Contents lists available at UGC-CARE

International Journal of Pharmaceutical Sciences and Drug Research

[ISSN: 0975-248X; CODEN (USA): IJPSPP]

Available online at www.ijpsdronline.com

Research Article

Protein Interaction Network Analysis of *β-catenin* to Map its Crucial Interacting Genes using Systems Biology Approach

Preema K. Andrade[1*], Manjunatha Hanumanthappa[2], Vijayalakshmi Venkataramanaiah[1], Nayana Prakash[2]

[1]*Department of PG Studies and Research in Biotechnology & Bioinformatics, Kuvempu University, Shivamogga, Karnataka, India*
[2]*Department of PG Studies and Research in Biochemistry, Jnana Bharathi, Bangalore University, Bangalore, Karnataka, India*

ARTICLE INFO

Article history:
Received: 31 December, 2022
Revised: 10 February, 2023
Accepted: 15 February, 2023
Published: 30 March, 2023

Keywords:
β-catenin, Wnt-signaling pathway, protein-protein interaction, Cytoscape, functional enrichment, Cytohubba

DOI:
10.25004/IJPSDR.2023.150209

ABSTRACT

β-catenin (*CTNNB1*), a scaffold protein, plays a vital role in embryonic development, adult tissue homeostasis, neurodegenerative diseases, bone diseases, chronic obstructive pulmonary diseases, wound healing, and pigment disorders. *β-catenin* is the crucial component in the *Wnt*-signaling pathway. *β-catenin* levels in the cytosol balance the *Wnt*-signaling pathway. The activated *Wnt*-signaling pathway allows the accumulation of *β-catenin* in the nucleus and further act as a promoter to initiate transcription of several oncogenes responsible for carcinogenesis. In the present study, we have used a systems biology approach to map *β-catenin* interactors. Primary protein-protein interaction databases such as PubChem, CTD, and STITCH databases were used to collect all the experimentally valid *β-catenin* interactors and their sub-cellular location was traced using UniProtKB database. The interactome was constructed and analyzed by utilizing STRING and Cytoscape tools. MCODE, Cytoscape tool, was utilized to construct and analyze the sub-networks. Correlation between modular seed proteins and *β-catenin* was studied using UALCAN database. Functional enrichment analysis was done using DAVID database. Cytohubba, a Cytoscape tool was utilized to identify the top gene interactors of *β-catenin.* Further, the expression and gene ontology of each *β-catenin* gene interactor were analyzed using UALCAN and CleuGO tools. The analysis reveals *β-catenin* interactors, *TP53, EP300, RPS27A, UBC, HDAC1, SRC, AKT1, EGFR, HSP90AA1*, and *CREBBP* as the first top gene interactors. This study uses a systems biology-driven approach and successfully identifies and understands the biological roles of the top *β-catenin* interactors.

Introduction

Systems biology is an integrated approach to study biology that unifies information from many different fields and gives a new system characteristic. Organisms continuously interact with their environment. This shapes the characteristics and evolutionary trajectories of the organism.[1] Some complex diseases involve many different proteins interacting through a complex network. In that case targeting a single gene will not be sufficient because the cell escapes from blockage via alternative molecular routes. Systems biology allows the researcher to understand and map the protein networks and provides computational, statistical, mathematical tools, and a variety of omic technologies to analyze these complex networks.[2] Understanding the biology of complex systems and disease mechanisms will open the door for developing efficient treatment strategies.

In general, Proteins perform their function in coordination with other proteins by forming a network. In other words, proteins activate their functionalities by their interaction with other proteins. These complex networks of proteins play a pivotal role in normal biological processes.[3,4] The disturbance in these complex networks result in pathological condition, including cancer.[5]

β-catenin (*CTNNB1*) is the multifunctional mammalian homolog of the Drosophila armadillo gene. Its high

***Corresponding Author:** Dr. Manjunatha Hanumanthappa

Address: Department of PG Studies and Research in Biotechnology & Bioinformatics, Kuvempu University, Shankaraghatta, Shivamogga, Karnataka, India

Email ✉: manjunatha75@gmail.com

Tel.: +91 8660456269

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

expression results in the initiation and prognosis of many kinds of cancer and targeting β-catenin signaling serves as the best way to treat cancer.[6] β-catenin is an important element in canonical *Wnt* signaling pathway where it serves as signal transducer along with disheveled protein (*Dvl),* glycogen synthase kinase 3 beta *(GSK3β),* adenomatous polyposis coli (*APC)*, axis inhibition (*AXIN)*, and T-cell factor/lymphoid enhancement factor (TCF/LEF). *Wnt* is the regulator of *β-catenin.* Upon its activation, *β-catenin* deposits in the cytoplasm and in the nucleus. Eventually, it attaches to *TCF/LEF1*, resulting in the transcription of Jun proto-oncogene (*Jun)*, Myc-proto-oncogene (*c-Myc),* and *CyclinD-1* (CCND1) which encode oncoproteins.[7,8]

In human, *β-catenin* consists of 781 amino acid residues with a central region consists of 141 to 664 amino acid residues which consists of 12 armadillo repeats flanked by amino- terminal domain (NTD) and carboxy-terminal domain (CTD). Structurally, amino- terminal domain and carboxy-terminal domain are flexible in nature, while the central region is a rigid scaffold. This scaffold allows the *β-catenin* interactors to bind the nucleus, cytoplasm, and the cell membrane.[9]

This study identified the top β-catenin interactors and their role in various molecular functions, biological processes and pathways. This helps to target those top genes to target complex diseases involving highly intertwined pathways.

Materials and Methods

Data Curation and Construction of PPI Network

All the interacting genes of *β-catenin* were retrieved by utilizing the primary protein-protein interaction databases such as STITCH v 5.0 (Search tool for interactions of chemical, http://stitch.embl.de/),[10] CTD (Comparative Toxicogenomics Database, http://ctdbase.org/),[11,12] and PubChem databases.[13] UniProtKB database was used to trace the curated genes' subcellular locations. Using STRING v 11.5 database (https://string-db.org/),[14] the genes were put through protein-protein interaction (PPI/interactome) network construction by setting the organism as "*Hamo sapiens"* and the minimum required interaction score was set up to "high confidence (0.70)". This database helps to integrate all the protein associations including functional association and physical interaction as well.[15] The obtained PPI networks were imported to construct interactome using Cytoscape v 2.8.0 and 3.7.2 and also utilized to visualize and analyze the protein interactions within the interactome.[16,17]

The interactome was subjected for network analyzer, a Cytoscape (v 3.7.2) tool to measure nodes in the PPI network,[18] including various topological parameters such as the number of nodes and edges, clustering co-efficient, network density and diameter, shortest paths, characteristic path length, average number of neighbors, neighborhood connectivity distribution, node degree distribution (P(k)), betweenness centrality (BC), closeness centrality (CC) and average clustering coefficient distribution.[19]

Construction of Sub-network

MCODE (Molecular Complex Detection), a plugin in Cytoscape software was adapted to analyze the significant modules or clusters in the interactome. This was carried out by keeping the node score cutoff value = 0.2, MCODE score >3, degree cutoff value = 2, k-core value = 4, number of nodes >4, and maximum depth = 100. Top ten modules with highest MCODE scores were presented that can analyze critically connected nodes in the interactome based on topological parameters such as score, nodes, and edges.[20,21]

Correlation Analysis of Modular Seed Proteins

The correlation of each modular seed protein with *β-catenin* was analyzed using UALCAN web resource.[22,23] The correlation was calculated using a statistically significant *p-value* <0.05. The dot plots were generated based on the *p-value* and R-value. A significant correlation between modular seed proteins and β-catenin may imply that the seed proteins are involved in biological functions and pathways mediated by *β-catenin.*

Functional Enrichment Analysis

Functional aspects, including cellular component (CC), molecular function (MF), and biological process (BP) of all the primary gene interactors of *β-catenin* were identified using DAVID database (https://david.ncifcrf.gov) (The Database for annotation, visualization and Integrated Discovery). The results were sorted out by considering *p-value* less than 0.05 (<0.05) as a critical criterion.[24-26] DAVID database was utilized to identify multifunctional regulatory proteins which regulate both cytoplasm and nucleus. These proteins were subjected for PPI network construction and visualization using Cytoscape v 2.8.0 and v 3.7.2.

Identification of *β-catenin* Gene Interactors

A Cytoscape plugin, cytohubba was utilized to explore crucial genes in the PPI network of *β-catenin* by using degree as topological analysis method. Cytohubba provides 11 topological analysis methods which involves degree, betweenness, bottleneck, closeness, EcCentricity, radiality, stress, maximum neighborhood component (MNC), density of maximum neighborhood component (DMNC), maximal clique centrality (MCC), edge percolated component (EPC)[27,28] The identified crucial genes were subjected to construct a co-expression network and identify internal associations in gene sets using GeneMANIA (http://www.genemania.org/) an online bioinformatic tool. GeneMANIA is an online source helps to analyze gene list, prioritize genes

for functional assays, and create hypothesis about the functions of the gene.[29,30]

Expression Analysis of *β-catenin* Gene Interactors

UALCAN (http://ualcan.path.uab.edu/) is a comprehensive and interactive web resource used for the study. This user-friendly web tool is built for the analysis of cancer OMICS data and helps to carry out *in-silico* validation of genes and the identification of biomarkers, also helps to carry out pan-cancer gene expression analysis.[31] This portal allows the analysis of the relative expression of a gene over tumor and normal samples in many different cancer types along with tumor stages and other clinicopathological characters. This study conducted the relative expression analysis of top ten β-catenin gene interactors based on clinicopathologic features of lung adenocarcinoma.[32]

Gene Ontology Enrichment Analysis by ClueGO

Gene ontology enrichment analysis was performed for the top 10 *β-catenin* interactors using ClueGO, a Cytoscape plugin. This tool can be used to compare modular functional annotations or visualize terms corresponding to the list of genes. The list of genes was analyzed by considering the *p-value* < 0.05 for the three functional annotations such as cellular component (CC), biological process (BP), and molecular function (MF).[33,34] KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways are identified for the top 10 *β-catenin* interactors by utilizing DAVID database.[35] This database helps reveal the pathways mediated by *β-catenin* interactors based on their *p-value*. The tool also provides information regarding functional annotations, gene ontology, general annotations, gene interactions, protein domains, tissue expressions, etc.

Results and Discussion

Protein-protein Interaction Network Analysis

There are 869 *β-catenin* interacting genes were obtained from primary protein-protein interaction databases and their sub-cellular locations were traced using UniProtKB database.[36] Primary protein-protein interaction data was subjected to STRING and Cytoscape (v 2.8.0 and 3.7.2) to construct and visualize *β-catenin* interactome (Fig. 1).[37]

Topological Analysis

The interactome was analyzed for its topological statistics using Network Analyzer, a Cytoscape tool.[38] As shown in Table 1, the interactome has 3113 nodes and 15449 edges. Here, nodes represent protein residues of the biological network and edge refers to the connections between the protein residues. Interactome has clustering coefficient of 0.632, where clustering coefficient is a major statistic of the interactome and refers to the measure of abundant connected triangles in an interactome. Since the interactome is highly complex, it has a small clustering coefficient, large shortest path length of 9687656 (100%),

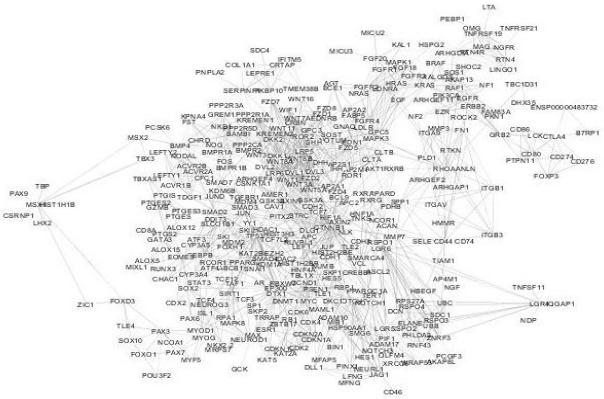

Figure 1: Interactome of *β-catenin* interactors constructed by Cytoscape v 3.7.1.

Table 1: Topological parameter of *β-catenin* interactome

Network Statistics	
Number of nodes/edges	3113/15449
Clustering co-efficient	0.632
Network centralization	0.079
Avg. number of neighbors	9.925
Network density	0.003
Shortest paths (%)	9687656 (100%)
Network diameter	13
Characteristic path length	4.614

and a characteristic path length of 4.614.[39] Complex networks refer to the complex biological systems which exists in real. All the other topological measures are listed out in Table 1.

The structure and biological activity of the proteins depend on the complex network of inter-residue interactions. Since the interactome has complex protein interaction, power-law distribution was performed for the nodes of the interactome.[40,41] The histograms were generated for closeness centrality, degree distribution, average connectivity, and shortest path length by considering a power law of the form $y = ax^b$ was fitted. Closeness centrality was measured to find the relationship between a single node and all the other nodes in the interactome (Fig. 2). The closeness centrality values range from 0 to 1. If the values are close to 1, it indicates that the node is close to other nodes. Whereas, the values are close to 0, that refers to the node is a long way from other nodes in the interactome. Thus, the closeness centrality values differ depending on the interactome size.[42] Node degree distribution of a node represents the number of edges, a node has with other nodes. In other words, it represents the number of connections between the other protein residues.[43] In addition to this, average neighborhood connectivity and the shortest path length distribution were also measured. Closeness centrality,

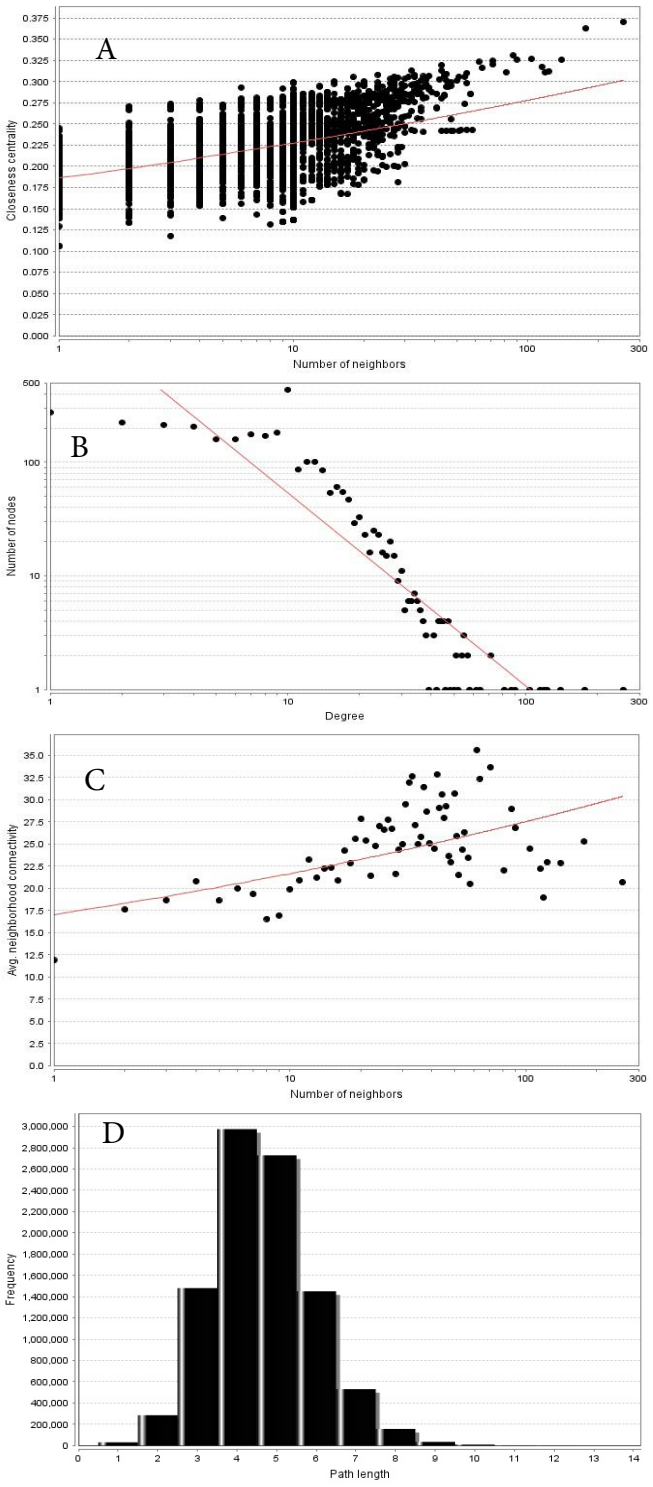

Fig. 2: Histograms representing a) closeness centrality; b) node degree distribution; c) average neighborhood connectivity; d) shortest path length of β-catenin interactome with power law fitted

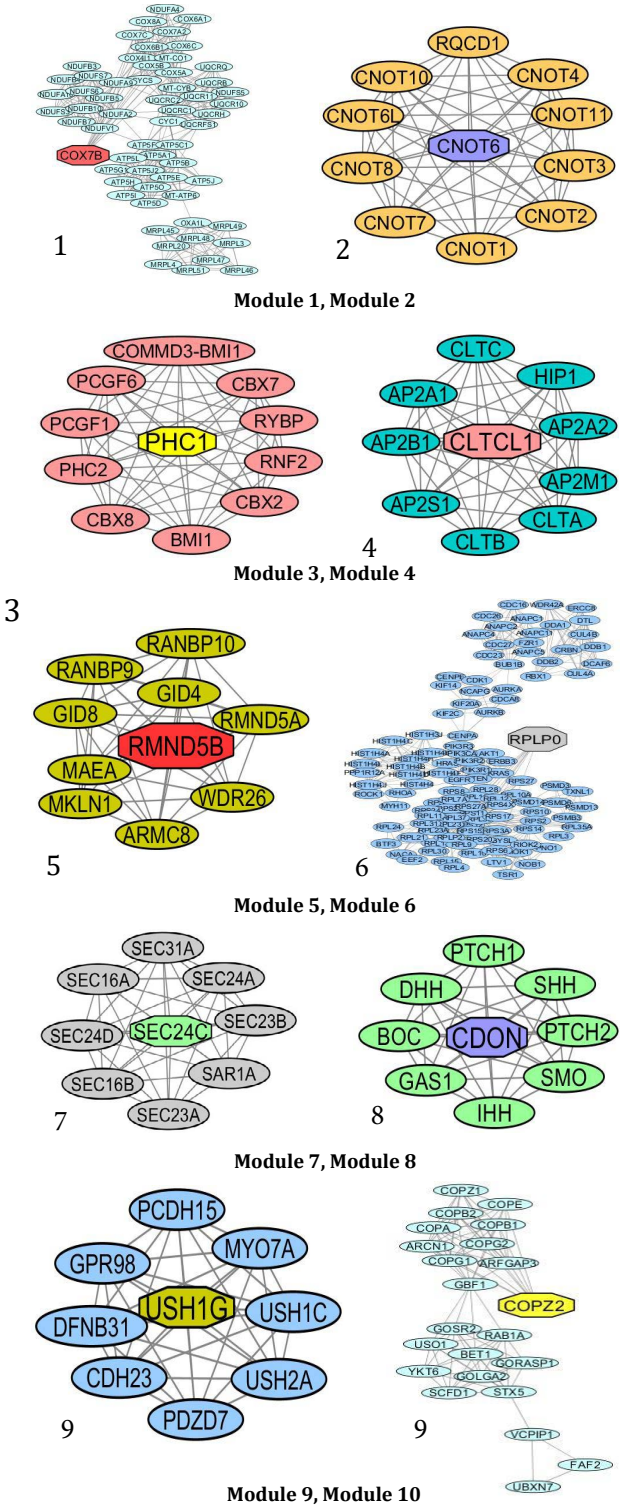

Figure 3: Modular analysis of the top 10 modules using MCODE Cytoscape plugin, containing the seed proteins highlighted in octagonal node shape

node degree distribution, and average neighborhood connectivity scores are found to be 0.186, 2668.2 and, 17.026, respectively. And R-squared = 0.233, 0.829, 0.313, respectively Fig.2.

Modular Analysis

The analysis provides 48 modules that are significantly involved in forming the interactome. Among them top ten modules were considered for the analysis. Each module represents the complex of proteins in a large interactome

and consists of a single protein, a crucial protein that connects all the other proteins in the module. This helps to study co-expression analysis and to target those seed proteins to inhibit certain diseases.[44]

Module 1 consists of a seed protein, COX7B, has a score value of 12.441 with 60 nodes (proteins), and 367 edges (interactions) (Fig. 3). COX7B (cytochrome c oxidase subunit 7B) is a small protein consisting of 80 amino acid sequence and belongs to the protein family that causes brain metastasis.[45] COX7B along with other proteins, take part in the pathogenesis of Parkinson's disease and in neurodegeneration by the regulation of oxidative stress, ATP production, and oxidative phosphorylation. This knowledge helps to understand the mechanism of Parkinson's disease.[46] CNOT6 is responsible to form module 2, has the score value 11, showing the interactions of 11 proteins with 55 interacting edges (Fig. 3). CNOT6 is a deadenylase subunit of the conserved CCR4-NOT complex. CNOT6 plays a vital role in carcinogenesis by overexpressing in androgen-independent prostate cancer, acute lymphoblastic leukemia, and acute myeloid leukemia. Thus, CNOT6 serves as the best therapeutic target for the cancer.[47] These subunits also involved in the inhibition of cell death and senescence.[48]

PHC1 (Polyhomeotic complex 1 homolog) is responsible for forming module 3, with a score 11 clustered by 11 interacting proteins with 55 interacting edges (Fig. 3). PHC1 protein is involved in brain growth and size. Mutations in these proteins result in primary microcephaly (MCPH). It is an autosomal recessive sporadic neurodevelopmental disorder where it involved in G2-M checkpoint arrest and DNA damage response.[49-51] Module 4 consists of a seed protein, CLTCL1 (clathrin, heavy chain-like 1) (Fig. 3), has the score value 10, and consists of 10 interacting proteins and 45 edges. CLTCL1 belongs to the clathrin family of genes which codes for the protein that consists of 1640 amino acid sequences. These proteins are over-expressed in muscle tissues and stabilizes spindle fibres during the mitotic phase of cell division.[52]

Similarly, module 5 (score= 9.556, node= 10, edges= 43) and module 6 (score= 9.518, node= 113, edges= 533) consists of seed proteins, RMND5B and RPLP0, respectively (Fig. 3). RMND5B is one of the members of CTLH complex and involved in E3-ligase enzymatic activity.[53] Biological processes mediated by CTLH complex are apoptosis, cell proliferation, cell adhesion, cell survival, cell migration and molecular functions such as WNT signaling pathway, NFkB signaling pathway, TGF β and P13-kinase signaling pathway. These signaling cascades play a crucial role in cancer cell plasticity.[54] RPLP0 are the ribosomal proteins termed as P proteins associated with normal function of fetal brain tissue, lymph nodes, spleen,[55] lung, placenta, bladder, skin, and other immune-mediated diseases. RPLP0 take active role in human cancers through its up-regulated activity. These P proteins can be used as prognosis markers against cancer specifically in gynecologic carcinomas.[56]

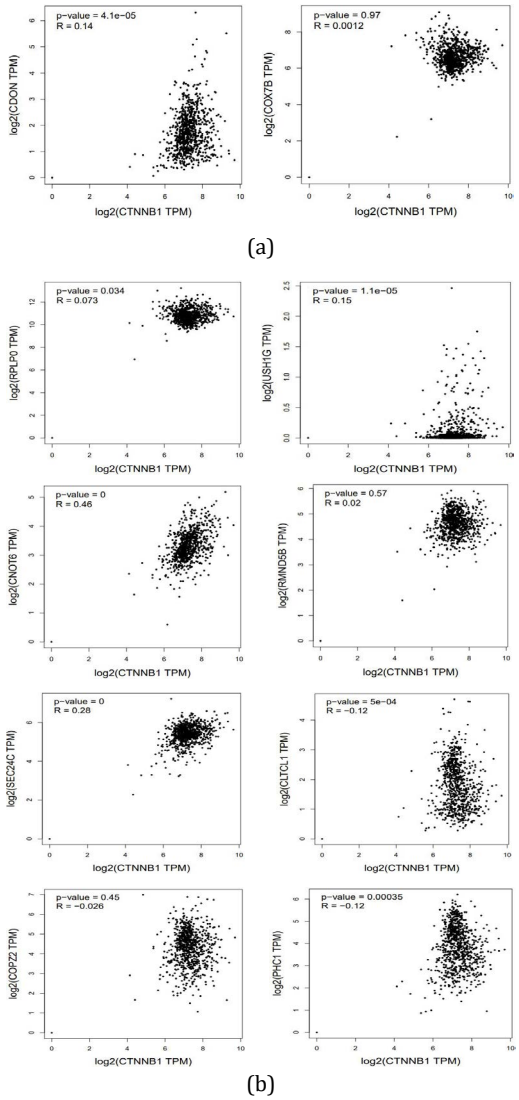

Fig. 4: Correlation expression analysis of modular seed proteins using UALCAN database. a) proteins showing significant correlation with β-catenin; b) proteins showing no significant correlation with β-catenin

RPLP0 take active part in acute myeloid leukemia (AML). AML may be targeted by downregulating RPLP0 gene along with other co-expressed genes [57].

The gene Sec24C is responsible for forming module 7, has the score value 9, showing the interactions of 9 proteins with 36 interacting edges (Fig. 3). Sec24C is associated with HIV-1 infection where it enhances the spreading of the infection. HIV involves a reverse transcription process where RNA copies DNA. This process can be targeted by reducing the stability of HIV-1 which can be achieved by the downregulation of Sec24C genes in HeLa cells.[58] Sec24C is phosphorylated by AKT which is over-expressed in cancer.[59] Overexpression of AKT can be seen in many types of cancer, including prostate, lung, gastric, pancreatic, breast, and ovarian cancer by increasing their survival and proliferation. AKT are the kinases involved in the phosphorylation of many proteins. Activated AKT

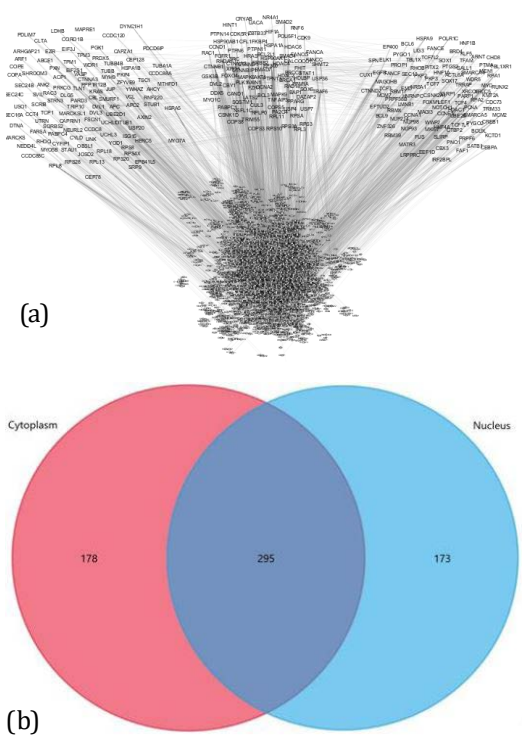

Fig. 5: PPI network and Venn diagram. a) extended PPI network of genes present in the cytoplasm, nucleus, and the genes present in both the cytoplasm and nucleus; b) Venn diagram showing 295 overlapped genes between the cytoplasm and the nucleus

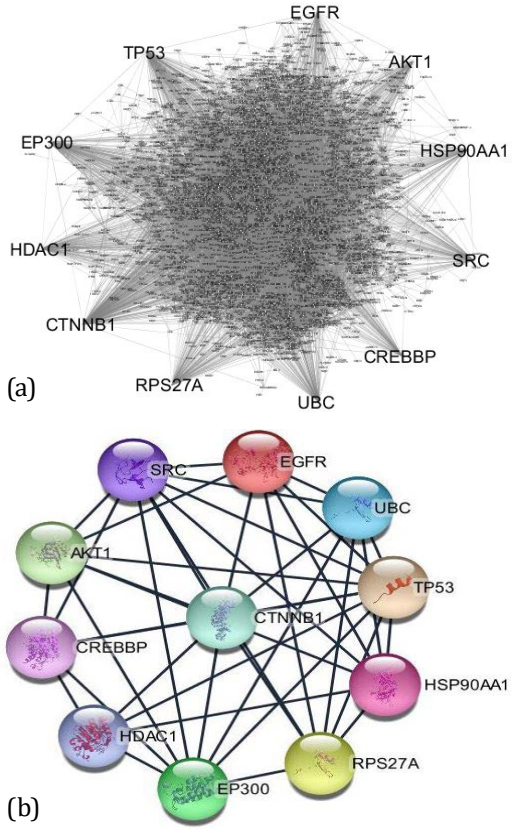

Figure 6: Top *β-catenin* interactors based on the degree method of Cytohubba. a) *β-catenin* interactome showing interactions for top *β-catenin* interacting genes b) stringified network of top *β-catenin* interactors

modulates the role of downstream proteins which are participated in cell survival, migration, proliferation, angiogenesis, and metabolism.[60,61]

Module 8 (score= 9, node= 9, edges= 36) consists of seed protein, Cdon (Fig. 3) is a transmembrane glycoprotein belongs to CAMs (cell adhesion molecules) superfamily. Mutated Cdon results in holoprosencephaly, characterized by forebrain midline defects.[62,63] By suppressing Wnt signaling, Cdon plays a vital role in the prevention of cardiac remodeling. An experiment with *Cdon*$^{-/-}$ mice shows altered expression of remodeling genes where these mice are characterized by cardiac dysfunction and fibrosis. Lack of Cdon results in malfunctioning and abnormal localization of connexin 43, a gap junction protein and causes hyperactivation of Wnt signaling. This information indicates that Cdon is an essential gene to prevent cardiac remodeling by controlling Wnt signaling.[64]

Similarly, module 9 has the score value 9, showing the interactions of 9 proteins with 36 interacting edges consisting of seed protein USH1G responsible for forming the module (Fig. 3). When USH1G (Usher syndrome 1 G) mutated, it is responsible for causing Usher syndrome. Usher syndrome is a genetic disorder distinguished by hearing disablement and loss of vision.[65,66] A study involves non-consanguineous Dutch family for the study, where the frameshift and heterozygous missense mutations in USH1G were reported which causes nonsyndromic sensorineural hearing loss.[67] Mutated USH1G is responsible to cause Usher syndrome type 1 along with, USH1C, PCDH15, CDH23, MYO7A, and CIB2 mutated genes. These genes perform the function in coordination, where they are interdependent with each other. There have been 11 studies carried out in 684 Usher syndrome patients. The biallelic disease-causing mutation rate of USH1G was found to be 1% (9/684).[68]

COPZ2 gene is responsible for forming module 10, which has a score 8.545. The module consists 23 proteins with 94 edges (Fig. 3). A study on human genetic association data, recognizes 12 SNPs (Single-nucleotide polymorphism) in COPZ2 genes, that actively participated in Alzheimer's disease, including COPZ1, COPA, COPD, COPB1, COPI, and PHLDB1.[69] COPZ2 plays a vital role in tumor metastasis and proliferation.[70] where these genes are downregulated in most of the cancer cell samples. COPZ1 genes are responsible to cause cancer. Hence, expression of COPZ2 protects tumor cells from killing by COPZ1 knockdown. This shows that tumor cells are COPZ1 dependent only because of COPZ2 silencing. Though COPZ2 has no tumor-suppressive activity, it protects microRNA 152, that is silenced in tumor cells concurrently with COPZ2 and ultimately take effect as a tumor suppressor *in vitro* and *in-vivo.* Downregulation of COPZ2 gene and microRNA 152

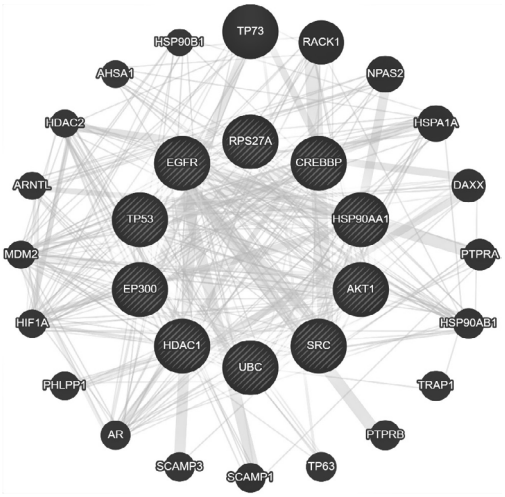

Figure 7: Co-expression network of top *β-catenin* interactors obtained by GeneMANIA

silencing in various different types of cancer uncovers the therapeutic opportunity to treat cancer.[71] All these genes in forming top modules in the β-catenin interactome open the door to uncover the disease diagnosis and therapeutic strategies.

Correlation between Modular Seed Proteins and *β-catenin*

Among ten modular seed proteins, CDON (R= 0.14; P= 0.000041), COX7B (R= 0.97; P= 0.0012), RPLP0 (R= 0.073; P= 0.035), and USH1G (R= 0.15; P= 0.000011) shows significant correlation with *β-catenin* (Fig. 4a). Where CNOT6 (R= 0.46; P= 0), RMND5B (R= 0.02; P= 0.57), SEC24C (R= 0.28; P= 0), CLTCL1 (R= -0.12; P= 0.0005), COPZ2 (R= -0.026; P= 0.45), and PHC1 (R= -0.12; P= 0.00035) has no significant correlation with *β-catenin* (Fig. 4b).

Functional Enrichment Analysis of *β-catenin* Interactors

All the primary *β-catenin* interacting gene sets were subjected to functional enrichment analysis using the DAVID database.[72] Functional aspects such as biological process (BP), molecular function (MF), and cellular component (CC) were annotated by considering *p-value* less than 0.05. The result shows that most of the *β-catenin* interactors are actively involved in biological processes such as transcription (191 genes), transcription regulation (187 genes), host-virus interaction (114 genes), and *Ubl* conjugation pathway (103 genes). The genes are also involved in biological rhythms, cell adhesion, *Wnt* signaling pathway, apoptosis, DNA repair, cell cycle, purine biosynthesis, mRNA splicing, mRNA processing, transport, DNA recombination, DNA recombination, etc growth regulation and all the other processes are illustrated in Table 2.

Results reveal that most of the *β-catenin* interactors participate in molecular function, DNA binding (137 genes) and also other functions such as RNA binding,

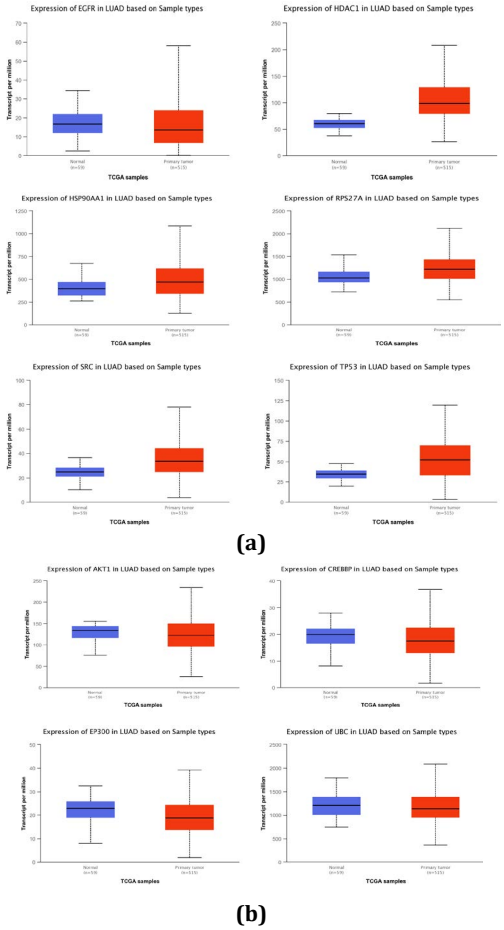

Figure 8: Expression analysis of top β-catenin gene interactors using UALCAN database. a) EGFR, HDAC1, HSP90AA1, RPS27A, SRC, and TP53 genes showing significantly high expression levels in lung adenocarcinoma (P value <0.05; num (N)= 59; num (T)= 515); b) The genes AKT1, CREBBP, EP300, and UBC showing less significant expression in lung adenocarcinoma (P value >0.05)

actin binding, rRNA binding, calmodulin binding and the genes belong to the protein groups such as activators, ribonucleoproteins, ribosomal proteins, viral nucleoproteins, transferases, repressors, chromatin regulators, motor proteins, protein phosphatase, myosin, hydrolases, helicases, kinases, proteases, ligases (Table 2). As shown in Table 2, *β-catenin* gene interactors are present in various cellular components, having *p-value* less than 0.05. Out of 869 *β-catenin* gene interactors, 474 genes are in cytoplasm and 467 genes in the nucleus.

The genes which are present in the cytoplasm, and nucleus are subjected for PPI network construction and visualized the genes present in both the cellular components using Cytoscape (v 3.7.1) (Fig. 5a). The results revealed that the number of genes present only in the nucleus is 173, and the number of genes present only in the cytoplasm is found to be 178, and the genes present in both the cellular components were found to be 295 (Table 3) and this is represented in Venn diagram using FunRich (v 3.1.3) online bioinformatic tool (Fig. 5b).

Generally, RNA processing and gene transcription events are taken place in nucleus. On the other hand, RNA decay and translation taken place in cytoplasm. But some proteins are not limited to a single event or single cellular component, rather, participate in the regulation of gene expression in the nucleus and the cytoplasm. These regulatory proteins are crucial, they can cause serious biological outcomes. These multifunctional proteins generally involved in diverse biological events such as cell-cycle regulation, surveillance of aberrant transcripts, sex determination, embryonic development, viral gene regulation and also perform its functions in immunological, muscular, reproductive, and neurological systems.[73]

Mapping of Crucial *β-catenin* Interactors

Out of 11 topological analysis methods of Cytohubba, degree method was utilized to identify *β-catenin* interactors in the interactome and the top 10 ranking genes are found to be tumor protein p53 (*TP53),* E1A binding protein p300 *(EP300),* ribosomal protein S27A *(RPS27A),* ubiquitin C *(UBC),* histone deacetylase 1 *(HDAC1),* SRC proto-oncogene *(SRC),* AKT serine/threonine kinase 1 *(AKT1),* epidermal growth factor receptor *(EGFR),* heat shock protein 90 alpha family class A member 1 *(HSP90AA1)*, and CREB binding protein (*CREBBP)* (Table 4) The interactome showing interactions for the top genes are shown in Fig. 6a. To retain the appearance of the network, Cytoscape-string plugin was used to highlight the top genes. (Fig. 6b)

Co-expression network was constructed and internal associations of top *β-catenin* gene interactors were identified with the help of GeneMANIA database. The co-expression of *β-catenin* PPI network are identified as, physical interactions of 54.99%, predicted of 8.84%, co-expression of 8.56%, co-localization of 6.51%, and pathway of 0.13% (Fig. 7).[74]

Validation of Top *β-catenin* Gene Interactors based on Clinicopathological Features

UALCAN online platform was utilized to validate the expression levels of ten β-catenin gene interactors in lung adenocarcinoma based on sample types.[75] The relative expression of six β-catenin gene interactors EGFR, HDAC1, HSP90AA1, RPS27A, SRC, and TP53 showed significantly high-expression levels (*p-value* < 0.05; num (N)= 59; num (T)= 515) in lung adenocarcinoma samples in the TCGA OMICS data (Fig. 8a) and relative expression of AKT1, CREBBP, EP300, and UBC were showing significantly less activity in lung adenocarcinoma tumor tissues (Fig. 8b). The results indicated that these six β-catenin gene interactors may participate in lung adenocarcinoma prognosis and can be a prognostic marker in lung adenocarcinoma.

Gene Ontology Enrichment Analysis of the Top *β-catenin* Interactors

The top 10 *β-catenin* interactors were evaluated for the pathways involved, using ClueGO, a Cytoscape plugin. The genes primarily participated in the biological processes, and molecular functions which involves the activation of innate immunity response (38.1%), regulation of telomerase activity (23.81%), nitric-oxide synthase regulator activity (19.05%), disordered domain specific binding (14.29%), and *p53* binding (4.76%) (Fig. 9).[76]

Table 2: Biological processes, molecular function, and cellular component of all the *β-catenin* interactors retrieved by david database

Biological Process	*P value*	*Molecular Function*	*P value*	*Cellular Component*	*p-value*
Host-virus interaction	1.31E-31	Activator	3.65E-16	Nucleus	1.08E-55
Biological rhythms	1.38E-25	Ribonucleoprotein	6.17E-17	Cytoplasm	1.07E-63
Cell adhesion	1.72E-05	Ribosomal protein	1.93E-13	Cell junction	7.55E-23
Wnt signaling pathway	4.02E-22	Tyrosine-protein kinase	2.58E-04	Cytoskeleton	1.21E-21
Ubl conjugation pathway	7.85E-19	Thiol protease	9.91E-13	Cell projection	7.76E-05
Transcription regulation		Serine/threonine-protein	0.009147	Spliceosome	2.72E-08
Transcription	6.97E-08	kinase		Mitochondrion inner	0.014633
Stress response	8.12E-08	DNA-binding	0.002999	membrane	
DNA damage	0.00819	RNA-binding	2.79E-08	Virion	0.009653
Apoptosis	9.82E-10	Viral nucleoprotein	4.74E-04	Synapse	0.002809
DNA repair	0.001458	Transferase	0.005992	Tight junction	5.90E-07
Cell division	4.29E-09	Repressor	3.89E-06	Chromosome	1.65E-04
Cell cycle	1.58E-06	Chromatin regulator	2.65E-07	Proteasome	1.36E-04
Purine biosynthesis	2.32E-08	Actin-binding	4.36E-05	Microtubule	7.63E-04
Translation regulation	0.031806	Helicase	8.11E-05	Synaptosome	0.00608
mRNA splicing	1.14E-04	Kinase	9.60E-04	Nuclear pore complex	0.007029
mRNA processing	9.17E-07	Protease	0.001104		
mRNA transport	1.24E-05	Motor protein	0.002155		
DNA recombination	7.16E-05	Protein phosphatase	0.004389		
Mitosis	3.96E-04	Myosin	0.004707		
Cell shape	0.001178	Hydrolase	0.01085		
Chromosome partition	0.001599	Ligase	0.014358		
Growth regulation	0.011122	Calmodulin-binding	0.017809		
	0.015365	rRNA-binding	0.044674		

Table 3: Genes associated with cytoplasm, nucleus, and in both the components which are obtained by david database

Genes present in Cytoplasm	*Genes present in nucleus*	*Genes present in Cytoplasm+Nucleus*
AHCY, ANK3, CTNNA3, EIF3J, JUP, PLEKHA4, RHOF, RPL8, ACLY, ACP1, CASP3, CBL, CORO1B, USP9X,ABCE1, UTRN, VASP, VCL, WDR1, YOD1, YWHAQ, YWHAZ, ZFYVE9, ACTG1, AKAP12, ANK2, APC, APC2, ARF1, ARFGEF2, ARHGAP21, ARHGAP32, ARHGAP5, AURKA, AXIN2, CALM1, CAPRIN1, CAPZA1, CCDC120, CCDC8, CCDC88A, CCDC88C, CCT2, CCT4, CDC37, CDC42, CEP128, CEP78, CFAP298, CIT, CLTA, CNOT2, COPA, COPE, CRYL1, CTNNA1, CYFIP1, CYLD, DARS1, DLG1, DLG5, DNAJC5, DTNA, DVL1, DVL3, DYNC1H1, EIF2S1, EIF3E, EPB41L5, EZR, FAM126A, FARSA, FLT1, FSCN1, GRB10, GRIP1, HERC5, HSPA1B, HSPA5, HSPH1, ILF3, ILK, ISG15, JOSD1, JOSD2, KIAA1217, KIF20A, KIF3A, KIF5B, KLC2, KRAS, LARS1, LDHB, LIMCH1, LZTS2, MAGI2, MAP1LC3B, MAPRE1, MARCKS, MARCKSL1, ME1, MPRIP, MSN, MTHFD1, MYH9, RAC2, RAC3, SCRIB, SEC16A, SEC24B, SEC24C, SHARPIN, SHROOM3, SIPA1L1, SLC9A3R1, SMURF1, SORBS2, SPECC1L, SPSB4, SPTAN1, SRCIN1, SRP9, STAU1, STRN3, STUB1, SVIL, TCP1, TES, TLN1, TMOD3, TOLLIP, TPM1, TPM3, TRIP10, TSC1, TRIM37, TUBA1A, TUBB, TUBB4B, UBE2D1, UCHL1, UCHL3, UNK, USO1, USP20, USP29, USP33, MYO5B, MYO7A, NEDD4L, NEURL2, NHS, NIPSNAP2, OBSL1, OTUB1, PABPC4, PARD3, PDCD6IP, PDLIM5, PDLIM7, PGK1, PICK1, PKP4, PLCH1, PLEKHA5, PLEKHA7, PPP1R12A, PPP1R12B, PRDX5, PRKCG, PXN, RARS1, RHOQ, RND1, RNF220, RNF8, RPL13, RPL18, RPS20, RPS28, RPS4X, RPS8,	ABCF2, ACTL6A, CBX5, EEF1D, FANCF, MEN1, RNF8, UBE2B, ZNF639, ADNP, ARNT, ASH2L, ASXL2, EFTUD2, EGFR, ELK1, EMD, ZMIZ2, ZNF326, BCL6, BCL9, BCL9L, BRD4, CA9, CNOT2, CREB1, CTBP2, CTNND2, CUX1, DUX4, CSNK2A1, CHD8, CEBPA, CBFB, CBX3, CCNA1, CCNE1, CDC73, CFTR, DIMT1, DNMT1, DOT1L, EP400, EPAS1, ERBIN, ESR1, EZH2, FAF1, FANCE ,FOXB1, FOXM1, FOXP1, FUS, H1-2, H3C1, H4C1, HDAC1, HNF1A, HNF1B, HNF4A, HNRNPC, HNRNPM, HSPA9, IRF2BPL, LEF1, LEO1, KCTD1, KLF5, KMT2A, KMT2D, LIG3, LMNA, LMNB1, LRPPRC, LUC7L2, MAGI3, MAGOHB, MATR3, MCM2, MCM7, MED9, NONO, NOTCH1, MKI67, NCOA2, MYC, NR5A1, NR5A2, NUDT5, NUP214, NUP93, NUP98, PARP1, PARP11, PCNA, PITX2, PKP2, PKP3, PNO1, POLR1C, PPIE, PRKDC, PROP1, PRPF38B, PRPF6, PYGO1, PYGO2, RALY, RANBP2, RHOB, PTGS2, PTMA, RBBP5, RBM17 ,RBM39, RBMX, RPA1, RPA2, SEIF3E, SALL1, SATB1, SEC13, SIRT6, SLC9A3R2, SLIRP, ZC3H11A, ZFP82, RUNX1, RUNX2, SMARCA4, SMARCA5, TBL1X, TBL1XR1, TBX3, TCERG1, TCF3, TCF4, TCF7, TCF7L1, TCF7L2, TEAD4, TFAM, TFAP2A, THRAP3, TJP2, TLE1, TLE3, TRIM28, TRIM33, USP49, UTP14A, XRCC5 ,XRCC6, ZBTB2, ZBTB25, SOCS1, SOX1, SOX17, USP17L2, VEZT, VRK3, WDR5, WWP2, UHRF2, SPECC1, SPN, SRRM2, SSRP1, SUV39H1, TADA2A, TADA3, TRRAP, TUB1, UBR5	ATXN2L, CACYBP, CHUK, CTNNA2, DDX1, EGLN3, FANCA, FHL2, FMR1, GAPDH, HDAC2, HDAC4, MCM5, NPM1, PTPN14, RAE1, RNF14, RXRA, RYK, STAT1, TARDBP, TMPO, TOP2A, USP15, ZBTB33 , ABL1, ACTB, ACTN4, AHR, AKT1, AMER1 , API5, APP, AR, ARFGEF1, ARMC8, ATF2, AURKB, AXIN1, BAP1, BCL2L1, BCL3, BRCA1, BRCC3, BTF3, BTRC, CAD, CALCOCO1, CAND1, CARM1, CASP8, CBY1, CCNA2, CCND1, CDC27, CDK1, CDK2, CDK5R1, CDK6, CDK9, CFL1, CHD4, CHMP4B, CITED1, COPS3, COPS5, COPS8, CPSF7, CRBN, CREBBP, CRYAB, CSNK1A1, CSNK1D, CTBP1 , CTNNB1, CTNNBIP1, CTNND1, CUL3, CTNNA2, CTBP1, DAZAP2, DDB1, DDX1, DDX17, DDX39B, DDX3X, DDX5, DHX9, COPS8, DVL2, ECT2, EEF2, EGLN3, EGR1 , EIF4A3, EIF4E, EIF6, ELAVL1, DST, CHUK, ERBB2, EP300, ESR2, EWSR1, FANCA, FANCC, ECT2, EGR1, ESR1, EWSR1, FOXO1, FOXO3, FOXO4, FANCG, PA2G4, PABPC1, PAN2, PARK7, KHDRBS1, KIF14, KIF22, FBXW11, FBXW7, FER, FERMT2, FGFR1, FHIT, FKBP4, FLII, FYN, FZR1, G3BP1, GEMIN5, GID8, GLIS2, GNA13, HSPB1, HTT, HUWE1, IKBKB, ILF3, IMPDH2, IQGAP1, GSK3B, HDAC6, HDAC7, HDLBP, HERC2, HIF1A, HINT1, HMBOX1, HMGB1, HMGB2, HNRNPA1, HNRNPA2B1, HNRNPK, HNRNPU, HRAS, HSP90AA1, HSP90AB1, HSPA1A, LGALS3, LGALS9, LIMD1, JADE1, KAT2A, KAT2B, KDR, KIF23, KPNB1, LATS2, MAP3K2, MAPK6, MAPK8, MAPK9, LRRFIP1, MCM5, MITF, LYN, MOV10, MUC1, MYBBP1A, MAP3K2 , NPM1, NR3C1, NR4A1, MKRN2, NUMA1, KIF23, MYO1C, NSD2, NSFL1C, NUDC, PCBP1, PCBP2, NCL, NDRG1, NXF1, NEK2, NF2, NFKB1, OTUD7B, NLK, PA2G4 , PABPC1, PHB2, PIN1, PKM, PLK1, PML, POLDIP3, POLR2A, POU5F1, PPARG, PPIA, PPM1A, PPP1CB, PPP1R13B, PPP2R1A, PRC1, PRKCA, PRKN, PRPF19, PSMA5, PSMB6, PSMC3, PSMC5, PSMC6, PTEN, PTPA, PTPN11, PTPN13, PTPN6, RAC1, RACK1, RAD18, RB1CC1, RHOA, RNF4, RNF6, RNPS1, RPL11, SERBP1, SET, SFN, SFPQ, SGK1, SHMT2, SIAH1, SIRT1, RPL3, RPLP0, RPS10, RPS27A, RPS3, RPS3A, RPSA, RAE1, RUVBL1, RUVBL2, RXRA, RYK, SMAD2, SMAD3, SMAD4, SMAD7, SND1, SNRPD1, SNRPD2, SORBS1, SET, SOX2, SOX6, SP1, SQSTM1, SRC, STAT1, TP53BP2, STIP1, SYNCRIP, TAX1BP3, TMPO, TNFAIP3, TOB1, TOP2A, TRAF4, TRAF6 , TRIM25, TXN, TXNL1, UACA, UBE2N, UBE3A, TRIM55, TRIM63, TSC22D1, UCHL5, USP18, USP2, USP21, USP36, USP4, USP10, USP11, USP15, USP7, VIRMA, WBP2 , YAP1, YBX1, XPO1, VCP, VCPIP1, WBP2, ZIC3, ZMIZ1, YWHAE, ZIC3 , ZBTB33, PTPN14, TARDBP, TAX1BP3, BTF3, LYN, STIP1, AMER1, API5, ATF2, ATXN2L, BCL2L1, POU5F1, PPP1CB, FHL2, CACYBP, GAPDH, HDAC2, HDAC4 , HINT1 , HSPA8, NCL

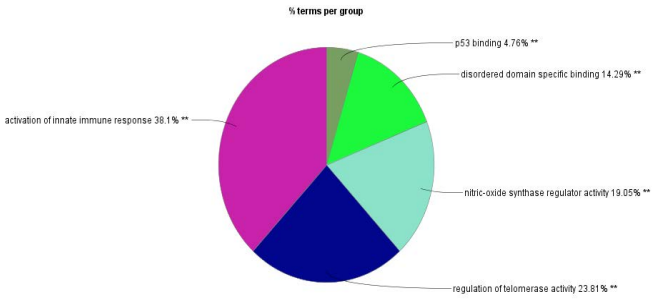

Fig. 9: Gene ontology enrichment analysis of *β-catenin* interactors obtained by CleuGO, a Cytoscape plugin

Table 4: Top *β-catenin* interactors along with the scores are obtained by degree method of cytohubba, a cytoscape plugin

Rank	Name	Score
1	TP53	177
2	EP300	140
3	RPS27A	123
4	UBC	119
5	HDAC1	115
6	SRC	104
7	AKT1	90
8	EGFR	87
9	HSP90AA1	81
10	CREBBP	71

Table 5: KEGG pathway enrichment analysis of the top *β-catenin* interactors using David database

KEGG pathway	*p-value*	*Genes*
Kaposi sarcoma-associated herpesvirus infection	1.32E-08	CREBBP, SRC, UBC, EP300, AKT1, RPS27A, TP53
Prostate cancer	2.59E-08	HSP90AA1, CREBBP, EP300, AKT1, TP53, EGFR
Thyroid hormone signaling pathway	7.91E-08	CREBBP, HDAC1, SRC, EP300, AKT1, TP53
Shigellosis	2.78E-06	SRC, UBC, AKT1, RPS27A, TP53, EGFR
Pathways in cancer	5.22E-06	HSP90AA1, CREBBP, HDAC1, EP300, AKT1, TP53, EGFR
Human papillomavirus infection	1.17E-05	CREBBP, HDAC1, EP300, AKT1, TP53, EGFR
Hepatitis B	1.74E-05	CREBBP, SRC, EP300, AKT1, TP53
Viral carcinogenesis	4.32E-05	CREBBP, HDAC1, SRC, EP300, TP53
Adherens junction	5.10E-05	CREBBP, SRC, EP300, EGFR
Mitophagy – animal	5.32E-05	SRC, UBC, RPS27A, TP53
Endocrine resistance	1.34E-04	SRC, AKT1, TP53, EGFR
HIF-1 signaling pathway	1.83E-04	CREBBP, EP300, AKT1, EGFR
MicroRNAs in cancer	2.21E-04	CREBBP, HDAC1, EP300, TP53, EGFR
Cell cycle	2.82E-04	CREBBP, HDAC1, EP300, TP53
FoxO signaling pathway	3.16E-04	CREBBP, EP300, AKT1, EGFR
Estrogen signaling pathway	3.68E-04	HSP90AA1, SRC, AKT1, EGFR
Fluid shear stress and atherosclerosis	3.76E-04	HSP90AA1, SRC, AKT1, TP53
JAK-STAT signaling pathway	5.90E-04	CREBBP, EP300, AKT1, EGFR
Tuberculosis	8.03E-04	CREBBP, SRC, EP300, AKT1
Bladder cancer	8.66E-04	SRC, TP53, EGFR
Proteoglycans in cancer	0.001172	SRC, AKT1, TP53, EGFR
Chemical carcinogenesis - receptor activation	0.001292	HSP90AA1, SRC, AKT1, EGFR
Lipid and atherosclerosis	0.001345	HSP90AA1, SRC, AKT1, TP53
Human T-cell leukemia virus 1 infection	0.001476	CREBBP, EP300, AKT1, TP53
Human cytomegalovirus infection	0.001534	SRC, AKT1, TP53, EGFR
Endometrial cancer	0.00173	AKT1, TP53, EGFR
Notch signaling pathway	0.001789	CREBBP, HDAC1, EP300
Renal cell carcinoma	0.002439	CREBBP, EP300, AKT1
Central carbon metabolism in cancer	0.00251	AKT1, TP53, EGFR
Melanoma	0.002653	AKT1, TP53, EGFR
Non-small cell lung cancer	0.002653	AKT1, TP53, EGFR
Glioma	0.002875	AKT1, TP53, EGFR
Pancreatic cancer	0.002951	AKT1, TP53, EGFR
Chronic myeloid leukemia	0.002951	HDAC1, AKT1, TP53
EGFR tyrosine kinase inhibitor resistance	0.003185	SRC, AKT1, EGFR
ErbB signaling pathway	0.003678	SRC, AKT1, EGFR
Huntington disease	0.003701	CREBBP, HDAC1, EP300, TP53
Colorectal cancer	0.003763	AKT1, TP53, EGFR
PI3K-Akt signaling pathway	0.005586	HSP90AA1, AKT1, TP53, EGFR
Glucagon signaling pathway	0.005769	CREBBP, EP300, AKT1
Growth hormone synthesis, secretion and action	0.00721	CREBBP, EP300, AKT1
Relaxin signaling pathway	0.008294	SRC, AKT1, EGFR
Breast cancer	0.01067	AKT1, TP53, EGFR
Gastric cancer	0.010951	AKT1, TP53, EGFR
Hepatitis C	0.012106	AKT1, TP53, EGFR
Hepatocellular carcinoma	0.013781	AKT1, TP53, EGFR
Wnt signaling pathway	0.014096	CREBBP, EP300, TP53
Influenza A	0.014254	CREBBP, EP300, AKT1
Neutrophil extracellular trap formation	0.017418	HDAC1, SRC, AKT1
Focal adhesion	0.019376	SRC, AKT1, EGFR
Epstein-Barr virus infection	0.019558	HDAC1, AKT1, TP53
Rap1 signaling pathway	0.021045	SRC, AKT1, EGFR
cAMP signaling pathway	0.023166	CREBBP, EP300, AKT1
Chemical carcinogenesis - reactive oxygen species	0.023561	SRC, AKT1, EGFR
Parkinson disease	0.03273	UBC, RPS27A, TP53
MAPK signaling pathway	0.039359	AKT1, TP53, EGFR

The KEGG pathway enrichment analysis reveals 56 pathways and diseases which are mediated by top 10 *β-catenin* interactors. The analysis uses the DAVID database to carry out KEGG pathway enrichment analysis[77] by considering the statistically significant *p-value*, less than 0.05.[78] The curated pathway includes Wnt signaling pathway, MAPK signaling pathway, Jak-STAT signaling pathway, P13K-Akt signaling pathway, ErbB signaling pathway, Notch signaling pathway, HIF-1 signaling pathway, cAMP signaling pathway, Rap1 signaling