STUDY ON COLORECTAL NEOPLASIA GENE EXPRESSION AND THEIR POSSIBLE ROLE IN RENAL CELL CARCINOMA

[1]Diksha Marwaha, [2]Ranjit Singh
[1]*Centre for Systems Biology and Bioinformatics, Panjab University, Chandigarh, India*
[2]*CGC Technical Campus, Jhanjeri, Mohali, Kharar, India*

Abstract Metastasis or secondary tumor, being advanced level of cancer is still incurable and unknown in some cancer types in terms of differential gene expression profiles. Therefore, the study and rational behind investigation was to identify group of differentially expressed (upregulated and downregulated) genes responsible for colorectal neoplasia to identify for purpose of biomarkers still unknown directly correlated with metastasis in kidney on basis of LogFC values (calculated in R platform) in CanEVOLVE tool to find DEGs in both diseases in different datasets, biological process annotation, functional enrichment analysis and false negative rates. All genes were collected from databases like MalaCards, CoReCG and renal cancer gene database. The connection between diseases was confirmed in Disease Connect database and relationship of genes with disease in GeneWays database. At the end, it was recognized that gene mutation was major cause of metastasis in kidney on basis of biological process annotation done in FIDEA tool and functional enrichment analysis in g:Profiler. Lastly, investigation was validated on basis of false discovery rate of functionally enriched biological pathways in STRING tool.

Keywords: Neoplasia, colorectal, renal cell carcinoma, metastasis, Systems Biology, differential expressed genes, biological process annotation.

I. INTRODUCTION

Cancer is an extremely complex, heterogeneous disease which could display a degree of complexity at the physiological, tissue and cellular levels. A cancer tissue or tumor often contain several distinct pathological featured cancer subtypes which can be recognized as cancer tissue complexity[1]. The co-existence of several cancer cells subtypes rely on the activation of different signaling pathways in one specific tumor which represent tissue complexity of cancers. Cancer have characteristics of uncontrolled cell growth, ability to invade surrounding tissues and finally to generate metastasis in distinct body parts. The cancer cells not only spread through blood stream but sometimes genes performing biological function in cells or tissues develop error due to genetic and epigenetic networks or uncontrolled division of cells in specific organ lead to stage called neoplasia. It occurs due to thousands of genes often interlinked with other organs and therefore tumor spreads from one organ to other organ and results in metastasis. Cancer cells can spread from one part to other part of body and develop into new tumor called metastasis or secondary tumor. Analysis of differential gene expression profiles in neoplasm or cancers to identify biomarkers are directly correlated with metastasis. This method was used in this research to study gene expression profile of colorectal neoplasia on basis of biological process annotation and signaling pathways due to phenotype mutation. In this review, the study of colorectal neoplasia in kidney neoplasia on basis of differential gene expression is hypothesized to identify if oncogene or group of oncogene responsible for colorectal neoplasia develop secondary tumor, to understand the ways in which differential gene expression profiles are correlated with metastasis and how biological process annotation and functional enrichment analysis could be helpful in outcome of biomarkers for diagnosis of neoplasia. Sub network biomarkers for cancer are always more reproducible then individual marker genes that are selected without network information, pathway enrichment analysis and annotation. In this approach, cancer microarray datasets of colorectal neoplasia and kidney cancer were chosen to study presence of gene list in both datasets and further upregulated and downregulated genes were analyzed on basis of biological process annotation and interaction network. This initial proposal will proceed towards with confirmation that cancer related genes not only function as network hub protein in many cellular processes but also upshot secondary cancer due to phenotypic mutation.

II. MATERIALS AND METHODS

1. **Data of cancer genes and phenotype neoplasm retrieval**
   To view relationship between colorectal neoplasia and renal cell carcinoma, DieaseConnect (available online: http://disease-connect.org) database [2] was preferred which is web server for analysis of diseases and genetic network visualization criteria. This procedure was considerate in terms of creating novel strategies for formulation of effective drugs and therapies for both diseases lead to similar treatment design. Neoplasm was similarly viewed in GeneWays (https:omictools.com) to confirm relationship between disease and the genes. GeneWays database provided detailed about any gene

query and to further know about co-regulator genes. The main goal of panorama was to understand relationship between neoplasm leading to carcinoma and genes at fault behind colorectal cancer and kidney cancer. Correspondingly, MalaCard database which mines and merges 44 data sources to generate 16919 human diseases and provide detailed data of both diseases and their correlation with respective genes [3]. Consequently, colorectal neoplasm, kidney cancer related genes were retrieved from MalaCard database.

2. Validation of cancer allied genes from corresponding databases

Colorectal Cancer Gene database (CoReCG) (Ims.snu.edu.in) [4] and renal cancer gene database (www.juit.ac.in) [5] were chosen considered to be single point source for confirmation of presence of oncogenes involved for the improvement of treatment and diagnosis of diseases. CoReCG database depicted information of colorectal neoplasia causing genes for development of biomarkers for early detection of colorectal cancer. All oncogenes associated with Renal cell carcinoma arise from renal tumor, were gained from omics technique and it is repository of all protein coding and non-coding genes that could efficiently provide with an estimate of different causes of occurrence of renal carcinoma including overexpression of oncogenes, downregulation or mutation.

3. Differential expression of desired genes under Cancer datasets

To analyze the overexpression and downregulation of oncogenes, CanEVOLVE tool has been used available online at (https://www.canevolve.org) [6]. Under options of primary analysis, Differential gene expression was chosen. After colorectal cancer dataset picked and 33 oncogenes (colorectal neoplasia and kidney cancer separate and common) under the "CANCER TYPE", the option "get results" was opted. Likewise, kidney cancer dataset was selected. For colorectal cancer, GEO series GSE5364 and for kidney cancer, GEO series GSE6344 were considered which are inbuilt in the CanEVOLVE tool. In the colorectal cancer series, there were 18 samples (9 colon cancer and 9 normal). On the other hand, within kidney cancer series, total samples were 20 (10 kidney cancer and 10 normal). All the query oncogenes were found in probeset ID and different statistical values including LogFC, p values or B values. Out of all the statistical terms, only LogFC was considered significant for the research hypothesis since the event was about expression of the different oncogenes in both diseases. LogFC (log fold change) which is ratio of normal condition verses reference condition in terms of expressed genes in different regions of body. Other values (t-values and p values not used further) were insignificant in statistical analysis for our hypothesis. The t-values were neglected and p-values(probability values) were to get more extreme values of no use). All statistical values were calculated in R platform. Total 80 probes or genes with LogFC values in colorectal neoplasia and 82 probes with LogFC values in renal cell carcinoma were separated. Both the diseases with different positive and negative values were set apart as upregulated and downregulated genes in both the diseases distinctly. Additionally, the probes with most upregulated (highest LogFC values) and the probes with most downregulated (lowest LogFC values) were enlisted in different categories in both diseases. The difference among LogFC values was calculated automatically and analyzed on basis of bar graphs. Large difference in LogFC values, genes expressing were rejected.

Only little or no difference in their LogFC values were significant statistically specifically if LogFC<0.01.

4. Analysis of eventual oncogenes expression and affirmation in metastasis

Resulting upregulated genes were investigated in online tools Functional Interpretation of Differential Expressed Analysis (FIDEA) [7], g:Profiler [8] and String [9]. In FIDEA tool, Genes were analyzed from static word cloud and interactive table under KEGG, interpro, gene ontology. g:Profiler was web server in which resulting genes were submitted that provided results graphically or in textual format. Gene ontology evidences were colored which depicted strength of evidence between gene and biological process and functional enrichment analysis. Final upregulated genes of colorectal neoplasia that showed expression in kidney cancer prediction was done on basis of false discovery rate which should be either less than 0.01 or 0.05 and considered statistically significant to be enriched in biological process. The endmost analysis was performed in form of bar graph for resulting gene between biological process annotation and false discovery rate. Therefore, the gene predicted with LogFC between 0.1 and 0.5 which indicated 50% increase in fold change or more than 50 times the gene overexpressed with very little difference in LogFC values in both colorectal neoplasia and renal cell carcinoma and therefore the gene not only responsible for neoplasia but as for the expression in kidney and indicated possibility of metastasis.

III. RESULTS

1. OMIM data retrieval of genes

OMIM genes of renal cell carcinoma and colorectal cancer were retrieved from Disease-Connect Database for disease-disease connection. Phenotype neoplasm was retrieved from Gene Ways to view disease-gene relationship. Phenotype Neoplasm along with affiliating genes responsible for colorectal cancer selected from Mala card Database for

further analysis in renal cell carcinoma. Similarly, genes list of renal cell carcinoma was received through same database. Finally, 33 genes were retrieved from their disease databases Colorectal Cancer Gene Database (CoReCG) and Renal Cancer Database (RCDB) to validate f these diseases were somehow related to each other on basis of differentially expressed genes. There were 10 genes common out of 33 genes when the data with information of genes was retrieved from their specific databases.

TABLE 1: LIST OF COLORECTAL NEOPLASIA AND KIDNEY NEOPLASIA GENES RECEIVED FROM DISEASE-CONNECT, GENE-WAYS AND MALA-CARD DATABASE

Colorectal neoplasia genes	Kidney neoplasia genes
APC (Adenomatosis polyposis coli) EPCAM(Tumor associated calcium signal transducer 1) TYMS(Thymidylate synthetase) MLH1(MutL homolog 1, colon cancer, nonpolyposis type 2) MLH3(MutL homolog 3) MSH2(MutS homolog 2, colon cancer, nonpolyposis type 1) MSH3(MutS homolog 3) MSH6(MutS homolog 6) MUTYH(mutY DNA glycolase) PMS2(PMS1 homolog 2) STK11(Serine Threonine kinase 11) KRAS(Kirsten rat sarcoma) PIK3CA(phosphatidylinositol 3-kinase)	DICER1(dicer1Rribonuclease III) FLCN(Folliculin) KIT(Tyrosine protein kinase kit) TSC1(Tuberous sclerosis 1) TSC2(Tuberous sclerosis complex 2) AMACR(Alpha methylacyl –coA racemase) PAX2(Paired box 2) MME(Membrane Metalloendopeptidase) CDH16(Cadherin 16) PAX8(Paired box 8)

TABLE 2: LIST OF GENES THAT WERE COMMON IN BOTH COLORECTAL AND KIDNEY NEOPLASIA

SOX9	Sex determining region Y)- box 9
SFRP1	Secreted Frizzled Related Protein 1
VHL	Von Hippel-Lindau tumor suppressor
ATM	Ataxia telangiectasia mutated
IL2	Interleukin 2V
VEGFA	Vascular endothelial growth factor A
MTOR	Mechanistic target of rapamycin serine kinase
TP53	Tumor protein
MET	MET proto oncogene
KDR	Kinase insert domain receptor

2. Differential Gene Expression Analysis

Cancer dataset was selected in canEVOLVE tool in which all 33 genes were submitted and for each disease, a specific cancer dataset was chosen for analysis of presence of genes and differential expression of genes. The first GEO series was GSE5364 with 18 samples of colon cancer in which there were 9 normal samples and 9 colon tumor samples. Similarly, another GEO series GSE6344 was selected from cancer type kidney with 20 samples out of which there were 10 normal samples and 10 samples of Kidney tumor. In the result, the output was Heatmap, probeset of 33 genes with statistical values that included logFC, t-value, p value, Adj. p value, B value for separate datasets. Out of all the statistical values, only logFC values were considered. Log fold change is ratio of expression value (treatment condition verses reference condition).If logFC value is zero, it means there was no change in the gene expression. If logFC value

is positive, it means genes are upregulated whereas if the logFC value is negative, it means genes are downregulated. All the values were calculated in R platform. Since, one unit of logFC translate to two fold change in expression and therefore, t-test, p value and B-value was neglected which is also similar to logFC.

gene symbol	logFC
MLH1	0.0155
PAX8	0.0468
MLH3	0.055
DICER1	0.0731
PAX8	0.1002
BAX	0.1149
PAX8	0.119
MTOR	0.1311
DICER1	0.1627
DICER1	0.213
MSH6	0.2251
MTOR	0.2361

Fig 1 differentially expressed genes (DEG) of colorectal neoplasia on basis of logFC value. All the values are near 0.05 to 0.1 for all the resulting genes. DEG identified with logFC <1 in canEVOLVE tool available at (http://www.canevolve.org/) are statistically significant that were calculated in R platform.

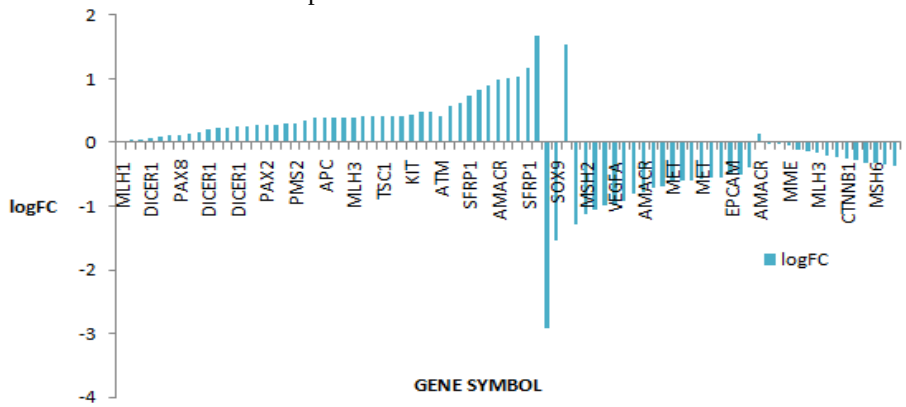

Fig 2: 2D graph plotted for gene symbol at horizontal axis and logFC values at vertical axis representing differentially expressed genes for colorectal neoplasia. *SFRP1* was most upregulated gene whereas *SOX9* was the most downregulated gene. This statistically significant graph was plotted in MS excel.

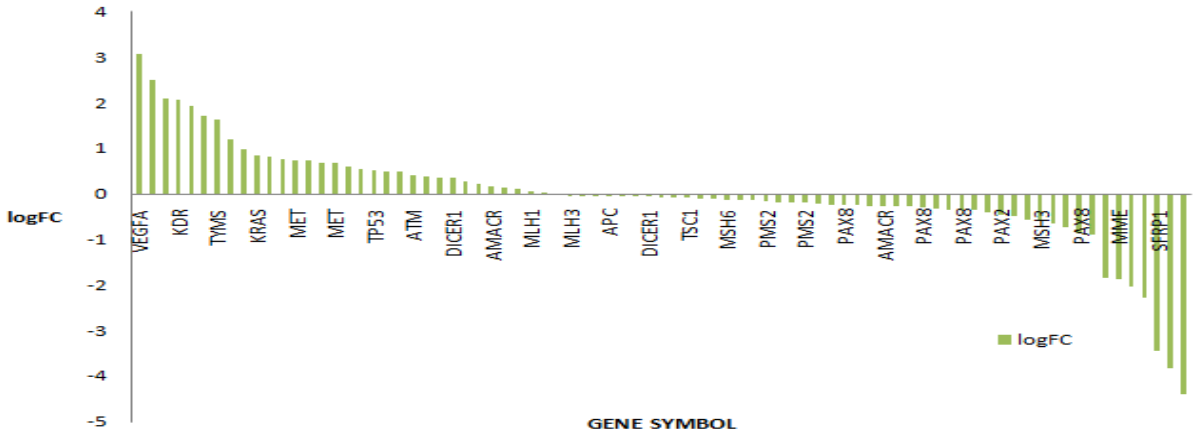

Fig 3: 2D graph plotted for gene symbol against logFC values for renal cell carcinoma. *VEGFA* gene was most overexpressed gene among all whereas *SFRP1* was significantly underexpressed gene. LogFC for DEG from kidney GEO dataset in canEVOLVE tool was selected for primary analysis.

1	genes symb	gene name	colorectal neoplasia logFC	kidney neoplasia logFC
2	MSH6	MutS homolog 6	0.2251	0.5353
3	APC	Adenomatosis polyposis coli	0.4746	0.4846
4	ATM	Ataxia telangiectasia mutated	0.4052	0.4221
5	DICER1	dicer 1,ribonuclease type III	0.3453	0.3457
6	AMACR	alpha-methyacyl-coA racemase	0.2177	0.1768
7	MLH3	MutL homolog 3	0.3956	0.1268

Fig 4: Common overexpressed genes in both colorectal neoplasia and kidney neoplasia were considered. Values near to zero were appraised significant such that logFC value of *APC* gene in both colorectal and kidney neoplasia were 0.4746 and 0.4846 respectively with little difference in their overexpression. Similarly, for gene *ATM*, logFC values were 0.4052 and 0.4221. *DICER1* gene has value 0.3453 in colorectal and 0.3457 in kidney with little difference in logFC values. some downregulated genes common in both carcinomas (*EPCAM*, *CDH16*, *PAX8*) were rejected due to difference in their negative logFC values which revealed that downregulated genes were not statistically significant in terms of logFC<1 and logFC difference between two carcinomas cutoff which should not exceed such as in *MLH3* (0.39 in colorectal and 0.12 in kidney neoplasia which was quite large).

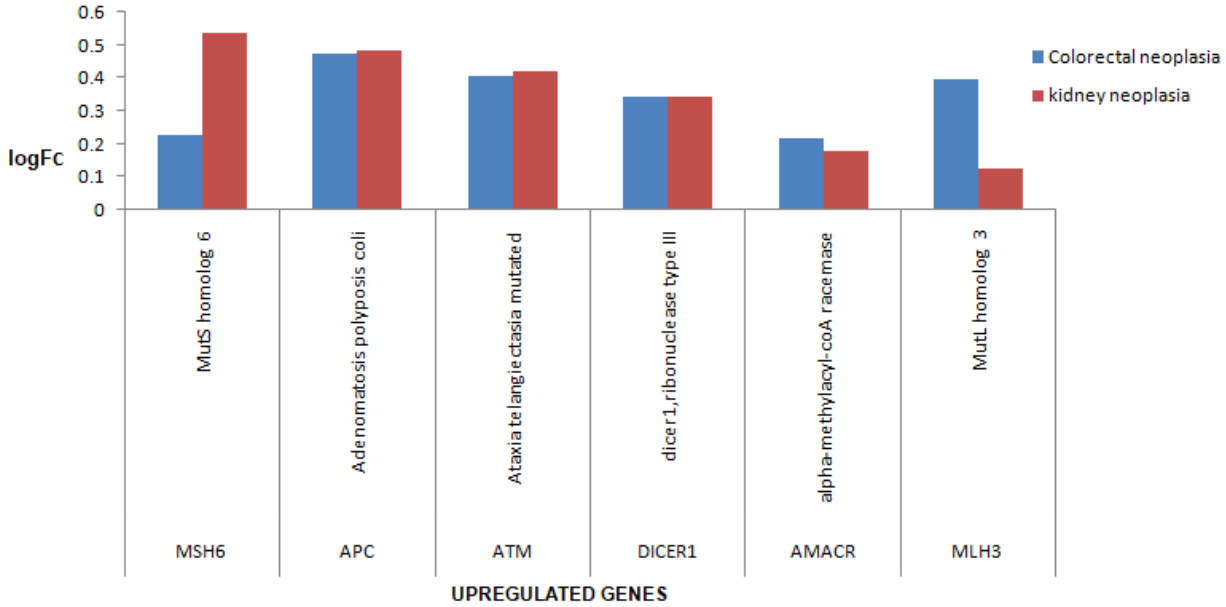

Fig 5: Result of plotted comparison graph for 3 common upregulated genes depicting close expression pattern in both phenotypes. *DICER1* has given very close expression pattern followed by *ATM* gene and then *APC* gene. Blue bar graph indicate gene of colorectal neoplasia logFC value and red bar graph indicate kidney neoplasia genes logFC values. *ATM* and *APC* genes were overexpressed in kidney neoplasia compared to colorectal neoplasia.

3. Gene Annotation for Upregulated Genes

Query genes were analyzed in FIDEA tool for functional enrichment analysis. Results of annotation were given by upregulated genes in gene ontology biological process only. Cellular component and molecular function gene ontology function of resulting genes were statistically insignificant. Outcomes of upregulated genes for gene ontology were in terms of static word cloud and interactive table. Word cloud analysis of biological process for three upregulated genes revealed statistical significant result with p-values in which "Mitotic spindle assembly checkpoint" was significant which considered to be enriched with one of the upregulated DEGs. In interactive table analysis of biological process annotation, p-value of mitotic spindle assembly checkpoint was 1.10e-5 that was more than rest of processes and further *APC* and *ATM* gene were predicted to be involved in metastasis of kidney neoplasia.

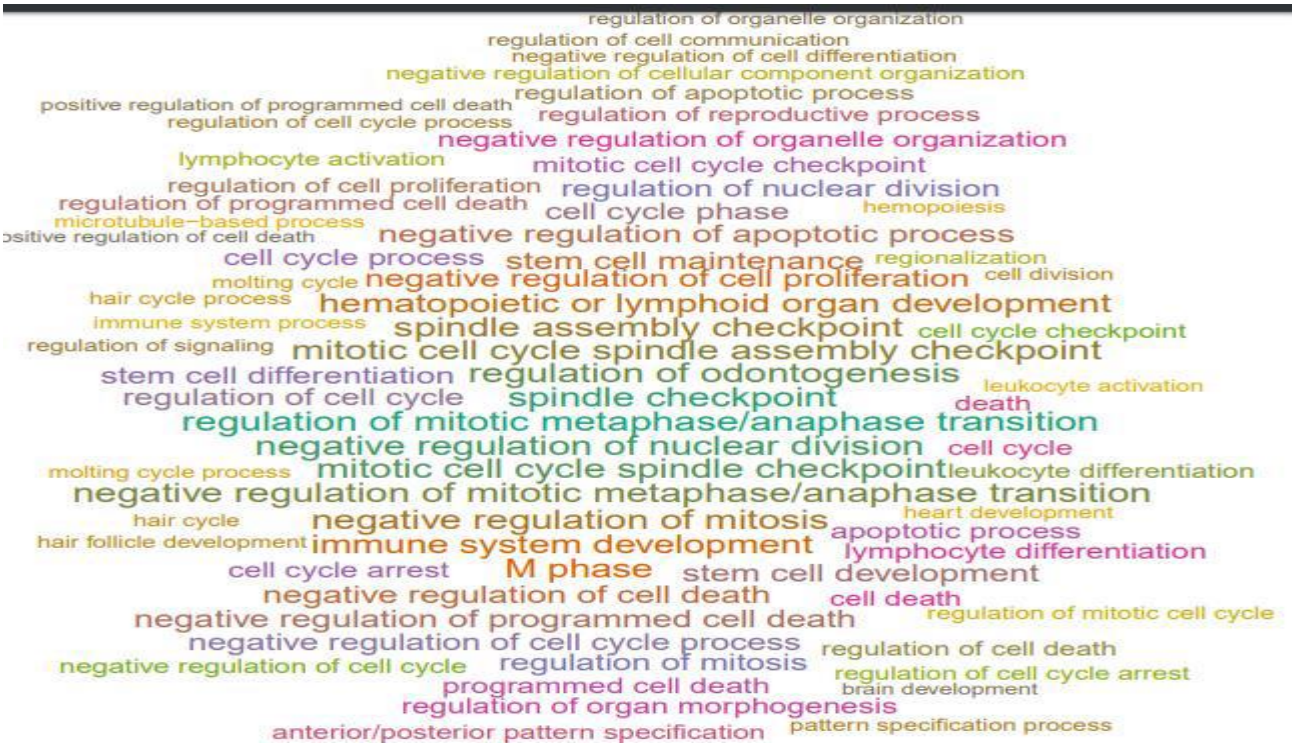

Fig 6: Word cloud analysis of gene ontology biological process. Larger the letters, smaller the p-value, significant is the result. In this fig, mitotic spindle assembly checkpoint was in largest letter among all processes followed by hematopoietic organ development, regulation of odontogenesis, and cell cycle arrest in larger letter.

ID Category	Category Name	Benjamini	P-Value	Fold Enrichment	N. Genes
GO:0030071	regulation of mitotic metaphase/anaphase transition	2.21e-3	2.01e-5	254.137	2
GO:0031577	spindle checkpoint	2.21e-3	1.52e-5	291.51	2
GO:0051784	negative regulation of nuclear division	2.21e-3	1.81e-5	267.874	2
GO:0071174	mitotic cell cycle spindle checkpoint	2.21e-3	1.34e-5	309.729	2
GO:0042481	regulation of odontogenesis	2.21e-3	7.49e-6	412.972	2
GO:0045841	negative regulation of mitotic metaphase/anaphase transition	2.21e-3	1.26e-5	319.72	2
GO:0007094	mitotic cell cycle spindle assembly checkpoint	2.21e-3	1.10e-5	341.77	2
GO:0071173	spindle assembly checkpoint	2.21e-3	1.18e-5	330.378	2
GO:0045839	negative regulation of mitosis	2.21e-3	1.81e-5	267.874	2
GO:0048534	hematopoietic or lymphoid organ development	2.61e-3	2.63e-5	33.56	3

Fig 7: Interactive table analysis of biological process depicting p-values and number of genes involved in specific category. In mitotic spindle assembly checkpoint, enriched DEG was *ATM* and *APC* gene.

4. Biological Process Annotation

To further validate the involvement of DEGs, interpretation of upregulated genes was performed in g: Profiler tool where cellular component and molecular function ontology were rejected. Therefore, biological process annotation was preferred.

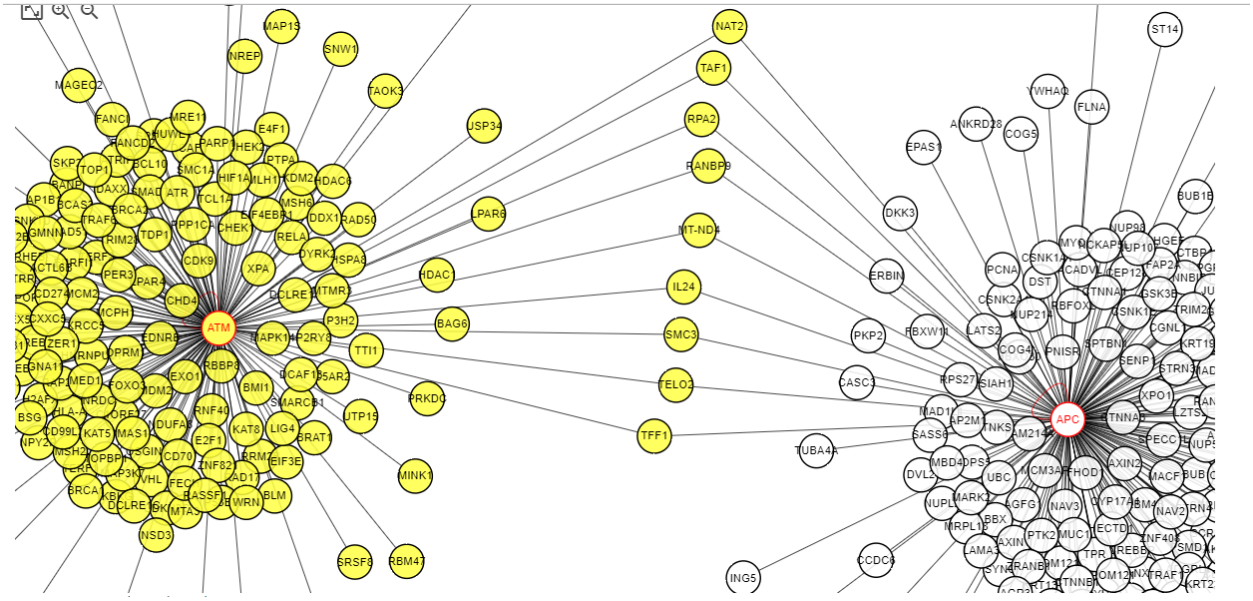

Fig 8: Enriched biological network with genetic interactions from BIOGRID database in g: Profiler tool is illustrated. Red nodes were the query genes (APC and ATM) whereas yellow nodes were selected neighboring genes. 9 neighboring genes were functionally significant in terms of connection of biological processes due to edges of query upregulated genes were associating those neighboring genes.

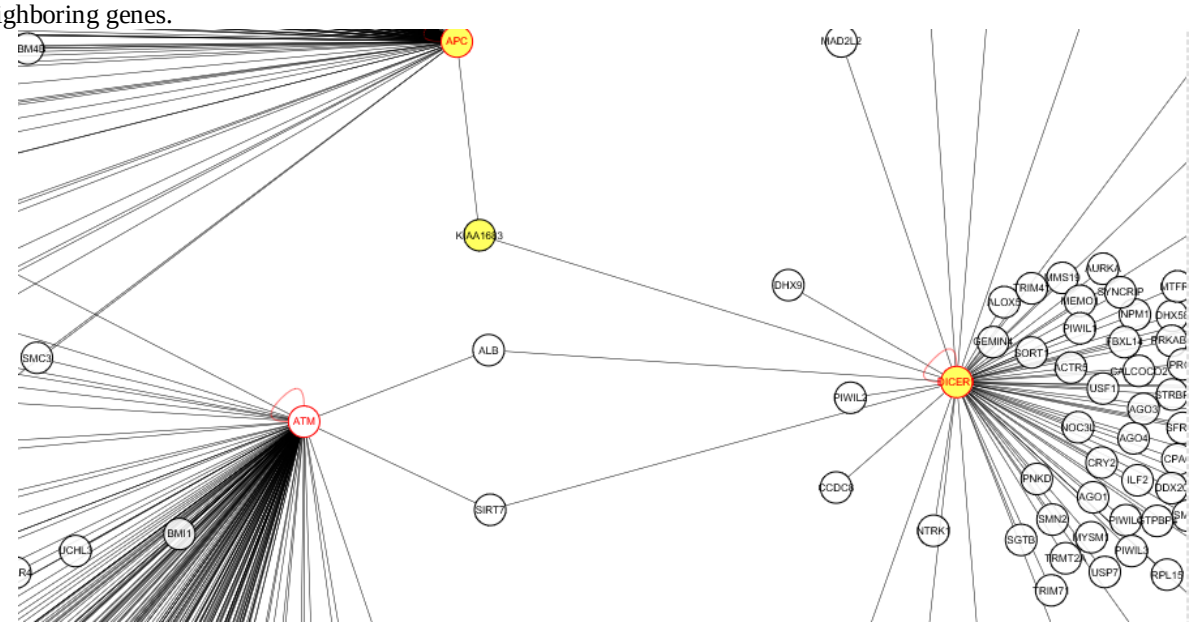

Fig 9: Enriched biological network delineating less number of neighboring yellow gene nodes between ATM and DICER1 (ALB and SIRT7) significantly enriched in biological process. KAA1683 common yellow gene node connecting APC and DICER1 query genes. APC gene hold responsible for colorectal neoplasia (retrieved from CoReCG database) whereas DICER1 influenced Kidney cancer. Genetic Network among these genes represented that ATM gene was significant in developing secondary cancer in view of fact that it was constituted to be common in both carcinomas and functionally enriched in biological processes as well as was mutant phenotype.

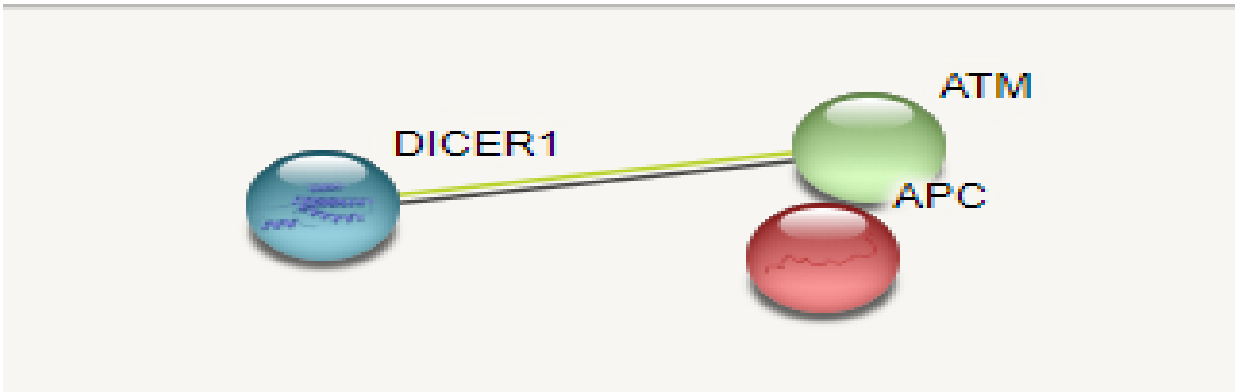

Fig 10: Query Nodes genes network analysis result using STRING online database. APC gene is in red color node which represent query protein, DICER1 in blue color filled node represent its 3D structure and ATM in green empty node revealed that its 3D structure was unknown. Edge between DICER1 and ATM illustrated that they belong to neighborhood genes. Edges manifested protein-protein interactions and protein contribute to biological process. Nodes displayed in different colors were proteins produced by single gene.

Biological Process (GO)			
pathway ID	*pathway description*	*count in gene set*	*false discovery rate*
GO:0007094	mitotic spindle assembly checkpoint	2	0.0132
GO:0042481	regulation of odontogenesis	2	0.0132
GO:0048534	hematopoietic or lymphoid organ development	3	0.0158
GO:0001942	hair follicle development	2	0.0215
GO:0006921	cellular component disassembly involved in execution phase of apoptosis	2	0.0215
GO:0044265	cellular macromolecule catabolic process	3	0.0215
GO:0097194	execution phase of apoptosis	2	0.0215
GO:0098773	skin epidermis development	2	0.0215
GO:0042633	hair cycle	2	0.0236
GO:0060429	epithelium development	3	0.0328
GO:0006915	apoptotic process	3	0.0379
GO:0007050	cell cycle arrest	2	0.0391
GO:0019827	stem cell population maintenance	2	0.0391
			(less ...)
KEGG Pathways			
pathway ID	*pathway description*	*count in gene set*	*false discovery rate*
05206	MicroRNAs in cancer	2	0.0416

Fig 11: Result given by biological processes functional enrichment of resulting upregulated genes on basis of false discovery rate. It was considered statistically significant when false discovery rate <0.05. It revealed significant interactions between query gene nodes and pathways common in these genes. APC and ATM genes were involved in pathway "mitotic spindle assembly checkpoint" with false discovery rate 0.0132. Lesser the discovery rate, more significant was the result. False discovery rate per pathway was decreasing downwards.

colorectal neoplasia	kidney neoplasia	biological process annotation	false discovery rate	term ID	string	((fidea)	(g.profile)
APC,ATM	ATM	mitotic spindle assembly checkpoint	0.0132	GO:0007094	present	Nil	present
APC	DICER1	regulation of odontogenesis	0.0132	GO:0042481	present	present	nil
APC,ATM	ATM,DICER1	hematopoietic or lymphoid organ development	0.0158	GO:0048534	present	present	nil
APC	DICER1	hair follicle development	0.0215	GO:0001942	present	present	nil
APC	DICER1	cellular component disassembly involved in execution phase Apoptos	0.0215	GO:0006921	present	nil	nil
APC,ATM	DICER1,ATM	cellular macromolecule catabolic process	0.0215	GO:0044265	present	nil	nil
APC	DICER1	execution phase of apoptosis	0.0215	GO:0097194	present	nil	present
APC	DICER1	skin epidermis development	0.0215	GO:0098773	present	nil	nil
APC	DICER1	hair cycle	0.0236	GO:0042633	present	present	nil
APC,ATM	DICER1.ATM	epithelium development	0.0328	GO:0060429	present	nil	nil
APC,ATM	DICER1,ATM	apoptotic process	0.0379	GO:0006915	present	nil	nil
APC,ATM	ATM	cell cycle arrest	0.0391	GO:0007050	present	present	nil
APC	DICER1	stem cell population maintenance	0.0391	GO:0019827	present	Nil	nil
ATM	DICER1,ATM	microRNA in cancer	0.0416	KEGG:05206	present	Nil	present

Fig 12: Analysis result of endmost upregulated genes on basis of biological process annotation. ATM gene concluded to be involved in biological processes in both carcinomas.

colorectal neoplasia	kidney neoplasia	biological process annotation for ATM gene	false discovery rate	term ID	STRING	fidea	g:Profiler
APC,ATM	ATM	mitotic spindle assembly checkpoint	0.0132	GO:0007094	present	nil	present
APC,ATM	ATM,DICER1	hematopoietic or lymphoid organ development	0.0158	GO:0048534	present	present	nil
APC,ATM	ATM,DICER1	cellular macromolecule catabolic process	0.0215	GO:0044265	present	nil	nil
APC,ATM	ATM,DICER1	epithelium development	0.0328	GO:0060429	present	nil	nil
APC,ATM	ATM,DICER1	apoptotic process	0.0379	GO:0006915	present	nil	nil
APC,ATM	ATM	cell cycle arrest	0.0391	GO:0007050	present	present	nil
ATM	ATM,DICER1	microRNA in cancer	0.0416	KEGG:05206	present	nil	present

Fig 13: Results of biological process annotation for ATM gene. Only ATM enriched biological processes were listed in both the phenotypes with their false discovery rate together with presence and absence in analysis tools (STRING, g: Profiler, FIDEA).

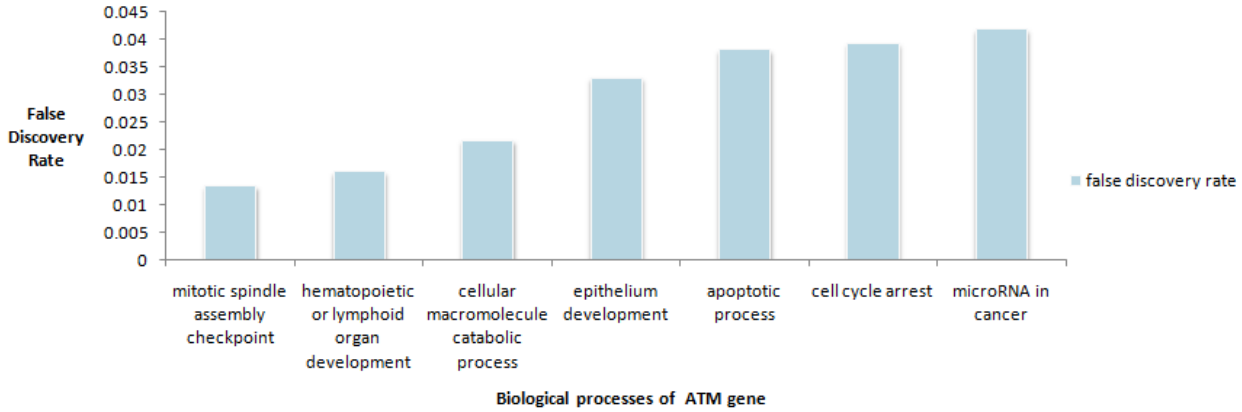

Fig 14: Plotted graph result for ATM gene on basis of biological process annotation and false discovery rate. Graph was arranged between biological processes at horizontal axis and false discovery rate at vertical axis. Mitotic spindle assembly checkpoint biological process with least false discovery rate (0.0132<0.05) in STRING tool analyzed and predicted that ATM gene mutation in mitotic spindle assembly checkpoint biological process perhaps advance in metastasis of kidney cancer.

IV. DISCUSSION

In System Biology Approach, any phenotype develops when thousands of molecules like genes, proteins, enzymes, signaling molecules etc interact with each other in multicellular complexes and in functional modules that make networks of these interacting molecules. Omics approach collects data of these interactions on basis of statistical scores like LogFC, false discovery rates, p-values etc to uncover functions of modules and assign features to genes and proteins which together or separately perform some biological functions in cells and tissues complexes in diseased as well as in normal body. In case if any disease like cancer develops in one part of body in any organ then it can be due to any reason like error in genetic & epigenetic networks, chromosomal instability, uncontrolled division of cells that can be due to alterations in the genes that might be performing some biological functions in those cells and due to mutations in genes, cells might lose their control of dividing that result in benign tumor that remain within the cells but with time those cells invade tissues and become malignant tumors and finally result in cancer. We have recently discussed about stages of cancer before becoming malignant that were dysplasia, hyperplasia and neoplasia. Till now scientists have detected neoplasia stage. Sometimes at genomic level, cancer complexity become major problem to study because cancer does not occur due to single gene mutation but due to thousands of genes and these genes are often interlinked with other organs cells therefore tumor spread from one organ to other organ and metastasis occurs and result in developing secondary tumors. To detect and handle such a complicated matter in terms of cancer at genomic level, differentially expressed gene profile from microarray dataset is really beneficial. Our aim and hypothesis was to identify if any gene or bunch of genes that was responsible for colorectal cancer express itself in kidney and lead to metastasis in kidney to develop secondary cancer therein on basis of any biological pathways or network common in both.

We retrieved data by giving renal carcinoma as input under disease view in Disease-connect database to retrieve genes that were common with colorectal cancer on basis of OMIM data and similarly phenotype neoplasm for both carcinomas was retrieved from GeneWays to view their connection. All the genes for colorectal cancer and kidney cancer were separately received from MalaCards database and to validate that genes belong to them and to retrieve their information, they were searched in their databases that is Colorectal Cancer Gene Database & Renal Cancer Gene Database. There were total 33 query genes studied and retrieved out of which 10 were found common in both neoplasm. STRING tool was used to build gene network of 33 query genes but it build of 15 genes only out of which 10 were common genes and 5 genes belong to colorectal cancer. The differential expression analysis of genes was done in canEVOLVE tool in cancer datasets to achieve

query genes differential expression profile on basis of statistical scores already calculated in R only LogFC values were taken into account to study upregulation and downregulation of 80 probes in separate datasets .the repeated genes list /probes with different LogFC values were separated as upregulated and downregulated genes and least value of LogFC for both colorectal cancer dataset and kidney cancer dataset genes were selected. Upregulated genes that were common in both cancers were separated and downregulated genes common in both were also selected with least LogFC .only upregulated genes were found significant when their graphs were plotted and downregulated genes were rejected because of large difference in their LogFC values. Among Upregulated genes, APC, DICER1 and ATM were giving close expression in both carcinomas and further they were validated by FIDEA tool where biological process "cell cycle assembly checkpoint" was common in both APC and ATM gene. In g:Profiler ,these genes were functionally enriched in network obtained from BIOGRID and validated that APC and ATM gene were of mutant phenotype in biological process "cell cycle assembly checkpoint".

The final prediction was done on basis of false Discovery rate of functionally enriched biological pathways in STRING that proved that "cell cycle assembly checkpoint" biological process with false discovery rate 0.0132<0.05 followed by all the biological processes that were also significant but less than assembly checkpoint and it proved that due to mutation in gene ATM that has some role in checking cell cycle mitotic assembly checkpoint result in metastasis of kidney neoplasia .ATM gene was mostly found responsible for breast and colorectal cancer but it was found common in both carcinomas and its 3D structure is unknown.

V. CONCLUSION

From this study, we predicted that colorectal cancer gene APC and kidney cancer gene DICER1 did not interact in functionally enriched biological network but common gene found in both the carcinomas, ATM showed much interactions with APC gene in network and were found more significant in biological process "mitotic spindle cell cycle assembly checkpoint" during their annotation with false discovery rate 0.0132<0.05 and these genes were of mutant type. It revealed that ATM (Ataxia telangiectasia mutated; Serine/threonine protein kinase which activates checkpoint signaling upon double strand breaks (DSBs), apoptosis and genotoxic stresses such as ionizing ultraviolet A light (UVA), thereby acting as a DNA damage sensor. Due to mutation in this gene, It lead to alteration in biological function "mitotic spindle assembly checkpoint" and it would not activate checkpoint signaling and lead to uncontrolled division of cells and metastasis of kidney neoplasia or develop secondary tumor. The 3D structure of ATM gene is unknown and its role is still uncertain and experimentally validated in kidney cancer.

VI. REFERENCES

[1]. Edwin Wang1, Anne Lenferink2, and Maureen O'Connor-McCourt2 http://www.academia.edu/5474451/*Cancer_systems_biology_explorin g_cancer-associated_genes_on_cellular_networks, academia.edu* [Accessed 2018]

[2]. Chun-Chi Liu Yu-Ting Tseng Wenyuan Li Chia-Yu Wu Ilya Mayzus, *DiseaseConnect: a comprehensive web server for mechanism-based disease–disease connections.*Nucleic Acid Research, 2014.p137-146

[3]. Noa Rappaport, Noam Nativ, Gil Stelzer, Michal Twik, *MalaCards: an integrated compendium for diseases and their annotation. The Journal of Biological Databases and Curation*, 2013

[4]. Rahul Agarwal, Binayak Kumar, Msk Jayadev, Dhwani Raghav And Ashutosh Singh, *CoReCG: a comprehensive database of genes associated with colon-rectal cancer.* The Journal of Biological Databases and Curation, 2015

[5]. Jayashree Ramana, *RCDB: Renal Cancer Gene Database.*BMC, Springer Nature,2012

[6]. Samur MK, Yan Z, Wang X, Cao Q, Munshi NC, Li C, Shah PK, canEvolve: a web portal for integrative oncogenomics. ReasearchGate, 2013

[7]. Daniel D'Andrea Luigi Grassi Mariagiovanna Mazzapioda Anna Tramontano, *FIDEA: a server for the functional interpretation of differential expression analysis.*Nucleic Acid Research, 2013.41:p8488

[8]. Jüri Reimand, Meelis Kull, Hedi Peterson, Jaanus Hansen, *g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments.* Nucleic Acid Research, 2007.35:p193200

[9]. Damian Szklarczyk, Andrea Franceschini, Stefan Wyder, Kristoffer Forslund, *STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acid Research*, 2015.l43:p447452