STM JOURNALS

International Journal of
Bioinformatics and Computational Biology

Volume 2, Issue 1, 2024
January—June
DOI (Journal): 10.37591/IJBCB

https://journals.stmjournals.com/ijbcb

Research

Gene Prioritization of Colorectal Cancer Using Computational Approach

Pankaj Kumar Tripathi[1], Chakresh Kumar Jain[2,*]

Abstract

Colorectal cancer (CRC) continues to be a major worldwide health issue, underscoring the need to pinpoint crucial genetic elements influencing its initiation and advancement. In this investigation, we utilize sophisticated computational methods to prioritize potential genes linked to CRC development. Through the utilization of various bioinformatics tools and comprehensive methodologies, we systematically examine extensive microarray datasets to identify potential genetic contributors. Initially, a vast amount of CRC microarray data (approximately more than 175 GSE datasets) was collected for gene expression analysis and applied filters such as unavailability of sample vs control, non-significant genes, untreated, non-published data, etc. After applying filters, we selected 25 datasets for further analysis. From these datasets, we find the common upregulated and downregulated genes and perform the network analysis to prioritize the genes based on their topological analysis, functional relevance, and biological interactions in CRC tissues. By integrating information from various sources, and applying different filters, we aim to identify key molecular players who may serve as biomarkers for CRC treatments. After applying our gene prioritizing method, we filter out the top ten genes (CDK1, CCNA2, BUB1B, CCNB1, KIF20A, BUB1, KIF11, CENPF, TOP2A, PLK1) have been identified, which are crucial for the development and proliferation of tumor cells, either directly or indirectly. Although CENPF gene role in CRC is unclear, this study shows it's significantly involved in cell division, spindle assembly checkpoint, chromosome segregation, microtubule binding, chromatin binding, and protein binding. Therefore, our finding proposed CENPF as a novel biomarker for CRC progression. This extensive computational gene prioritization study enhances the current comprehension of the molecular landscape of CRC, offering valuable insights for subsequent experimental validation and potential clinical applications.

Keywords: Colorectal cancer, microarray data, bioinformatics, gene identification, pathway analysis, network analysis

***Author for Correspondence**
Chakresh Kumar Jain
E-mail: ckj522@yahoo.com

[1]Research Scholar, Department of Biotechnology, Jaypee Institute of Information Technology, Uttar Pradesh, India
[2]Assistant Professor, Department of Biotechnology, Jaypee Institute of Information Technology, Uttar Pradesh, India

Received Date: April 30, 2024
Accepted Date: May 13, 2024
Published Date: June 15, 2024

Citation: Pankaj Kumar Tripathi, Chakresh Kumar Jain. Gene Prioritization of Colorectal Cancer Using Computational Approach. International Journal of Bioinformatics and Computational Biology. 2024; 2(1): 39–52p.

INTRODUCTION

Colorectal cancer is a critical public health problem because it is the third most frequent type of cancer and the second most common cause of cancer-related deaths globally [1]. It leads to substantial physical and psychosocial morbidity, contributing to high cancer-related mortality and morbidity rates [2, 3]. The association between serum biomarkers and CRC risk remains an area of active research, with some reported associations lacking statistical significance [4, 5]. Identifying reliable biomarkers for early detection and prognosis of CRC is crucial for public health. Biomarker analysis is essential for improving the

diagnosis and prognosis of illnesses. Numerous potential biomarkers that hold promise for advancing CRC screening, diagnosis, and therapeutic strategies have been identified. Additionally, genome-wide research has revealed potential biomarkers linked to the molecular pathways involved in CRC development, including elevated microsatellite instability (MSI), MLH1 inactivation, and mutations in PIK3CA, KRAS, APC, TP53, and SMAD4 genes [6]. Identification of biomarkers in the gut microbiome is of utmost importance for the early detection and prevention of colorectal cancer [7]. These biomarkers aid in early detection, personalized therapy, and assessment of cancer risk. A comprehensive analysis of these biomarkers and their association with various factors provides valuable information for the diagnosis, prognosis, and personalized treatment of CRC [8].

For this study, we collected many GSE datasets for CRC based on tumor versus normal tissue. We analyzed gene expression using R Packages to explore genes that are highly expressed during CRC progression. We constructed a network of protein-protein interactions (PPI) using the STRING database and Cytoscape (v3.10.1) plugins to identify important key genes (hub genes) [9]. Network topological analysis and centrality measurements were used to ascertain the top ten hub genes. Additionally, we performed pathway analysis and functional annotation to gain insight into the roles of these genes in cell proliferation, metastasis, migration, and oncogenesis [10]. We also performed an in silico gene expression validation using the GEPIA database.

METHODOLOGY

Data Exploration

Gene Expression Omnibus, commonly referred to as GEO, is a publicly available data repository that houses large sets of high-throughput and functional genomics data, facilitating scientific research by allowing researchers to deposit and access valuable genetic information. The CRC expression data were retrieved from the NCBI GEO repository [11] with the following search parameters: (1) disease, colorectal cancer; (2) organism selected, Homo sapiens; (3) type of study, gene expression profile by array; (4) attribute name, tissue; and (5) type of entry, series. After applying the specific parameters, we retrieved 175 GSE datasets for CRC. We further filtered out the 175 GSE datasets based on availability of CRC vs. normal/control, data published in journals, non-treated, data belonging to only mRNA, and we selected only 25 GSE datasets for further study.

Analysis of Differential Gene Expression for CRC

Gene expression analysis for all 25 GSE datasets was performed using R software. The R packages Affy, limma, and other required packages were utilized to identify Differentially Expressed Genes (DEGs) within GEO series data. A p-value ≤ 0.05 and $|\log FC| \geq 1.00$ were utilized to evaluate the statistical significance of each dataset [12]. After performing differential gene expression analysis, we identified 9,645 upregulated and 11,455 downregulated genes from the complete datasets. We selected genes for further analysis that are common in at least five GSE datasets, resulting in 733 upregulated approved genes (HGNC) and 1,088 downregulated genes.

PPI Interaction Networks and Hub Genes Analysis

We constructed a PPI network to study the protein interaction information and identified hub genes using the StringApp plugin in Cytoscape. (v 3.10.1). To scrutinize statistically significant results, the high confidence score was set as ≥ 0.09 for the String Interactome (range of scores between 400 and 1000) [13]. The PPI network's topological analysis, node distribution, and centrality measurement were computed using the Cytoscape plugin Network Analyzer [14] and CytoHubba [15] to identify the top ten hubs inside the network [16]. Subsequently, key PPI network clusters were examined using MCODE analysis in Cytoscape. The advanced options were configured with a K-Core of 2, 0.2 as the node score threshold, and 2 as the degree cutoff [17].

Functional Annotation of DEGs

Functional annotation was performed to gain meaningful insights from a set of DEGs, highlighting the key pathways influenced by changes in gene expression in diseased samples. We utilized the

DAVID database (http://david.ncifcrf.gov) for our analysis, which categorizes the input genes according to the most impacted activities, such as biological activities (BP), Molecular Functions (MF), and cellular components (CC) [18]. Furthermore, we conducted KEGG pathway analysis using the EnrichR database, a comprehensive resource of molecular pathways offering detailed information on functional and biological systems. A p-value filter of ≤ 0.05 was applied to ensure the exclusion of results with low confidence levels [19].

Validation of Putative Key Genes (Expression Analysis)

To ascertain the levels of hub gene expression in COAD patients and healthy colons, GEPIA 2.0 (http://gepia2.cancer-pku.cn/) was used. GEPIA, a database focused on analyzing the cancer-specific data from both the Genotype-Tissue Expression (GTEx) and TCGA databases, offers a feature known as "expression DIY," enabling users to dynamically plot gene expression profiles based on their chosen samples and methodologies, and provides the results in dot plots or box plots [20].

RESULT AND DISCUSSION

Data Explored for Analysis

We explored the gene expression data of CRC using specific search parameters and extracted 175 GSE datasets for CRC. We further applied filters based on the availability of CRC versus normal/control comparisons, publication in journals, non-treated conditions, and data limited to mRNA. Ultimately, 25 GSE datasets were selected for further analyses (Table 1).

Table 1. All 25 analyzed dataset with the number of DEGs.

S.N.	GSE Data Set	DEGs	Upregulated Genes	Downregulated Genes
1.	GSE198383	1325	153	1172
2.	GSE185770	1034	210	824
3.	GSE164191	162	70	92
4.	GSE161533	1580	804	776
5.	GSE118949	3966	1605	2361
6.	GSE145626	640	520	120
7.	GSE110224	822	359	463
8.	GSE110223	593	245	348
9.	GSE122183	557	192	365
10.	GSE97689	25	14	11
11.	GSE113513	1863	758	1105
12.	GSE81804	392	120	272
13.	GSE73360	3590	629	2961
14.	GSE73883	1	0	1
15.	GSE41328	1321	608	713
16.	GSE37283	2337	409	1928
17.	GSE32323	2606	1252	1354
18.	GSE29638	321	263	58
19.	GSE22598	2550	1217	1333
20.	GSE33114	7746	5434	2312
21.	GSE24551	1610	370	1240
22.	GSE24514	783	425	358
23.	GSE20916	2551	1189	1362
24.	GSE8671	3409	1010	2399
25.	GSE100179	638	361	277

Analysis of Differentially Expressed Genes

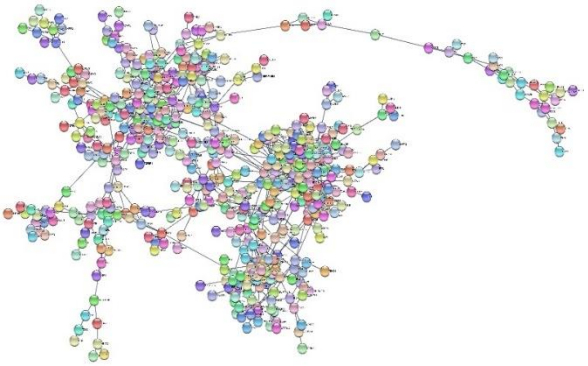
Expression analysis was performed with the standards of False Discovery Rate (FDR) <0.05, |logFC| ≥1, and 21,100 DEGs were obtained from twenty-five GSE datasets. Of the 21,100 differentially expressed genes, 9,645 were upregulated and 11,455 were downregulated. Of the 25 GSE datasets, DEGs common to at least five GSE datasets were selected for further analysis, resulting in the discovery of 1821 DEGs. Among these, 1,088 genes were downregulated, and 733 genes were upregulated, all of which are approved genes in the HUGO gene nomenclature.

PPI Interaction Result

Using the StringApp of Cytoscape, we examined protein-protein interactions to better understand the functional relationships between genes and proteins. For 1821 common DEGs, a PPI network was built with a confidence score of >0.90. After curating the network by removing isolated nodes, a protein-protein interaction network was established with 1,421 nodes and 2,296 edges (Figure 1(a)). A statistical summary of the PPI network is shown in Figure 1(b). Because the network was extremely dense, the clusters of the network were created using the MCODE plugin of Cytoscape, and the top three are shown in Figure 1(c), (d), and (e).

Topology and Hub Genes Analysis of Upregulated Genes

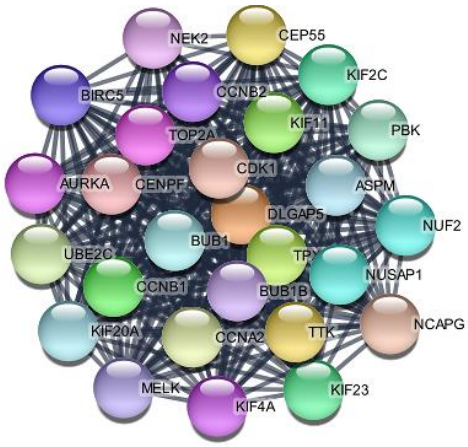
To evaluate the topological properties of the network and perform hub gene analysis, we used Cytoscape plugins such as CytoHubba and Network Analyzer. We examined metrics such as closeness centrality, degree distribution, eigenvector centrality, and betweenness centrality to analyze the network topology [14, 15]. The detailed topological parameters with statistical values are listed in Table 2. Through the evaluation of these topological measures, we identified the top ten hub genes (Figure 1 (F)) and elucidated their involvement in biological functions, molecular interactions, and cellular processes [15]. Additionally, this analysis provides insights into potential biomarkers for colorectal cancer.



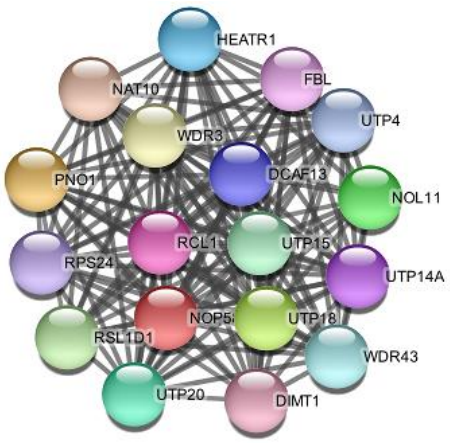
(a) PPI Interaction Network

String interaction with a confidence score of 0.90 summary statistics	
Number of nodes	1421
Number of edges	2296
Avg. number of neighbours	7.042
Network diameter	23
Network radius	12
Characteristic path length	6.566
Clustering coefficient	0.409
Network density	0.011
Network heterogeneity	1.237
Network centralization	0.082
Connected components	711
Analysis time (sec)	0.079

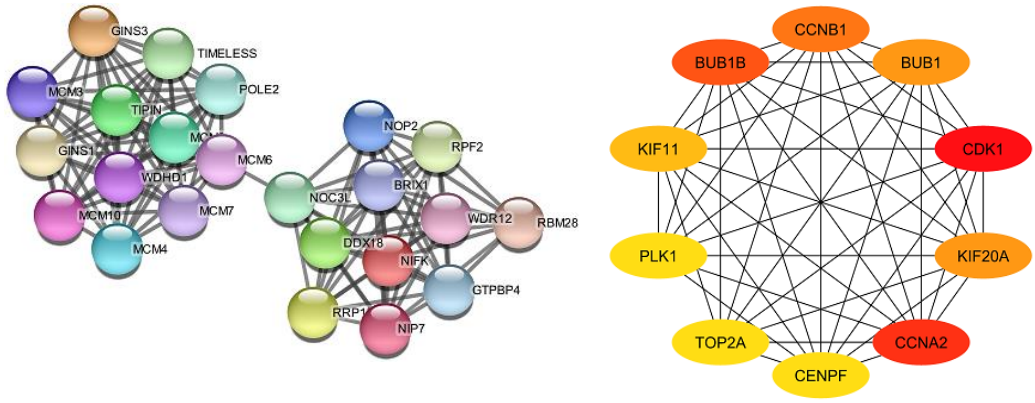
(b) PPI Interaction summary statistics



(c) Cluster 1



(d) Cluster 2



(e) Cluster 3 (f) Hub Genes Degree Sorted

Figure 1. PPI interaction network with clusters and hub gene network analysis performed using Cytoscape, where (a) Represents the PPI Interaction network, (b) Summary Statistics of PPI network, (c), (d), and (e) represents Cluster 1, 2, 3 simultaneously, and (f) Represents the top ten hub gene degree sorted network.

Table 2. Topological analysis with centrality measurement.

Gene symbol name	Degree	Betweenness centrality	Closeness centrality	Eigenvector centrality	Clustering coefficient
CDK1	58	0.113112056	0.232758621	0.212498993	0.319419238
CCNA2	54	0.068450193	0.231543624	0.201394111	0.341719078
BUB1B	50	0.026788153	0.210152284	0.196012035	0.371428571
CCNB1	48	0.036988748	0.224674385	0.203411773	0.432624113
KIF20A	43	0.011264352	0.216451725	0.203274056	0.548172757
BUB1	43	0.02274789	0.216075157	0.189900324	0.473975637
KIF11	41	0.010615632	0.21630094	0.198740348	0.57804878
CENPF	38	0.015312738	0.213548831	0.182375491	0.566145092
TOP2A	38	0.009420871	0.215325936	0.189079791	0.611664296
PLK1	38	0.038037204	0.205221414	0.139257342	0.332859175

Gene Ontology Enrichment Analysis

Gene ontology was performed using the DAVID database for the top ten hub genes, which revealed enriched functional categories for these genes. The most significant GO term in BP was cell division, with the highest number of genes involved and the lowest p-value (4.28618E-11). Some genes are associated with processes such as the mitotic spindle assembly checkpoint, G2/M transition of the mitotic cell cycle, chromosome segregation, and the mitotic cell cycle, among others. Regarding CC, the genes were associated with functions including condensed chromosome outer kinetochore, spindle, centrosome, nucleus, spindle pole, kinetochore, midbody, spindle microtubule, nucleoplasm, and the cyclin A2-CDK1 complex. The genes included in MF are engaged in various activities such as protein kinase binding, chromatin binding, microtubule motor activity, cyclin-dependent protein serine/threonine kinase regulator activity, and protein kinase activity [21]. The top 10 hub genes (p-value < 0.05) are listed in the comprehensive gene ontology table; these genes are denoted as MF in Table 3, BP in Table 4, and CC in Table 5.

KEGG Pathway Analysis

To identify the key pathways and molecular components of the top ten hub genes, we utilized EnrichR software. Enriched pathway analysis identified genes majorly associated with the P53 signaling pathway, cell cycle, progesterone-mediated oocyte maturation, viral carcinogenesis, and FoxO signaling pathway. As demonstrated in Figure 2, a p-value<0.05 is the cutoff condition for the significance of the enriched pathway analysis.

Table 3. Molecular function (MF).

S.N.	GO-MF-Id	Term	Count	%	p-value	Genes
1.	GO:0005524	ATP binding	7	70	1.97319E-05	TOP2A, PLK1, CDK1, BUB1B, KIF20A, KIF11, BUB1
2.	GO:0019901	Protein kinase binding	5	50	4.81005E-05	CCNA2, CCNB1, PLK1, KIF20A, KIF11
3.	GO:0008017	Microtubule binding	4	40	0.000232858	CENPF, PLK1, KIF20A, KIF11
4.	GO:0004672	Protein kinase activity	4	40	0.000644734	PLK1, CDK1, BUB1B, BUB1
5.	GO:0004674	Protein serine/threonine kinase activity	4	40	0.000654545	PLK1, CDK1, BUB1B, BUB1
6.	GO:0016538	Cyclin-dependent protein serine/threonine kinase regulator activity	2	20	0.015152024	CCNA2, CCNB1
7.	GO:0003682	Chromatin binding	3	30	0.021192356	TOP2A, CENPF, CDK1
8.	GO:0003777	Microtubule motor activity	2	20	0.027777442	KIF20A, KIF11
9.	GO:0005515	Protein binding	10	100	0.028271299	TOP2A, CCNA2, CENPF, CCNB1, PLK1, CDK1, BUB1B, KIF20A, KIF11, BUB1
10.	--	--	--	--	--	--

Table 4. Biological processes (BP).

S.N.	GO-BP-Id	Term	Count	%	p-value	Genes
1.	GO:0051301	Cell division	8	80	4.28618E-11	CCNA2, CENPF, CCNB1, PLK1, CDK1, BUB1B, KIF11, BUB1
2.	GO:0007094	Mitotic spindle assembly checkpoint	4	40	2.84829E-07	CENPF, PLK1, BUB1B, BUB1
3.	GO:0000086	G2/M transition of mitotic cell cycle	4	40	1.36862E-06	CCNA2, CCNB1, PLK1, CDK1
4.	GO:0007052	Mitotic spindle organization	3	30	0.00029516	CCNB1, PLK1, KIF11
5.	GO:0007059	Chromosome segregation	3	30	0.000865542	TOP2A, CENPF, BUB1
6.	GO:0000278	Mitotic cell cycle	3	30	0.001750783	CENPF, PLK1, KIF11
7.	GO:0044772	Mitotic cell cycle phase transition	2	20	0.01023776	CCNA2, CCNB1
8.	GO:0051754	Meiotic sister chromatid cohesion, centromeric	2	20	0.001868382	BUB1B, BUB1
9.	GO:0007091	Anaphase transition of mitotic cell cycle/Metaphase	2	20	0.003267632	PLK1, BUB1B
10.	GO:1905448	Positive regulation of mitochondrial ATP synthesis coupled electron transport	2	20	0.001868382	CDK1, CCNB1

Table 5. Cellular component (CC).

S.N.	GO-CC-Id	Term	Count	%	p-value	Genes
1.	GO:0000940	Condensed chromosome outer kinetochore	5	50	5.62126E-12	CENPF, CCNB1, PLK1, BUB1B, BUB1
2.	GO:0005819	Spindle	5	50	3.35867E-07	CENPF, PLK1, BUB1B, KIF20A, KIF11
3.	GO:0005813	Centrosome	6	60	3.58647E-06	CCNA2, CENPF, CCNB1, PLK1, CDK1, BUB1B
4.	GO:0005634	Nucleus	10	100	1.6047E-05	TOP2A, CCNA2, CENPF, CCNB1, PLK1, CDK1, BUB1B, KIF20A, KIF11, BUB1
5.	GO:0000922	Spindle pole	4	40	2.17291E-05	CENPF, CCNB1, PLK1, KIF11
6.	GO:0000776	Kinetochore	4	40	3.50044E-05	CENPF, PLK1, BUB1B, BUB1
7.	GO:0030496	Midbody	4	40	6.00768E-05	CENPF, PLK1, CDK1, KIF20A
8.	GO:0005876	Spindle microtubule	3	30	0.000125638	PLK1, CDK1, KIF11
9.	GO:0005654	Nucleoplasm	8	80	0.000252046	TOP2A, CCNA2, CENPF, CCNB1, PLK1, CDK1, KIF20A, BUB1
10.	GO:0097122	Cyclin A2-CDK1 complex	2	20	0.000876979	CCNA2, CDK1

Gene Expression Validation

The key gene expression levels in CRC patients and healthy individuals were investigated using GEPIA. Our findings were confirmed by the GEPIA results, which showed that the expression levels of all 10 hub genes were significantly higher in COAD samples than in healthy tissues. Notably, these findings were statistically significant ($P < 0.05$) (Figure 3). The results are depicted in box plots that visually represent the expression trends. Overall, the GEPIA expression analysis results offer valuable insights into the expression patterns and potential functions of these pivotal hub genes in CRC.

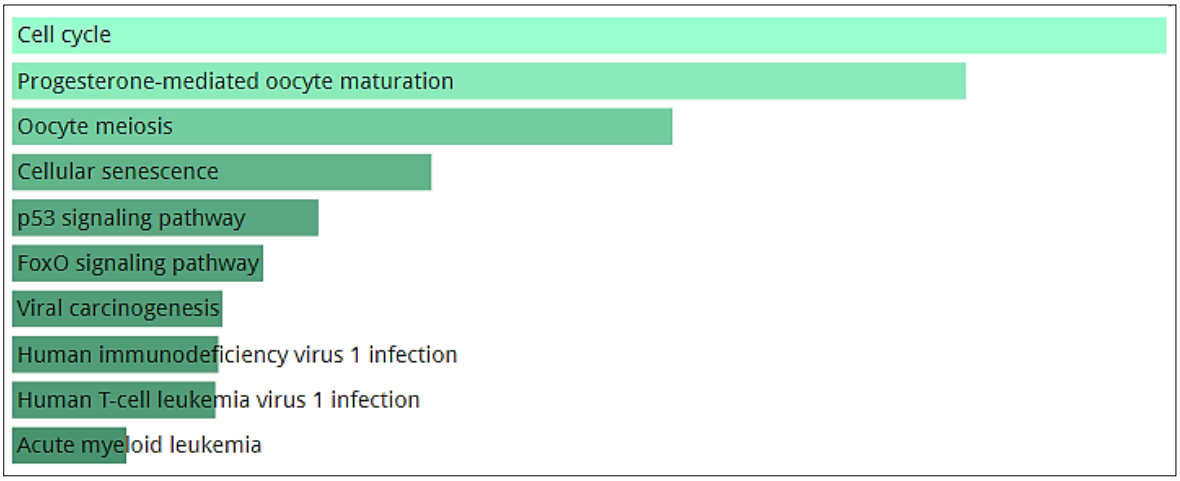
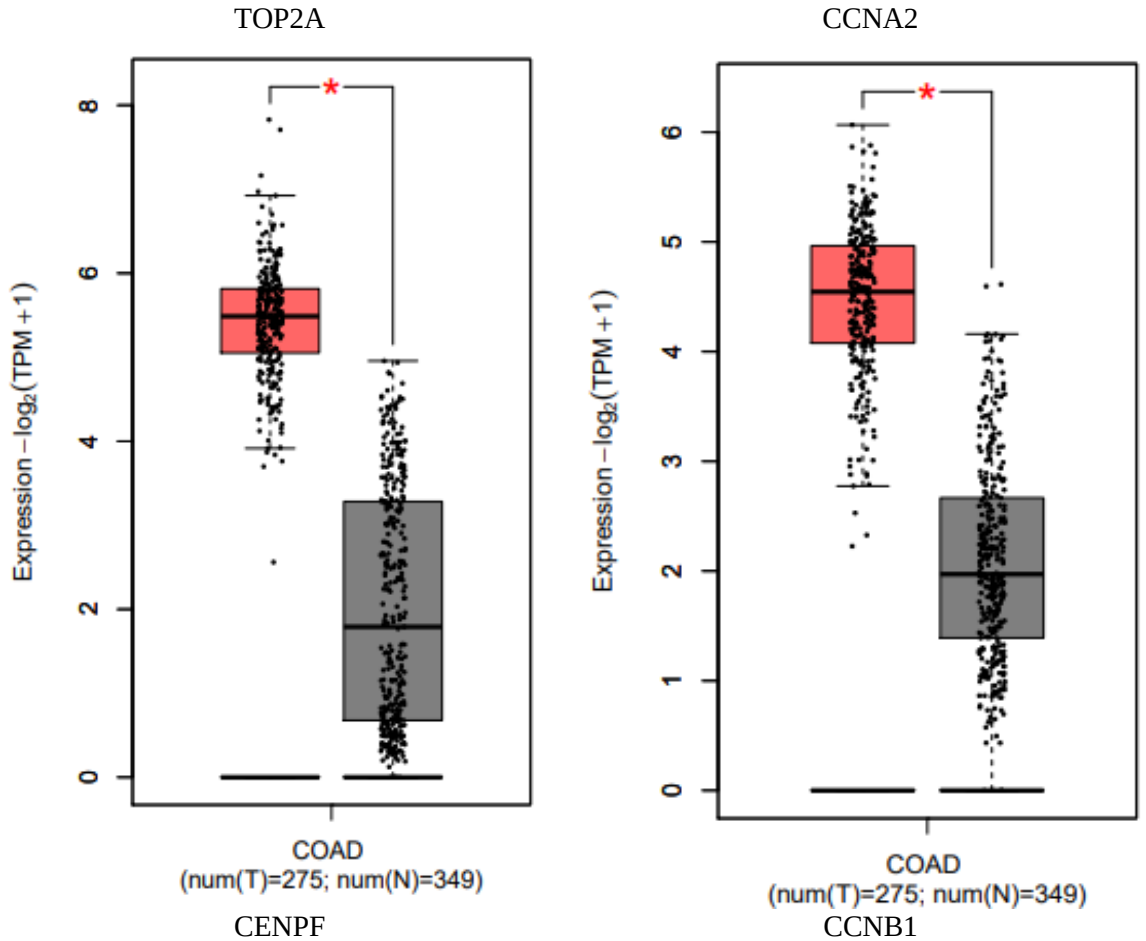
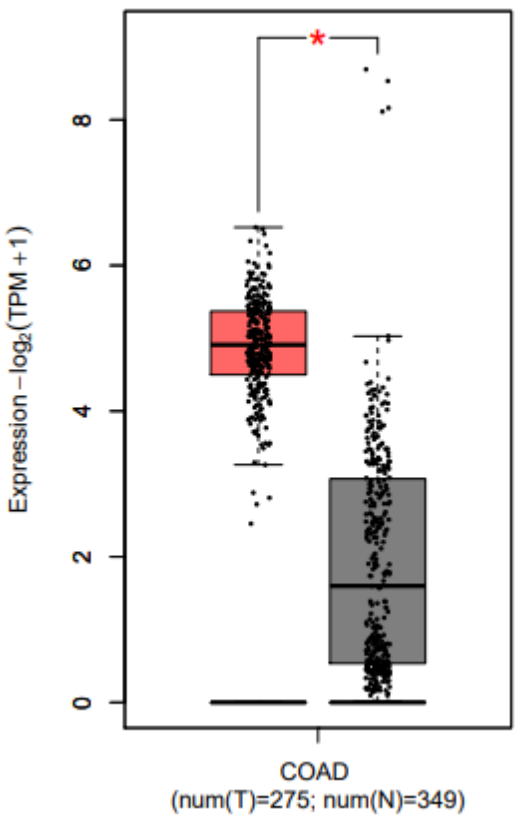
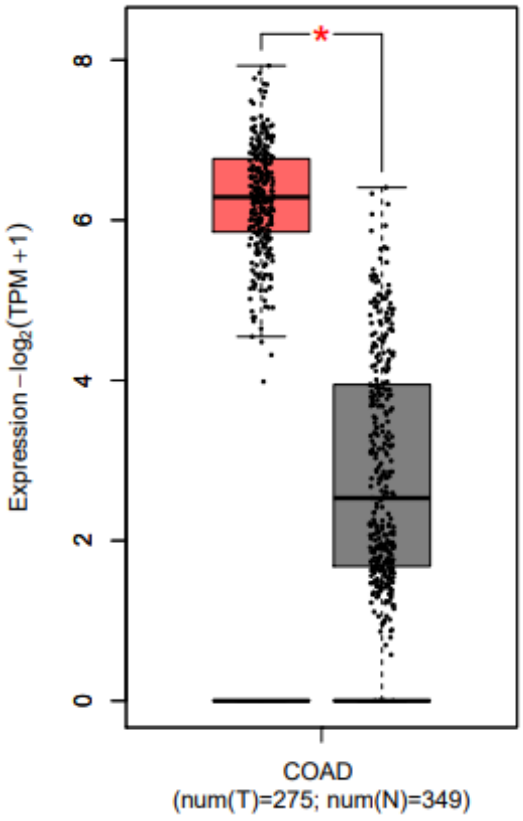


Figure 2. Unveiling molecular pathways: exploring hub genes through KEGG analysis with EnrichR.

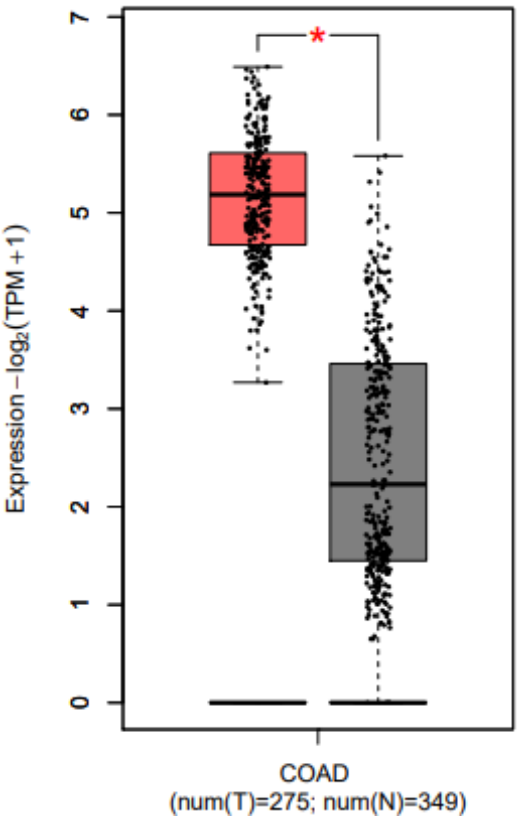




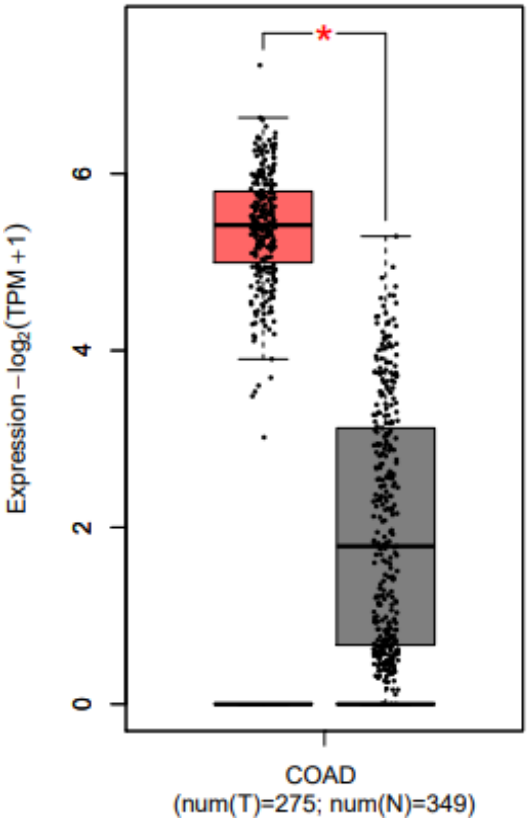
PLK1



CDK1



BUB1B



KIF20A

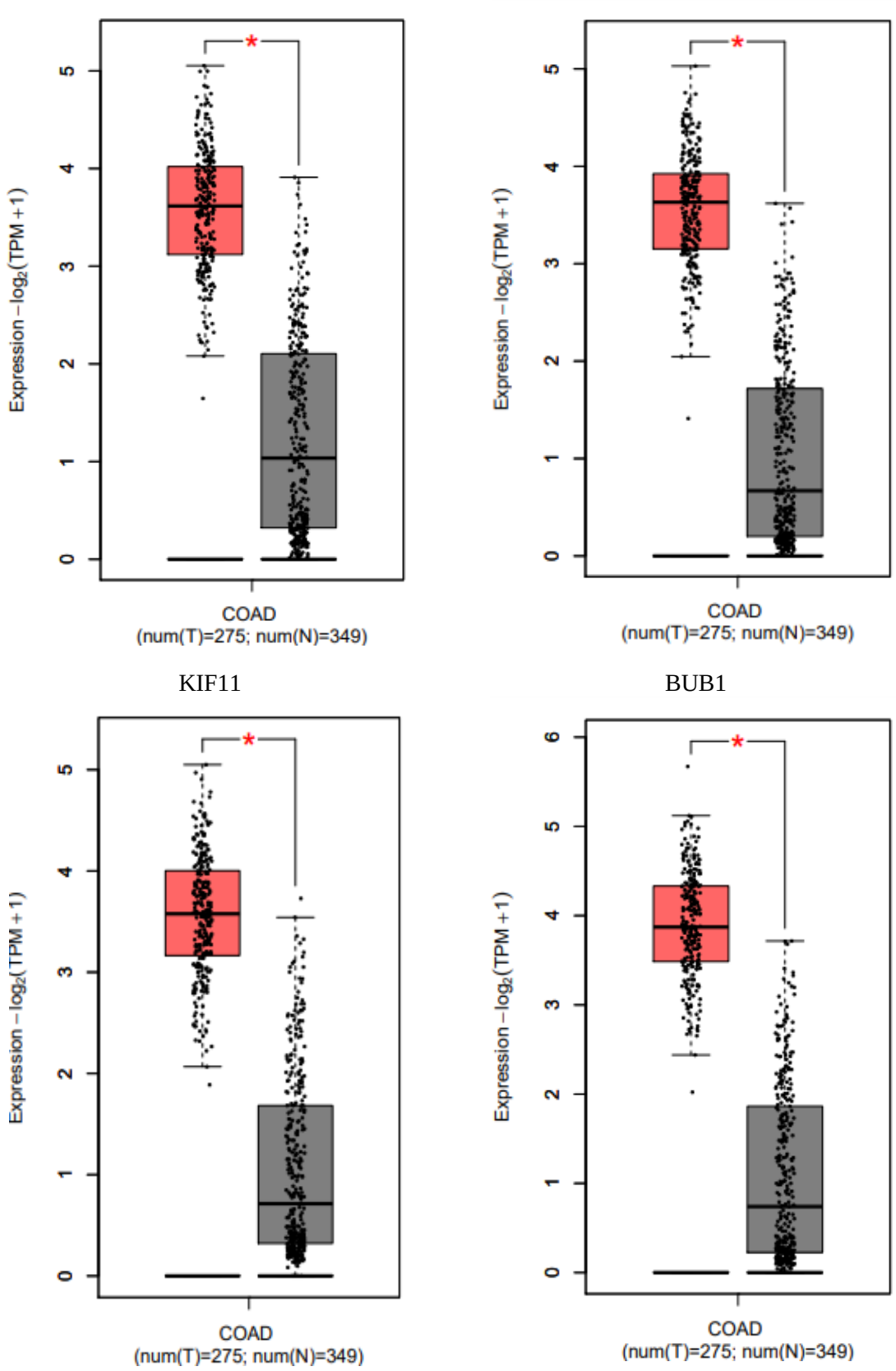


Figure 3. Validation of hub gene expression in COAD patients vs. healthy controls using TCGA and GTEx data in GEPIA. Remarkably, all hub genes exhibited significant overexpression in 275COAD patients compared to 349 healthy controls.

DISCUSSION

Colorectal cancer (CRC) presents a substantial public health challenge, underscoring the importance of identifying biomarkers for its early detection. Numerous studies have explored potential CRC biomarkers using diverse methodologies [22]. In this study, we explored potential biological markers using an in silico bulk study of microarray data. Here, we performed DGE analysis for 25 microarray GSE datasets and explored the differentially expressed genes that were common in at least five GSE datasets. Further PPI interaction networks were analyzed, and the top ten hub genes CDK1, CCNA2, BUB1B, CCNB1, KIF20A, BUB1, KIF11, CENPF, TOP2A, and PLK1, which are functionally enriched in the cell cycle, cellular science, P53 signaling pathway, were found to be responsible for CRC progression. Several studies have emphasized the importance of these hub genes in the prognosis, diagnosis, and treatment of CRC. Zhang et al. (2018) demonstrated that CDK1 plays a role in apoptosis resistance in CRC, implying its participation in cancer cell survival and treatment resistance [23]. A study reported that CDK1 is linked to the advancement of cell proliferation and survival in CRC by phosphorylating and inhibiting the FOXO1 transcription factor, underscoring its significance in cancer progression [24]. CCNA2, belonging to the cyclin family, has been consistently found to be upregulated in CRC tissues, suggesting its potential involvement in the transformation and progression of cancer. Reducing CCNA2 levels substantially hinders CRC cell growth by disrupting cell cycle advancement and prompting apoptosis [25, 26]. BUB1B is linked to unfavorable survival and metastasis in CRC, and its low expression correlates with poor outcomes in human colon adenocarcinomas. Moreover, BUB1B has been identified as a potential therapeutic target and prognostic factor in CRC, highlighting its importance in this disease [27]. Studies have demonstrated that CCNB1 drives cell proliferation, tumor expansion, and cancer recurrence. It is also linked to unfavorable overall survival and disease-free survival in patients with hepatocellular carcinoma [28]. CCNB1 has also been identified as a promoter of cell proliferation and tumor growth in human colorectal cancer, emphasizing its crucial involvement in the development and progression [29]. KIF20A is crucial for promoting cell proliferation, chemotherapy resistance, and tumor advancement in colorectal cancer [30]. Elevated expression of KIF20A has been linked to adverse clinical outcomes and diminished survival across different cancer types, suggesting its potential as a prognostic indicator [31, 32]. BUB1 is a mitotic checkpoint serine/threonine kinase and overexpression of BUB1 is associated with several types of cancer, including CRC. BUB1 correlates with cancer progression and stem cell potential, with mutations present in a subset of CRC cell lines showing chromosomal instability and progression [33, 34]. KIF11 is an oncogene that promotes invasion, proliferation, and self-renewal in glioblastoma, breast cancer, and other cancer types [35]. KIF11 is a chromosome instability gene in colorectal cancer cell lines, indicating that it may play a role in fostering genomic instability. Furthermore, KIF11 contributes to the maintenance of cancer stem cells by improving the ability of breast cancer cells to self-renew via engagement in the Wnt/β-catenin pathway [36, 37]. TOP2A plays a significant role in cancer progression and prognosis, particularly in colorectal cancer. Studies have shown its correlation with the metastasis-free interval and treatment response in breast cancer [38]. Moreover, TOP2A's increased expression is associated with better treatment responses in CRC and breast malignancies, suggesting its potential as a biomarker for treatment efficacy [39]. Its importance in controlling cancer cell development and death underscores its potential as a biomarker for CRC detection and treatment [40]. Shin et al. (2016) demonstrated PLK1's essential role in mitosis timing and CDK1-cyclin B complex activity, which is crucial for cell entry into the M phase [41]. PDK1 phosphorylates PLK1, which in turn phosphorylates MYC, leading to oncogenic transformation in CRC [42]. PLK1 knockdown synergistically inhibits CRC cell growth. Additionally, PLK1 drives MEK inhibitor resistance in KRAS-mutant colon cancer cells [43]. The CENPF gene has not been clearly described for its role in CRC. However, in this study, CENPF was significantly enriched in gene ontology analysis with functions such as cell division, mitotic spindle assembly checkpoint, chromosome segregation, microtubule binding, chromatin binding, and protein binding. Therefore, our findings revealed the possibility of using this gene through network analysis in the progression of CRC.

CONCLUSION

The primary objective of this study was to identify potential biomarkers from a comprehensive dataset and prioritize candidate genes for further investigation. Through integrated bioinformatics analysis, we identified 1,821 DEGs. Among these, ten hub genes, CDK1, CCNA2, BUB1B, CCNB1, KIF20A, BUB1, KIF11, CENPF, TOP2A, and PLK1were highlighted, offering promising avenues for understanding the pathogenesis, prognosis, and therapeutic strategies for CRC. Of the ten hub genes identified, nine were recognized as potential markers in various cancer types, including CRC. However, the remaining hub gene, CENPF, has not been investigated in the context of CRC. Our study proposes CENPF as a novel candidate biomarker that is potentially implicated in CRC progression. While these findings suggest that these hub genes may be useful as prognostic and diagnostic indicators for CRC, it is essential to acknowledge that our results are based solely on computational analysis. Therefore, subsequent validation using animal models and clinical trials is imperative to corroborate these observations.

Abbreviation

GEPIA: gene expression profiling interactive analysis

CRC: colorectal cancer

COAD: colon adenocarcinoma

PPI: protein-protein interaction

MCODE: molecular complex detection

Data Availability

The data used for the study was obtained from publicly available databases, such as GEO, NCBI, and publicly accessible datasets.

Acknowledgments

We are thankful for the amenities provided by the Department of Biotechnology at the Jaypee Institute of Information Technology in Noida, India.

Funding

None.

Declaration

Ethics: Not Applicable
Conflict of Interest: No interest.

REFERENCES

1. Jayarathna DK, Rentería ME, Kho PF, Batra J, Gandhi NS. Dehydroepiandrosterone sulfate and colorectal cancer risk: a Mendelian randomization analysis. Twin Res Hum Genet. 2022;25:180–6. DOI: 10.1017/thg.2022.31.
2. Dunn J, Lynch B, Rinaldis M, Pakenham K, McPherson L, Owen N, et al. Dimensions of quality of life and psychosocial variables most salient to colorectal cancer patients. Psychooncology. 2006;15:20–30. DOI: 10.1002/pon.919.
3. Han W, Xing W, Wang K, Wang B, Bai K. Alisol A attenuates malignant phenotypes of colorectal cancer cells by inactivating PI3K/Akt signaling. Oncol Lett. 2022;24:249. DOI: 10.3892/ol.2022. 13369.
4. Woolcott CG, Wilkens LR, Nomura AMY, Horst RL, Goodman MT, Murphy SP, et al. Plasma 25-hydroxyvitamin D levels and the risk of colorectal cancer: the multiethnic cohort study. Cancer Epidemiol Biomarkers Prev. 2010;19:130–4. DOI: 10.1158/1055-9965.EPI-09-0475.
5. Guertin KA, Li XS, Graubard BI, Albanes D, Weinstein SJ, Goedert JJ, et al. Serum trimethylamine N-oxide, carnitine, choline, and betaine in relation to colorectal cancer risk in the alpha tocopherol, beta carotene cancer prevention study. Cancer Epidemiol Biomarkers Prev. 2017;26:945–52. DOI: 10.1158/1055-9965.EPI-16-0948.

6. Zhou JJ, Zheng S. Colorectal cancer: basic and translational research. Gastrointest Tumors. 2014;1:18–24. DOI: 10.1159/000354994.
7. Liang X, Hendryx M, Qi L, Lane D, Luo J. Association between prediagnosis depression and mortality among postmenopausal women with colorectal cancer. PLoS One. 2020;15. DOI: 10.1371/journal.pone.0244728.
8. Corbo C, Cevenini A, Salvatore F. Biomarker discovery by proteomics-based approaches for early detection and personalized medicine in colorectal cancer. Proteomics Clin Appl. 2017;11. DOI: 10.1002/prca.201600072.
9. Janani B, Vijayakumar M, Priya K, Kim JH, Geddawy A, Shahid M, et al. A network-based pharmacological investigation to identify the mechanistic regulatory pathway of andrographolide against colorectal cancer. Front Pharmacol. 2022;13. DOI: 10.3389/fphar.2022.967262.
10. Qiang R, Zhao Z, Tang L, Wang Q, Wang Y, Huang Q. Identification of 5 hub genes related to the early diagnosis, tumour stage, and poor outcomes of hepatitis B virus-related hepatocellular carcinoma by bioinformatics analysis. Comput Math Methods Med. 2021;2021:1–20. DOI: 10.1155/2021/9991255.
11. Barrett T, Edgar R. Mining microarray data at NCBI's gene expression omnibus (GEO). In: Bina M, editor. Gene mapping, discovery, and expression. Methods Mol Biol. Vol 338. Humana Press; 2006. p. 175–90. DOI: 10.1385/1-59745-097-9:175.
12. Kolur V, Vastrad B, Vastrad C, Kottursheti S, Tengli A. Identification of candidate biomarkers and therapeutic agents for heart failure by bioinformatics analysis. BMC Cardiovasc Disord. 2021;21:329. DOI: 10.1186/s12872-021-02146-8.
13. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: network analysis and visualization of proteomics data. J Proteome Res. 2019;18:623–32. DOI: 10.1021/acs.jproteome.8b00702.
14. Assenov Y, Ramírez F, Schelhorn S-E, Lengauer T, Albrecht M. Computing topological parameters of biological networks. Bioinformatics. 2008;24:282–4. DOI: 10.1093/bioinformatics/btm554.
15. Chin C-H, Chen S-H, Wu H-H, Ho C-W, Ko M-T, Lin C-Y. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014;8. DOI: 10.1186/1752-0509-8-S4-S11.
16. Yan X, Liu XP, Guo ZX, Liu TZ, Li S. Identification of hub genes associated with progression and prognosis in patients with bladder cancer. Front Genet. 2019;10. DOI: 10.3389/fgene.2019.00408.
17. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics. 2003;4:2. DOI: 10.1186/1471-2105-4-2.
18. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol. 2003;4. DOI: 10.1186/gb-2003-4-9-r60.
19. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44:W90–7. DOI: 10.1093/nar/gkw377.
20. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res. 2019;47:W556–60. DOI: 10.1093/nar/gkz430.
21. Greene AN, Nguyen ET, Paranjpe A, Lane A, Privette Vinnedge LM, Solomon MB. In silico gene expression and pathway analysis of DEK in the human brain across the lifespan. Eur J Neurosci. 2022;56:4720–43. DOI: 10.1111/ejn.15791.
22. Lind GE, Danielsen SA, Ahlquist T, Merok MA, Andresen K, Skotheim RI, et al. Identification of an epigenetic biomarker panel with high sensitivity and specificity for colorectal cancer and adenomas. Mol Cancer. 2011;10:85. DOI: 10.1186/1476-4598-10-85.
23. Zhang P, Kawakami H, Liu W, Zeng X, Strebhardt K, Tao K, et al. Targeting CDK1 and MEK/ERK overcomes apoptotic resistance in BRAF-mutant human colorectal cancer. Mol Cancer Res. 2018;16:378–89. DOI: 10.1158/1541-7786.MCR-17-0404.

24. Liu P, Kao TP, Huang H. CDK1 promotes cell proliferation and survival via phosphorylation and inhibition of FOXO1 transcription factor. Oncogene. 2008;27:4733–44. DOI: 10.1038/onc.2008.104.
25. Ding X, Duan H, Luo H. Identification of core gene expression signature and key pathways in colorectal cancer. Front Genet. 2020;11. DOI: 10.3389/fgene.2020.00045.
26. Li Z, Zhang Y, Zhou Y, Wang F, Yin C, Ding L, et al. Tanshinone IIA suppresses the progression of lung adenocarcinoma through regulating CCNA2-CDK2 complex and AURKA/PLK1 pathway. Sci Rep. 2021;11:23681. DOI: 10.1038/s41598-021-03166-2.
27. Burum-Auensen E, DeAngelis PM, Schjølberg AR, Røislien J, Mjåland O, Clausen OPF. Reduced level of the spindle checkpoint protein BUB1B is associated with aneuploidy in colorectal cancers. Cell Prolif. 2008;41:645–59. DOI: 10.1111/j.1365-2184.2008.00539.x.
28. Zhuang L, Yang Z, Meng Z. Upregulation of BUB1B, CCNB1, CDC7, CDC20, and MCM3 in tumor tissues predicted worse overall survival and disease-free survival in hepatocellular carcinoma patients. Biomed Res Int. 2018;2018:1–8. DOI: 10.1155/2018/7897346.
29. Fang Y, Yu H, Liang X, Xu J, Cai X. Chk1-induced CCNB1 overexpression promotes cell proliferation and tumor growth in human colorectal cancer. Cancer Biol Ther. 2014;15:1268–79. DOI: 10.4161/cbt.29691.
30. Xiong M, Zhuang K, Luo Y, Lai Q, Luo X, Fang Y, et al. KIF20A promotes cellular malignant behavior and enhances resistance to chemotherapy in colorectal cancer through regulation of the JAK/STAT3 signaling pathway. Aging. 2019;11:11905–21. DOI: 10.18632/aging.102505.
31. Zhang W, He W, Shi Y, Gu H, Li M, Liu Z, et al. High expression of KIF20A is associated with poor overall survival and tumor progression in early-stage cervical squamous cell carcinoma. PLoS One. 2016;11. DOI: 10.1371/journal.pone.0167449.
32. Li H, Zhang W, Sun X, Chen J, Li Y, Niu C, et al. Overexpression of kinesin family member 20A is associated with unfavorable clinical outcome and tumor progression in epithelial ovarian cancer. Cancer Manag Res. 2018;Volume 10:3433–50. DOI: 10.2147/CMAR.S169214.
33. Jiang N, Liao Y, Wang M, Wang Y, Wang K, Guo J, et al. BUB1 drives the occurrence and development of bladder cancer by mediating the STAT3 signaling pathway. J Exp Clin Cancer Res. 2021;40:378. DOI: 10.1186/s13046-021-02179-z.
34. Jiang W, Yu Y, Bhandari A, Hirachan S, Dong X, Huang X, et al. Budding uninhibited by benzimidazoles 1 might be a poor prognosis biomarker promoting the progression of papillary thyroid cancer. Environ Toxicol. 2023;38:2047–56. DOI: 10.1002/tox.23812.
35. Zhou J, Chen WR, Yang LC, Wang J, Sun JY, Zhang WW, et al. KIF11 functions as an oncogene and is associated with poor outcomes from breast cancer. Cancer Res Treat. 2019;51:1207–21. DOI: 10.4143/crt.2018.460.
36. Pei Y, Li G, Ran J, Wan X, Wei F, Wang L. Kinesin family member 11 enhances the self-renewal ability of breast cancer cells by participating in the Wnt/β-catenin pathway. J Breast Cancer. 2019;22:522. DOI: 10.4048/jbc.2019.22.e51.
37. Neska-Długosz I, Buchholz K, Durślewicz J, Gagat M, Grzanka D, Tojek K, et al. Prognostic impact and functional annotations of KIF11 and KIF14 expression in patients with colorectal cancer. Int J Mol Sci. 2021;22:9732. DOI: 10.3390/ijms22189732.
38. Brase JC, Schmidt M, Fischbach T, Sültmann H, Bojar H, Koelbl H, et al. ERBB2 and TOP2A in breast cancer: a comprehensive analysis of gene amplification, RNA levels, and protein expression and their influence on prognosis and prediction. Clin Cancer Res. 2010;16:2391–401. DOI: 10.1158/1078-0432.CCR-09-2471.
39. Heestand GM, Schwaederle M, Gatalica Z, Arguello D, Kurzrock R. Topoisomerase expression and amplification in solid tumours: analysis of 24,262 patients. Eur J Cancer. 2017;83:80–7. DOI: 10.1016/j.ejca.2017.06.019.
40. Ding X, Duan H, Luo H. Identification of core gene expression signature and key pathways in colorectal cancer. Front Genet. 2020;11. DOI: 10.3389/fgene.2020.00045.
41. Weng Ng WT, Shin JS, Roberts TL, Wang B, Lee CS. Molecular interactions of polo-like kinase 1 in human cancers. J Clin Pathol. 2016;69:557–62. DOI: 10.1136/jclinpath-2016-203656.

42. Tan J, Li Z, Lee PL, Guan P, Aau MY, Lee ST, et al. PDK1 signaling toward PLK1–MYC activation confers oncogenic transformation, tumor-initiating cell activation, and resistance to mTOR-targeted therapy. Cancer Discov. 2013;3:1156–71. DOI: 10.1158/2159-8290.CD-12-0595.
43. Yu C, Luo D, Yu J, Zhang M, Zheng X, Xu G, et al. Genome-wide CRISPR-Cas9 knockout screening identifies GRB7 as a driver for MEK inhibitor resistance in KRAS mutant colon cancer. Oncogene. 2022;41:191–203. DOI: 10.1038/s41388-021-02077-w.

