Master Degree Project

UNIVERSITY OF SKÖVDE

To identify novel oncogenes for the design of novel tools for diagnosis and treatment of cancer

Master Degree Project in infection Biology
60 ECTS
Spring term Year

Selva Kumar Mandiramoorthy
A14selma@student.his.se
19930713-3513
Course responsible

Diana Tilvek
Diana.tilevik@his.se
School of Bio science
University of skÖvde, Sweden

Examiner
Patric Nilsson
Patric.nilsson@his.se
School of Bio-science
University of skÖvde, Sweden

Supervisor
Annica K. B. Gad, Ph.D.
Senior/ Leading Researcher (R4)
annica.gad@staff.uma.pt
CQM- Madeira Chemistry Research Centre / Centro de Química da Madeira
University of Madeira
Campus da Penteada
9020-105 Funchal, PORTUGAL

ABSTRACT

Cancer is a disease caused by an uncontrolled cell growth that destroys the healthy tissue of the body. It is one of the deadliest diseases in the world that alters many cellular mechanisms and features. In this report, a list of 22 upregulated oncogenes is studied to identify the novel oncogene. The need to determine the novel oncogene is to develop the anti-cancer agent. To determine the novel oncogene, gene enrichment analysis (GEA) was performed. It is a method to identify classes of genes that are over-represented in the large set of genes to determine the phenotypes of the organisms. DAVID and PANTHER are the methods used to carry on this study as it has Gene Ontology (GO) embedded in it. The GEA uses fishers exact test to determine the enriched gene by the standard p-value of 0.05. To further study the oncogene Network Enrichment Analysis was performed with EVINET. We found that microtubule was significantly enriched in NEA. The genes significantly enriched for GO microtubule were studied. The significantly enriched microtubule in NEA might then be used as a target for anti-cancer agent and used to develop the drug in the future.

POPULAR SCIENTIFIC ABSTRACT

Cancer is a disease caused by an uncontrolled cell growth that destroys the healthy tissue of the body. Cancer composition alters many cellular mechanisms and functions. According to the World Health Organization (WHO), cancer is the second leading cause of death globally, and it is estimated to be about 9.6 million deaths in 2018. Cancer is caused mainly due to infections such as hepatitis and human papillomavirus (HPV) mostly in the middle – low-income countries. Cancer is a genetic disease that can affect any part of the body. The main feature of a tumor is that it can grow abnormal cells beyond the usual boundaries and then destroy the specific organ and spread to the other parts of the body in a process called metastasis. The ability of the abnormal cells to spread to the other organs is known as metastasizing cells.

Gene Enrichment analysis it is a method to identify classes of genes and protein that are over-represented in the large set of genes by the user to define the phenotypes of the organisms. In this study, it is used because it helps to understand the direct interest to the understanding of the functional mechanism in a cell also specific pathway activated in a given data. GEA works on the statistical analysis it used basic Fisher exact test to provide the raw p-value and multiple hypotheses or FDR are calculated by Benjamini-Hochberg method or Bonferroni method. The technique used to perform this study is by DAVID and PANTHER. They are the online software tool for accomplishing this Enrichment Analysis. The technology works with human data as background collected in a group and stored in a database known as Gene Ontology (GO). The result from this study is provided in three distributions; Cellular Component, Biological Pathway, and Molecular Functions.

Network Enrichment analysis was performed to determine the topological information and the network links between the gene-protein interaction. NEA performs the quick binomial test as described in methods and quantifies enrichment in each AGS (Altered Gene Set) FGS (Functional Gene Set) pair with some statistics: chi-squared score, z-score, p-value, q-value (the p-value Adjusted for false discovery rate). EVINET is an online tool used to perform the NEA.

The result from this study gave the understanding of 22 upregulated genes. The significantly enriched genes are KIF18A, KIF14, KIF13B include known as kinesin protein family members. These genes are significantly enriched for the microtubule-based process. Microtubule is the significant component of the cytoskeleton. Microtubule is a tubular polymer that forms the structure and shape to the cytoplasm of eukaryotic cells and regulates cell adhesions. These are results helpful in constructing the anti-cancer agent. Microtubule is said to be a potent as it is essential in a cellular process, cell growth, cytokinesis, transport, and mobility in the cell used as the development of anti-cancer agent.

LIST OF ABBREVIATIONS

AGS – Altered Gene Set

FGS – Functional Gene Set

GSEA – Gene Set Enrichment Analysis.

ICGC - International Cancer Genomics Consortium.

TCGA - Cancer Genome Atlas.

GO – Gene Ontology.

NCBI - National Center for Biotechnology Information.

PSA - prostate-specific antigen.

IBD - Inflammatory Bowel Disease.

DAVID - The Database for Annotation, Visualization, and Integrated Discovery.

PANTHER - Protein Analysis Through Evolutionary Relationships.

KEGG - Kyoto Encyclopedia of Genes and Genomes

FAP - Familial Adenomatous Polyposis.

HNPCC - hereditary nonpolyposis colorectal cancer.

FDR – False Discovery Rate.

PIP3- phosphatidylinositol-trisphosphate.

HA- Hyaluronan

Contents

INTRODUCTION

Cancer

Cancer is a multistep process in the human body which causes more than 7 million deaths worldwide every year. Cancer is one of the deadliest diseases in the world which involves the successive genetic and epigenetic alterations of the genes that perform homeostatic control and suppress the survival of aberrantly proliferating cells. In most cases, cancer arises in epithelial cells that show the carcinomas in organs such a lung, skin breast, liver, and pancreas. This alteration of the standard cell to a cancer cell in the epithelial cells is one of the "hallmarks of cancer"[1]. If the genetic and epigenetic changes can be identified, then these genes can be used as biomarkers for various cancer cell types. In the last two to three decades, there has been evidence stating that cancer is a disease of the genome. The development of the high throughput sequencing in recent years has witnessed an excellent data exploration and systematic study of the cancer genome[2].

Oncogenes

The mutation of the normal cells into cancer cells when altered by the external stress that leads to abnormal growth of cell are termed as oncogenes. The cells generally undergo apoptosis known as programmed cell death, but this process can be blocked by the oncogenes and lead to cancer. Oncogene is activated by mutation of the encoded protein that enhances its transforming activity. It is said that oncogenes are associated with uncontrolled cell proliferation and cancer[3]. The activation of the oncogene can be through the structural alterations resulting from gene fusion, by transcription factor to enhancer elements, or by amplification or by chromosomal rearrangements[4].

Oncogene list

The work was carried out previously by identifying almost 200 genes that are upregulated in metastatic BJ-RAS-cells as compared to normal, primary BJ cell[24]. To find novel genes upregulated in cancer, a previous student analyzed the expression of these genes in three different cancer types, prostate, colon, and breast in the human protein atlas project. The list of 22 oncogenes obtained was CD83, CLGN, CPE, CSE1L, CXADR, FAM84A, FKBP4, GABBR2, HMMR, KIF13B, KIF14, KIF18A, LDOC1, MOK, NCAPH, NEDD4L, PCOLCE2, PPM1E, PRUNE2, SLC29A1, STMN3, TCOF1.

Breast cancer

Breast cancer is one of the leading cause of cancer mortality among females around the world. It is a complex and heterogeneous malignancy that were estimated to be 1.7 million cases and 521.900 deaths in 2012. The breast cancer incidence rates are higher in developed countries like North America, UK, western Europe, Australia, and New Zealand than in Asia and Africa[5]. Breast cancer is classified into four subtypes based on the DNA microarray-based expression

profiling. Luminal A subtype was identified as having the most favorable outcome; Luminal B is considered to be aggressive with a higher metastatic risk, human epidermal growth (HER) factor 2 -positive here the HER2 gene is overexpressed, and Basel like breast cancer is predominantly triple negative. The last two subtypes are associated with poor prognosis [6,7]. In this study, the gene list is checked for the enrichment to determine the oncogenes that increase the spread of breast cancer.

Prostate cancer

Prostate cancer is one of dangerous disease which affects men. Increase in smoking and a decrease in exercise increase the risk factor of prostate cancer. PSA (prostate-specific antigen) screening is critical to detect early prostate cancer[8]. It is known that prostate cancer is higher in Europe, North America compared to Asia. Also, Asians have a better survival rate[9]. Prostate cancer is said to found higher in older adults, but nowadays it's rapidly increasing in the younger generation. Prostate cancer is considered to be that less than 5% of prostate cancer is hereditary. Approximately 95% of prostate cancer are adenocarcinomas. Roughly 4% of all prostate malignancies arise from the transitional epithelium of the urethra or ducts as transitional cell carcinoma. It is also known as tumors of other organs may spread into the prostate. Histological recognition of prostate cancer depends on the overall assessment of the architecture and upon the cytology of individual cells. The prostate cancer cell cytoplasm may contain large amounts of acid phosphatase and prostate-specific antigen (PSA). Using immunostaining, it is possible to differentiate prostatic carcinoma cells from other tumor cells[10]. This study tries to uncover the genes driving the spread of prostate cancer.

Colorectal cancer

Colorectal cancer is considered as third most common cancer all over the world and fourth most death causing disease. It is the primary cause for the morbidity and mortality across the globe. It is common cancer affects both men and women. Developed countries like North America, Australia, western Europe are more affected by this cancer compared to Asia and Africa[11]. Inflammatory bowel disease (IBD) is a combination of ulcerative colitis and Crohn disease. Ulcerative colitis causes inflammation in colon and rectum whereas Crohn disease causes inflammation in of the bowel wall and also in the digestive tract from mouth to the anus. These conditions have a significant chance of increasing colorectal cancer in patients[12,13]. Colorectal cancer is considered to be a hereditary disease. It is known as familial adenomatous polyposis FAP, and hereditary nonpolyposis colorectal cancer (HNPCC) are inherited condition for colorectal cancer. It is found that HNPCC is mutated in the DNA pathway by the gene *MLH1* and *MSH2*, and FAP is mutated by tumor suppressor gene *APC*[14].

Cancer in System Biology

The complete understanding of complex biological systems beyond the molecular level was developed by the new way approach called as systems biology. Systems biology is the study of organisms, and it's related and interacting network of genes, proteins and biochemical reactions which gave give rise to life. System biology helps us to analyze and focus all components in the organisms. It is useful to learn about the disease, individual gene, protein. The essential

importance of system biology at present is because it enables the development of the drugs against cancer[15]. How the intracellular networks of standard cells are perturbed during carcinogenesis can be used to develop useful predictive models where scientists and clinicians can validate the new therapies and drugs. Cancer systems biology approaches are based on the use of computational and mathematical methods for understanding the complexity in cancer heterogeneity. There are many databases builds on the success of human genome projects such as The Cancer Genome Atlas (TCGA) and International Cancer Genomics Consortium (ICGC). In cancer systems biology there are many multiple data types used and integrated along with the clinical data as this reason for various data is getting wider. So, the computational approaches used in cancer systems biology get increased, by the new mathematical and computational algorithms that reflect the dynamic interplay between experimental biology and the quantitative sciences. The emerging challenges for the extreme heterogeneity and genomic instability of cancer cells are to distinguish the 'driver' aberrations, that directly alter the behavior of a tumor and that represent the potential targets or known as biomarkers[16]. Computational tools can understand the complexity of cancer by integrating new datasets with prior knowledge. It is known as Ben Raphael described the HotNet2 algorithm, that combines cancer mutational data with known protein interactions. The development of software protocols has provided the means to automatically document analyses for purposes of analysis the gene, protein, and biomarker also to understand the cellular component thereby making way for new drugs in the new world[17].

Gene Ontology

The development in the high throughput techniques has put cancer biology more exciting and new ways to develop drugs for cancer and many other diseases. In the modern cancer system biology, it is easy to get information of every detail of the cancer gene by just a click on the data. There are many databases which save the information about the genes and proteins; Gene Ontology, KEGG pathway, Reactome pathway, String Gene Atlas. From all these database gene ontologies plays an essential role in the identification of gene and protein details. GO is a community-based bioinformatics resource that supplies information about gene product function using ontologies to represent biological knowledge[18]. The Gene Ontology (GO) project provides a comprehensive source for functional genomics. This GO was developed in a collaborative effort that creates evidence-supported annotations to describe the biological roles of individual genomic products (e.g., genes, proteins, ncRNAs, complexes) by classifying them using ontologies. The three categories of ontologies are a Biological process, Cellular component, Molecular functions. A biological process in GO refers to the biological objective to which the gene or gene product contributes. The biological process is accomplished via one or more ordered assemblies of molecular functions. Examples of higher-level biological process terms are 'cell growth and maintenance' or 'signal transduction.' Examples of (lower level) process terms are 'translation,' 'pyrimidine metabolism' or 'cAMP biosynthesis.' The molecular function is defined as the biochemical activity (including specific binding to ligands or structures) of a gene product. Molecular functions describe what is done without specifying where or when the event occurs. Examples of broad functional terms are 'enzyme,' 'transporter' or 'ligand.' Examples of narrower functional terms are 'adenylate cyclase' or 'Toll receptor

ligand.' A cellular component refers to the place in the cell where a gene product is active. These terms reflect the understanding of eukaryotic cell structure. The cellular component includes such terms as 'ribosome' or 'proteasome,' specifying where multiple gene products would be found. It also includes terms such as 'nuclear membrane' or 'Golgi apparatus' These GO ontologies are developed for a generic eukaryotic cell; accordingly, specialized organs or body parts are not represented. GO terms are connected into nodes of a network. Thus, the connections between its parents and children are known and form what is technically described as directed acyclic graphs[19]. The ontologies are dynamic, in the sense that they exist as a network that is changed as more information accumulates but have sufficient uniqueness and precision so that databases based on the ontologies can automatically be updated as the ontologies mature. The ontologies are flexible in another way so that they can reflect the many differences in the biology of the diverse organisms, such as the breakdown of the nucleus during mitosis. In this way, the GO Consortium has built up a system that supports a common language with specific, agreed-on terms with definitions and supporting documentation (the GO ontologies) that can be understood and used by a wide biological community[20.21].

Gene Enrichment Analysis

Gene Enrichment Analysis it is a method to identify classes of genes and protein that are over-represented in the large set of genes by the user to determine the phenotypes of the organisms[22]. Gene Set Enrichment Analysis allows one to questions that are of direct interest to the understanding of the functional mechanism in a cell (i.e.) is a specific pathway activated in a given tissue under some treatment, is the path active more than other pathways[23]. Annotation enrichment or GEA or Pathway Enrichment Analysis otherwise called by has become the go-to secondary analysis undertaken on collections of genes identified by high-throughput genomic methods owing to its ability to provide valuable insight into the standard biological function underlying a list of genes[24]. By systematically mapping, genes and proteins to their associated biological annotations (such as gene ontology [GO] terms or pathway membership) and then comparing the distribution of the conditions within a gene set of interest with the background distribution of these terms (e.g. all genes represented on a microarray chip), enrichment analysis can identify gene terms which are statistically over- or under-represented within the list of interest[25].

Network Enrichment Analysis (NEA)

Gene Set Enrichment Analysis is the method used to determine the biological characterization of the experimentally altered gene (AGS). By GSEA it is useful to obtain much genetic and protein information also gene interaction is available, but lack of topological details and the network link is not answered. For this, a need for a new high throughput technique is developed known as NEA (Network Enrichment Analysis)[26]. GSEA only counts the number of genes shared between a preliminary list while NEA considers network edges between any genes of the two sets in the global network[27]. The idea of NEA is that the presence of enrichment between two sets of gene says A and B, can be assessed by comparing the number of links connecting nodes in A and B with a reference distribution, which models the number of links between the same two sets in the absence of enrichment[28]. NEA determines the input regarding "genes," a

range of functional nodes of a biological network that can be analyzed with this algorithm, such as protein molecules, genomic regions that encode proteins, microRNAs, promoters, and enhancers, etc. Nodes listed in AGS, FGS, and NET should employ the same ID format. There are many online tools to perform the NEA namely FunGeneNT, EVINET, R programming, NetGen, etc. NEA performs the quick binomial test as described in Methods and quantifies enrichment in each AGS FGS pair with some statistics: chi-squared score, z-score, p-value, q-value (the p-value Adjusted for false discovery. The rating number of network edges that exist between any nodes of AGS and FGS and a respective number of AGS-FGS edges expected by chance, calculated with the binomial formula[29].

AIM

The project aims to study the list of 22 upregulated oncogenes that was obtained from the previous student about:

1) overrepresentation of the gene in GO terms by using a various suitable method known as GEA.

2) To check for the network enrichment analysis for the oncogene in EVINET.

MATERIALS AND METHODS

Materials: Functional annotation clustering was performed with the web-based tool DAVID (Database for Annotation, Visualization, and Integrated Discovery, version 6.8, http://david.abcc.ncifcrf.gov/home.jsp). GSEA was performed with other method web-based tool known as PANTHER (http://www.pantherdb.org). The Network Enrichment Analysis (NEA) was performed using the web-based tool EVINET (https://www.evinet.org/).

DAVID

A high-throughput and integrated data-mining using DAVID were done to analyse a list of 22 genes. First, the gene list was entered in DAVID. Secondly, a gene identifier was selected using the official gene symbol. Thirdly, an annotation analysis was done on the gene list. Fourthly, the species was chosen, here human gene list is used, so *Homo sapiens* was selected. The gene list was saved as list 1 in list manager and was used for analysing the results. The output of this enrichment analysis was obtained at several options and viewed using the functional annotation tool. The production of the enriched gene list and its p-value is discussed in the results.

PANTHER

The curated list of genes CD83, CLGN, CPE, CSE1L, CXADR, FAM84A, FKBP4, GABBR2, HMMR, KIF13B, KIF14, KIF18A, LDOC1, MOK, NCAPH, NEDD4L, PCOLCE2, PPM1E, PRUNE2, SLC29A1, STMN3, TCOF1 was uploaded to the panther online software tool. Firstly, upload the gene list and select gene list ID. Secondly, species were selected as *Homo sapiens* to be used as background for statistical analysis. Thirdly, the selection of the statistical analysis over-representation of the test was chosen.

In Gene ontology, there are many cellular components for which gene can be distributed based on their cellular localization. The Panther method was used to check for the enrichment of genes in the gene list for all the Cellular components terms, Molecular function and Biological process in the gene ontology database. The cutoff result was based on the enrichment in the cellular component, molecular function, and Biological process. The less the p-value for a cellular component, molecular function and Biological process the more significantly enriched the gene list is in the component.

Statistical analysis:

$$p = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}} = \frac{\binom{a+b}{b}\binom{c+d}{d}}{\binom{n}{b+d}} = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{a!\;b!\;c!\;d!\;n!}$$

DAVID and PANTHER analysis uses Fisher exact test to determine the p-value. The p-value generates weather the evidence from the sample does not affect the population. The lower the p-value, the higher the gene list enriched [45]. The significance level used in DAVID was 0.05, which means the result shown for the gene list are strongly enriched. The 22-gene list uses the human genome as background. If the threshold frequency was less than two, then it can be said that the genes of the pathway are not enriched — the significant benefit with DAVID analysis that it shows only the threshold frequency higher than 2. The results for the data were extracted and saved as Excel files.

EVINET

The altered list of genes CD83, CLGN, CPE, CSE1L, CXADR, FAM84A, FKBP4, GABBR2, HMMR, KIF13B, KIF14, KIF18A, LDOC1, MOK, NCAPH, NEDD4L, PCOLCE2, PPM1E, PRUNE2, SLC29A1, STMN3, TCOF1 was uploaded to the EVINET online software tool. The altered 22 gene list is the gene from the upregulated oncogenes from the transcription experiment in colorectal, breast and prostate cancer. Firstly, upload the altered gene list and select the project title. Secondly, species were selected as *Homo sapiens* to be used as background for statistical analysis. Thirdly, the selection of the network was made. Here the pathway common is the database selected. Fourthly, the background of the human data link is selected, here the GO cellular component is selected. Finally, the check and submit was done and calculate network enrichment is done by selecting, and the result is shown. The results display all the GO terms that are linked with the functional gene sets. The significantly enriched gene was checked, the table was selected and tested for the symbol "*" which means the cellular term is enriched considerably. The significantly enriched cellular component is selected, and the genes that are shared by this term is checked via subnetwork.

RESULTS

Results from DAVID

In order to determine the GSEA, the DAVID analysis was first used. The result from the DAVID analysis showed the effect based on the cellular component, molecular function, and biological pathway, according to the gene ontology.

The Enrichment for the list of 22 genes was checked for cellular component enrichment, and the results that are obtained, are seen to be significantly enriched in GO terms.

The results are obtained for the enriched annotated cellular component term based on the p-value 0.05. In order to understand the effects, the graph was created for the percentage of genes involved in enriched terms and the chart based on the p-value (Figure 1). From the chart, it was seen that cellular components cytoplasm, neuron projection, and kinesin complex are significantly enriched in GO terms. Based on the enriched terms it was determined that the genes that are involved in the cellular components kinesin complex and neuron projection have three genes each. The following genes were significantly enriched in the indicated components KIF14, KIF18A, KIF13B in the kinesin complex, STMN3, GABBR2, and CXADR in neuron projection, PRUNE2, STMN3, TCOF1, KIF18A, GABBR2, NEDD4L, CXADR, and KIF13B in the cytoplasm component.

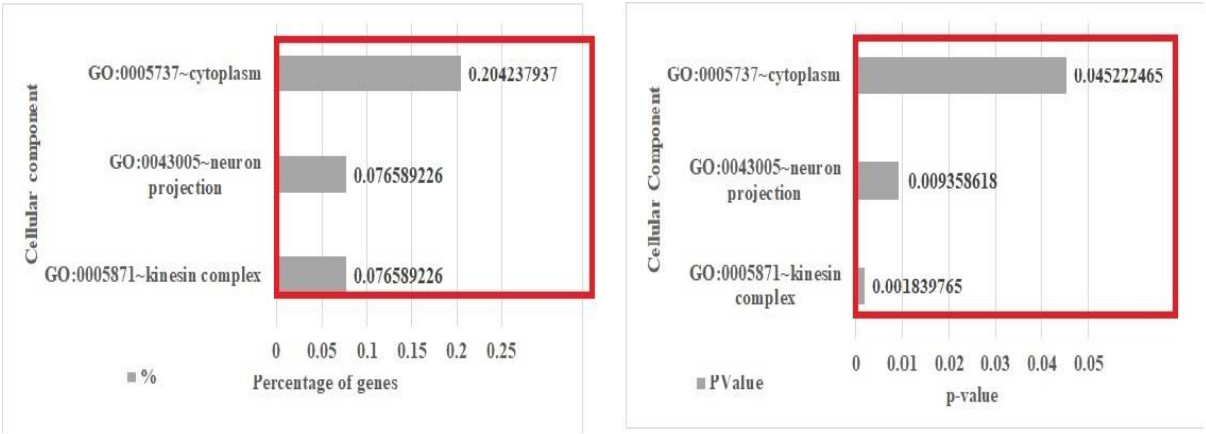

Figure 1: Enriched gene terms in GO cellular component for percentage of gene and p-value. Based on the lower p-value and gene involved more than two the more enriched the gene list is. The red box indicates that GO terms kinesin complex, neuron projection, and cytoplasm are significantly enriched, and gene involved in these GO terms is more than two-fold enriched.

The result was obtained for the enriched molecular function for the 22-gene list from the GO category (Figure 2).

The result obtained was made into two based on the percentage of the genes and the p-value. The GO molecular functions terms are determined for the enriched gene list based on the p-value less the 0.05 and gene involved more than two. ATP dependent microtubule motor activity, plus end-directed was significantly enriched, and the genes that are involved in the enriched term are KIF14, KIF18A.

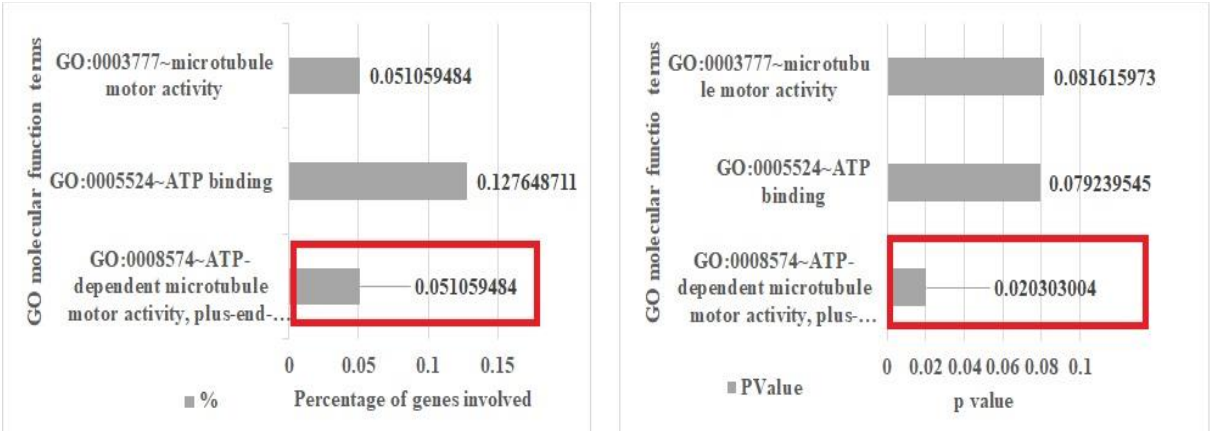

Figure 2: Enriched gene terms in GO molecular function for a percentage of gene and p-value. Based on the lower p-value and gene involved more than two the more enriched the gene list is. The red box indicates that GO terms ATP dependent microtubule motor activity, plus end-directed are significantly enriched, and gene involved in these GO terms is more than two-fold enriched.

The result for the 22-gene list from the GO category for the molecular function is obtained for only enriched terms (Figure 3).

The result obtained was made into two based on the percentage of the genes and the p-value. The GO biological pathways terms are determined for the enriched gene list based on the p-value less the 0.05 and gene involved more than two. GO terms mitotic metaphase plate congression, cytoskeleton-dependent intracellular transport, microtubule depolymerization, microtubule-based movement are significantly enriched. The genes that are involved in the enriched biological pathways are a microtubule-based moment in KIF14, KIF18A, KIF13B, microtubule depolymerization in STMN3, KIF18A, cytoskeleton-dependent intracellular transport in KIF14, KIF13B, mitotic metaphase plate congression in KIF14, KIF18A are the list of genes that are significantly enriched in the list of 22 genes.

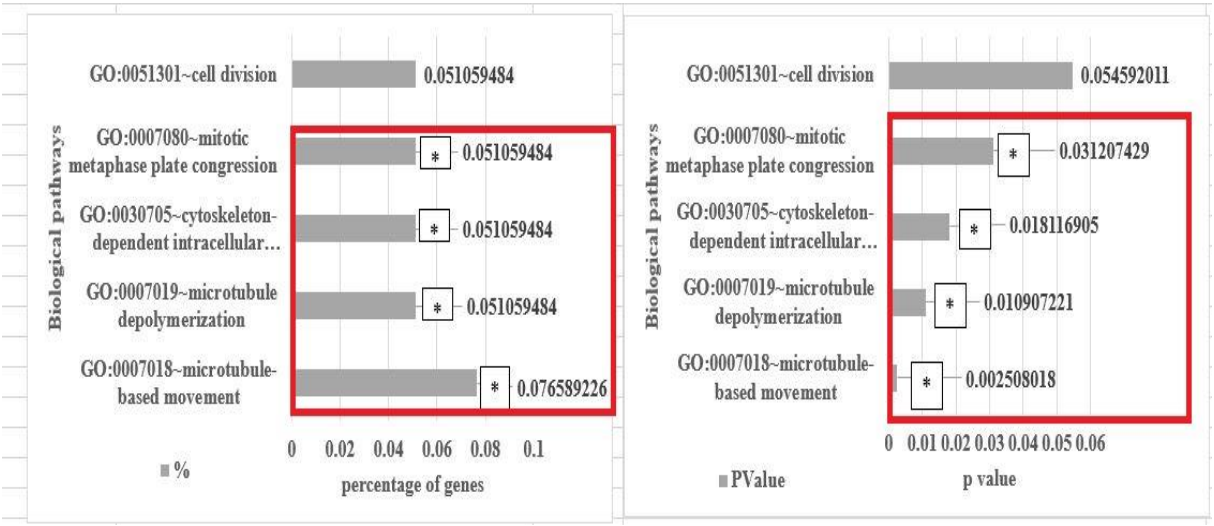

Figure 3: Enriched gene terms in GO biological pathways for a percentage of gene and p-value. Based on the lower p-value and gene involved more than two the more enriched the gene list is. The red box indicates that GO terms mitotic metaphase plate congression, cytoskeleton-dependent intracellular organization, microtubule depolymerization, microtubule-based movement are significantly enriched, and gene involved in these GO terms is more than two-fold enriched.

Results from PANTHER

In order to determine the GSEA, a different method called PANTHER analysis was used. The result from the panther analysis showed the effect based on the cellular component, molecular function, and biological pathway, according to the gene ontology. PANTHER online tool shows only the result from GO terms.

The Enrichment for the list of 22 genes was checked for cellular component enrichment, and the results obtained were significantly enriched in GO terms.

To check for the cellular component enriched terms, a graph based on the p-value 0.05 was created (Figure 4). The Fisher exact test which is used as statistical analysis gave the results for enriched terms. Here most of the cellular components terms are enriched, based on the concept lower the p-value the more the gene is enriched. The GO terms kinesin complex and microtubule-associated complex are significantly enriched.

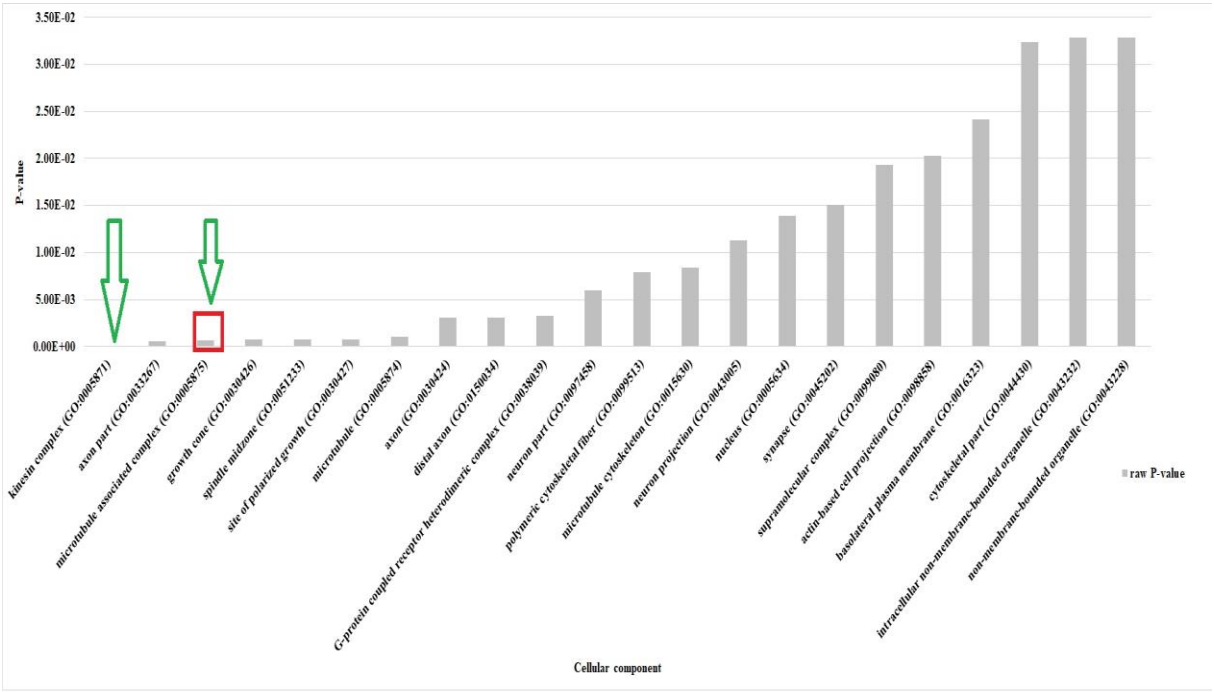

Figure 4: Enriched gene terms in GO cellular component for p-value. Based on the lower p-value more enriched the gene list is. The red box with a green arrow indicates that GO terms kinesin complex, microtubule-associated complex are significantly enriched, and gene involved in these GO terms is more than two-fold enriched.

Table 1: Enriched gene symbol for the least p-value with respective genes that are involved in the GO cellular component with a number of genes involved in it.

Cellular component GO terms.	Number of genes involved	P value	Gene Symbol
kinesin complex (GO:0005871)	3	3.55E-05	KIF18A, KIF14, KIF13B.
microtubule associated complex (GO:0005875)	3	6.21E-04	KIF18A, KIF14, KIF13B.

The genes that are enriched in the GO cellular component terms for the least p-value with gene involved more than two and above are determined (Table 1). The p-value along with the gene involved in the GO terms and the related gene from the study of 22 gene list are seen. The complete enriched GO cellular component along with gene involved and p-value is seen (Table S1). It can be seen that KIF18A, KIF14, KIF13B was enriched for both kinesin complex and microtubule-associated complex.

The Enrichment for the list of 22 genes was also checked for Molecular Function enrichment, and the results that are significantly enriched in GO terms are determined.

A graph was created to determine the significance level of each enriched molecular function terms (Figure 5). Here most of the molecular functions terms are enriched, based on the concept lower the p-value the more the gene is enriched. The data show that microtubule motor activity, tubulin activity, motor activity, microtubule binding, cytoskeleton protein binding are significantly enriched.

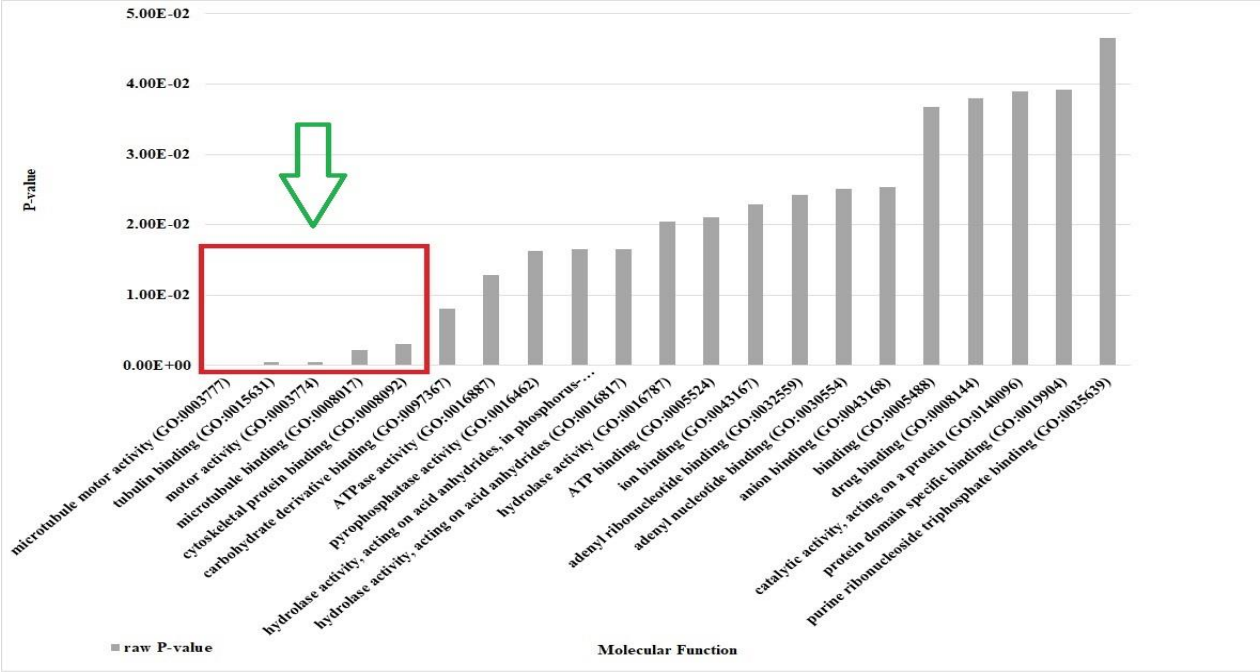

Figure 5: Enriched gene terms in GO molecular functions for p-value. Based on the lower p-value more enriched the gene list is. The red box indicates that GO terms microtubule motor activity, tubulin activity, motor activity, microtubule binding, cytoskeleton protein binding is significantly enriched, and gene involved in these GO terms is more than two-fold enriched.

Table 2: Enriched gene symbol for the least p-value with respective genes that are involved in the GO molecular function with a number of genes involved in it.

GO molecular function complete	No of genes involved	P-value	Gene Symbol
motor activity (GO:0003774)	3	5.01E-04	KIF18A,KIF14,KIF13B
microtubule binding (GO:0008017)	3	2.24E-03	KIF18A,KIF14,KIF13B
tubulin binding (GO:0015631)	4	4.33E-04	STMN3,KIF18A, KIF14,KIF13B
microtubule motor activity (GO:0003777)	3	1.21E-04	KIF18A,KIF14,KIF13B
cytoskeletal protein binding (GO:0008092)	5	3.04E-03	STMN3,KIF18A, KIF14,KIF13B,FKBP4

The genes that are enriched in the GO molecular functions terms for the least p-value with a gene involved more than two, and above are determined (Table 2). The p-value along with the gene involved in the GO terms and the related gene from the study of 22 gene list are seen. The complete enriched GO molecular function along with gene involved and p-value are seen (Table S2). It can be seen that KIF18A, KIF14, KIF13B was enriched for motor activity, microtubule

binding, and microtubule motor activity, STMN3, KIF18A, KIF14, KIF13B for tubulin binding, and STMN3, KIF18A, KIF14, KIF13B, FKBP4 for cytoskeletal protein binding are enriched.

The Enrichment for the list of 22 genes was checked for biological pathways, and the results obtained are seen to significantly enriched in GO terms.

A graph to determine the significance level of each enriched biological pathways terms was created (Figure 6). Here most of the biological pathways are enriched, based on the concept lower the p-value the more the gene is enriched. The GO terms microtubule-based movement, regulation of microtubule-based process, regulation of cytoskeleton organization is significantly enriched for the p-value less than 0.05 gene count equal and more than two. The detailed view of the GO biological pathway terms along with p-value and genes involved is shown in table 3.

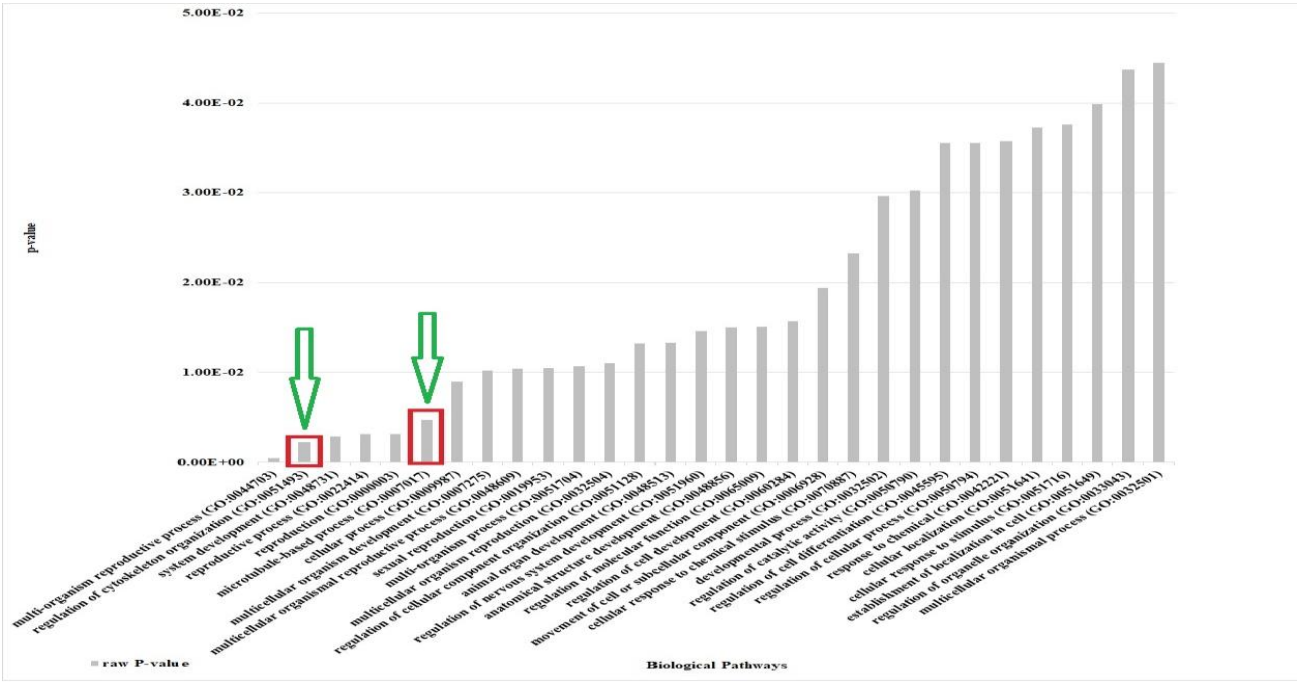

Figure 6: Enriched gene terms in GO biological pathway for p-value. Based on the lower p-value more enriched the gene list is. The red box with green arrow indicates that GO terms regulation of cytoskeleton organization and microtubule based process are significantly enriched, and gene involved in these GO terms is more than two-fold enriched.

Table 3: Enriched gene symbol for the least p value with respective genes that are involved in the GO biological process with number of genes involved in it.

GO biological process complete	Gene involved	P-value	Gene symbol
microtubule-based process (GO:0007017)	4	4.72E-03	STMN3, KIF18A, KIF14, KIF13B

regulation of cytoskeleton organization (GO:0051493)	4	2.22E-03	STMN3, KIF18A, FKBP4, PPM1E

The genes that are enriched in the GO biological pathways terms for the least p value with gene involved more then two and above are determined (Table 3). The p-value along with the gene involved in the GO terms and the related gene from the study of 22 gene list are seen. The complete enriched GO biological functions along with gene involved and p value are seen (Table S3). It can be seen that STMN3,KIF18A, KIF14,KIF13B for microtubule based process and STMN3,KIF18A,FKBP4,PPMIE for regulation of cytoskeleton organization are enriched.

Results from EVINET

In order to determine the network enrichment analysis, EVINET was used for a network enrichment results for the list of 22 gene set. The EVINET tool produces the result based on the FDR for the GO enriched term. It is known that lower the FDR value the higher is the network enriched. The significance value used here was <0.01. This data obtained showed that there are four significantly enriched FGS from GO cellular component (Table 4)

Table 4: The Network Enrichment Analysis for the enriched terms in GO cellular component.

AGS	#links AGS	FGS	#links FGS	#links_AGS, #links_FGS, and #linksAGS2FGS	FDR	Shared Genes
USERS_LIST_AGS	**1810**	**GO:0005874; MICROTUBULE**	**40081**	**97**	**2.60 E-14**	**4***
USERS_LIST_AGS	**1810**	**GO:0005828;KINETOCHORE_MICROTUBULE**	**1316**	**7**	**1E-04**	**1***
USERS_LIST_AGS	**1810**	**GO:0005871;KINESIN_COMPLEX**	**6167**	**17**	**0.0015**	**3***
USERS_LIST_AGS	**1810**	**GO:0000796;CONDENSIN_COMPLEX**	**727**	**4**	**0.0057**	**1***

Here AGS= users gene list
Links AGS=Total number of network links produced in the current networks.
FGS= Functional Annotation.
links FGS= Total number of links produced by the FGS in the current network.
links_AGS links_FGS and linksAGS2FGS= Number of links between AGS and FGS.
FDR= False Discovery Rate of the network analysis.
Shared genes= Network enrichment links showing how many genes are shared by the two sets FGS and AGS based on the enrichment indicating "*."

The EVINET shows the result based on the FDR value. It shows the topological information and the interaction between the AGS and FGS. The GO terms Microtubule, kinetochore-microtubule, kinesin complex, and condensin complex is found to be less the <0.01 for FDR values and are significantly enriched. The enriched GO terms have genes shared in them (Fig S2, S3, S4, S5). The common gene link between AGS and FGS for the GO terms in Microtubule, kinetochore-microtubule, kinesin complex. is KIF18A. The AGS shared with FGS in microtubule are KIF13B, KIF14, FKBP4, and HMMR. In Kinetochore microtubule the AGS shared with FGS is CSEIL. In kinesin complex, the AGS shared with FGS is KIF14, KIF13B, and CSEIL. In condensin complex, the AGS shared with FGS is NCAPH.

Other Methods

The two different methods for the GSA was used. They are EnrichR and Functional rich enrichment analysis. The EnrichR and FunRich methods also gave the result for the list of 22 gene sets. These two methods have the same concept as DAVID and PANTHER, but the output was not as expected from in mentioned methods. So, the further analyzing of the results from EnrichR and FunRich analysis was stopped.

DISCUSSION

DAVID, PANTHER, FUNSET analysis and EnrichR the methods are most frequently used and user-friendly online tools[30,31,32,33]. Indeed these methods have a distinct difference in them based on their output. These methods give a common output known as a raw p-value. The raw p-value is crucial because it tells the gene are significantly enriched. The raw p values in each of these methods are generated by Fisher exact test. But in FUN RICH enrichment analysis, the raw p values are calculated by a hypergeometric test. The other output such as the FDR is also called as multiple hypothesis test series in these methods. In fun rich enrichment analysis, the various hypotheses are corrected by the Bonferroni method. In EnrichR it is even more complicated as it has raw p-value calculated by Fisher exact test, along with it it has adjusted p-value, old p-value, old adjusted p-value, and Z score. In Enrichr by considering these results, it's very complicated to know the genes that are enriched in GO terms. EnrichR and Funrich enrichment analysis results were not completely satisfied as it had the complicated ways to select for the enriched terms of the 22-gene list.

DAVID and Panther are handy tools and are recently updated and frequently used method for the enrichment analysis. DAVID and PANTHER have to GO embedded in them. DAVID used Fisher exact test for the raw p-value. DAVID also calculate the multiple hypothesis tests using Bonferroni, Benjamini Hochberg, and False discovery rate. But DAVID gives us the default result of the enriched gene for GO terms based on the raw p-value for CC, MF, BP. PANTHER on the other side has the same properties as DAVID, but PANTHER is an active tool (i.e.) updated frequently and has GO embedded in it. PANTHER gives the result based on the binomial calculation. It also provides FDR, the vital part of the PANTHER is that contains the GO terms if the gene list is under-represented or over-represented.

The enrichment of the gene was obtained by checking the raw p-values. The raw p-value in DAVID and PANTHER are generated by the Fisher exact test. The raw p-values were used to

filter the enriched genes for the p-value less than 0.05. The lower the p-value, the lower the gene list are significantly enriched in the GO terms[45]. Fisher exact test is a high-throughput statistical experiment developed by R.A fisher. This method examines whether regulated genes are over-represented in a given gene set by an analysis of independence in a two-by-two contingency table, where the test statistic can be constructed based on $\chi 2$, hypergeometric, or binomial distribution[46]. There is a new method developed to cut off the criterion for the gene list known as FDR. The FDR takes considers p-value as input and gives the output. The primary use of FDR is increasing the sensitivity of gene set enrichment analysis by removal of background noise[47]. The foremost reason to avoid the FDR in this study is that the gene list used here is minimal. FDR analysis performs very well when there is a big gene list of more than 50. But the gene list used here is 22 oncogene list. The background selected was GO which has human genes database. So, when comparing with these large human backgrounds for the list of 22 gene list, it is good to be not considering the FDR values and go with a raw p-value. The raw p-value tells the gene list are enriched for the GO terms. The DAVID analysis gave the result based on the raw p-value less than 0.05 and with more than two genes involved in a GO component.

From the DAVID, PANTHER the microtubule-based process of GO component that is significantly enriched are determined. Microtubule is the significant component of the cytoskeleton. Microtubule is a tubular polymer that forms the structure and shape to the cytoplasm of eukaryotic cells. Microtubule is composed of α- and β-tubulin heterodimers arranged ahead to tail in protofilaments that are aligned in parallel to form long, slender tubes. It is known as microtubules play an essential role in several eukaryotic cells such as cellular process, cell growth, cytokinesis, transport, and mobility. Microtubules, together with microfilaments and intermediate filaments, form the cell cytoskeleton. Microtubule plays a crucial role in the super targets of anti-cancer drugs[34]. The microtubule as cancer target will continue to be the essential chemotherapeutic agents. The polymerization dynamics of microtubules regulate the biological functions in all eukaryotic cells[35]. Microtubule importance in the mitosis and cell division makes them a super target for a group of chemically diverse anticancer drugs. Microtubule recently from marine sources have grabbed much attention due to unique tubulin binding features and remarkable ability to reduce tumor progression[36].

An existing review from regarding the microtubule-targeting drugs has categorized into two major categories based on the stabilization. Microtubule stabilizing agents such as paclitaxel, docetaxel, epothilones, and discodermolide that bind to tubulin polymer and stabilize the microtubules. Microtubule destabilizing agents such as vinca alkaloids, colchicine, and combretastatin that binds to tubular dimers and cause destabilization. These categories ultimately alter the equilibrium between tubulin and microtubule resulting in the disruption of mitotic spindle thereby affecting a critical transition in the cell cycle leading to cell death[37,38]. In addition to the effects in mitosis, microtubule plays a vital role in the cellular function along each phase of the cell cycle. Some complex proteins are involved in the oncogenesis includes p53, c-Myc, BRCA1, androgen receptor, APC and Src are known to associate with and traffic along microtubules. One-third of MAP kinase protein is associated with microtubules. MTAs has the abilities to interrupt functions of interphase microtubules which thereby would be

expected to weaken the activity of these proteins in cancer maintenance and progression. MTAs were recently shown to inhibit the translocation of DNA repair proteins to the nucleus and significantly to have the clinical efficacies of combining MTAs and DNA damaging agents[39].

In the results for GEA, the microtubule-based process was the enriched GO terms in common. The new high throughput method known as Network Enrichment Analysis was used to check for the network analysis. NEA is very efficient because it has a functional gene set as background and it analysis the network link for the altered gene, i.e., the user's gene list which obtained by the previous experiment. The EVINET uses FDR to check for the enriched gene and give the result for the p-value less than 0.01. There were 147 recognized network enrichments as per EVINET. The common gene link was found between AGS and FGS for the GO terms in Microtubule, kinetochore-microtubule, kinesin complex. The gene that links AGS and FGS is KIF18A. The AGS shared with FGS in microtubule are KIF13B, KIF14, FKBP4, and HMMR. In Kinetochore microtubule the AGS shared with FGS is CSEIL. In kinesin complex, the AGS shared with FGS is KIF14, KIF13B, and CSEIL. In condensin complex, the AGS shared with FGS is NCAPH.

In the results, gene list KIF18A, KIF14, KIF13B that were enriched for the microtubule-based process are obtained. The results obtained are common for GO biological terms from DAVID and PANTHER. These genes are of kinesin family, which are motor proteins and involved in the microtubule-based process. Microtubules have a motor protein that moves along them known as kinesin. The strong affinity of the kinesin to the microtubule is due to the addition of ATP in the development of the kinesin motion[24]. It has been shown that KIF genes are associated with breast cancer by accumulating the levels at the microtubule plus ends and ends up disturbing the dynamics of the cytoskeleton and mitotic spindle formation. As a result, cells overexpressing KIF are often multinucleated[40]. The overexpression of the KIF is linked to higher proliferation rates with increased migration and increased invasiveness in both colorectal and breast cancer, and that these features can be reduced via siRNA[41]. KIFα has its main function in transporting phosphatidylinositol-trisphosphate (PIP3) in the cell membrane in neuronal cells which then contributes to cell polarity[42]. KIF genes promote chromosome congregation by attenuating chromosome oscillation magnitudes and is indispensable for the depolymerization of microtubules during cellular division[41]. It is seen that microtubule-dependent protein KIF genes have its essential function during cytokinesis of normal cell division. If there is overexpression of this protein can lead to genetic instabilities and later induce tumor formation[43]. It has, quite recently, been seen that if KIF expression is suppressed in hepatocellular carcinoma, the prognosis was more favorable indicating that KIF is strongly linked to tumorigenesis. The same report also showed that overexpression of KIF might be connected with increased migratory behaviour[44].

The other enriched gene from EVINET obtained was NCAPH, HMMR, FKBP4, and CSEIL. Surprisingly it was found HMMR are enriched for microtubule. Hyaluronan (HA) is generally one of the cell binding sites. HA is implicated in a variety of disease, homeostatic and developmental process[48]. There is more research going in study related to HMMR. Increased migration and mobile behavior have been observed in cells transformed with RAS oncogene

before, and the association between hyaluronic acid and increased motility has been made. It is also postulated that the binding of hyaluronic acid to CD44 activates downstream simulation of Rho GTPases, which is the cause for the increased migration HMMR overexpression has been strongly associated with many advanced cancer types due to its stimulation via hyaluronic acid, which is secreted from stromal fibroblasts when tumors are present[49].

NCAPH is enriched for condensing complex in EVINET. NCAPH is a subunit of Condensin which is a multi-protein complex that places a significant role in chromosome assembly segregation during mitosis and meiosis[50]. It has been demonstrated that condensing complex is excluded from the nucleus until nuclear envelope breakdown. The inhibition of prophase chromosome condensation occurs after condensing I knockdowns expression leads to colon cancer[51].

ETHICAL ASPECTS

This study was performed mainly in system biology, so there was no any animals or mammals were affected. The 22-gene list used here are from the transcriptome gene which is upregulated. So, there is no need any permission from any department related to ethical aspects.

CONCLUSION

The primary objective of this study was to determine if the list of 22 genes are enriched in particular GO terms. The two methods DAVID and PANTHER for gene enrichment analysis and EVINET for network enrichment analysis were used. From the result for gene enrichment, the raw p-value less than 0.05, suggesting that microtubule-based process might be significantly enriched. The FDR values in the network enrichment analysis in EVINET showed that microtubule was indeed enriched. The genes that were enriched for the GO terms were KIF18A, KIF14, KIF13B, also known as the Kinesin family that regulated microtubule function. From the result, it can be concluded that alteration in the KIF genes can lead to breast and colorectal cancer and that these genes can be novel oncogenes. We also find that the microtubule which plays an essential role in the cell division is an essential anti-cancer target. Also, HMMR is enriched in microtubule, and there is a lot of research work going as it regulates of cell locomotion.

FUTURE WORK

Based on the result from this study microtubule and its related genes are significantly enriched. The genes of kinesin family KIF18A, KIF14, KIF13B are significantly enriched or the microtubule-based process. The overexpression of the KIF gene in microtubule can lead to tumor formation. By this way, microtubule is used as a target for anti-cancer agents. There is a lot of research going on for microtubule to be used as anticancer agents. KIF genes contribute to the cell polarity by this it may be of interest for future study and experiments may reveal new elements in migratory mechanisms in both healthy and malignant cells.

ACKNOWLEDGMENT

First and foremost, I want to thank my supervisor **Annica Gad**. Thank you for your wisdom and knowledge in this field and the patience you have bestowed upon me during our time. I want to thank **Diana Tilevik** for her wise words and waffles during my time at the University of Madeira. I would also like to thank **Maciej Motyka** for his help and guidance through my work. Lastly, I would like to thank **Benjamin Ulfenborg** for discussions and help when I needed it. This research was supported financially by Fundação para a Ciência e a Tecnologia (FCT) with funds from the Portuguese Government (PEst-OE/QUI/UI0674/2013) and by the Agência Regionalparao Desenvolvimento da Investigaçaõ Tecnologia e Inovação (ARDITI) through project M1420-01-0145-FEDER-000005—Centro de Química da Madeira—CQM (Madeira 14-20).

REFERENCE

1. Sever, R., & Brugge, J. S. (2015). Signal transduction in cancer. *Cold Spring Harbor Perspectives in Medicine*, *5*(4).
2. Yang, Y., Dong, X., Xie, B., Ding, N., Chen, J., Li, Y., … Fang, X. (2015). Databases and Web Tools for Cancer Genomics Study. *Genomics, Proteomics & Bioinformatics*, *13*(1), 46–50.
3. Xing, Z., Chu, C., Chen, L., & Kong, X. (2016). The use of Gene Ontology terms and KEGG pathways for analysis and prediction of oncogenes. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1860*(11), 2725–2734.
4. Croce, C. M. (2008). Oncogenes and Cancer. *New England Journal of Medicine*, *358*(5), 502–511.
5. Wang, J., Du, Q., & Li, C. (2017). Bioinformatics analysis of gene expression profiles to identify causal genes in luminal B2 breast cancer. *Oncology Letters*, *14*(6), 7880–7888.
6. Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., … Botstein, D. (2000). Molecular portraits of human breast tumours. *Nature*, *406*(6797), 747–752.
7. Hu, Z., Fan, C., Oh, D. S., Marron, J., He, X., Qaqish, B. F., … Perou, C. M. (2006). The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics*, *7*(1), 96.
8. Cuzick, J., Thorat, M. A., Andriole, G., Brawley, O. W., Brown, P. H., Culig, Z., … Wolk, A. (2014). Prevention and early detection of prostate cancer. *The Lancet. Oncology*, *15*(11), e484-92.
9. Goggins, W. B., & Wong, G. (2009). Cancer among Asian Indians/Pakistanis living in the United States: low incidence and generally above average survival. *Cancer Causes & Control*, *20*(5), 635–643.

10. Mazhar, D., & Waxman, J. (2002). Prostate cancer. *Postgraduate Medical Journal*, *78*(924), 590–595.

11. Brawer, M. K. (n.d.). Prostate-specific antigen: current status. *CA: A Cancer Journal for Clinicians*, *49*(5), 264–281.

12. Haggar, F. A., & Boushey, R. P. (2009). Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clinics in Colon and Rectal Surgery*, *22*(4), 191–197.

13. Cancer of the colon and Rectum [Internet]. [Cited July 2006]. Available from: https://www.mycareplusonline.com/sites/default/files/cmgfiles/WYNTK_Colon_Cancer.pdf

14. Papadopoulos, N., Nicolaides, N. C., Wei, Y. F., Ruben, S. M., Carter, K. C., Rosen, C. A., … Adams, M. D. (1994). Mutation of a mutL homolog in hereditary colon cancer. *Science (New York, N.Y.)*, *263*(5153), 1625–1629.

15. Systems Biology - Its Importance and Current Scenario. (n.d.). Retrieved November 29, 2018, from https://www.biotecharticles.com/Biology-Article/Systems-Biology-Its-Importance-and-Current-Scenario-575.html.

16. Werner, H. M. J., Mills, G. B., & Ram, P. T. (2014). Cancer Systems Biology: a peek into the future of patient care? *Nature Reviews. Clinical Oncology*, *11*(3), 167–176.

17. Archer, T. C., Fertig, E. J., Gosline, S. J. C., Hafner, M., Hughes, S. K., Joughin, B. A., … Shajahan-Haq, A. N. (2016). Systems Approaches to Cancer Biology. *Cancer Research*, *76*(23), 6774–6777.

18. Gene Ontology Consortium, T. G. O. (2015). Gene Ontology Consortium: going forward. *Nucleic Acids Research*, *43*(Database issue), D1049-56.

19. Tripathi, S., Christie, K. R., Balakrishnan, R., Huntley, R., Hill, D. P., Thommesen, L., … Laegreid, A. (2013). Gene Ontology annotation of sequence-specific DNA binding transcription factors: setting the stage for a large-scale curation effort. *Database*, *2013*(0), bat062-bat062.

20. Dönitz, J., & Wingender, E. (2012). The ontology-based answers (OBA) service: a connector for embedded usage of ontologies in applications. *Frontiers in Genetics*, *3*.

21. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, *25*(1), 25–29.

22. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., … Mesirov, J. P. (2005). *Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles*.

23. Ackermann, M., & Strimmer, K. (2009). A general modular framework for gene set enrichment analysis.

24. Danielsson, F., Skogs, M., Huss, M., Rexhepaj, E., O'Hurley, G., Klevebring, D., … Lundberg, E. (2013). Majority of differentially expressed genes are down-regulated during malignant transformation in a four-stage model. *Proceedings of the National Academy of Sciences*, *110*(17), 6853–6858.

25. Tipney, H., & Hunter, L. (n.d.). *An introduction to effective use of enrichment analysis software*. Retrieved from https://humgenomics.biomedcentral.com/track/pdf/10.1186/1479-7364-4-3-202.

26. Alexeyenko, A., Lee, W., Pernemalm, M., Guegan, J., Dessen, P., Lazar, V., … Pawitan, Y. (2012). Network enrichment analysis: extension of gene-set enrichment analysis to gene networks. *BMC Bioinformatics*, *13*(1), 226.

27. Jeggari, A., & Alexeyenko, A. (2017). NEArender: an R package for functional interpretation of 'omics' data via network enrichment analysis. *BMC Bioinformatics*, *18*(S5), 118.

28. Signorelli, M., Vinciotti, V., & Wit, E. C. (2016). NEAT: an efficient network enrichment analysis test. *BMC Bioinformatics*, *17*(1), 352.

29. Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, *100*(16), 9440–9445.

30. Huang, D. W., Sherman, B. T., Tan, Q., Collins, J. R., Alvord, W. G., Roayaei, J., … Lempicki, R. A. (2007). The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biology*, *8*(9), R183.

31. Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., & Thomas, P. D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Research*, *45*(D1), D183–D189.

32. Pathan, M., Keerthikumar, S., Ang, C.-S., Gangoda, L., Quek, C. Y. J., Williamson, N. A., … Mathivanan, S. (2015). FunRich: An open access standalone functional enrichment and interaction network analysis tool. *PROTEOMICS*, *15*(15), 2597–2601.

33. Kuleshov, M. V, Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., … Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, *44*(W1), W90-7.

34. Pilhofer, M., Ladinsky, M. S., McDowall, A. W., Petroni, G., & Jensen, G. J. (2011). Microtubules in bacteria: Ancient tubulins build a five-protofilament homolog of the eukaryotic cytoskeleton. *PLoS Biology*, *9*(12), e1001213.

35. Mukhtar, E., Adhami, V. M., & Mukhtar, H. (2014). Targeting microtubules by natural agents for cancer therapy. *Molecular Cancer Therapeutics*, *13*(2), 275–284.

36. De, O., & Chatterji, B. P. (2017). Marine Derived Anticancer Drugs Targeting Microtubule. *Recent Patents on Anti-Cancer Drug Discovery*, *12*(2), 102–127.

37. Parker, A. L., Kavallaris, M., & McCarroll, J. A. (2014). Microtubules and their role in cellular stress in cancer. *Frontiers in Oncology*, *4*, 153.

38. Gan, P. P., McCarroll, J. A., Byrne, F. L., Garner, J., & Kavallaris, M. (2011). Specific β-Tubulin Isotypes Can Functionally Enhance or Diminish Epothilone B Sensitivity in Non-Small Cell Lung Cancer Cells. *PLoS ONE*, *6*(6), e21717.

39. Risinger, A. L., Dybdal-Hargreaves, N. F., & Mooberry, S. L. (2015). Breast Cancer Cell Lines Exhibit Differential Sensitivities to Microtubule-targeting Drugs Independent of Doubling Time. *Anticancer Research*, *35*(11), 5845–5850.

40. Zhang, C., Zhu, C., Chen, H., Li, L., Guo, L., Jiang, W., & Lu, S. H. (2010). Kif18A is involved in human breast carcinogenesis. *Carcinogenesis*, *31*(9), 1676–1684.

41. Nagahara, M., Nishida, N., Iwatsuki, M., Ishimaru, S., Mimori, K., Tanaka, F., … Mori, M. (2011). Kinesin 18A expression: Clinical relevance to colorectal cancer progression. *International Journal of Cancer*, *129*(11), 2543–2552.

42. Horiguchi, K., Hanada, T., Fukui, Y., & Chishti, A. H. (2006). Transport of PIP3 by GAKIN, a kinesin-3 family protein, regulates neuronal cell polarity. *The Journal of Cell Biology*, *174*(3), 425–436.

43. Madhavan, J., Coral, K., Mallikarjuna, K., Corson, T. W., Amit, N., Khetan, V., … Kumaramanickavel, G. (2007). High Expression of *KIF14* in Retinoblastoma: Association with Older Age at Diagnosis. *Investigative Opthalmology & Visual Science*, *48*(11), 4901.

44. Yang, T., Zhang, X.-B., & Zheng, Z.-M. (2013). Suppression of KIF14 expression inhibits hepatocellular carcinoma progression and predicts favorable outcome. *Cancer Science*, *104*(5), 552–557.

45. Ji, R.-R., Ott, K.-H., Yordanova, R., & Bruccoleri, R. E. (2011). FDR-FET: an optimizing gene set enrichment analysis method. *Advances and Applications in Bioinformatics and Chemistry : AABC*, *4*, 37–42.

46. Khatri, P., & Draghici, S. (2005). Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics*, *21*(18), 3587–3595.

47. Goeman, J. J., & Buhlmann, P. (2007). Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics*, *23*(8), 980–987.

48. Turley, E. A., Austen, L., Vandeligt, K., & Clary, C. (1991). Hyaluronan and a cell-associated hyaluronan binding protein regulate the locomotion of ras-transformed cells. *The Journal of Cell Biology*, *112*(5), 1041–1047.

49. Yeo, T. K., Nagy, J. A., Yeo, K. T., Dvorak, H. F., & Toole, B. P. (1996). Increased hyaluronan at sites of attachment to mesentery by CD44-positive mouse ovarian and breast tumor cells. *The American Journal of Pathology*, *148*(6), 1733–1740.

50. Yin, L., Jiang, L.-P., Shen, Q.-S., Xiong, Q.-X., Zhuo, X., Zhang, L.-L., … Chen, Y.-B. (2017). NCAPH plays important roles in human colon cancer. *Cell Death & Disease*, *8*(3), e2680.

51. Hirota, T., Gerlich, D., Koch, B., Ellenberg, J., & Peters, J.-M. (2004). Distinct functions of condensin I and II in mitotic chromosome assembly. *Journal of Cell Science*, *117*(26), 6435–6445.

APPENDIX: SUPPLEMENTARY FIGURES

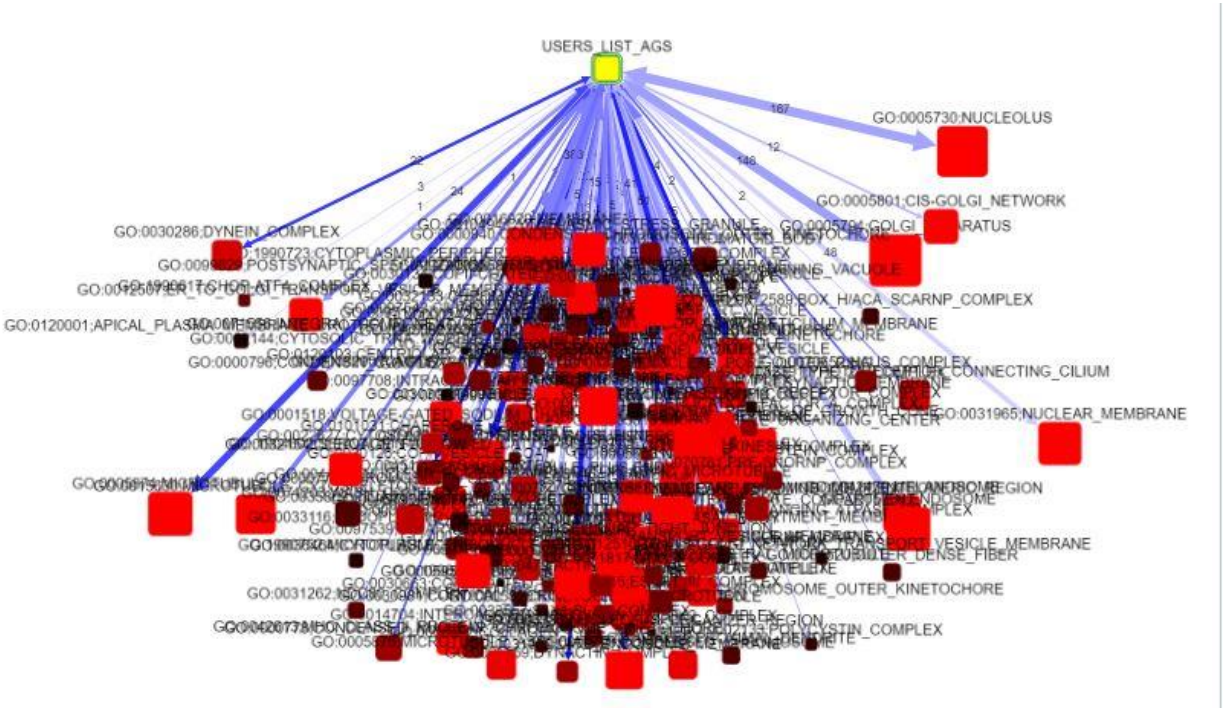

Figure S1: Network map of EVINET for the list of 22 genes. The figure is showing the links between AGS and FGS. Yellow - Altered gene list (Users gene list). Red - Functional Gene Set.

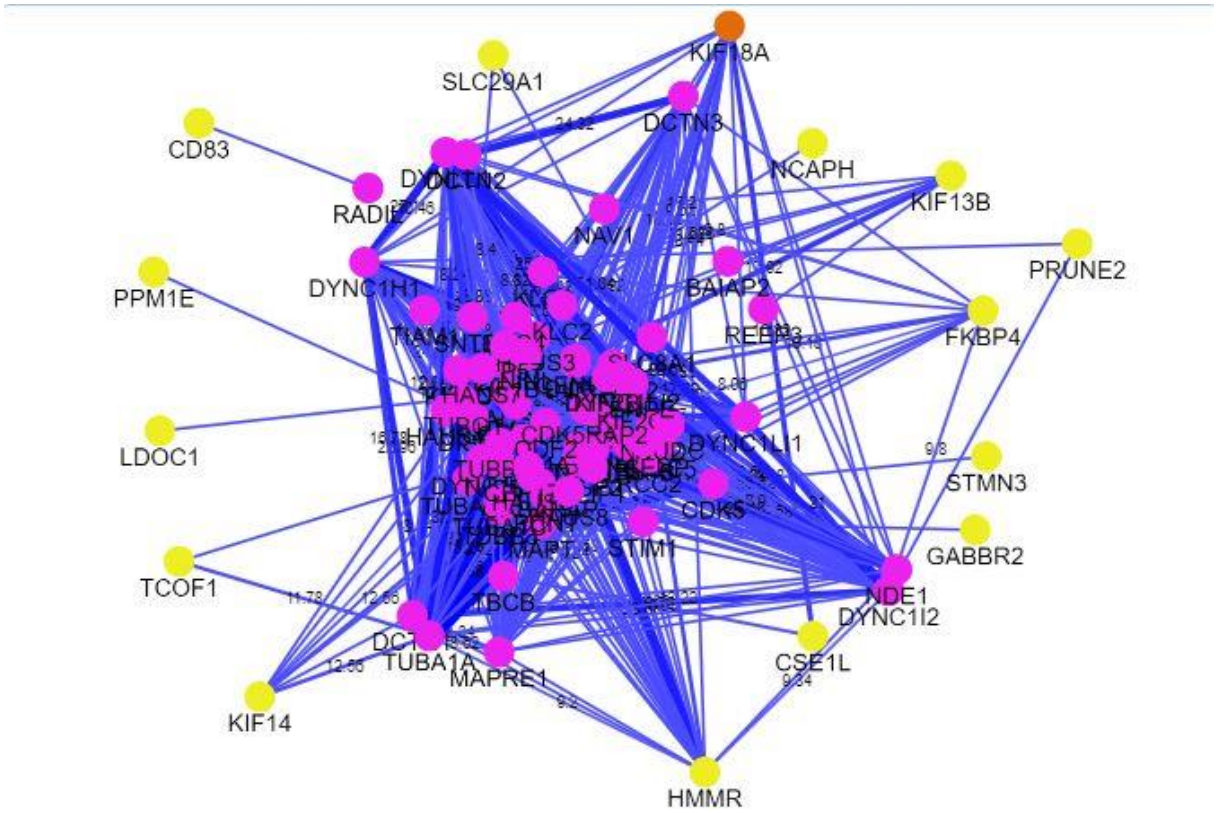

Figure S2: Network map for the enriched GO term microtubule. The figure shows us the link for the AGS and FGS for the shared genes. Yellow - Altered gene set (user gene list), Pink - Functional Gene Set. Orange - The common gene sharing FGS and AGS.

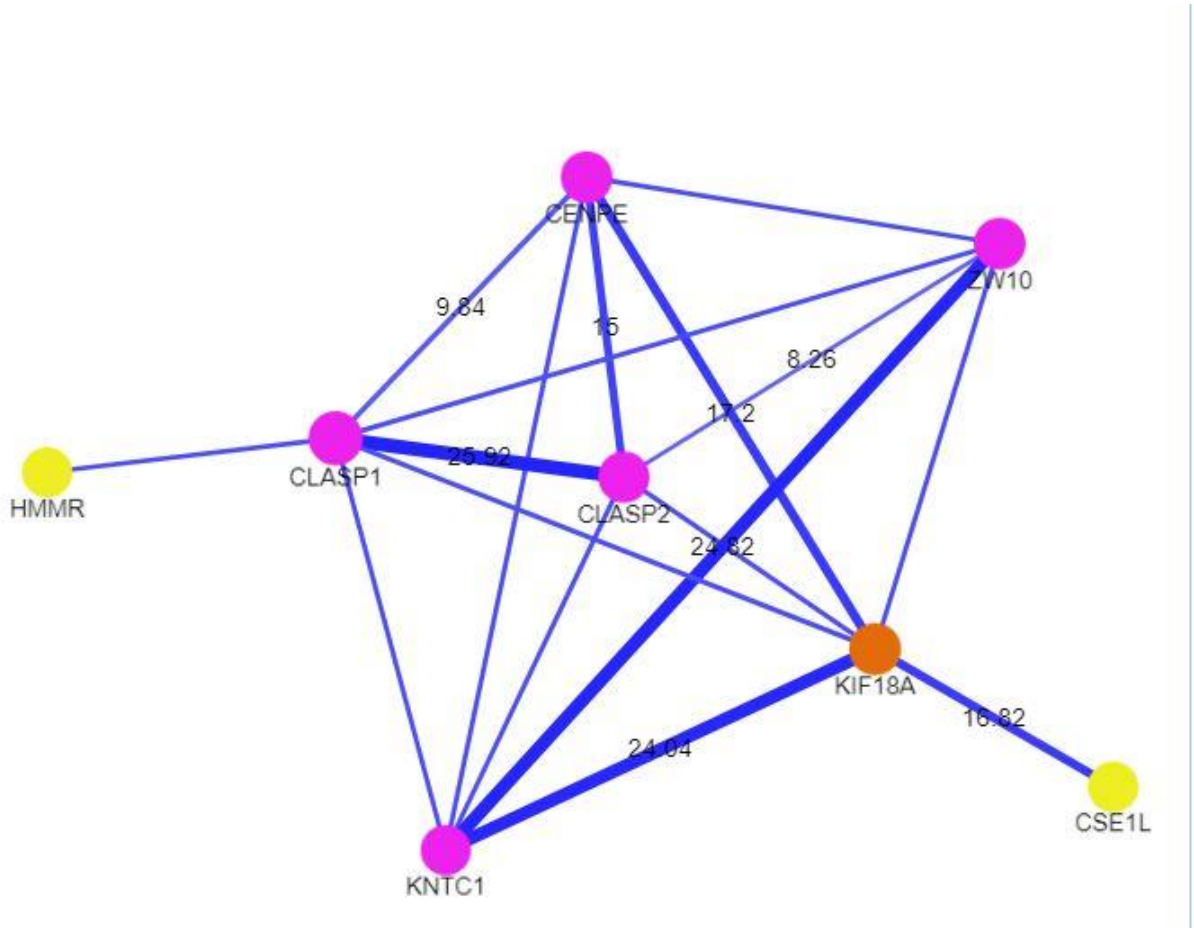

Figure S3: Network map for the enriched GO term kinetochore_microtubule. The figure shows us the link for the AGS and FGS for the shared genes. Yellow -Altered gene set (user gene list), Pink - Functional Gene Set. Orange - The common gene pooling FGS and AGS.

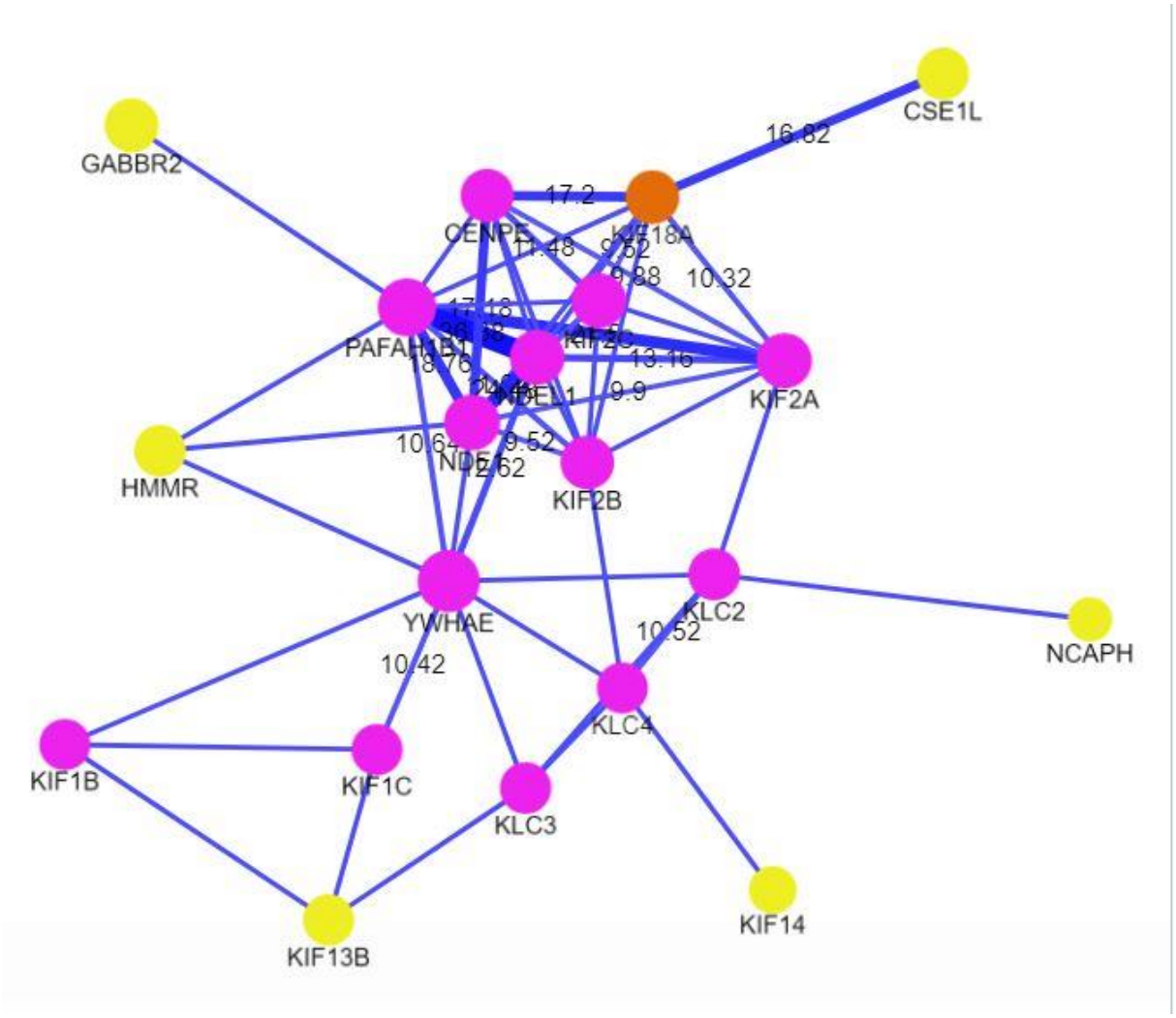

Figure S4: Network map for the enriched GO term kinesin_complex. The figure shows us the link for the AGS and FGS for the shared genes. Yellow=Altered gene set (Users gene list), Pink= Functional Gene Set. Orange= The common gene pooling FGS and AGS.

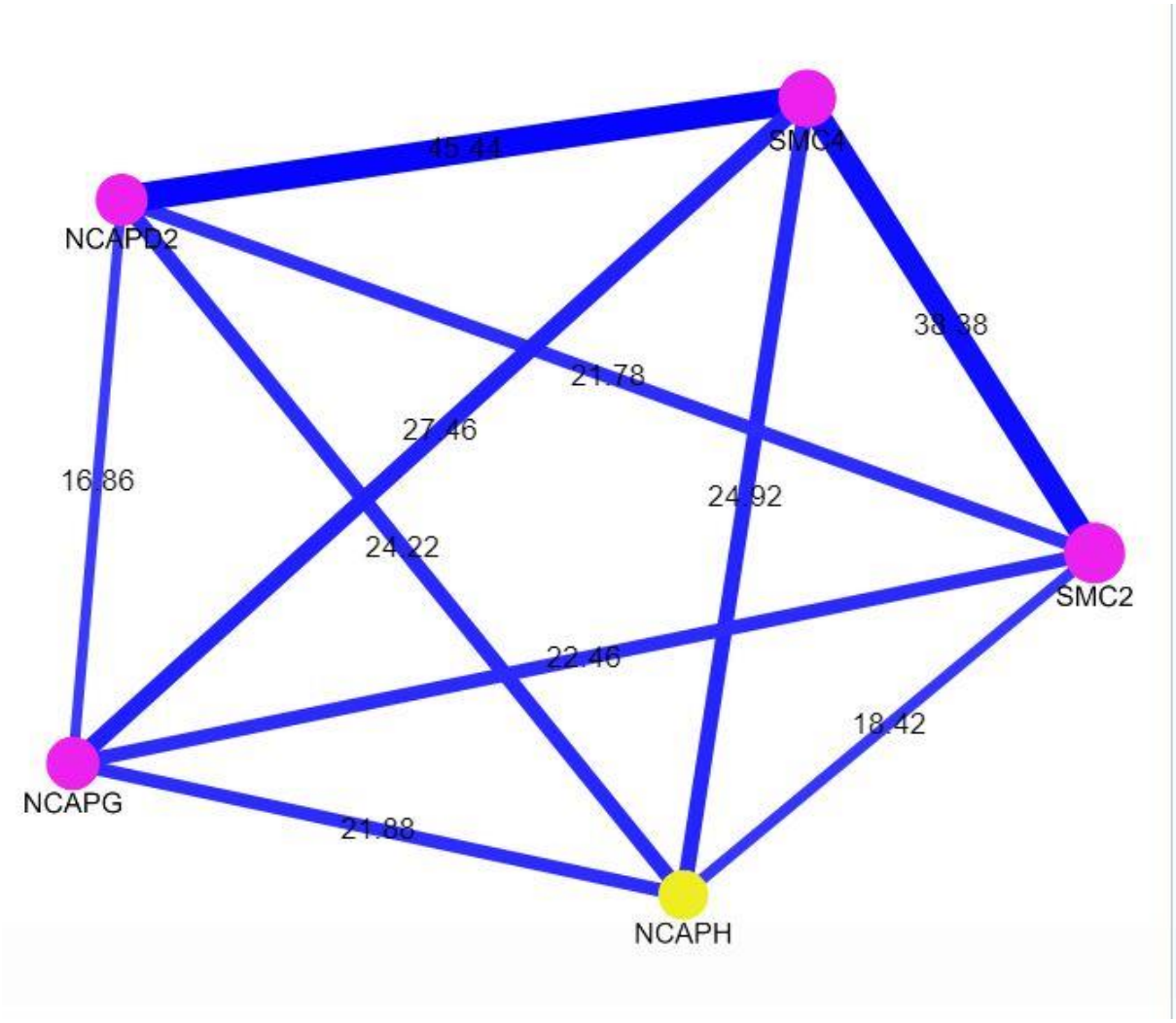

Figure S5: Network map for the enriched GO term is condensing: complex. The figure shows us the link for the AGS and FGS for the shared genes. Yellow=Altered gene set (Users gene) Pink= Functional Gene Set.

Cellular component GO terms.	Number of genes involved	P value	Gene Symbol
spindle midzone (GO:0051233)	2	7.42E-04	KIF18A, KIF14
kinesin complex (GO:0005871)	3	3.55E-05	KIF18A, KIF14, KIF13B.
microtubule associated complex (GO:0005875)	3	6.21E-04	KIF18A, KIF14, KIF13B.
growth cone (GO:0030426)	3	7.32E-04	STMN3, FKBP4, CXADR.
the site of polarized growth (GO:0030427)	3	7.98E-04	STMN3, FKBP4, CXADR.
distal axon (GO:0150034)	3	3.08E-03	STMN3, FKBP4, CXADR.
axon part (GO:0033267)	4	6.17E-04	STMN3, FKBP4, CXADR, KIF13B.

actin-based cell projection (GO:0098858)	2	2.03E-02	KIF13B, CXADR.
microtubule (GO:0005874)	4	1.06E-03	KIF18A, KIF14, FKBP4, KIF13B.
basolateral plasma membrane (GO:0016323)	2	2.42E-02	SLC29A1, CXADR.
axon (GO:0030424)	4	3.04E-03	STMN3, FKBP4, CXADR, KIF13B
polymeric cytoskeletal fiber (GO:0099513)	4	7.89E-03	KIF18A, KIF14, FKBP4, KIF13B.
synapse (GO:0045202)	4	1.51E-02	SLC29A1, GABBR2, CPE, CXADR.
microtubule cytoskeleton (GO:0015630)	5	8.44E-03	HMMR, KIF18A, KIF14, FKBP4, KIF13B.
supramolecular complex (GO:0099080)	4	1.93E-02	KIF18A, KIF14, FKBP4, KIF13B.
neuron projection (GO:0043005)	5	1.13E-02	STMN3, FKBP4, GABBR2, KIF13B, CXADR.
neuron part (GO:0097458)	6	6.03E-03	STMN3, SLC29A1, FKBP4, GABBR2, KIF13B, CPE, CXADR.
cytoskeletal part (GO:0044430)	5	3.24E-02	HMMR, KIF18A, KIF14, FKBP4, KIF13B.
intracellular non-membrane-bounded organelle (GO:0043232)	9	3.29E-02	HMMR, KIF18A, LDOC1, KIF14, NCAPH, FKBP4, KIF13B, PPMIE, TCOF1.
non-membrane-bounded organelle (GO:0043228)	9	3.29E-02	HMMR, KIF18A, LDOC1, KIF14, NCAPH, FKBP4, KIF13B, PPMIE, TCOF1
nucleus (GO:0005634)	14	1.39E-02	MOK, KIF18A, LDOC1, KIF14, NCAPH, FKBP4, PPMIE, PPMIE(M), TCOF1, CLGN, CSE1L, NEDD4L, CPE, CXADR.
Cellular component GO terms.	Number of genes involved	P value	Gene Symbol
spindle midzone (GO:0051233)	2	7.42E-04	KIF18A, KIF14
kinesin complex (GO:0005871)	3	3.55E-05	KIF18A, KIF14, KIF13B.
microtubule associated complex (GO:0005875)	3	6.21E-04	KIF18A, KIF14, KIF13B.
growth cone (GO:0030426)	3	7.32E-04	STMN3, FKBP4, CXADR.
the site of polarized growth (GO:0030427)	3	7.98E-04	STMN3, FKBP4, CXADR.

distal axon (GO:0150034)	3	3.08E-03	STMN3, FKBP4, CXADR.
axon part (GO:0033267)	4	6.17E-04	STMN3, FKBP4, CXADR, KIF13B.
actin-based cell projection (GO:0098858)	2	2.03E-02	KIF13B, CXADR.
microtubule (GO:0005874)	4	1.06E-03	KIF18A, KIF14, FKBP4, KIF13B.
basolateral plasma membrane (GO:0016323)	2	2.42E-02	SLC29A1, CXADR.
axon (GO:0030424)	4	3.04E-03	STMN3, FKBP4, CXADR, KIF13B
polymeric cytoskeletal fiber (GO:0099513)	4	7.89E-03	KIF18A, KIF14, FKBP4, KIF13B.
synapse (GO:0045202)	4	1.51E-02	SLC29A1, GABBR2, CPE, CXADR.
microtubule cytoskeleton (GO:0015630)	5	8.44E-03	HMMR, KIF18A, KIF14, FKBP4, KIF13B.
supramolecular complex (GO:0099080)	4	1.93E-02	KIF18A, KIF14, FKBP4, KIF13B.
neuron projection (GO:0043005)	5	1.13E-02	STMN3, FKBP4, GABBR2, KIF13B, CXADR.
neuron part (GO:0097458)	6	6.03E-03	STMN3, SLC29A1, FKBP4, GABBR2, KIF13B, CPE, CXADR.
cytoskeletal part (GO:0044430)	5	3.24E-02	HMMR, KIF18A, KIF14, FKBP4, KIF13B.
intracellular non-membrane-bounded organelle (GO:0043232)	9	3.29E-02	HMMR, KIF18A, LDOC1, KIF14, NCAPH, FKBP4, KIF13B, PPMIE, TCOF1.
non-membrane-bounded organelle (GO:0043228)	9	3.29E-02	HMMR, KIF18A, LDOC1, KIF14, NCAPH, FKBP4, KIF13B, PPMIE, TCOF1
nucleus (GO:0005634)	14	1.39E-02	MOK, KIF18A, LDOC1, KIF14, NCAPH, FKBP4, PPMIE, PPMIE(M), TCOF1, CLGN, CSE1L, NEDD4L, CPE, CXADR.

Table S1: Overall enriched gene list distributed among the various cellular components. The gene symbols of the genes involved in multiple cellular components and the p values for each element are also indicated.

GO molecular function complete	No of genes involved	P-value	Gene Symbol
cell adhesive protein binding involved in AV node cell-bundle of His cell communication (GO:0086082)	1	2.18E-03	CXADR
tubulin-dependent ATPase activity (GO:0070463)	1	2.18E-03	KIF18A
DNA topoisomerase (ATP-hydrolyzing) activator activity (GO:0072587)	1	2.18E-03	NCAPH
DNA topoisomerase (ATP-hydrolyzing) regulator activity (GO:0072586)	1	2.18E-03	NCAPH
copper-dependent protein binding (GO:0032767)	1	4.36E-03	FKBP4
G protein-coupled GABA receptor activity (GO:0004965)	1	4.36E-03	GABBR2
exopolyphosphatase activity (GO:0004309)	1	6.53E-03	PRUNE2
DNA topoisomerase binding (GO:0044547)	1	6.53E-03	NCAPH
RNA polymerase I core binding (GO:0001042)	1	7.62E-03	TCOF1
connexin binding (GO:0071253)	1	8.70E-03	CXADR
RNA polymerase II CTD heptapeptide repeat phosphatase activity (GO:0008420)	1	8.70E-03	PPM1E
sodium channel inhibitor activity (GO:0019871)	1	8.70E-03	NEDD4L
nucleoside transmembrane transporter activity (GO:0005337)	1	8.70E-03	SLC29A1
FK506 binding (GO:0005528)	1	9.79E-03	FKBP4
macrolide binding (GO:0005527)	1	9.79E-03	FKBP4
nuclear export signal receptor activity (GO:0005049)	1	1.09E-02	CSE1L
protein binding involved in heterotypic cell-cell adhesion (GO:0086080)	1	1.19E-02	CXADR
potassium channel inhibitor activity (GO:0019870)	1	1.30E-02	NEDD4L
ATP-dependent microtubule motor activity, plus-end-directed (GO:0008574)	2	4.24E-04	KIF18A,KIF14
neurexin family protein binding (GO:0042043)	1	1.52E-02	CPE
glucocorticoid receptor binding (GO:0035259)	1	1.73E-02	FKBP4
microtubule plus-end binding (GO:0051010)	1	1.84E-02	KIF18A
tau protein binding (GO:0048156)	1	2.16E-02	FKBP4
nucleocytoplasmic carrier activity (GO:0140142)	1	2.16E-02	CSE1L
hyaluronic acid binding (GO:0005540)	1	2.16E-02	HMMR
ATP-dependent microtubule motor activity (GO:1990939)	2	1.15E-03	KIF18A,KIF14
GABA receptor activity (GO:0016917)	1	2.48E-02	GABBR2
protein binding involved in protein folding (GO:0044183)	1	2.59E-02	CLGN
14-3-3 protein binding (GO:0071889)	1	3.01E-02	KIF13B

ATPase activator activity (GO:0001671)	1	3.01E-02	NCAPH
metallocarboxypeptidase activity (GO:0004181)	1	3.12E-02	CPE
microtubule motor activity (GO:0003777)	3	1.21E-04	KIF18A,KIF14,KIF13B
cyclin-dependent protein serine/threonine kinase activity (GO:0004693)	1	3.54E-02	MOK
cyclin-dependent protein kinase activity (GO:0097472)	1	3.65E-02	MOK
sodium channel regulator activity (GO:0017080)	1	4.17E-02	NEDD4L
nucleobase-containing compound transmembrane transporter activity (GO:0015932)	1	4.17E-02	SLC29A1
Ran GTPase binding (GO:0008536)	1	4.17E-02	CSE1L
ion channel inhibitor activity (GO:0008200)	1	4.17E-02	NEDD4L
peptidase activator activity (GO:0016504)	1	4.28E-02	PCOLCE2
RNA polymerase core enzyme binding (GO:0043175)	1	4.28E-02	TCOF1
channel inhibitor activity (GO:0016248)	1	4.28E-02	NEDD4L
protein serine/threonine phosphatase activity (GO:0004722)	2	3.42E-03	PPM1E,PPM1E
carbohydrate derivative transmembrane transporter activity (GO:1901505)	1	4.38E-02	SLC29A1
peptidyl-prolyl cis-trans isomerase activity (GO:0003755)	1	4.49E-02	FKBP4
G protein-coupled neurotransmitter receptor activity (GO:0099528)	1	4.49E-02	GABBR2
molecular carrier activity (GO:0140104)	1	4.59E-02	CSE1L
ATPase regulator activity (GO:0060590)	1	4.80E-02	NCAPH
carboxypeptidase activity (GO:0004180)	1	4.91E-02	CPE
cis-trans isomerase activity (GO:0016859)	1	4.91E-02	FKBP4
PDZ domain binding (GO:0030165)	2	4.69E-03	KIF14,CXADR
motor activity (GO:0003774)	3	5.01E-04	KIF18A,KIF14,KIF13B
microtubule binding (GO:0008017)	3	2.24E-03	KIF18A,KIF14,KIF13B
tubulin binding (GO:0015631)	4	4.33E-04	STMN3,KIF18A, KIF14,KIF13B
phosphoprotein phosphatase activity (GO:0004721)	2	1.75E-02	PPM1E,PPM1E
glycosaminoglycan binding (GO:0005539)	2	2.44E-02	HMMR,PCOLCE2
phosphatase activity (GO:0016791)	2	3.71E-02	PPM1E,PPM1E
ATPase activity (GO:0016887)	3	1.29E-02	KIF18A,KIF14,KIF13B
cytoskeletal protein binding (GO:0008092)	5	3.04E-03	STMN3,KIF18A, KIF14,KIF13B,FKBP4
pyrophosphatase activity (GO:0016462)	4	1.63E-02	KIF18A,KIF14,KIF13B,PRUNE2
hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides (GO:0016818)	4	1.65E-02	KIF18A,KIF14,KIF13B,PRUNE2
hydrolase activity, acting on acid anhydrides (GO:0016817)	4	1.65E-02	KIF18A,KIF14,KIF13B,PRUNE2
protein domain specific binding (GO:0019904)	3	3.92E-02	STMN3,KIF14,CXADR
ATP binding (GO:0005524)	5	2.11E-02	MOK, KIF18A,KIF14,FKBP4,KIF13B
adenyl ribonucleotide binding (GO:0032559)	5	2.43E-02	MOK, KIF18A,KIF14,FKBP4,KIF13B
adenyl nucleotide binding (GO:0030554)	5	2.51E-02	MOK, KIF18A,KIF14,FKBP4,KIF13B

carbohydrate derivative binding (GO:0097367)	7	8.04E-03	MOK, KIF18A,KIF14,FKBP4,KIF13B, HMMR, PCOLCE2
drug binding (GO:0008144)	5	3.80E-02	MOK,KIF18A, KIF14,FKBP4,KIF13B
purine ribonucleoside triphosphate binding (GO:0035639)	5	4.65E-02	MOK,KIF18A, KIF14,FKBP4,KIF13B
hydrolase activity (GO:0016787)	7	2.05E-02	KIF18A,KIF14,PPM1E,KIF13B,PPM 1E,PRUNE2,CPE
catalytic activity, acting on a protein (GO:0140096)	6	3.89E-02	MOK,FKBP4,PPM1E,PPM1E,NEDD 4L,CPE
anion binding (GO:0043168)	7	2.54E-02	MOK,HMMR, KIF18A,KIF14, PCOLCE2,FKBP4,KIF13B
ion binding (GO:0043167)	12	2.29E-02	MOK,HMMR, KIF18A,KIF14, PCOLCE2,FKBP4,KIF13B, PPM1E,CLGN,PRUNE2,CPE,PPM1E
binding (GO:0005488)	21	3.67E-02	MOK,HMMR, KIF18A,KIF14, PCOLCE2,FKBP4,KIF13B, PPM1E,CLGN,PRUNE2,CPE,PPM1E, STMN3,LDOC1,NCAPH,GABBR2,T COF1,CSE1L,NEDD4L,CD83,CXAD R

Table S2: Overall enriched gene list distributed among the various molecular functions. The gene symbols of the genes involved in the different biological process are also indicated along with the p-value. The table shows the result based on the p-value 0.05.

GO biological process complete	Gene involved	P-value	Gene symbol
microtubule depolymerization (GO:0007019)	2	1.93E-04	STMN3, KIF18A
protein depolymerization (GO:0051261)	2	7.42E-04	STMN3, KIF18A
mitotic metaphase plate congression (GO:0007080)	2	1.10E-03	KIF18A, KIF14
microtubule polymerization or depolymerization (GO:0031109)	2	1.52E-03	STMN3, KIF18A
metaphase plate congression (GO:0051310)	2	1.69E-03	KIF18A, KIF14
mitotic sister chromatid segregation (GO:0000070)	3	2.27E-04	KIF18A, KIF14, NCAPH
establishment of chromosome localization (GO:0051303)	2	2.94E-03	KIF18A, KIF14
chromosome localization (GO:0050000)	2	3.01E-03	KIF18A, KIF14
regulation of microtubule polymerization or depolymerization (GO:0031110)	2	3.34E-03	STMN3, FKBP4
sister chromatid segregation (GO:0000819)	3	4.33E-04	KIF18A, KIF14, NCAPH
mitotic nuclear division (GO:0140014)	3	5.11E-04	KIF18A, KIF14, NCAPH

regulation of microtubule cytoskeleton organization (GO:0070507)	3	1.04E-03	STMN3, KIF18A, FKBP4
negative regulation of protein complex assembly (GO:0031333)	2	9.58E-03	KIF14, FKBP4
cellular protein complex disassembly (GO:0043624)	2	9.71E-03	STMN3, KIF18A
nuclear chromosome segregation (GO:0098813)	3	1.58E-03	KIF18A, KIF14, NCAPH
regulation of microtubule-based process (GO:0032886)	3	1.62E-03	STMN3, KIF18A, FKBP4
meiotic nuclear division (GO:0140013)	2	1.20E-02	KIF18A, NCAPH
regulation of G2/M transition of the mitotic cell cycle (GO:0010389)	2	1.21E-02	HMMR, KIF14
male sex differentiation (GO:0046661)	2	1.25E-02	KIF18A, FKBP4
viral life cycle (GO:0019058)	2	1.45E-02	NEDD4L, CXADR
regulation of cell cycle G2/M phase transition (GO:1902749)	2	1.47E-02	HMMR, KIF14
meiotic cell cycle process (GO:1903046)	2	1.47E-02	KIF18A, NCAPH
microtubule-based movement (GO:0007018)	3	2.62E-03	KIF18A, KIF14, KIF13B
chromosome segregation (GO:0007059)	3	2.89E-03	KIF18A, KIF14, NCAPH
nuclear division (GO:0000280)	3	3.45E-03	KIF18A, KIF14, NCAPH
female pregnancy (GO:0007565)	2	1.77E-02	LDOC1, FKBP4
organelle fission (GO:0048285)	3	4.37E-03	KIF18A, KIF14, NCAPH
protein dephosphorylation (GO:0006470)	2	2.10E-02	PPM1E, PPM1E
multi-multicellular organism process (GO:0044706)	2	2.38E-02	LDOC1, FKBP4
protein folding (GO:0006457)	2	2.42E-02	FKBP4, CLGN
T cell activation (GO:0042110)	2	2.44E-02	KIF13B, CXADR
regulation of Ras protein signal transduction (GO:0046578)	2	2.50E-02	STMN3, KIF14
the cellular process involved in reproduction in a multicellular organism (GO:0022412)	3	5.81E-03	KIF18A, NCAPH, CXADR
meiotic cell cycle (GO:0051321)	2	2.52E-02	KIF18A, NCAPH
protein-containing complex disassembly (GO:0032984)	2	2.54E-02	STMN3, KIF18A
regulation of cytoskeleton organization (GO:0051493)	4	2.22E-03	STMN3, KIF18A, FKBP4, PPM1E
sex differentiation (GO:0007548)	2	3.28E-02	KIF18A, FKBP4
reproductive structure development (GO:0048608)	3	1.02E-02	KIF18A, LDOC1, FKBP4

regulation of membrane potential (GO:0042391)	3	1.04E-02	SLC29A1, NEDD4L, CXADR
reproductive system development (GO:0061458)	3	1.04E-02	KIF18A, LDOC1, FKBP4
proteasome-mediated ubiquitin-dependent protein catabolic process (GO:0043161)	2	4.00E-02	KIF14, NEDD4L
regulation of cell morphogenesis involved in differentiation (GO:0010769)	2	4.18E-02	KIF13B, NEDD4L
dephosphorylation (GO:0016311)	2	4.25E-02	PPM1E, PPM1E
proteasomal protein catabolic process (GO:0010498)	2	4.64E-02	KIF14, NEDD4L
microtubule-based process (GO:0007017)	4	4.72E-03	STMN3, KIF18A, KIF14, KIF13B
regulation of neuron projection development (GO:0010975)	3	1.48E-02	FKBP4, KIF13B, NEDD4L
multi-organism reproductive process (GO:0044703)	6	4.88E-04	KIF18A, LDOC1, FKBP4, NCAPH, CLGN, CXADR
regulation of small GTPase mediated signal transduction (GO:0051056)	2	4.88E-02	STMN3, KIF14
cellular response to drug (GO:0035690)	2	4.96E-02	LDOC1, PPM1E
cellular response to lipid (GO:0071396)	3	1.87E-02	KIF18A, LDOC1, FKBP4
mitotic cell cycle process (GO:1903047)	3	2.60E-02	KIF18A, KIF14, NCAPH
multicellular organismal reproductive process (GO:0048609)	4	1.04E-02	KIF18A, LDOC1, NCAPH, CXADR
sexual reproduction (GO:0019953)	4	1.05E-02	KIF18A, NCAPH, CXADR, CLGN
multicellular organism reproduction (GO:0032504)	4	1.10E-02	KIF18A, LDOC1, NCAPH, CXADR
regulation of neuron differentiation (GO:0045664)	3	3.07E-02	FKBP4, KIF13B, NEDD4L
regulation of plasma membrane-bounded cell projection organization (GO:0120035)	3	3.18E-02	FKBP4, KIF13B, NEDD4L
regulation of cell projection organization (GO:0031344)	3	3.29E-02	FKBP4, KIF13B, NEDD4L
developmental process involved in reproduction (GO:0003006)	3	3.34E-02	KIF18A, LDOC1, FKBP4
gamete generation (GO:0007276)	3	3.47E-02	KIF18A, NCAPH, CXADR
regulation of nervous system development (GO:0051960)	4	1.46E-02	KIF14, FKBP4, KIF13B, NEDD4L
regulation of cell development (GO:0060284)	4	1.57E-02	KIF14, FKBP4, KIF13B, NEDD4L
mitotic cell cycle (GO:0000278)	3	3.71E-02	KIF18A, KIF14, NCAPH
reproductive process (GO:0022414)	6	3.11E-03	KIF18A, LDOC1, NCAPH, FKBP4, CLGN, CXADR
reproduction (GO:0000003)	6	3.14E-03	KIF18A, LDOC1, NCAPH, FKBP4, CLGN, CXADR

movement of cell or subcellular component (GO:0006928)	5	1.94E-02	KIF18A, KIF14, KIF13B, NEDD4L, CXADR
regulation of organelle organization (GO:0033043)	4	4.37E-02	STMN3, KIF18A, FKBP4, PPM1E
multi-organism process (GO:0051704)	7	1.07E-02	KIF18A, LDOC1, NCAPH, FKBP4, CLGN, NEDD4L, CXADR
regulation of cell differentiation (GO:0045595)	5	3.56E-02	KIF14,FKBP4,KIF13B,NEDD4L,CD83
regulation of cellular component organization (GO:0051128)	7	1.32E-02	STMN3, KIF18A, KIF14, FKBP4, KIF13B, PPM1E, NEDD4L
establishment of localization in cell (GO:0051649)	5	3.99E-02	KIF18A, SLC29A1, KIF14, KIF13B, CSE1L
regulation of catalytic activity (GO:0050790)	6	3.02E-02	STMN3, KIF14, PCOLCE2, NCAPH, GABBR2, PPM1E
animal organ development (GO:0048513)	8	1.33E-02	KIF18A, SLC29A1, LDOC1, KIF14, FKBP4, TCOF1, CPE, CXADR
system development (GO:0048731)	11	2.84E-03	STMN3, KIF14, PCOLCE2, NCAPH, GABBR2, PPM1E, KIF18A, SLC29A1, LDOC1, FKBP4, KIF13B, TCOF1, NEDD4L, CPE, CXADR
cellular response to chemical stimulus (GO:0070887)	7	2.32E-02	KIF18A, SLC29A1, LDOC1, FKBP4, PPM1E, CXADR, PCOLCE2
cellular localization (GO:0051641)	6	3.73E-02	KIF18A, SLC29A1, KIF14, KIF13B, CSEIL, CPE
regulation of molecular function (GO:0065009)	8	1.51E-02	STMN3, LDOC1, KIF14, PCOLCE2, NCAPH, GABBR2, PPM1E, NEDD4L
multicellular organism development (GO:0007275)	11	1.02E-02	STMN3, KIF14, SLC29A1, CPE, KIF18A, LDOC1, FKBP4, KIF13B, TCOF1, NEDD4L, CPE, CXADR
anatomical structure development (GO:0048856)	11	1.50E-02	STMN3, KIF14, SLC29A1, CPE, KIF18A, LDOC1, FKBP4, KIF13B, TCOF1, NEDD4L, CPE, CXADR
response to chemical (GO:0042221)	9	3.58E-02	KIF18A, SLC29A1, LDOC1, PCOLCE2, FKBP4, PPM1E, NEDD4L, CD83, CXADR
developmental process (GO:0032502)	11	2.96E-02	STMN3, KIF14, SLC29A1, CPE, KIF18A, LDOC1, FKBP4, KIF13B, TCOF1, NEDD4L, CPE, CXADR
cellular response to stimulus (GO:0051716)	12	3.76E-02	MOK, KIF18A, SLC29A1, LDOC1, PCOLCE2, FKBP4, GABBR2, KIF13B, PPM1E, CD83, CPE, CXADR
multicellular organismal process (GO:0032501)	12	4.45E-02	STMN3, KIF18A, SLC29A1, LDOC1, KIF14, NCAPH, FKBP4, KIF13B, TCOF1, NEDD4L, CPE, CXADR
regulation of cellular process (GO:0050794)	17	3.56E-02	MOK, KIF18A, SLC29A1, LDOC1, PCOLCE2, FKBP4, GABBR2, KIF13B, PPM1E, CD83, CPE, CXADR, HMMR, STMN3, KIF14, NEDD4L, TCOF1
cellular process (GO:0009987)	22	8.95E-03	CD83 CLGN CPE CSE1L CXADR FAM84A FKBP4 GABBR2 HMMR KIF13B KIF14 KIF18A LDOC1 MOK NCAPH NEDD4L PCOLCE2 PPM1E PRUNE2 SLC29A1 STMN3 TCOF1

Table S3: Overall enriched gene list distributed among the various biological pathways. The gene symbols of the genes involved in the different biological process are also indicated along with the p-value ≤0.05.