

Integrative bioinformatics analysis for identification of hub genes and pathways responsible for early pathogenesis of retinoblastoma

Aditi Karmakar

Indian Institute of Chemical Biology CSIR

Md. Maqsood Ahamad Khan

University of Allahabad

Nidhi Kumari

Indian Institute of Chemical Biology CSIR

Senthil Kumar (✉ skumar@iicb.res.in)

Indian Institute of Chemical Biology CSIR <https://orcid.org/0000-0001-6040-4814>

Primary research

Keywords: Retinoblastoma, Gene Expression Omnibus (GEO), Differentially Expressed Genes (DEGs), Hub genes, Co-expression Network Analysis, Biomarkers

Posted Date: June 1st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-30876/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

Retinoblastoma (Rb) is the most common childhood malignancy in which intra-ocular tumors developed at a very young age. It is very crucial to detect Rb at an earlier stage and start appropriate therapy to prevent further metastasis. With the recent advancement of multi-omics analysis of microarray data and sophisticated bioinformatics tools, we could possibly identify potential early diagnostic biomarkers and novel therapeutic targets for retinoblastoma.

Methods

Microarray datasets (DNA methylation-GSE57362, miRNA-GSE7072, & mRNA-GSE110811) were utilized from NCBI-GEO. The GEO2R were employed to discover Differentially Expressed Genes (DEGs) in Retinoblastoma. Further, an integrated analysis of these genes was performed and a co-expression network was formed to identify hub genes. GEPIA server was used to validate these genes which are responsible for the progression of retinoblastoma.

Results

Differentially expressed genes were identified on the basis of $P\text{-value} \leq 0.05$ and $\log_2\text{fc} \geq 2$. In GSE57362 a total no. of 267 genes methylated status, in GSE7072 a total no. of 265 gene targets of miRNAs and in GSE110811 a total no. of 770 genes were shown to be differentially expressed. Further, 10 hub genes, 5 bottleneck genes, and 3 common genes were identified by constructing the co-expression network. Survival analysis was done to validate the identified candidate genes using the GEPIA web server. Pathway enrichment analysis and gene ontology demonstrated that these genes were mostly enriched in biological processes such as regulation of cell proliferation and involved in multiple pathways like p53 pathway, cell cycle, and apoptosis.

Conclusion

This study suggested that these hub genes possibly will play vital roles in the onset and progression of retinoblastoma and could be serve as potential biomarkers to facilitate the diagnosis and treatment of this disease in the future.

Background

Retinoblastoma (Rb), a rare form of childhood malignancy, is the most common ocular malignancy found in children under age five which causes a great impact on the overall life of the affected child. Both, heritable and sporadic forms of retinoblastoma were observed. Retinoblastoma usually initiates with the biallelic mutation in the retinoblastoma gene (RB1) leading to $RB1^{-/-}$, but this mutation alone does not result in retinoblastoma. The mutation initially forms benign retinoma and then further genetic or epigenetic modifications are required for the development of retinoblastoma (Dimaras et al. 2015). Globally, the incidence of retinoblastoma is 1 per 16,000–18,000 live births with the prediction of about 8,000 cases per year. About 69% of all those cases are reported in middle-income countries, 20% in low-income, and 11% in high-income countries (Dimaras et al. 2015). There is a vast gap in the retinoblastoma treatment facilities among these countries. Consequently, the high-income countries have a patient survival rate of >

95% whereas it is only about 30% in low-income countries (Dimaras et al. 2015). The major reasons behind the poor outcome of retinoblastoma cases are the late diagnosis, recurrence, and metastasis, lack of availability of resources, refusal to perform enucleation.

The pRB protein (encoded by RB1 gene) is found to interact with at least 110 cellular proteins (Morris and Dyson 2001) but the exact molecular mechanism by which it normally suppresses retinoblastoma is still elusive. Studies have shown that RB can act in E2F dependent as well as E2F independent manner (Dick et al. 2013). In E2F dependent manner, RB negatively regulates E2F transcription factor thus suppressing cell cycle progression, but RB closely related genes p107 and p130 also suppress E2F, though to a lower extent, and it not mutated in cancers thus contradicting the above perception to some extent and indicating the differences in their biological processes (Classon & Harlow, 2002; Dick et al., 2013). E2F independent action of RB is achieved through upregulation of p27 which results in suppression of cell cycle and apoptosis of cell, but the decrease in expression of p27 with an increase in expression of pRB during maturation of cone precursor cell, cut out any consideration of p27 mediated suppression of retinoblastoma (Dimaras et al. 2015). However, the pathway of p53 is directly linked with the RB pathway through the p14/Arf tumor suppressor gene (Mendoza and Grossniklaus 2015) and can be responsible for RB1 mediated retinoblastoma suppression. It is observed that large deletions including RB1 in children result in fewer tumors than that of common null mutation indicating the role of adjacent unknown deleted genes in survival of tumor cells (Dimaras et al. 2012). Further, there are approx 1.5% other cases of unexplained unilateral sporadic retinoblastoma where both RB1 and MYCN genes are normal. Hence, there are various factors which are interlinked with each other to initiate and progress the disease. To understand the molecular mechanism of retinoblastoma and pRB action, it is essential to find out the key genes playing a critical role in retinoblastoma development, functional pathways and biological processes by systematic analysis of multiplexed microarray data.

The main purpose of this study is the identification of hub genes and molecular pathways involved in the onset and progression of retinoblastoma by employing potential integrative bioinformatics analysis.

Methods

Raw data collection

Three sets of raw microarray data consisting of transcriptomics profile of retinoblastoma were downloaded from Gene Expression Omnibus Database (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>). They belong to genome-scale DNA methylation profiling in Retinoblastoma by using Illumina Human Methylation 450 Bead Chip platform and samples from the normal human eye and five ocular diseases (GSE57362, 25 Samples) (Berdasco et al. 2017), Profiling of miRNAs in human retinoblastoma by using 2 k custom array and RNA was extracted from 2 retinoblastoma and 2 matched normal retina samples (GSE7072, 4 Samples) (Huang et al. 2007), Distinct Gene Expression Profiles Define Anaplastic Grade in Retinoblastoma by using Affymetrix Human Gene 2ST Array and RNA was extracted from the 28 retinoblastoma and 3 matched normal retina samples (GSE110811, 31 Samples) (Hudson et al. 2018). Microarray datasets contain cancer and normal samples. The details of the datasets are given in Table 1.

Table 1
Summary of the microarray datasets and differentially expressed genes in each dataset

Serial No.	Cancer type	Accession No.	Datasets	Sample	Upregulated genes	Downregulated genes	Total DEGs*
1	Retino - blastoma	GSE57362	DNA Methylation	Normal Retina = 10 Retinoblastoma = 15	100	167	267
2		GSE7072	miRNA	Normal Retina = 2 Retinoblastoma = 2	161	104	265
3		GSE110811	mRNA	Normal Retina = 3 Retinoblastoma = 28	33	737	770

*Selection of the significant DEGs were made totally on the basis of P.value \leq 0.05 and $|\log_2FC| \geq 2$

Data Processing

GEO2R online software (<https://www.ncbi.nlm.nih.gov/geo/geo2r>) is used to analyze the raw micro-array data for quality check and identify the differentially expressed genes (DEGs). A quality check of raw reads is necessary before going to the expression analysis and network construction. In GEO2R, the value distribution tab is used to assess the distributions of values across Samples (median-centered) which generally indicate that the data are normalized and each dataset was separately normalized by median normalization and the quality of the data accessed through box-plot. A brief outline of the workflow representing the datasets and methods used to conduct this study has been given in Fig. 1.

Identification Of Differentially Expressed Genes

GEO2R is a web tool that allows differentiation among groups of samples which are differentially expressed across experimental conditions. The genes which are differentially expressed between retinoblastoma samples and control samples were identified by the LIMMA package (Linear Models for Microarray data) using GEO2R. An adjusted p-value (p_{adj}) \leq 0.05 and logarithmic fold change (\log_2FC) \geq 2 were used as the threshold to find DEGs and control the statistical significance of the expressed genes.

Pathway Enrichment Analysis And Gene Ontology Of DEGs

Gene ontology is used for an integrated research annotation and visualization which allows a complete set of annotation methods understanding the cellular component, molecular function, and biological process concept behind these genes. In this study PANTHER database (Protein Analysis Through Evolutionary Relationships) (<http://www.pantherdb.org/tools/>) was used to perform functional annotation and ShinyGO v0.61

(<http://bioinformatics.sdstate.edu/go/>) for KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analysis of differentially expressed genes.

PPI Network Construction And Hub Genes Identification

The PPIs co-expression network of DEGs were created by using STRING (Search Tools For Retrieval of Interacting Genes/Proteins) database (<https://string-db.org/>). The database currently contains 24,584,628 proteins from 5090 organisms. The network with a confidence score ≥ 0.7 was selected for PPI network construction, and the network was visualized and merged by using the Cytoscape (<http://cytoscape.org/>) software. In Cytoscape, each gene is represented as “node” and the strength of the co-expression relationship between genes is represented as “edge” and the number of interactions between the specific gene and other genes is represented as “degree”. Node with a degree > 10 was selected as hub nodes. There are several plugins in Cytoscape which are used to analyze network and find the hub genes such as Network analyzer which is used for determining the topological parameters of the networks and cytoHubba plugins were used for identifying the hub genes and bottleneck genes.

Mining And Analysis Of Module In PPIs Network

MCODE (Molecular Complex Detection) plugin in Cytoscape was used to extract and analyze modules/clusters from the merge network. The significant cluster with MCODE score ≥ 6 and node density cutoff = 1, node score cutoff = 0.2, k -core = 2 used as statistical significance.

Validation Of Core Genes By TCGA

GEPIA (<http://gepia.cancer-pku.cn/index.html>) (Gene Expression Profiling Interactive Analysis) is a web server for expression profiling of the candidate genes in retinoblastoma and their respective control sample. A statistical significance threshold $\log_2FC \geq 2$, $P\text{-value} \leq 0.05$, and $FDR < 1\%$ is used for further validation. The survival analysis was executed to calculate the expression levels of hub genes and their survival time of patients with retinoblastoma. The hazard ratio and the P -value were calculated using log-rank tests. GEPIA provides tumor and normal differential expression analysis by using RNA sequencing expression data of 9,736 tumors and 8,587 normal samples from the TCGA.

Results

Identification of differentially expressed genes (DEGs)

We identified DEGs between normal and tumor samples individually for each dataset by LIMMA package using GEO2R. Significantly differentially expressed genes were selected by qualifying the criteria $P\text{-value} \leq 0.05$ and $|\log_2fc| \geq 2$. We henceforth found a total of 267 Differentially Methylated Genes (DMGs) in GSE57362, 265 Differentially Expressed miRNAs (DE-miRNAs) targets in GSE7072 and 770 DEGs in GSE110811. The total numbers of upregulated and downregulated genes in each dataset are summarized in Table 1. All three datasets were normalized and shown in **Fig. 2(A-C)**. Volcano plots are drawn for DEGs in each of the three datasets shown in **Fig. 3(A-C)**. Our computational analysis unveiled that 3 common genes found in the upregulated-downregulated miRNA dataset, upregulated-downregulated mRNA dataset and upregulated-downregulated DNA methylation dataset. Although no common gene is present among the downregulated miRNA dataset, upregulated mRNA dataset and hypo-DNA methylated dataset.

We also drew Venn diagrams comprising genes of mRNA and miRNA (both upregulated and downregulated) datasets as well as mRNA and DNA methylation (both upregulated and downregulated) datasets but found no common genes in the overall aspect shown in **Fig. 4(A-F)**.

Gene Ontology And Pathway Enrichment Analysis

The PANTHER and ShinyGO v0.61 were utilized to unwrap the biological activities and key regulatory pathways associated with the DEGs in DNA methylation (267 genes), miRNA (265 target genes) and mRNA (770 genes). Significantly enriched KEGG pathways and GO terms were identified based on $P\text{-value} \leq 0.05$. The top 20 significant biological processes, molecular functions, and cellular components associated with the DEGs in the three datasets are shown in **Fig. 5 (A-I)**. Table 2 shows hub genes involved in the top 17 significant molecular pathways in DNA methylation, miRNA and mRNA datasets of Retinoblastoma. The result showed that the p53 pathway, cell cycle and apoptosis pathway are the common pathways.

Table 2
Top 17 molecular pathways involved in retinoblastoma

Serial No.	Designation	Description	Count	P.value
1	hsa04110	Cell cycle	8	3.84E-10
2	hsa05203	Viral carcinogenesis	8	1.32E-08
3	hsa05200	Pathways in cancer	8	1.16E-06
4	hsa04919	Thyroid hormone signaling pathway	5	3.31E-05
5	hsa04115	p53 signaling pathway	4	1.82E-04
6	hsa04722	Neurotrophin signaling pathway	4	0.001015304
7	hsa04068	FoxO signaling pathway	4	0.001398004
8	hsa05205	Proteoglycans in cancer	4	0.004386821
9	hsa04662	B cell receptor signaling pathway	4	0.006133462
10	hsa04012	ErbB signaling pathway	3	0.009611516
11	hsa04660	T cell receptor signaling pathway	3	0.012558596
12	hsa04668	TNF signaling pathway	3	0.014290638
13	hsa04151	PI3K-Akt signaling pathway	3	0.019617316
14	hsa04310	Wnt signaling pathway	3	0.023116873
15	hsa04024	cAMP signaling pathway	3	0.045005064
16	hsa04010	MAPK signaling pathway	3	0.059756486
17	hsa04150	mTOR signaling pathway	2	0.056688768

Co-expression Network Construction

A single gene does not always contribute to disease progression by simply altering its' expression level rather multiple genes interact with each other in a complex manner to cause the progression of any disease. Expression of a particular gene is the main criterion for the identification of DEGs and the interaction between multiple genes is not usually taken into consideration for that (Singh and Som 2019). The co-expression association between several genes is widely studied by constructing gene co-expression networks. We have found that the DNA methylation co-expression network has 63 nodes and 83 interactions, the miRNA co-expression network has 129 nodes and 424 interactions and mRNA co-expression network has 243 nodes and 471 interactions as described in Table 3 and shown in **Fig. 6 (A-C)**. After constructing PPI networks on Cytoscape, we found 5 hub genes and 3 bottleneck genes in each of the three datasets as shown in Table 4.

Table 3
Topological parameters used in the co-expression network

Dataset	Total DEGs	Nodes	Edges	Clustering-coefficient	Correlation	R ²
GSE57362	267	63	83	0.771	0.840	0.781
GSE7072	265	129	424	0.691	0.966	0.761
GSE110811	770	243	471	0.789	0.975	0.872

Table 4
Hub genes and Bottleneck genes found in each dataset by constructing PPI networks

Dataset	Hub Genes	Degree	Bottleneck Genes	Degree
GSE57362	POLR2A	10	POLR2A	42
	CCND1	7	MAPK11	42
	KCNAB2	6	ESR1	42
	CLEC4D	5		
	CD47	5		
GSE7072	TP53	27	PIK3CA	26
	HDAC1	26	HIF1A	16
	MAPK1	26	TP53	15
	STAT3	24		
	RB1	23		
GSE110811	CCNB1	22	AURKA	73
	AURKA	18	SMC3	45
	RPS6	17	GIT1	33
	RPS6	16		
	RPL10A	16		

All three datasets have some correlated gene pairs that were represented by an intersection network formed by combining all the three individual networks (**Fig. 7**). From this study, we further stated that individual genes were

differentially expressed rather than being co-expressed. Thus we found that the progression of retinoblastoma might be caused by two factors 1) coordinated expression of the genes and 2) the interactions among the genes.

After that, we used the MCODE plugin in Cytoscape which is used as MCODE algorithm to mine and analyze modules/clusters from the co-expression network. Firstly we set the parameter of MCODE plugin such as node density cut off 1, node score cut off to 0.2, k-core 2 and the maximum depth to 100. Three modules/clusters were identified and shown in **Fig. 8**. Further analysis of module we found that 9 hub genes in cluster 1 (MCODE Score-10.607) (*CDK1, CDK2, CCNB1, CUL1, RB1, TP53, HDAC1, JUN, MAPK1*) and 2 common genes (*E2F3, ESR1*). No hub genes were present in clusters 2 and 3. A total of 6 bottleneck genes (genes with high betweenness centrality) were also found in the co-expression network (shown in Table 5).

Table 5
Hub genes, Common genes and Bottleneck genes found in the co-expression network

Type of genes	Name of genes	Degree
Hub Genes	CDK1	53
	MAPK1	44
	PIK3CA	40
	TP53	38
	HDAC1	37
	RB1	32
	CDK2	32
	CCNB1	32
	CUL1	31
JUN	30	
Common Genes	E2F3	40
	ESR1	42
	UNC5D	30
Bottleneck Genes	MAPK1	46
	PIK3CA	44
	PSME3	40
	NR3C1	32
	NHP2	27
	ITGB1	27

Validation Of Expressed Core Genes

The expression level of the 12 critical genes obtained from DNA methylation, miRNA, and mRNA datasets of both normal and tumor samples is checked by using the GEPIA tool. After performing the whole analysis, we observed that there is significant differential expression of these genes in malignant conditions in comparison to normal conditions by considering the level of significance of $P\text{-value} \leq 0.05$, $\log_2FC > 2$ and $FDR < 1\%$ as shown in **Fig. 9**. We also performed stage and grade boxplots to validate the 10 hub genes and 3 common genes based on the three datasets (DNA methylation-GSE57362, miRNA-GSE7072, & mRNA-GSE110811) [**Fig. 10 (A-M)**].

Survival Analysis:

To validate the hub genes of Retinoblastoma, a total of 512 cases of LGG (Lower Grade Glioma) patients, 166 cases of GBM (Glioblastoma Multiforme) patients, and 1092 cases of BRCA patients were registered in this learning as unavailability of Rb samples. The patients were grouped into two categories: high expression groups and low expression groups on the basis of gene expression profile, and expression levels higher than the median. Overall survival analysis was determined on the basis of gene expression. Later on, survival analysis was executed to ascertain the correlation between gene expression level and the overall survival of the patient. We further analyzed the data to validate our findings in the GEO group. Kaplan-Meier and long-rank test analysis revealed that six hub genes were significant in the survival analysis from 10 hub genes and 3 common genes, which were linked with patient prognosis (Fig.). To summarize, high expression of *CCNB1*, *CDK1*, *TP53*, *RB1*, *JUN* and low expression of *HDAC1* was shown to predict a worse prognosis in patients with retinoblastoma. Survival analysis graphs for each gene were shown in **Fig. 11 (A-M)**.

Discussion

Retinoblastoma (Rb) is the most common ocular pediatric malignancy caused by the inactivation of the *RB* genes. The early diagnosis of Rb will provide the possibility to control the primary tumor with effective therapies which will ultimately improve the clinical outcomes. In this study, we identified hub and common candidate genes for retinoblastoma by integrative bioinformatics analysis and these genes could be further served as potential therapeutic targets. We compared the gene expression profiles of DNA methylation, miRNA, and mRNA for both healthy and tumor samples using individual microarray datasets available in NCBI-GEO. Overall 3 common key genes and 10 hub genes were identified for all the three datasets by comparing the DEGs for all the datasets taken into account.

In the present study, a total no. of 10 hub genes was found after merging the co-expression networks. In this intersection network, connections were established among these 10 hub genes shown to be over/under-expressed. This represented a conserved co-expression association that was established across the three datasets. The 10 hub genes are *CDK1*, *CDK2*, *CCNB1*, *CUL1*, *RB1*, *TP53*, *HDAC1*, *JUN*, *PIK3CA* and *MAPK1* (shown in Table 5). Further, we have also identified 3 potential common genes (*E2F3*, *ESR1* and *UNC5D*) (also shown in Table 5). We have further validated these key genes involved in retinoblastoma by GEPIA tool and survival analysis and found six potential candidate genes(*CCNB1*, *CDK1*, *TP53*, *RB1*, *JUN* and *HDAC1*).

The previous studies were found that those 10 real hub genes were involved in the process of cell proliferation and tumorigenesis. CDK1 gene encodes a protein that belongs to the Ser/Thr protein kinase family and serves as a catalytic subunit of M-phase promoting factor (MPF) which plays a major role in both G1/S and G2/M phase transitions of the eukaryotic cell cycle. CDK1 was identified as a hub gene responsible for the prognosis and progression of bladder cancer (Yan et al. 2019). The protein encoded by the CDK2 gene is also a member of a

serine/threonine-protein kinase family that participates in the regulation of cell cycle and activity of this protein is critically essential during the G1 to S phase transition of the eukaryotic cell cycle (Stelzer et al. 2016). The cyclin E–Cdk2 may have the ability to block the function of the C-terminal region of pRb expressed in trans through a mechanism different from that of cyclin D–Cdk4/6 which does not depend on C-terminal phosphorylation of pRb (Harbour et al. 1999).

CCNB1 encodes a protein that is required for the proper regulation of the G2/M transition phase of the cell cycle (Stelzer et al. 2016). This protein forms complexes with p34 (cdc2) to produce the MPF. The CCNB1 gene was shown to be upregulated in gastric cancer (GC) (Begnami, Fregnani, Nonogaki, & Soares, 2010; D. Wang et al., 2015) overexpression of this gene was related to different clinical stages, such as lymph node metastasis (Begnami et al. 2010), thus contributing to poor survival in GC patients (Yasuda et al. 2002). Growth inhibition of GC cell by CCNB1 along with CCNB2 may give an explanation for poor prognosis in patients who suffered in GC (Li, Wang, and Lin 2018). CCNB1 was also identified as a hub gene and responsible for the prognosis and progression of bladder cancer (Yan et al. 2019). CUL1 is the core component of SCF (SKP1-CUL1-F-box protein) E3 ubiquitin-protein ligase complexes, which takes an integral part in various cellular events like a progression of cell cycle, transcription and signal transduction by mediating ubiquitination of various proteins involved in these processes. It has been observed that Epstein-Barr nuclear antigen 3C (EBNA3C) binds to pRb and triggers its degradation by recruiting the ubiquitin ligase complex Skp1/Cul1/F-box (SCFSkp2) (Knight, Sharma, & Robertson, 2005; Macdonald & Dick, 2012).

RB1 is a key regulator of the cell cycle and its' active, hypophosphorylated form binds the transcription factor, E2F1 thus repressing the transcriptional activity of E2F1 target genes. The pRb adaptor has been deployed to regulate DNA replication and DNA repair, either indirectly through modulating the chromatin structure or directly by modulating the protein complexes that are involved in these biochemical processes (Meenalakshmi Chinnam and David W. Goodrich 2013). Mutation in this gene causes Retinoblastoma, Osteogenic sarcoma, and Bladder cancer. The highest rate of RB1 mutation (about 90%) is also observed in small cell lung cancer (SCLC) (Meenalakshmi Chinnam and David W. Goodrich 2013). TP53 gene encodes a tumor suppressor protein possessing numerous domains for transcriptional activation, DNA binding, and oligomerization process. It has been shown that TP53 mutations have a consistent association with poor prognosis in different types of cancers like leukemia, breast, head and neck and colorectal (Petitjean et al., 2007; Olivier, Hollstein, & Hainaut, 2010) while somatic TP53 mutations is evident in colorectal, ovarian, larynx, esophageal, lung cancers and head and neck cancers (rate: 38–50%) (Olivier, Hollstein, and Hainaut 2010).

The protein encoded by the HDAC1 gene is a component of the histone deacetylase complex. HDAC1 deacetylates p53 by interacting with metastasis-associated protein-2 thereby modulating the effect of p53 on apoptosis and cell growth. The loss of HDAC1 activity could result in the induction of expression of genes exclusively regulated by RB and the suppression of their protective role in tumor development and cancer progression (Ropero and Esteller 2007). JUN gene is the putative transforming gene originated in the avian sarcoma virus 17 and the encoded protein directly interacts with specific target sequences of DNA to control the expression of several genes (Stelzer et al. 2016). JUN gene binds to the promoter region of USP28 in colorectal cancer (CRC) cells and activates USP28 through activation of the KRAS-mediated transcription process (Serra et al. 2014).

The gene Phosphoinositide-3-kinase (PI3K) is engaged in cell signaling pathways. Mutations in the PIK3CA gene hyperactivate the kinase activity of the PI3K pathway, thereby causing the deregulation of cell proliferation and formation of most solid malignant tumors such as breast, colon, lung, ovarian, urothelial, cervical cancers (Alqahtani, Ayes, and Halawani 2019). MAPK1 gene encodes a protein belonging to the MAP kinase family and it has been recognized as a moonlighting protein as it is able to perform mechanistically different functions. MAP kinases are

known to amalgamate multiple biochemical signaling pathways at one point and play a significant role in a vast number of cellular processes such as transcription regulation, proliferation, differentiation and developmental process (Stelzer et al. 2016).

In this study we are reporting three common genes were found among DNA methylation, miRNA and mRNA dataset. E2F3 gene encodes a transcription factor specifically binding to RB1 and it occurs in a cell-cycle dependent manner. Progression of cell-cycle from G1 to S phase is controlled by the proper functioning of the DRTF1/E2F complex (Stelzer et al. 2016). According to our study, E2F3 might be playing a crucial role in the progression of retinoblastoma and can be used as a potential therapeutic target. E2F3 along with KIF14 may represent important pro-oncogenes that have been seen to be dysregulated in most Rb tumors (Madhavan et al. 2009).

ESR1 gene encodes an estrogen receptor is associated with hormone therapy resistant in breast cancer patients (Clatot, Augusto, & Fiore, 2017) while the studies conducted so far have shown that the main reason was the occurrence of different types of mutations in the ligand-binding domain of ESR1. The clinical trials also in a process to develop therapeutics that invigorate the degradation of ESR1 protein (e.g. Fulvestrant) for patients who are non-responsive to hormonal therapy (Lei et al. 2019).

UNC5D gene product serves as a netrin NTN4 receptor that is involved in cell migration and promotes the survival of the neuronal cells. When UNC5D is not associated with netrin ligand, it acts as a dependence receptor to induce the process of apoptosis (H. Wang, Zhang, et al. 2014). Lu et al. reported that UNC5D was not often present in various cancer cell lines and its' expression became downregulated in primary renal cell carcinoma (RCC) (Lu et al. 2013). Another study revealed that UNC5D in co-operation with E2F3 and p53 is involved in the apoptosis caused due to deficiency in Nerve Growth Factor (NGF) in netrin-1 deficient human neuroblastoma cells (Zhu et al. 2013).

The most of the available studies were reports conculation of either methylation or gene expression or miRNA microarray data. Therefore we attempted to report the more accurate and reliable retinoblastoma candidate genes by complying multiple datasets. The existing evidence strongly supports that the identified candidate and hub genes in this study were involved in numerous ways to the onset and progression of various cancers. Although, there are some limitations in this study that we conducted so far. Considering the data, we could conclude that all these 10 hub genes were significantly differentially expressed in retinoblastoma but we further need to conduct more molecular biological experiments using preclinical models to validate the roles of these genes in the onset and progression of retinoblastoma.

Conclusion

This integrative bioinformatics approach has made it possible for us to identify 10 hub genes in three GEO datasets i.e. DNA methylation, miRNA, and mRNA whose co-expression might cause tumor formation and metastasis in children affected with Retinoblastoma. The biological functions and molecular pathways of these 10 predicted hub genes (CDK1, CDK2, CCNB1, CUL1, RB1, TP53, HDAC1, PIK3CA, JUN, and MAPK1) which have been shown to be differentially expressed in Rb and detailed and further studies need to be conducted to validate their roles in Rb progression as well as diagnosis. These genes are known to regulate the functions of several other downstream genes and in the near future exploring these elements could help the researchers to interpret the possible reasons for dysregulation of these hub genes in this cancer. These genes could also be served as potential therapeutic markers to develop promising medications to treat retinoblastoma in the forthcoming years.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets were used in the present study are available in the NCBI-GEO (<https://www.ncbi.nlm.nih.gov/geo/>)

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by institutional grant No P07-MLP-120 from CSIR- Indian Institute of Chemical Biology, Kolkata, India.

Authors' contributions

A. K., M.A.K, and S. K. G. have given an equal contribution in framing research topics, collecting study materials, performing the entire research work, analyzing the data, preparing the manuscript, proofreading, and so forth. N. K. has written the introduction section of the manuscript and involved in the data analysis. All authors read and approved the final manuscript.

Acknowledgements

AK thanks UGC- Award No: 16-[Dec-2017/2018(NET/CSIR)] and NK thanks CSIR-Award No: 1121732018 India for their fellowships. MAK thanks the University of Allahabad and CSIR-IICB for giving him the internship opportunity. SKG acknowledges Director, CSIR-IICB for his support.

References

1. Alqahtani A, Hazem SK, Ayesha, and Hafez Halawani. "PIK3CA Gene Mutations in Solid Malignancies: Association with Clinicopathological Parameters." no. Figure. 2019;1:1–18.
2. Begnami MD, José Humberto TG, Fregnani S, Nonogaki, Fernando AS. Evaluation of Cell Cycle Protein Expression in Gastric Cancer: Cyclin B1 Expression and Its Prognostic Implication. *Hum Pathol.* 2010;41(8):1120–27. <https://doi.org/10.1016/j.humpath.2010.01.007>.
3. 10.1038/s41598-017-12084-1
Berdasco M, Gómez A, Rubio MJ, Jaume Català-mora, Miguel Lopez, Cristina Hernández, Shigeo Yoshida, et al. 2017. "DNA Methylomes Reveal Biological Networks Involved in Human Eye Development, Functions and Associated Disorders," no. September: 1–15. <https://doi.org/10.1038/s41598-017-12084-1>.

4. Caligiuri I, Toffoli G, Giordano A, Rizzolio F. 2013. "PRb Controls Estrogen Receptor Alpha Protein Stability and Activity" 4 (6): 875–83.
5. Classon M, and Ed Harlow. 2002. "THE RETINOBLASTOMA TUMOUR SUPPRESSOR IN DEVELOPMENT AND CANCER" 2 (December): 910–17. <https://doi.org/10.1038/nrc950>.
6. Clatot F, Laetitia Augusto, and Frédéric Di Fiore. n.d. "ESR1 Mutations in Breast Cancer" 9: 9–10. <https://doi.org/10.1200/JCO.2016.67.3061>.
7. Dean JL, Thangavel C, McClendon AK, Reed CA, Knudsen ES. Therapeutic CDK4 / 6 Inhibition in Breast Cancer: Key Mechanisms of Response and Failure. *Oncogene*. 2010;29(28):4018–32. <https://doi.org/10.1038/onc.2010.154>.
8. Dick FA, Seth M, Rubin RC, Program, and Santa Cruz. 2013. "Molecular Mechanisms Underlying RB Protein Function" 14 (5): 297–306. <https://doi.org/10.1038/nrm3567.Molecular>.
9. Dimaras H, Corson TW, Cobrinik D, White A, Zhao J. 2015. "Retinoblastoma," no. August. <https://doi.org/10.1038/nrdp.2015.21>.
10. Dimaras H, Kimani K, Elizabeth AO, Dimba P, Gronsdahl A, White, Helen SL, Chan, Brenda LG. Retinoblastoma Is the Most Common Intraocular Cancer of Childhood. It Is Initiated by Mutation of The". 2012;379:1436–46. [https://doi.org/10.1016/S0140-6736\(11\)61137-9](https://doi.org/10.1016/S0140-6736(11)61137-9).
11. Harbour JW, Luo RX, Santi AD, Antonio A, Postigo DC, Dean, St Louis, and Phosphorylation S-. 1999. "Intramolecular Interactions That Progressively Block Rb Functions as Cells Move through G1" 98: 859–69.
12. "Using Expression Profiling Data to Identify Human <https://doi.org/10.1038/NMETH1130>
Huang JC, Tomas Babak TW, Corson G, Chua S, Khan BL, Gallie TR, Hughes BJ, Blencowe BJ, Frey, Quaid DM. 2007. "Using Expression Profiling Data to Identify Human MicroRNA Targets" 4 (12): 1045–49. <https://doi.org/10.1038/NMETH1130>.
13. Hudson L, Mendoza P, Hudson W, Ziesel A, Hubbard GB, Wells J, Dwivedi B, et al. Distinct Gene Expression Profiles Define Anaplastic Grade in Retinoblastoma. *Am J Pathol*. 2018. <https://doi.org/10.1016/j.ajpath.2018.06.013>.
14. Knight JS, Nikhil Sharma, Erle SR. 2005. "Epstein – Barr Virus Latent Antigen 3C Can Mediate the Degradation of the Retinoblastoma Protein through an SCF Cellular Ubiquitin Ligase," no. Track II.
15. Lan Hsin-chieh, Li Hua-jung, Lin G, Lai Pao-yen, and Bon-chu Chung. 2007. "Cyclic AMP Stimulates SF-1-Dependent CYP11A1 Expression through Homeodomain-Interacting Protein Kinase 3-Mediated Jun N-Terminal" 27 (6): 2027–36. <https://doi.org/10.1128/MCB.02253-06>.
16. Lei JT, Xuxu Gou S, Seker MJ, Ellis S, Smith, and Cellular Biology. 2019. "ESR1 Alterations and Metastasis in Estrogen Receptor Positive Breast Cancer Jonathan." <https://doi.org/10.20517/2394-4722.2019.12.ESR1>.
17. Li Shu-yu, Wang Jian-ping, Lin J. 2018. "Clinical Significance and Biological Roles of Cyclins in Gastric Cancer," 6673–85.
18. Liu H, and Ming Zhou. 2017. "Antitumor Effect of Quercetin on Y79 Retinoblastoma Cells via Activation of JNK and P38 MAPK Pathways," 1–8. <https://doi.org/10.1186/s12906-017-2023-6>.
19. Lovicu FJ, John W, Mcavoy. "FGF-Induced Lens Cell Proliferation and Differentiation Is Dependent on MAPK (ERK1 / 2. Signalling". 2001;5084:5075–84.
20. Lu D, Dong D, Zhou Yu, Lu M, Pang Xue-wen, Li Y, Tian Xiao-jun, Zhang Yu, and Jun Zhang. 2013. "The Tumor-Suppressive Function of UNC5D and Its Repressed Expression in Renal Cell Carcinoma," 2883–93. <https://doi.org/10.1158/1078-0432.CCR-12-2978>.

21. Macdonald JI, Frederick AD. 2012. "Posttranslational Modifications of the Retinoblastoma Tumor Suppressor Protein as Determinants of Function," 619–33. <https://doi.org/10.1177/1947601912473305>.
22. Madhavan J, Mitra M, Mallikarjuna K, Pranav O, Srinivasan R, Nagpal A, Venkatesan P, Kumaramanickavel G. 2009. "KIF14 and E2F3 mRNA Expression in Human Retinoblastoma and Its Phenotype Association," no. January: 235–40.
23. Marcel V, Perrier S, Aoubala M, Ageorges S, Groves MJ, Diot A, Fernandes K, Tauro S, Jean-christophe, Bourdon. D 160p53 Is a Novel N-Terminal P53 Isoform Encoded by D 133p53 Transcript. *FEBS Lett.* 2010;584(21):4463–68. <https://doi.org/10.1016/j.febslet.2010.10.005>.
24. Chinnam M, Goodrich DW. 2013. *RB1, Development and Cancer*. <https://doi.org/10.1016/B978-0-12-380916-2.00005-X.RB1>.
25. 10.1016/bs.pmbts.2015.06.012
Mendoza PR, Hans EG. 2015. *The Biology of Retinoblastoma. Molecular Biology of Eye Disease*. 1st ed. Elsevier Inc. <https://doi.org/10.1016/bs.pmbts.2015.06.012>.
26. "Retinoblastoma Protein Partners
617
Morris EJ, Nicholas JD. 2001. "Retinoblastoma Protein Partners," no. 617.
27. Olivier M, Hollstein M, Hainaut P. 2010. "TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use," 1–17.
28. Paulo JA, Annemieke De Jong H, Ovaa AF, Alpi, and J Wade. 2017. "Two Distinct Types of E3 Ligases Work in Unison to Regulate Substrate Ubiquitylation" 166 (5): 1198–1214. <https://doi.org/10.1016/j.cell.2016.07.027.Two>.
29. Petitjean A, M I W Achatz AL, Borresen-dale P, Hainaut M, Olivier, and Sao Paolo. TP53 Mutations in Human Cancers: Functional Selection and Impact on Cancer Prognosis and Outcomes. no. 2007;191170:2157–65. <https://doi.org/10.1038/sj.onc.1210302>.
30. Ropero S, and Manel Esteller. 2007. "The Role of Histone Deacetylases (HDACs) in Human Cancer" 1: 19–25. <https://doi.org/10.1016/j.molonc.2007.01.001>.
31. Serra RW, Minggang Fang SM, Park L, Hutchinson, and Michael R Green. 2014. "A KRAS-Directed Transcriptional Silencing Pathway That Mediates the CpG Island Methylator Phenotype" 2: 1–22. <https://doi.org/10.7554/eLife.02313>.
32. Singh R, and Anup Som. 2019. "Identification of Common Candidate Genes and Pathways for Progression of Ovarian, Cervical and Endometrial Cancers." *Meta Gene*, 100634. <https://doi.org/10.1016/j.mgene.2019.100634>.
33. "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses
<https://doi.org/10.1002/cpbi.5>
Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Stein TI, et al. 2016. "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses," no. June: 1–33. <https://doi.org/10.1002/cpbi.5>.
34. Upadhyaya D, Ogata M, Lixing WR. 2013. "MAPK1 Is Required for Establishing the Pattern of Cell Proliferation and for Cell Survival during Lens Development" 1582: 1573–82. <https://doi.org/10.1242/dev.081042>.
35. Wang Da-guang, Chen G, Wen Xiao-yu, Wang D, Cheng Zhi-hua, Si-qiao S. "Identification of Biomarkers for Diagnosis of Gastric Cancer by Bioinformatics". 2015;16:1361–65.
36. Wang H, Wu Q, Li S, Zhang BIN, Chi Z, and Liangchun Hao. 2014. "Unc5D Regulates P53 – Dependent Apoptosis in Neuroblastoma Cells," 2411–16. <https://doi.org/10.3892/mmr.2014.2100>.

37. Wang H, Zhang B, Gu M, Li S. 2014. "Overexpression of the Dependence Receptor UNC5H4 Inhibits Cell Migration and Invasion, and Triggers Apoptosis in Neuroblastoma Cell." <https://doi.org/10.1007/s13277-014-1706-6>.
38. Xie L, Overbeek PA, Lixing WR. 2006. "Ras Signaling Is Essential for Lens Cell Proliferation and Lens Growth during Development" 298: 403–14. <https://doi.org/10.1016/j.ydbio.2006.06.045>.
39. Yan X, Liu Xiao-ping, Guo Zi-xin, Liu Tong-zu, Li S. 2019. "Identification of Hub Genes Associated With Progression and Prognosis in Patients With Bladder Cancer" 10 (May): 1–12. <https://doi.org/10.3389/fgene.2019.00408>.
40. Yasuda M, Takesue F, Inutsuka S, Honda M, Nozoe T, and Daisuke Korenaga. 2002. "Overexpression of Cyclin B1 in Gastric Cancer and Its Clinicopathological Significance: An Immunohistological Study," 412–16. <https://doi.org/10.1007/s00432-002-0359-9>.
41. [10.1038/ncb801](https://doi.org/10.1038/ncb801)
Yin Y, Stephen CW, Gloria Luciani M, Fåhraeus R. 2003. "P53 Stability and Activity Is Regulated by Mdm2-Mediated Induction of Alternative P53 Translation Products" 4 (June 2002). <https://doi.org/10.1038/ncb801>.
42. Zhang Man-man, Cui FSun,B, Zhang Le-le, Ya Fang, and Li Y. 2017. "Tumor-Suppressive Function of UNC5D in Papillary Thyroid Cancer" 8 (56): 96126–38.
43. Zhao H, Yang T, Madakashira BP, Cornelius A, Thiels A, Chad CM, Garcia H, Zhang, et al. 2009. "Fibroblast Growth Factor Receptor Signaling Is Essential for Lens Fiber Cell Differentiation" 318 (2): 276–88. <https://doi.org/10.1016/j.ydbio.2008.03.028.Fibroblast>.
44. Zhu Y, Li Y, Haraguchi S, Yu M, Ohira M, Ozaki T, Nakagawa A, et al. 2013. "Dependence Receptor UNC5D Mediates Nerve Growth Factor Depletion – Induced Neuroblastoma Regression" 123 (7). <https://doi.org/10.1172/JCI65988DS1>.

Figures

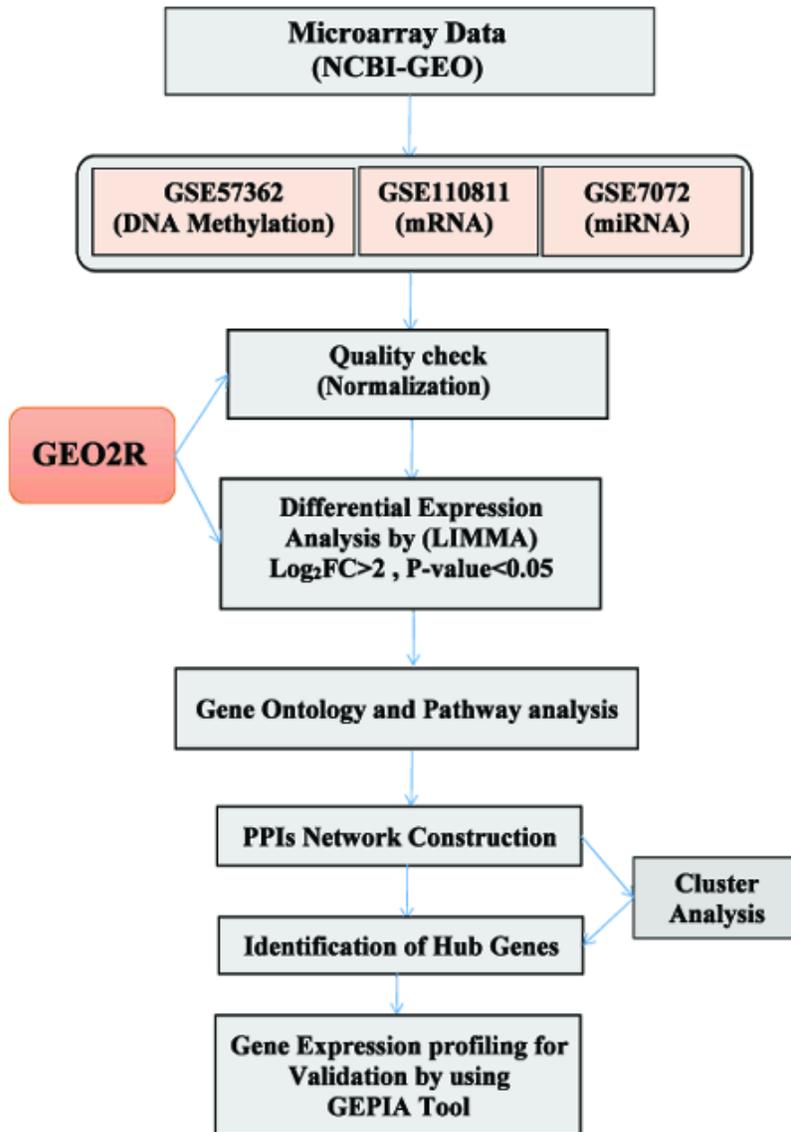


Fig.1. Outline of the workflow executed in microarray studies

Figure 1

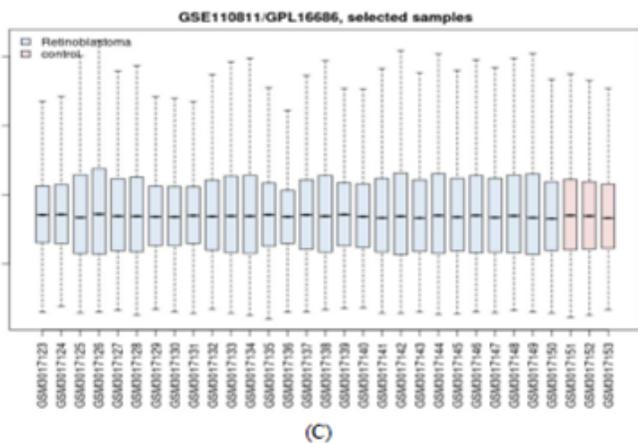
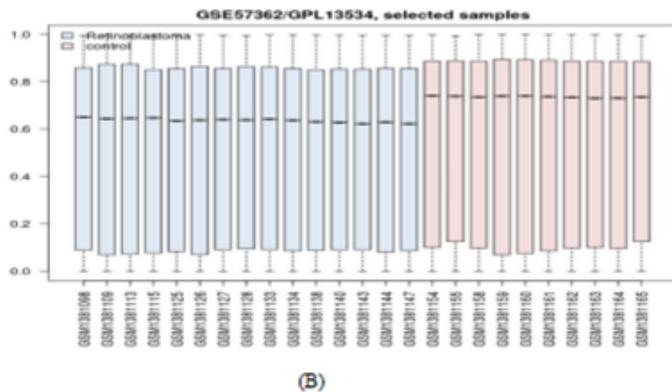
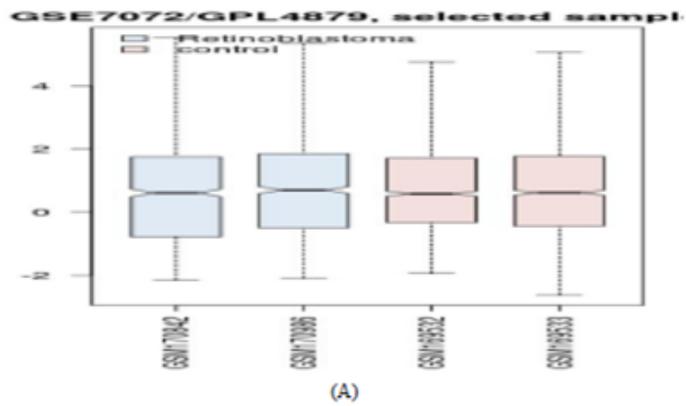


Fig. 2. Normalized pattern of all three datasets (A-C)

Figure 2

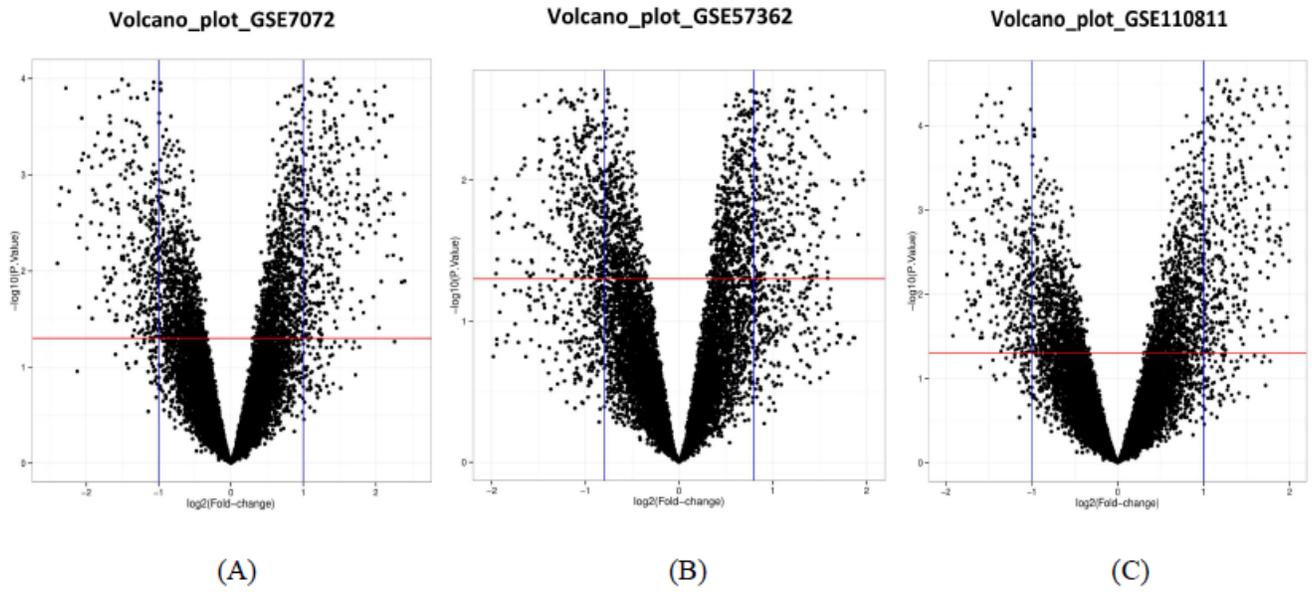


Fig. 3. Volcano plot of the DEGs in the three datasets (A-C)

Figure 3

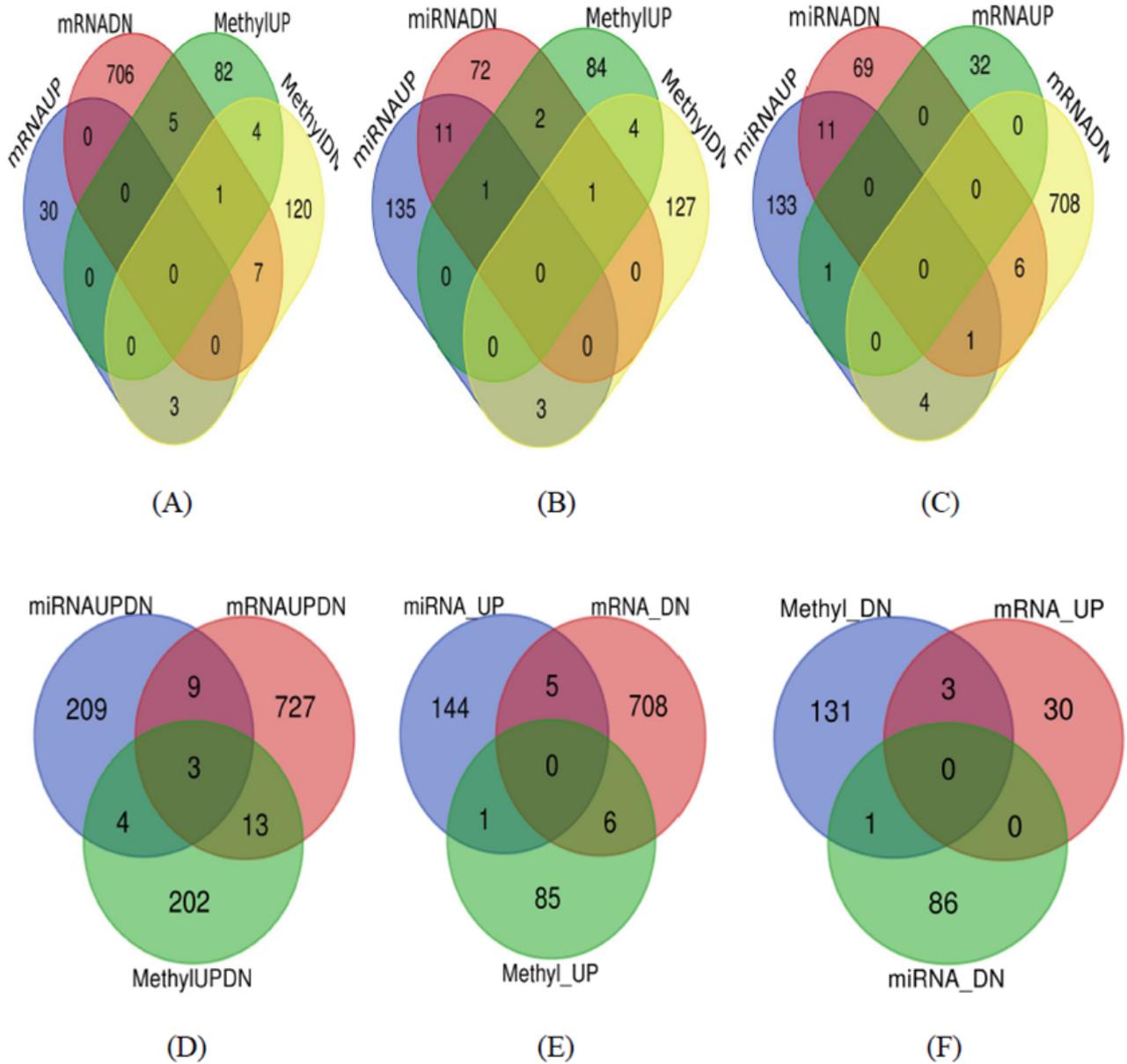


Fig. 4. Venn diagram of the DEGs in the three datasets (A-F). (D) demonstrate three genes are common in upregulated-downregulated miRNA dataset, upregulated-downregulated mRNA dataset and upregulated-downregulated DNA methylation dataset.

Figure 4

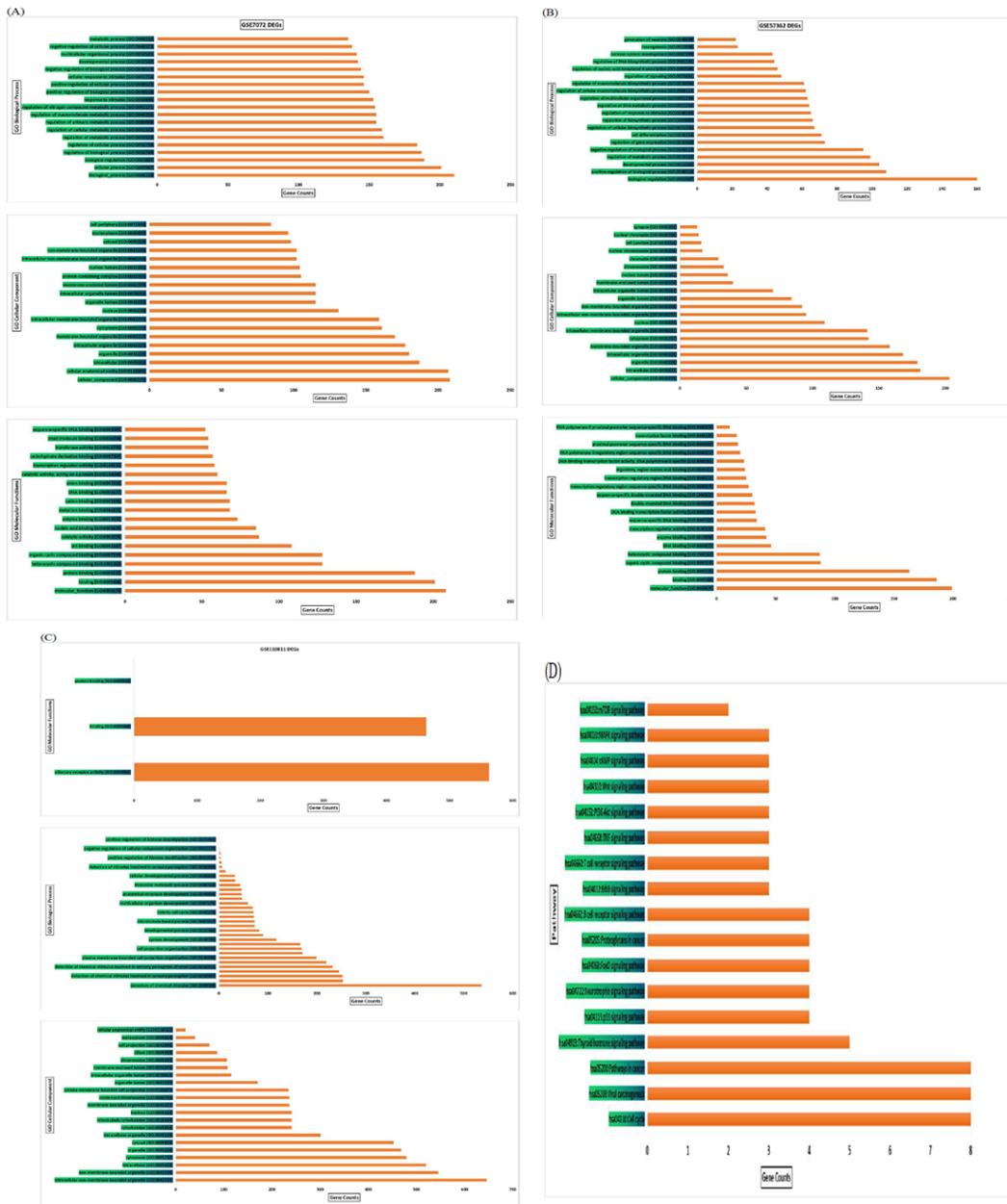


Fig. 5. Functional enrichment analysis. (A), (B), (C) demonstrate enriched terms for the biological process, molecular function and cellular component of the DEGs in the three datasets. (D) demonstrate top 17 significant pathway enrichment terms of the hub genes in the three datasets.

Figure 5

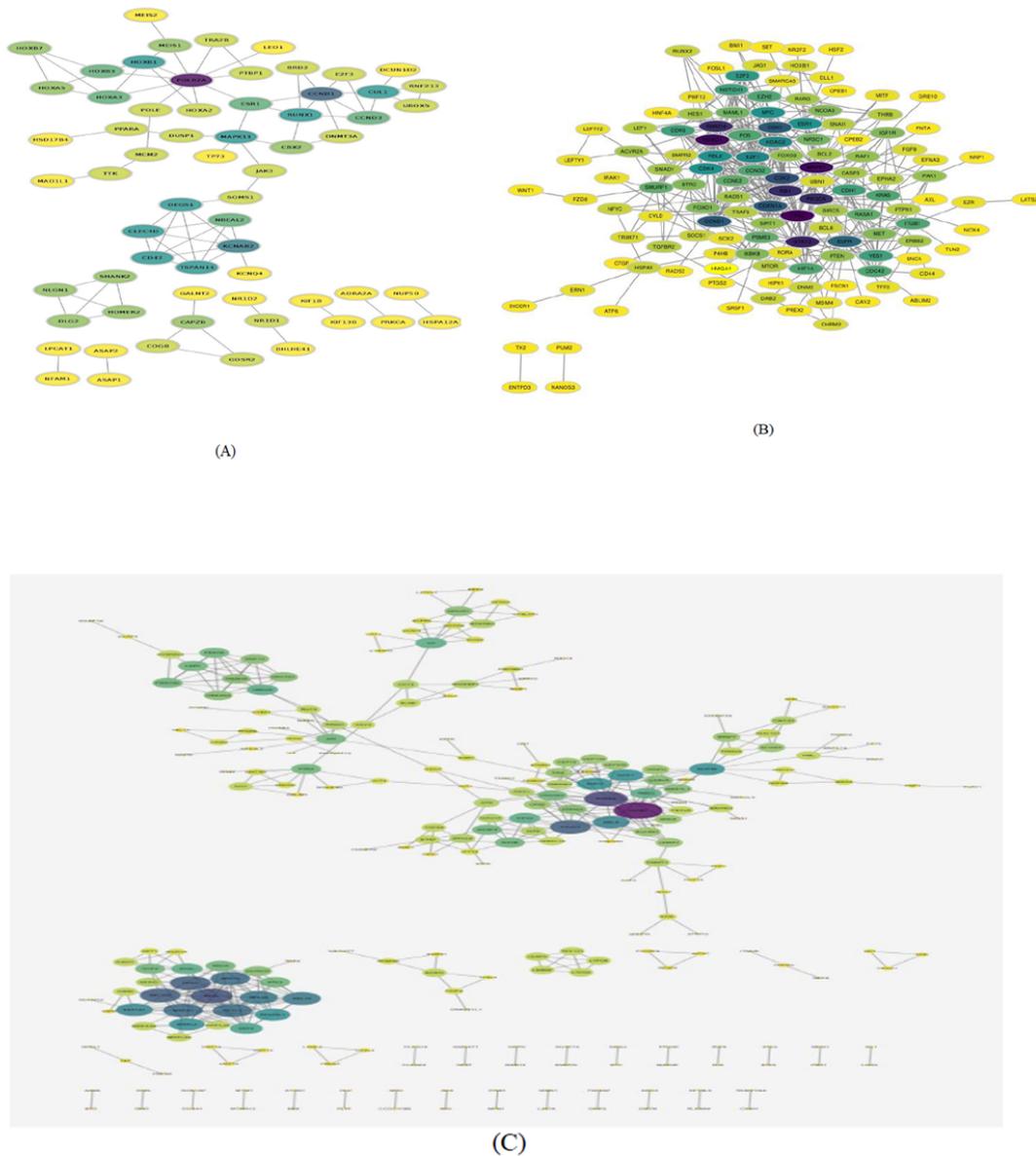


Fig. 6. Co-expression network of DEGs in the three datasets. (A) DNA methylation co-expression network has 63 nodes and 83 interactions, (B) miRNA co-expression network has 129 nodes and 424 interactions and (C) mRNA co-expression network has 243 nodes and 471 interactions.

Figure 6

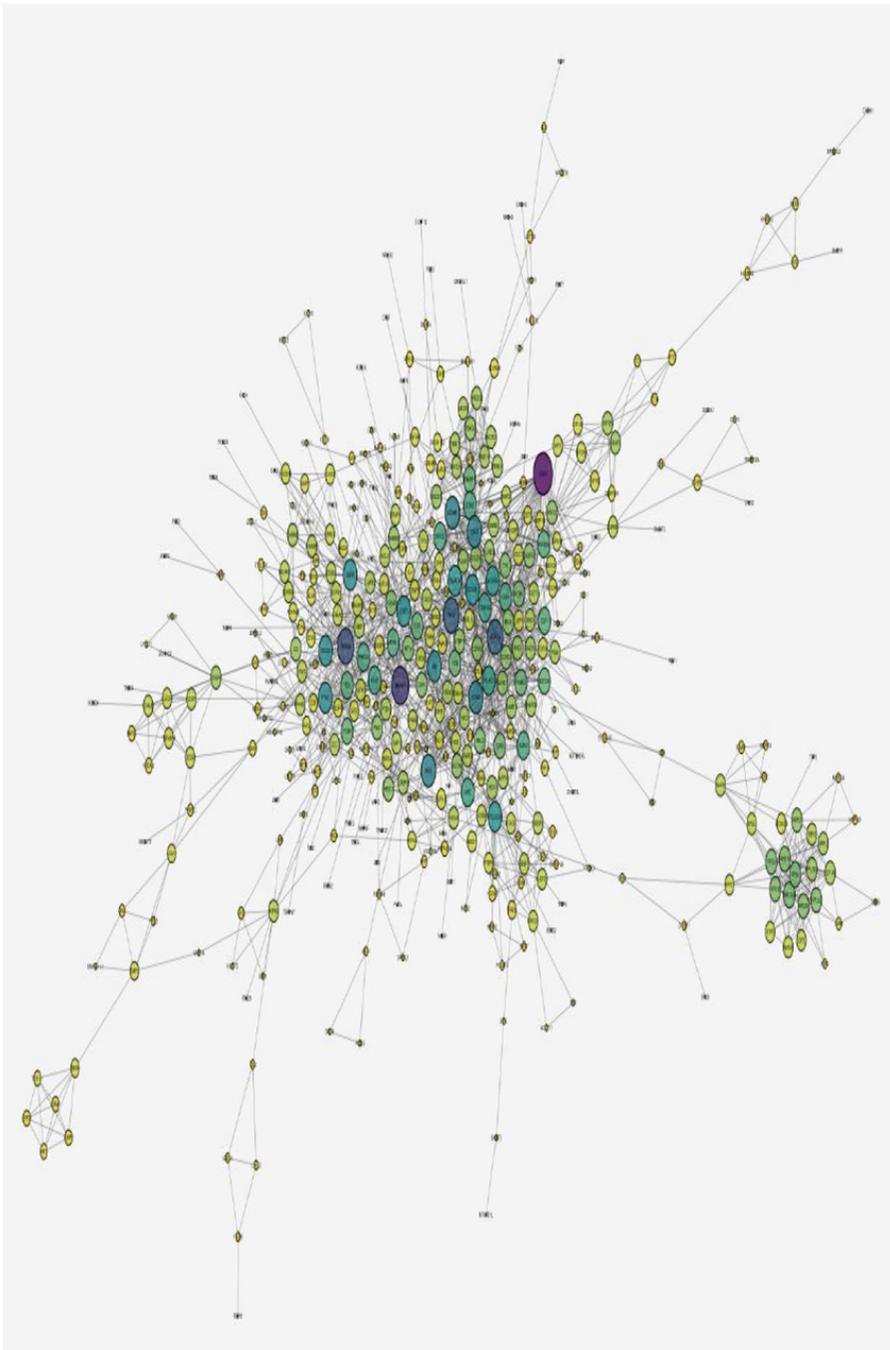


Fig. 7. Intersection network formed by merging all the three individual networks

Figure 7

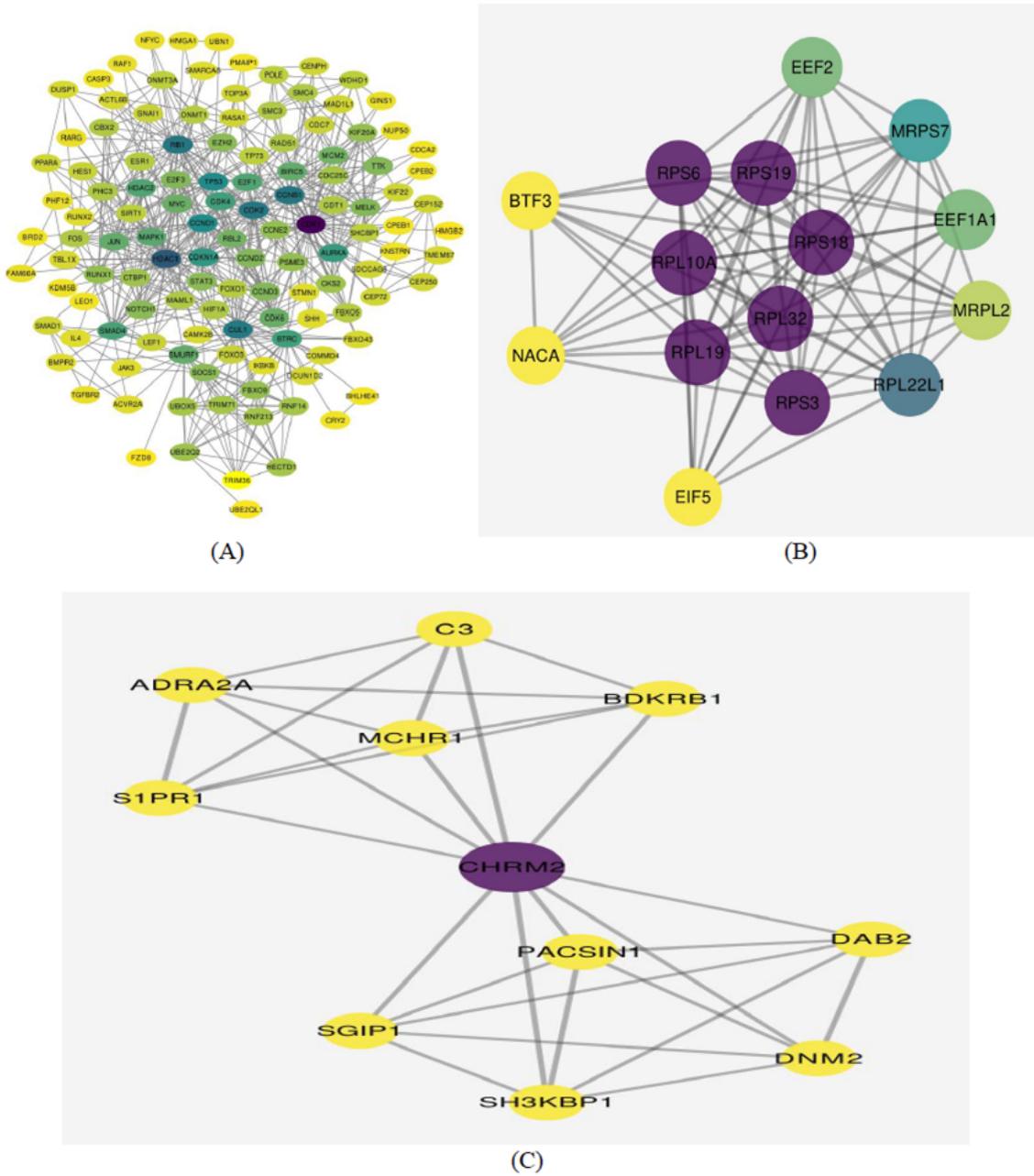


Fig. 8. (A), (B) and (C) represent MCODE clusters

Figure 8

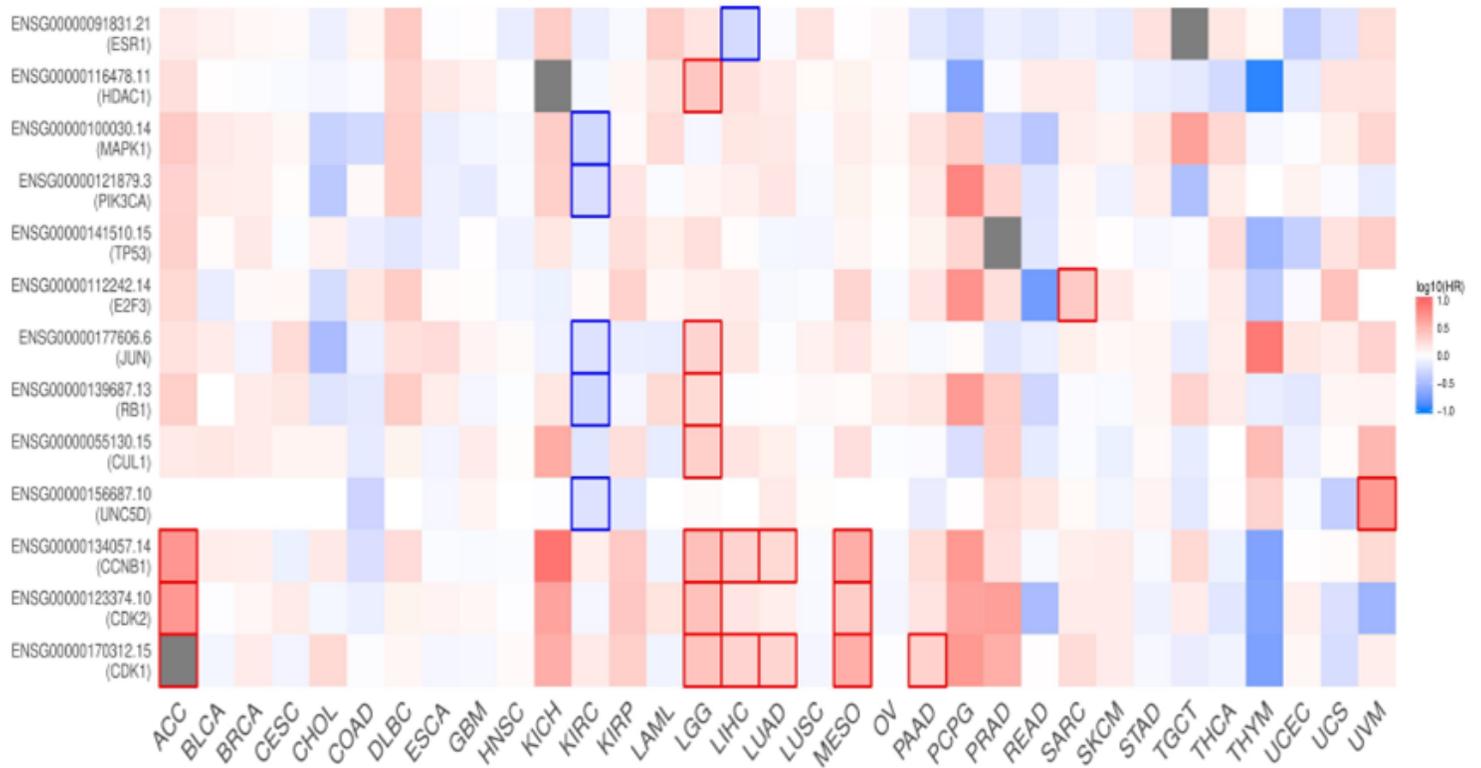


Fig. 9. Survival maps of the hub genes. It was observed that there is significant differential expression of these genes in malignant conditions in comparison to normal conditions.

Figure 9

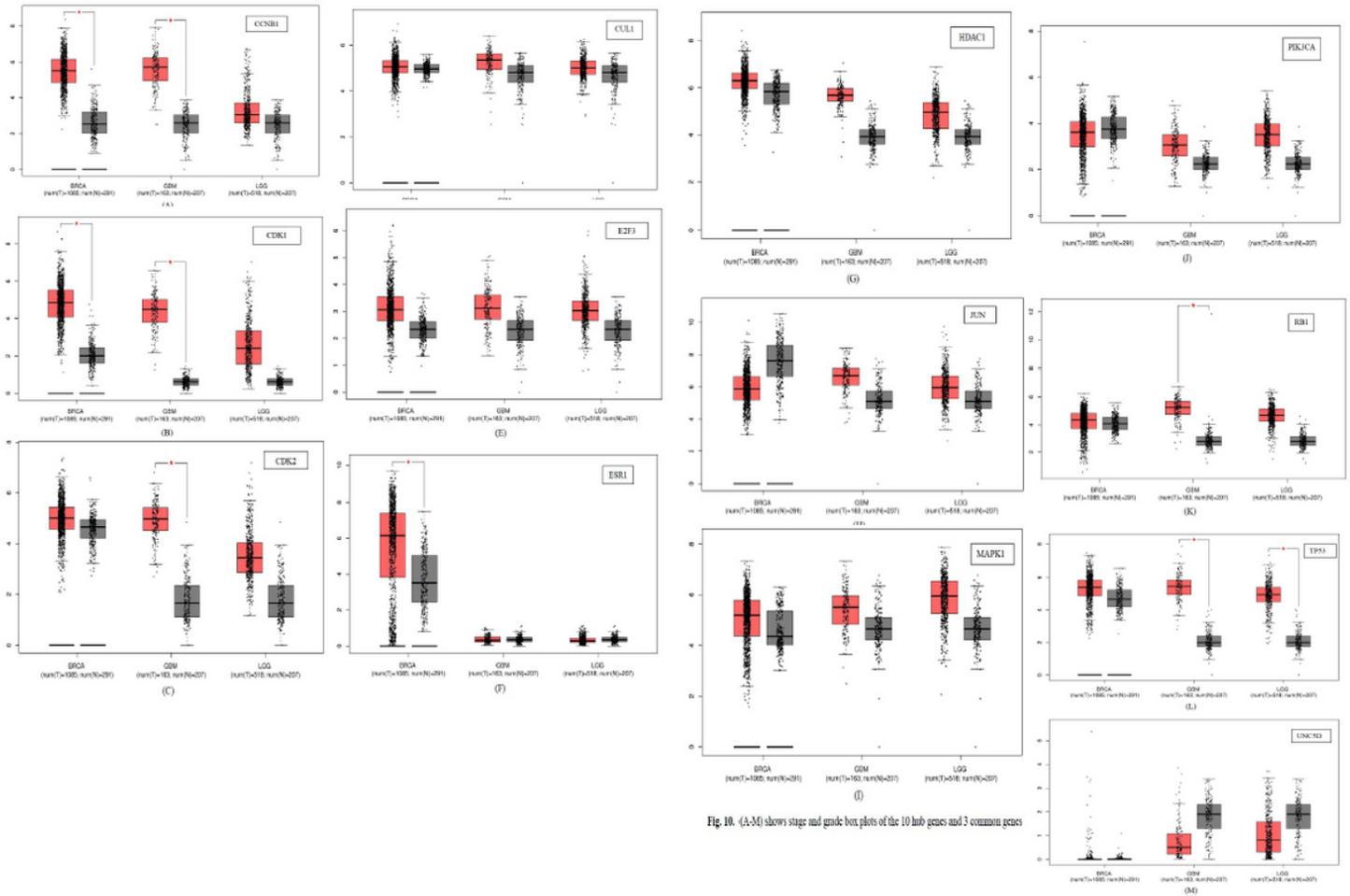


Fig. 10. (A-M) shows stage and grade box plots of the 10 hub genes and 3 common genes

Figure 10

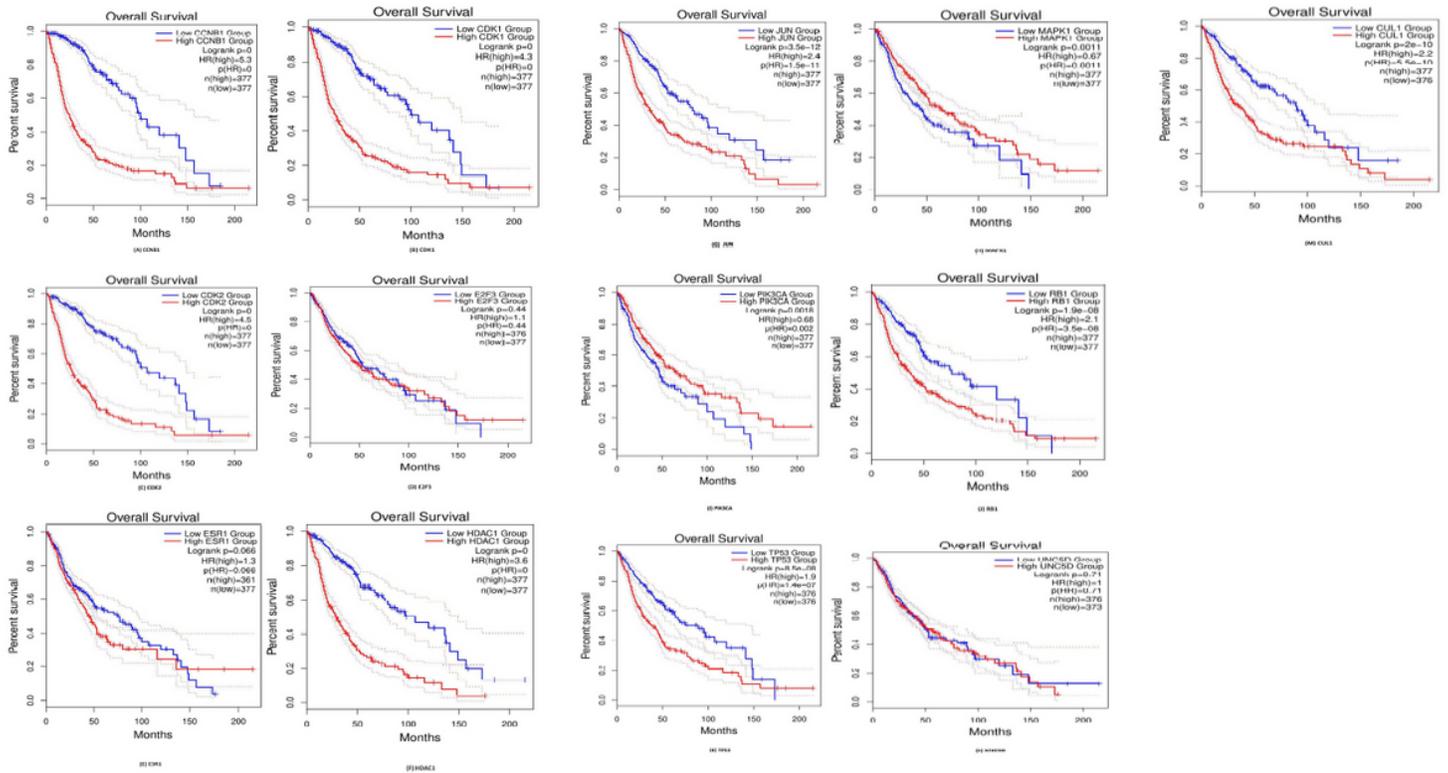


Fig. 11. (A-M) shows survival analysis graphs for each of the 10 hub genes and 3 common genes

Figure 11