genes linked by known disorder–gene associations offers a platform to explore in a single graph-theoretic framework all known phenotype and disease gene associations, indicating the common genetic origin of many diseases. The Online Mendelian Inheritance in Man (OMIM) is used as the data source for disease-gene relations in Diseasome. Mouse is the primary model organism to study mammalian genetics.The genome of mouse is incisively and specifically modified and controlled to study the mutations in the human genome, to discover the molecular mechanisms of various complex human diseases such as cancers, diabetes, hereditary and neurological disorders. Researchers have already identified that, essential human genes are likely to encode hub proteins and are expressed widely in most tissues, suggesting that disease genes also would play a central role in the human interactome. In our present study we have constructed and classified the human diseasome network for cancer for the better understanding of disease gene association in different types of cancer.This project aims to map the human cancer disease network onto the mouse genotype/phenotype data partaining to different types of cancer, by generating multi-partite networks of human cancer vs – human/mouse genes – phenotypic abnormalities observed in targeted knock-out-mouse models in cancer. The resulting networks will enrich the effort to curate specific symptoms and effects of different types of cancer to improve medical diagnosis.

Key words: Cancer, Human, Mouse, OMIM, Disease, Diseasome

I .Introduction:

Diseasome is a compilation of networks between human diseases with the genes causing disease (Goh et al., 2007). It is a network based study that relates human genetic disorders with the corresponding genes(Martignoni M et al., 2006)..Genes associated with similar disorders show both higher likelihood of physical interactions between their products and higher expression profiling similarity for their transcripts, supporting the existence of distinct disease-specific functional modules(Hulbert AJ, 2008). The Online Mendelian Inheritance in Man (OMIM) is used as the data source for disease-gene relations in Diseasome(Joanna S. Amberger et al., 2004). Mouse is the primary model organism to study mammalian genetics.The Genetic resemblance between mouse and human organisms is the reason behind using mouse as a model organism to study human diseases (McKusick ,V.A., 1998) . More than 90% of the mouse and human genomes can be divided into related conserved syntonic regions, which show the gene order in the genomes (Robert L. Perlman., 2016). The Diseasome mapping consists of multiple networks namely: the human disease network (HDN), the disease genes network (DGN) and the bi-partite human disease and gene network. In the study of , Goh et al. It was proposed that the disorders can be associated with each other using the shared disease-causing genes. The main list of Diseasome contained 1,284 disorders and 1,777 disease genes and all diseases are categorized based on 22 distinct disease classes(Olson H et al., 2000). Diseasome particularly focuses on the molecular relationships between genetic variation and phenotypic information, and it is a seminal work in terms of discovering the mechanisms of complex diseases. It is important here to note that, revealing complex disease mechanisms is one of the most crucial problems in biomedical research, currently (Botstein and Risch, 2003, Kann, 2009). It had already been stated in the literature that many human diseases occur due to the factors related to genetic variations (Hirschhorn and Daly, 2005). Up to date, various databases are constructed for annotating the relations between genes and diseases of human such as OMIM (Hamosh et al., 2005), CTDTM (Davis et al., 2010) and NHGRI-EBI GWAS catalog (Welter et al., 2013). Due to the nature of database curation process the associations are not complete, so the integration of multiple existing resources usually leads to more comprehensive view of the current biomedical knowledge.

Cancer, is one of leading cause of death worldwide. There are several analysis that have been performed , which enables the better understanding towards cancer proteomics by deciphering genetic and epigenetic data for gene regulatory networks analysis. These data has uncovered many important protein-protein interaction (PPI) pairs which are integrative part of the cancer network.

In our present study we have created a diseasome network for different types of cancer, which will eventually help us to better understand the cancer processes to identify biomarkers and therapeutic targets, and predict the prognosis of cancer in acute cancer patients.

II. Materials and methods:

2.1 DATA DOWNLOAD AND PROCESSING :

Curated morbid map file was being downloaded from Online Mendelian Inheritance in man (OMIM). Morbid Map (MM) of the OMIM is one of the most comprehensive and highly curated disorder gene association database. The OMIM MM shows the cytogenetic map location of disease genes in OMIM. The link www.omim.org is being opened up, wherein download tab in menu is selected and registration for download is done.
Two different datasets about human and mouse organisms were extracted. Dataset 1 contains the Human disease – human gene relation information and downloaded from Diseasome resource and the Dataset 2 contains Mouse affected system (phenotype) – mouse gene information derived from MGI and Human data were downloaded from OMIM. Mouse genes attribute was chosen as a foreign key, to relate these two sets.

2.2 DATASET DOWNLOAD FROM DISEASOME & DATA PROCESSING:

Diseasome dataset was created from curated OMIM data, that has been used as the source to constitute Dataset 1. It includes disease ID, disease name, disorder class, size (s) that show the number of associated genes, degree (k) shows number of disorder classes it connects to, class degree (K) is the number of distinct disorder classes it connects to and genes written as comma delimited at the last column.
The curated table contains the Disease ID, Disorder name, Human Gene Symbols, OMIM ID, Chromosome Position of the related gene and Disorder Class information. Disorder names were aligned in an alphabetical order and distinct consecutive numbers are given in ascending order starting from 1. These numbers are called as Disease ID and assigned for analysis in Gephi. Disorder names are distinctly ordered with their related human genes and in accordance OMIM Ids are retrieved. If a disorder has more than one genes related to it, these genes are separated with comma.
Mouse orthologues of human genes were converted and extracted with the online converter tool called as HCOP: Orthologue Predictions Search.

2.3 DATASET DOWNLOAD FROM MGI & DATA PROCESSING:

Mouse affected systems information (i.e. phenotypes) was collected from the MGI database. Collected mouse orthologue genes with HCOP were imported to the MGI batch summary tool for creating Dataset 2.
Only the targeted null/knock-out mouse genes were taken into consideration during the generation of Dataset 2.
The dataset 2 includes affected system information with unique "Mammalian phenotype ID" of all recorded mouse genes with marker symbols in that database. It also provides unique MGI IDs for these genes, allele type and allele attribute information.

2.4 Collection of Cancer data: Curated Morbid map from dataset 1 was searched regorously for different types of cancer information. A total set of 74 different types of cancer with the involvement of 202 gene was identified.
Dataset 2 was constructed for this set following the process in 2.3.

2.5 INTEGRATION OF DATA & GENERATING THE NETWORKS:

The data integration was based on connecting human diseases and mouse affected systems (i.e. phenotypes) by using mouse/human orthologous genes. We have followed the genes as nodes stretgy to generate the networks.

Human diseases are indirectly connected to the mouse phenotypes (i.e. affected systems) while using mouse/human orthologous genes as the mediator. Relations in-between genes-diseases-phenotypes. Human diseases are indirectly connected to the mouse phenotypes (i.e. affected systems) while using mouse/human orthologous genes as the mediator.

III Result and Discussion:

3.1 Creation of Dataset 1:

Dataset 1 was created with required disease name corresponding disease ID, human genes, OMIM ID, Chromosome Position of the related gene and Disorder Class information. Disorder names were aligned in an alphabetical order and distinct consecutive numbers are given in ascending order starting from 1. These numbers are called as Disease ID and assigned for analysis in Gephi.The Figure 3.1 shows a screenshot the curated dataset1

	Disease ID	Disorder name	Gene symbols	OMIM ID	Chromosome	Class
2	427	Diabetes mellitus, type II, 125853 (3)	AKT2	164731	19q13.1-q13.2	Endocrine
3	427	Diabetes mellitus, type II, susceptibility to, 125853 (3)	IPF1	600733	13q12.1	Endocrine
4	918	Lung cancer, 211980 (3)	KRAS2, RASK2	190070	12p12.1	Cancer
5	918	Lung cancer, 211980 (3)	PPP2R1B	603113	11q22-q24	Cancer
6	918	Lung cancer, 211980 (3)	SLC22A1L, BWSCR1A, IMPT1	602631	11p15.5	Cancer
7	918	Lung cancer, somatic, 211980 (3)	MAP3K8, COT, EST, TPL2	191195	10p11.2	Cancer
8	1324	Rheumatoid arthritis, progression of, 180300 (3)	IL10, CSIF	124092	1q31-q32	Connective tissue
9	1324	Rheumatoid arthritis, susceptibility to, 180300 (3)	MHC2TA, C2TA	600005	16p13	Connective tissue
10	1324	Rheumatoid arthritis, susceptibility to, 180300 (3)	NFKBIL1	601022	6p21.3	Connective tissue
11	1324	Rheumatoid arthritis, susceptibility to, 180300 (3)	PADH, PADI5, PAD	605347	1p36	Connective tissue
12	1324	Rheumatoid arthritis, susceptibility to, 180300 (3)	PTPN8, PEP, PTPN22, LYP	600716	1p13	Connective tissue
13	1324	Rheumatoid arthritis, susceptibility to, 180300 (3)	RUNX1, CBFA2, AML1	151385	21q22.3	Connective tissue
14	1324	Rheumatoid arthritis, susceptibility to, 180300 (3)	SLC22A4, OCTN1	604190	5q31	Connective tissue
15	1324	Rheumatoid arthritis, systemic juvenile, susceptibility to,	MIF	153620	22q11.2	Connective tissue
16	1304	Rapid progression to AIDS from HIV1 infection (3)	CX3CR1, GPR13, V28	601470	3pter-p21	Immunological
17	1305	Rapp-Hodgkin syndrome, 129400 (3)	TP73L, TP63, KET, EEC3, SHFM4, LMS, RHS	603273	3q27	multiple
18	70	AIDS, delayed/rapid progression to (3)	KIR3DL1, NKAT3, NKB1, AMB11, KIR3DS1	604946	19q13.4	Immunological
19	70	AIDS, rapid progression to, 609423 (3)	IFNG	147570	12q14	Immunological
20	70	AIDS, resistance to (3)	CXCL12, SDF1	600835	10q11.1	Immunological
21	1043	Mycobacterium tuberculosis, suceptibility to infection by	NRAMP1, NRAMP	600266	2q35	Immunological
22	1400	Sjogren-Larsson syndrome, 270200 (3)	ALDH3A2, ALDH10, SLS, FALDH	609523	17p11.2	Metabolic
23	1454	Stroke, susceptibility to, 1, 606799 (3)	PDE4D, DPDE3, STRK1	600129	5q12	Cardiovascular
24	1454	Stroke, susceptibility to, 601367 (3)	ALOX5AP, FLAP	603700	13q12	Cardiovascular
25	1533	Tuberculosis, susceptibility to (3)	IFNGR1	107470	6q23-q24	Respiratory

FIGURE 3.1: Sample Dataset 1

For our present study, we have identified 74 different types of cancer, with 202 human genes with their OMIM Id, chromosome location.

	A	B	C	D	E	F
1	DISEASE ID	DISEASE NAME	HUMAN GENE	OMIM ID	CHROMOSOME	CLASS
2	207	Bladder cancer, 109800 (3)	FGFR3, ACH	134934	4p16.3	Cancer
3	207	Bladder cancer, 109800 (3)	KRAS2, RASK2	190070	12p12.1	Cancer
4	207	Bladder cancer, 109800 (3)	RB1	180200	13q14.1-q14.2	Cancer
5	207	Bladder cancer, somatic, 109800 (3)	HRAS	190020	11p15.5	Cancer
6	228	Breast and colorectal cancer, susceptibility to (3)	CHEK2, RAD53, CHK2, CDS1, LFS2	604373	22q12.1	Cancer
7	228	Breast cancer, 114480 (3)	PIK3CA	171834	3q26.3	Cancer
8	228	Breast cancer, 114480 (3)	PPM1D, WIP1	605100	17q22-q23	Cancer
9	228	Breast cancer, 114480 (3)	SLC22A1L, BWSCR1A, IMPT1	602631	11p15.5	Cancer
10	228	Breast cancer, 114480 (3)	TP53, P53, LFS1	191170	17p13.1	Cancer
11	228	Breast cancer-1 (3)	BRCA1, PSCP	113705	17q21	Cancer
12	228	Breast cancer 2, early onset (3)	BRCA2, FANCD1	600185	13q12.3	Cancer
13	228	Breast cancer (3)	TSG101	601387	11p15.2-p15.1	Cancer
14	228	Breast cancer, early-onset, 114480 (3)	BRIP1, BACH1, FANCJ	605882	17q22	Cancer
15	228	Breast cancer, invasive intraductal (3)	RAD54L, HR54, HRAD54	603615	1p32	Cancer

Figure 3.2: Dataset 1 for cancer diseases

3.2 Creation of Dataset 2:

Dataset 2 consists of phenotype terms with their MP ID's and targeted knock-out mouse orthologues of human genes. Human gene column again was added for the ease of understanding. This dataset is based on mouse data. Mouse affected systems information (i.e. phenotypes) was collected from the MGI database.An example dataset 2 is described in figure 3.3

	MP ID	AFFECTED SYSTEM (PHENOTYPE)	MOUSE GENE	HUMAN GENE	DISEASE ID	HUMAN DISEASE	DISORDER CLASS
2	MP:0002078	abnormal glucose homeostasis	Akt2	AKT2	427	Diabetes mellitus, type II	Endocrine
3	MP:0014169	decreased brown adipose tissue mass	Akt2	AKT2	427	Diabetes mellitus, type II	Endocrine
4	MP:0009356	decreased liver triglyceride level	Akt2	AKT2	427	Diabetes mellitus, type II	Endocrine
5	MP:0030022	decreased muscle cell glucose uptake	Akt2	AKT2	427	Diabetes mellitus, type II	Endocrine
6	MP:0014146	decreased white adipose tissue mass	Akt2	AKT2	427	Diabetes mellitus, type II	Endocrine
7	MP:0005378	growth/size/body region phenotype	Akt2	AKT2	427	Diabetes mellitus, type II	Endocrine
8	MP:0000316	cellular necrosis	Kras2	KRAS2	918	Lung cancer	Cancer
9	MP:0010856	dilated respiratory conducting tubes	Kras2	KRAS2	918	Lung cancer	Cancer
10	MP:0005379	endocrine/exocrine gland phenotype	Kras2	KRAS2	918	Lung cancer	Cancer
11	MP:0010771	integument phenotype	Kras2	KRAS2	918	Lung cancer	Cancer
12	MP:0010768	mortality/aging	Kras2	KRAS2	918	Lung cancer	Cancer
13	MP:0002006	neoplasm	Kras2	KRAS2	918	Lung cancer	Cancer
14	MP:0001954	respiratory distress	Kras2	KRAS2	918	Lung cancer	Cancer
15	MP:0010768	mortality/aging	Kras2	KRAS2	918	Lung cancer	Cancer
16	MP:0005387	immune system phenotype	Pad	PAD	1324	Rheumatoid arthritis	Connective tissue
17	MP:0004924	abnormal behavior	Pad	PAD	1324	Rheumatoid arthritis	Connective tissue
18	MP:0003935	abnormal craniofacial development	Tp73l	TP73L	1305	Rapp-Hodgkin syndrome	multiple
19	MP:0000428	abnormal craniofacial morphology	Tp73l	TP73L	1305	Rapp-Hodgkin syndrome	multiple
20	MP:0001672	abnormal embryo development	Tp73l	TP73L	1305	Rapp-Hodgkin syndrome	multiple
21	MP:0010942	abnormal respiratory epithelium morphology	Tp73l	TP73L	1305	Rapp-Hodgkin syndrome	multiple
22	MP:0000549	absent limbs	Tp73l	TP73L	1305	Rapp-Hodgkin syndrome	multiple
23	MP:0005380	embryo phenotype	Tp73l	TP73L	1305	Rapp-Hodgkin syndrome	multiple
24	MP:0010771	integument phenotype	Tp73l	TP73L	1305	Rapp-Hodgkin syndrome	multiple
25	MP:0010768	mortality/aging	Tp73l	TP73L	1305	Rapp-Hodgkin syndrome	multiple

Figure 3.3: Sample dataset for dataset 2

For this project 1567 mouse affected systems information (i.e. phenotypes) was collected from MGI report along with their MP ID's.These mouse affeted systems were aligned to their corresponding human orthologs and dataset 3 was created for the construction of cancer Diseasome.Figure 3.4 shows the glimps of cancer diseasome.

MOUSE GENE	HUMAN GENE	MP ID	phenotype
Mdm2	MDM2	MP:0005384	cellular phenotype
Mdm2	MDM2	MP:0006043	decreased apoptosis
Mdm2	MDM2	MP:0001262	decreased body weight
Mdm2	MDM2	MP:0000333	decreased bone marrow cell number
Mdm2	MDM2	MP:0001698	decreased embryo size
Mdm2	MDM2	MP:0002875	decreased erythrocyte cell number
Mdm2	MDM2	MP:0005378	growth/size/body region phenotype
Egfr	EGFR	MP:0002161	abnormal fertility/fecundity
Egfr	EGFR	MP:0000763	abnormal filiform papillae morphology
Egfr	EGFR	MP:0006257	abnormal fungiform papillae morphology
Egfr	EGFR	MP:0008325	abnormal gonadotroph morphology
Braf	BRAF	MP:0002182	abnormal astrocyte morphology
Braf	BRAF	MP:0000297	abnormal atrioventricular cushion morphology
Braf	BRAF	MP:0010029	abnormal basicranium morphology
Braf	BRAF	MP:0001614	abnormal blood vessel morphology
Braf	BRAF	MP:0001777	abnormal body temperature homeostasis
Braf	BRAF	MP:0004259	small placenta
Braf	BRAF	MP:0001209	spontaneous skin ulceration
Braf	BRAF	MP:0005426	tachypnea
Erbb2	ERBB2	MP:0000920	abnormal myelination
Erbb2	ERBB2	MP:0003871	abnormal myelin sheath morphology

Figure3.4 : dataset 2 for Cancer

3.4 Construction of Dataset 3:

Dataset 1 and Dataset 2 were merged by integrating the human and mouse data tables to create Dataset 3. A link was established between human and mouse data using the targeted knock-out mouse orthologues of human genes.A sample datset3 is shown in figure 3.5.

MP ID	AFFECTED SYSTEM (PHENOTYPE)	MOUSE GENE	HUMAN GENE	DISEASE ID	HUMAN DISEASE	DISORDER CLASS
MP:0002078	abnormal glucose homeostasis	Akt2	AKT2	427	Diabetes mellitus, type II	Endocrine
MP:0014169	decreased brown adipose tissue mass	Akt2	AKT2	427	Diabetes mellitus, type II	Endocrine
MP:0009356	decreased liver triglyceride level	Akt2	AKT2	427	Diabetes mellitus, type II	Endocrine
MP:0030022	decreased muscle cell glucose uptake	Akt2	AKT2	427	Diabetes mellitus, type II	Endocrine
MP:0014146	decreased white adipose tissue mass	Akt2	AKT2	427	Diabetes mellitus, type II	Endocrine
MP:0005378	growth/size/body region phenotype	Akt2	AKT2	427	Diabetes mellitus, type II	Endocrine
MP:0000316	cellular necrosis	Kras2	KRAS2	918	Lung cancer	Cancer
MP:0010856	dilated respiratory conducting tubes	Kras2	KRAS2	918	Lung cancer	Cancer
MP:0005379	endocrine/exocrine gland phenotype	Kras2	KRAS2	918	Lung cancer	Cancer
MP:0010771	integument phenotype	Kras2	KRAS2	918	Lung cancer	Cancer
MP:0010768	mortality/aging	Kras2	KRAS2	918	Lung cancer	Cancer
MP:0002006	neoplasm	Kras2	KRAS2	918	Lung cancer	Cancer
MP:0001954	respiratory distress	Kras2	KRAS2	918	Lung cancer	Cancer
MP:0010768	mortality/aging	Kras2	KRAS2	918	Lung cancer	Cancer
MP:0005387	immune system phenotype	Pad	PAD	1324	Rheumatoid arthritis	Connective tissue
MP:0004924	abnormal behavior	Pad	PAD	1324	Rheumatoid arthritis	Connective tissue
MP:0003935	abnormal craniofacial development	Tp73l	TP73L	1305	Rapp-Hodgkin syndrome	multiple
MP:0000428	abnormal craniofacial morphology	Tp73l	TP73L	1305	Rapp-Hodgkin syndrome	multiple
MP:0001672	abnormal embryo development	Tp73l	TP73L	1305	Rapp-Hodgkin syndrome	multiple
MP:0010942	abnormal respiratory epithelium morphology	Tp73l	TP73L	1305	Rapp-Hodgkin syndrome	multiple
MP:0000549	absent limbs	Tp73l	TP73L	1305	Rapp-Hodgkin syndrome	multiple
MP:0005380	embryo phenotype	Tp73l	TP73L	1305	Rapp-Hodgkin syndrome	multiple
MP:0010771	integument phenotype	Tp73l	TP73L	1305	Rapp-Hodgkin syndrome	multiple
MP:0010768	mortality/aging	Tp73l	TP73L	1305	Rapp-Hodgkin syndrome	multiple

Figure 3.5: Sample Dataset 3

Again, we have created a dataset 3 for the construction of cancer diseasome with name of different types of cancer, disease ID(3 is specified for cancer), phenotype, mouse and human gene, MP ID's . Figure 3.6 shows the cancer diseasome

MP ID	MOUSE GENE	HUMAN GENE	HUMAN DISEASE	phenotype	CLASS
MP:0005384	Mdm2	MDM2	Accelerated tumor formation, susceptibility to (3)	cellular phenotype	CANCER
MP:0006043	Mdm2	MDM2	Accelerated tumor formation, susceptibility to (3)	decreased apoptosis	CANCER
MP:0001262	Mdm2	MDM2	Accelerated tumor formation, susceptibility to (3)	decreased body weight	CANCER
MP:0000333	Mdm2	MDM2	Accelerated tumor formation, susceptibility to (3)	decreased bone marrow cell number	CANCER
MP:0001698	Mdm2	MDM2	Accelerated tumor formation, susceptibility to (3)	decreased embryo size	CANCER
MP:0002875	Mdm2	MDM2	Accelerated tumor formation, susceptibility to (3)	decreased erythrocyte cell number	CANCER
MP:0005378	Mdm2	MDM2	Accelerated tumor formation, susceptibility to (3)	growth/size/body region phenotype	CANCER
MP:0002161	Egfr	EGFR	Adenocarcinoma of lung, response to tyrosine kinase inhit	abnormal fertility/fecundity	CANCER
MP:0000763	Egfr	EGFR	Adenocarcinoma of lung, response to tyrosine kinase inhit	abnormal filiform papillae morphology	CANCER
MP:0006257	Egfr	EGFR	Adenocarcinoma of lung, response to tyrosine kinase inhit	abnormal fungiform papillae morphology	CANCER
MP:0008325	Egfr	EGFR	Adenocarcinoma of lung, response to tyrosine kinase inhit	abnormal gonadotroph morphology	CANCER
MP:0002182	Braf	BRAF	Adenocarcinoma of lung, somatic, 211980 (3)	abnormal astrocyte morphology	CANCER
MP:0000297	Braf	BRAF	Adenocarcinoma of lung, somatic, 211980 (3)	abnormal atrioventricular cushion morphol	CANCER
MP:0010029	Braf	BRAF	Adenocarcinoma of lung, somatic, 211980 (3)	abnormal basicranium morphology	CANCER
MP:0001614	Braf	BRAF	Adenocarcinoma of lung, somatic, 211980 (3)	abnormal blood vessel morphology	CANCER
MP:0001777	Braf	BRAF	Adenocarcinoma of lung, somatic, 211980 (3)	abnormal body temperature homeostasis	CANCER
MP:0004259	Braf	BRAF	Adenocarcinoma of lung, somatic, 211980 (3)	small placenta	CANCER
MP:0001209	Braf	BRAF	Adenocarcinoma of lung, somatic, 211980 (3)	spontaneous skin ulceration	CANCER
MP:0005426	Braf	BRAF	Adenocarcinoma of lung, somatic, 211980 (3)	tachypnea	CANCER
MP:0000920	Erbb2	ERBB2	Adenocarcinoma of lung, somatic, 211980 (3)	abnormal myelination	CANCER
MP:0003871	Erbb2	ERBB2	Adenocarcinoma of lung, somatic, 211980 (3)	abnormal myelin sheath morphology	CANCER

Figure 3.6: Cancer Dataset 3

3.5 Clustering and analysis of the network:

The network visualisation was done in gephi tool. The disease names were taken as nodes and the edges was for assigned for the phenotypes.The network for different types of cancer was created, which has been shown in figure 3.7. In this figure the genes partening to different types of cancer is used as the nodes, and phenotypes for cancer is being used as the edges. After analysis of the network it is being observed a very high modularity in cancer phenotypes.Specially the genes, that are present at the core of the network shows similar kind of phenotypes rather than the nodes present on the surface of the network.

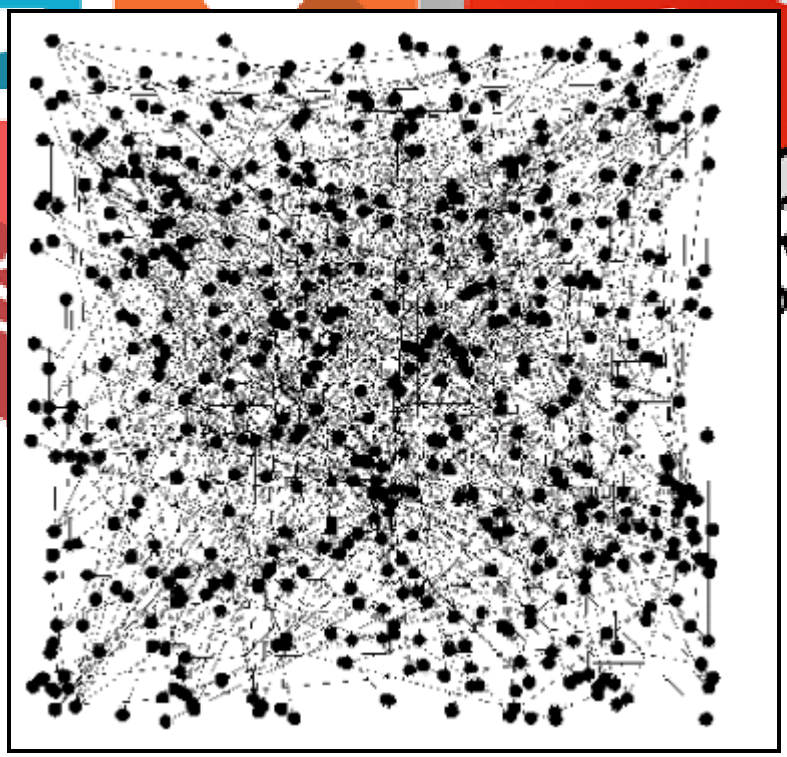


Figure 3.7 Cancer Diseasome.

So, to analyze the gnes present in the network, we have reconstructed the network only with the genes as nodes without edges.This time we have used an unique colour codes for the genes. The colour code is shoen in figure 3.8(a). The network is reconstructed using the colour code in 3.8(b). In this new network it is being observed the KRAS2 is the gene that shows majority of the phenotypes in cancer(Fig 3.8 a)

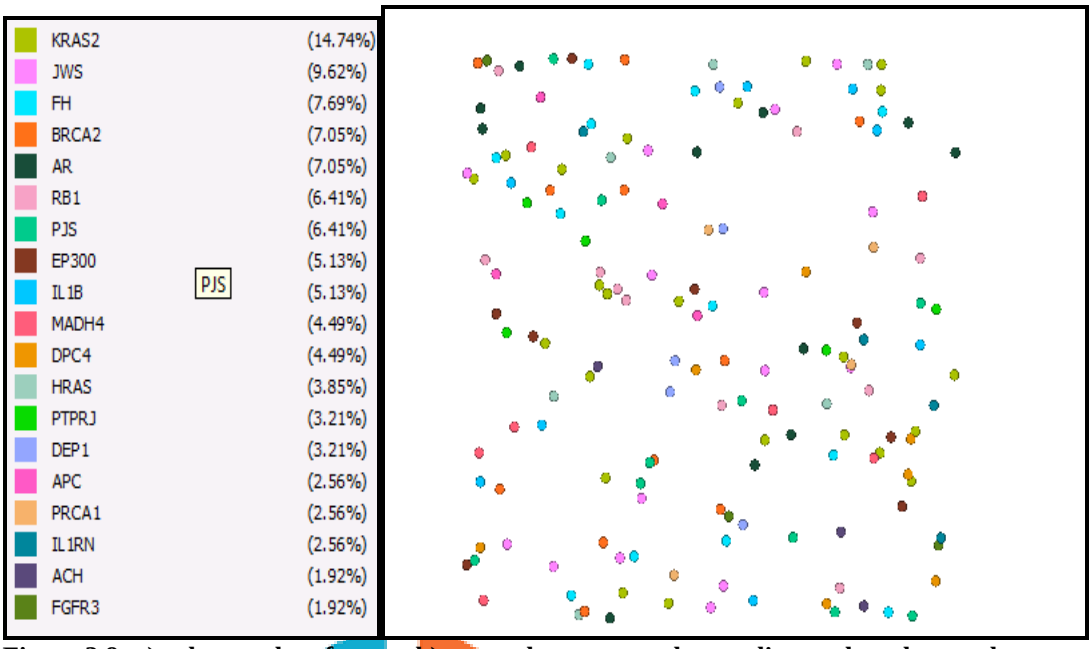


Figure 3.8: a) colour codes of genes. b) network constructed according to the colour code.

So, to analyze, the core part of the network more intensely as well as the phenotypes associated with it, we decided to break the main network into different modules. Figure 3.9 shows the modularity in the cancer network. After analysis of the modularity in the cancer network, it has been found that there are 813 different modules that are present and the most dominant phenotypes in cancer are nervous system phenotype modularity 34, abnormal dermis papillary layer morphology modularity 27, abnormal embryo development modularity 21 etc.

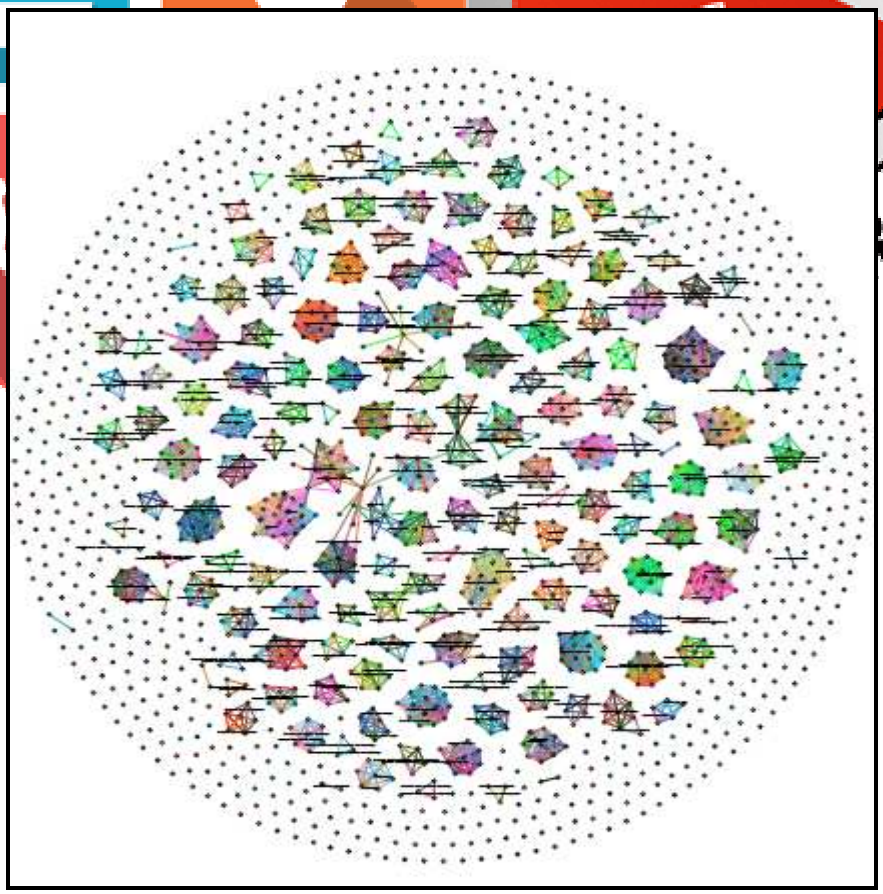
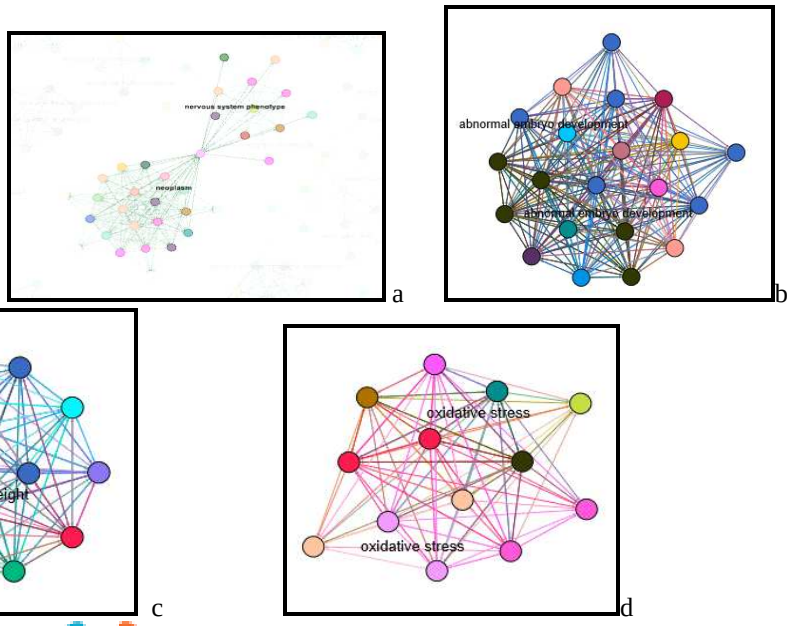


Fig 3.9: Modularity in cancer network.

so, to make our conclusion more apprehensive, we have given detailed diagram of some of the modularity in the following figure 3.10.



a b c d

Figure 3.10: a) Cluster of neoplasm and nervous system phenotype b) Cluster of abnormal embryo development c) Cluster of decreased body weight d) Cluster of oxidative stress.

In our present study , we have even listed out the major phenotypes that occur during different types across different stages of cancer.Table 3.1 shows a detailed list of mostly occur phenotypes in cancer.

phenotype	modularity_class	No of Occurance
nervous system phenotype	119	34
abnormal dermis papillary layer morphology	298	27
abnormal embryo development	771	21
ventricular septal defect	328	16
decreased incidence of tumors by chemical induction	108	16
enlarged thymus	62	16
increased tumor incidence	565	15
decreased embryo size	291	15
premature death	302	14
decreased body weight	35	14
prenatal lethality, incomplete penetrance	741	13
no abnormal phenotype detected	562	13
abnormal thyroid-stimulating hormone level	307	13
abnormal bone marrow cell morphology/development	339	12
increased apoptosis	146	12
increased thyrotroph cell number	83	12
respiratory distress	512	11
abnormal cell physiology	340	11
cardiovascular system phenotype	332	11
abnormal hematopoietic system morphology/development	299	11
increased or absent threshold for auditory brainstem response	281	11
abnormal long bone diaphysis morphology	173	10
decreased cardiac muscle contractility	153	10
abnormal coat appearance	121	10
abnormal myocardium layer morphology	9	10

Table 3.1: The common phenotypes of all cancer.

In our present study, we have created diseasome for all types of cancer in human. From the diseasome, we have identified the major phenotypes that occur from different types to diffreent stages of cancer.This analysis of diseasome has been carried out using knockout mouse model will help in characterization of different types of cancer. This study will have a significant impact on the development of methodological approaches toward precise identification of pathological cells and would allow for more effective detection of cancer-related changes.

IV Reference:

1. **KWANG-IL GOH, MICHAEL E. CUSICK, DAVID VALLE, BARTON CHILDS, MARC VIDAL, AND ALBERT-LÁSZLÓ BARABÁSI, 2007. THE HUMAN DISEASE NETWORK**
2. McKusick,V.A. (1998) Mendelian Inheritance in Man. A Catalog of Human Genes and Genetic Disorders, 12th edn. Johns Hopkins University Press, Baltimore, MD.
3. Ada Hamosh*, Alan F. Scott, Joanna S. Amberger, Carol A. Bocchini and Victor A. McKusick, 2004 Online Mendelian Inheritance in Man (OMIM), a knowledgebase ofhumangenes and genetic disorders.
4. Olson H, Betton G, Robinson D et al., 2000 Concordance of the toxicity of pharmaceuticals in humans and in animals.; 32:56–67.
5. Martignoni M, Groothuis GM, de Kanter R. 2006 Species differences between mouse, rat, dog, monkey and human CYPmediated drug metabolism, inhibition and induction. Expert Opin Drug Metab Toxicol ;2:875–94.
6. Suarez RK, Darveau CA. 2005 Multi-level regulation and metabolic scaling. J Exp Biol;208:1627–34.
7. Hulbert AJ 2008. The links between membrane composition, metabolic rate and lifespan. Comp Biochem Physiol A Mol Integr Physiol ;150:196–203.
8. Berry RJ, Bronson FH 1992. Life history and bioeconomy of the house mouse. Biol Rev Camb Philos Soc ;67:519–50.
9. Phelan JP, Rose MR.2005 Why dietary restriction substantially increases longevity in animal models but won't in humans. Age Res Rev ;4:339–50.
10. Nguyen TL, Vieira-Silva S, Liston A et al.2015 How informative is the mouse for human gut microbiota research? Dis Model Mech ;8:1–16.
11. Mestas J, Hughes CC.2004 Of mice and not men: differences between mouse and human immunology. J Immunol ;172:2731–8.
12. Nachman MW, Searle JB. 1995 Why is the house mouse karyotype so variable? Trends Ecol Evol ;10:397–402.
13. Austad SN.2002 A mouse's tale. Nat Hist ;111:64–70.
14. Paigen K. 2003 One hundred years of mouse genetics: an intellectual history the classical period (1902–1980). Genetics 2003;163:1–7.

Acknowledgement:

We sincerely thank the management of The Oxford College of Engineering and VGST for providing the opportunity to work.