Identification of Hub Genes for Early Diagnosis and Predicting Prognosis in Colon Adenocarcinoma: A Bioinformatics Analysis

Shuo Xu (✉ sxu95@cmu.edu.cn)

Shengjing Hospital of China Medical University <https://orcid.org/0000-0003-2457-3519>

Dingsheng Liu

Shengjing Hospital of China Medical University

Mingming Cui

Shengjing Hospital of China Medical University

Yao Zhang

Shengjing Hospital of China Medical University

Yu Zhang

Shengjing Hospital of China Medical University

Hong Zhang

Shengjing Hospital of China Medical University

Research Article

Keywords: colon adenocarcinoma, hub genes, biomarker, prognosis, bioinformatics analysis

Posted Date: March 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-265035/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Colon adenocarcinoma (COAD) is among the most common digestive system malignancies worldwide. Although some molecular analyses of colon cancer were previously performed, the pathogenesis and gene signatures remain unclear. This study explored the genetic characteristics and molecular mechanisms underlying colon cancer development.

Methods: Three gene expression data sets were obtained from the Gene Expression Omnibus (GEO) database. GEO2R was used to determine differentially expressed genes (DEGs) between COAD and normal tissues. Then, the intersection of the data sets was obtained. Metascape was used to perform the functional enrichment analyses. Next, STRING was used to build protein-protein interaction (PPI) networks. Hub genes were identified and analysed using Cytoscape. Next, survival analysis of the hub genes was performed. To explore the early diagnostic value of the hub genes, UALCAN was used to analyse expression characteristics. Finally, alterations in the hub genes were predicted and analysed by cBioPortal.

Results: Altogether 436 DEGs were detected. The DEGs were mainly enriched in cell cycle phase transition, nuclear division, meiotic nuclear division, cytokinesis. Based on PPI networks, 20 hub genes were selected. Among them, 6 hub genes (*CCNB1*, *CCNA2*, *AURKA*, *NCAPG*, *DLGAP5*, and *CENPE*) showed significant prognostic value in colon cancer ($p < 0.05$), while 5 hub genes (*CDK1*, *CCNB1*, *CCNA2*, *MAD2L1*, and *DLGAP5*) were associated with early colon cancer diagnosis.

Conclusions: In conclusion, integrated bioinformatics analysis was used to identify hub genes that reveal the potential mechanism of carcinogenesis and progression of colon cancer. The hub genes might be novel biomarkers for early diagnosis, treatment, and prognosis of colon cancer.

Background

Colon adenocarcinoma (COAD) is among the most common digestive system malignancies worldwide. There were 1,096,601 new colon cancer cases and 551,269 deaths worldwide in 2018(1). In the last decade, both the incidence and mortality of colon cancer increased in rapidly transitioning countries including the Baltic countries, Russia, China, and Brazil(2). As previously reported, the 5-year survival rate was more than 90% for patients diagnosed with stage I, but only 12% for patients diagnosed with stage IV(3). Thus, early diagnosis and surgical resection of colon cancer will greatly improve disease prognosis. The current early screening tests include non-invasive tests of stool and blood-based tests, radiologic tests and invasive test like colonoscopy. However, the compliance rate of screening is low, mainly due to the variable sensitivity and specificity of non-invasive tests and low acceptance of the invasive tests as well as the expensive cost(4). As a result, only 39 % of tumours are diagnosed at an early stage, and the colon cancer remains a serious health burden worldwide(5). Thus, it is essential to uncover the molecular mechanism and to explore key biomarkers for early colon cancer diagnosis.

RNA sequencing and gene expression microarrays are widely applied in cancer studies. Bioinformatics analysis of these data can be used to identify significant biomarkers which may improve cancer early diagnosis, predict prognosis, and inform therapeutic responses.(6, 7). Although there are some previous studies of gene expression in colon cancer, but few studies involve multiple gene expression files and focus on an early diagnosis of the disease. Hence, we performed this study in order to

deepen the understanding of the underlying mechanism and provide novel biomarkers for early diagnosis and prognosis of the disease.

In this study, we selected and downloaded three gene sets (GSE110224, GSE44076, and GSE47063) from the Gene Expression Omnibus (GEO) database. Then, GEO2R and a Venn diagram webtool was used to identify differentially expressed genes (DEGs) between COAD and normal tissues in the three data sets and the intersection between the data sets respectively. Third, Metascape was used for functional DEG annotation. Subsequently, a protein-protein interaction (PPI) network was established, and the most significant module was isolated and analysed. Hub genes were identified by CytoHubba, and their prognostic value was analysed by Gene Expression Profiling Interactive Analysis (GEPIA). Moreover, we explored the early diagnostic value of the hub genes. Finally, genetic alterations of hub genes were predicted and analysed using cBioPortal. Collectively, this study analysed the genetic, mRNA, and protein changes that occur during development of COAD, attempted to reveal the underlying molecular mechanism, and explore effective COAD biomarkers.

Methods

Microarray data.

We first searched the GEO database (8, 9) and identified three microarray data sets (GSE110224, GSE44076, and GSE47063)(10–12) describing gene expression differences between COAD and normal colon tissue. GSE110224 is based on platform GPL570 ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array), GSE44076 is based on platform GPL13667 ([HG-U219] Affymetrix Human Genome U219 Array), and GSE47063 is based on platform GPL6102 (Illumina human-6 v2.0 expression beadchip). All data are freely available online.

DEG identification

GEO2R is commonly used to process sample information from GEO Series and to identify DEGs among user-defined groups. After screening the sample information in the three data sets, only the COAD samples and the corresponding normal tissues were included. After GEO2R analysis, DEGs were obtained by intersecting genes with an adjusted $P < 0.05$ and $|\log FC| \geq 1$ in each data set using a Venn diagram.

Gene ontology and pathway enrichment analysis of DEGs

Metascape(13, 14) is an open access online tool for comprehensive gene list annotation and analysis. In this study, DEG pathway and process enrichment analyses were performed using Metascape. The

parameters were set as follows: 3 for Min Overlap, 1.5 for Min Enrichment, and P-Value Cutoff of 0.05. The enrichment results are presented as bar charts. Corresponding network graph nodes with similarity degree more than 0.3 were connected with curved edges. Edge thickness was positively correlated with the degree of similarity.

PPI network construction and module analysis

The Search Tool for the Retrieval of Interacting Genes (STRING) database (15, 16) was used to construct the PPI network with an interaction score > 0.4. Then, Cytoscape (Version 3.7.2)(17) software was used to visualise and analyse PPI networks. Molecular Complex Detection (MCODE)(Version 1.6)(18), a Cytoscape plugin, was used to identify the most significant gene module in colon cancer. Then, we annotated the function of the module genes using Metascape.

Hub gene selection and analysis

CytoHubba (Version 0.1)(19), a Cytoscape plugin, was used to identify the network hub genes. We used a degree-ranked method to identify hub genes with a criterion of degree no less than 67. ClueGO(20) is another Cytoscape plugin that can create and visualise functionally grouped networks of biological terms and pathways. The CluePedia(21) Cytoscape plugin is a functional extension of ClueGO and a search tool for new markers potentially associated with pathways. In our study, ClueGO (Version 2.5.6) and CluePedia (Version 1.5.6) were used to analyse the biological processes and pathway enrichment of hub genes.

Analysis of prognostic value of hub genes

GEPIA (22, 23) is an integrated bioinformatics analysis tool which was designed for transforming genomic big data into intuitive graphics. It contains data from 33 cancer types from more than 18,000 samples. Thus, users can conduct various data mining for specific genes and cancer species. In this study, GEPIA was used to perform survival analysis based on gene expression. $P \leq 0.05$ was considered statistically significant.

Hub gene expression analysis

UALCAN(24, 25) is a comprehensive interactive online resource which contains clinical data from 31 cancer types from the TCGA database. The database is widely used to analyse cancer OMICS data. We used UALCAN to perform differential expression analysis of the hub genes and their association with clinicopathological parameters of COAD patients. Moreover, the Human Protein Atlas (26, 27) is a website for users to freely access data for exploration of the human proteome, which contains transcriptome data from 17 main cancer types using data from nearly 8000 patients. The immunohistochemistry (IHC) images were manually annotated using fixed guidelines. Basic annotation parameters include staining intensity, fraction of stained cells and subcellular localization. In this study, IHC data of the hub genes were downloaded and used for direct comparison the protein expression.

Analysis of alterations of hub genes

cBioPortal (28, 29) is a free web server for interactively exploring cancer genomics datasets. In this study, cBioPortal was utilised to predict the genetic alterations of eight hub genes in 378 COAD samples (TCGA, PanCancer Atlas) which contained mutations and putative copy-number alterations from GISTIC and mRNA Expression z-Scores (RNASeq V2 RSEM) with a z-score threshold ± 2.0 .

Statistical analysis

Microarray data analysis was performed by using GEO2R. GEOquery R package was used to transform the original data into R data structure, and then the statistical test of limma (Linear Models for Microarray Analysis) R package was used to identify DEGs. Survival analysis was performed by using GEPIA and log-rank test. The transcripts per million (TPM) expression value and *t*-test were used for analysis of the relationship between hub genes expression and clinicopathological parameters. $P \leq 0.05$ was considered statistically significant.

Results

DEGs in colon cancer

Among the three datasets (GSE110224, GSE44076, and GSE47063), there were 127 COAD tissues and 117 normal tissues. After GEO2R analysis, we screened 1617 DEGs (745 upregulated, 872 downregulated) from GSE110224, 4450 DEGs (2095 upregulated, 2355 downregulated) from GSE44076, and 2259 DEGs (1056 upregulated, 1203 downregulated) from GSE47063. Then, 436 DEGs were obtained by overlapping the three dataset results, including 267 downregulated genes (Fig. 1A) and 169 upregulated genes (Fig. 1B).

DEG gene ontology (GO) and pathway enrichment in colon cancer

GO and pathway enrichment analysis were conducted using Metascape to predict DEG function. The top 20 GO items were divided into 3 categories: biological processes (14 items), cellular components (4 items), and molecular functions (2 items; Table 1 and Fig. 2A-2B). The DEGs were mainly enriched in cell cycle, transcriptional regulation, and ion transport. Enriched biological processes included cell cycle phase transition, nuclear division, meiotic nuclear division, cytokinesis, DNA replication, negative regulation of cell proliferation, regulation of reproductive process, regulation of MAPK cascade, positive regulation of transferase activity, bicarbonate transport, inorganic ion homeostasis, cellular response to organic cyclic compound, cellular response to nitrogen compound, and mesenchymal cell differentiation. Cellular component analysis showed that the DEGs were significantly enriched in the apical part of the cell, spindle, microvillus, and basolateral plasma membrane. Molecular functions of these genes were histone kinase activity and activity of hydrolase acting on ester bond.

The top 20 Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome pathways are shown in Table 1 and Fig. 2C-2D. DEGs were mainly enriched for terms associated with the cell cycle, reversible

hydration of carbon dioxide, proximal tubule bicarbonate reclamation, transport of small molecules, cyclin A/B1/B2 associated events during G2/M transition, and regulation of TP53 activity through phosphorylation pathway.

DEG PPI network and modules

We use the STRING database to predict protein interactions between the DEGs. A PPI network composed of 369 nodes and 2708 edges was constructed (Fig. 3). Then, MCODE was used to isolate the significant network modules. We selected the most significant module with the highest degree (Fig. 4A) and functionally annotated the involved genes (Table 4). GO enrichment analysis showed that the genes were mainly enriched in biological processes, including chromosome segregation, cell cycle phase transition, positive regulation of cell cycle, DNA replication, meiotic cell cycle, attachment of spindle microtubules to kinetochore, DNA conformation change, signal transduction by p53 class mediator, positive regulation of transferase activity, sister chromatid cohesion, cytokinetic process, and protein localisation to cytoskeleton. Cellular component analysis showed that these genes were mainly enriched in the spindle, midbody, kinesin complex, and intercellular bridge. Molecular function analysis showed that these genes were mainly enriched in catalytic activity, acting on DNA, and chromatin binding. Pathway analysis revealed that these genes were mainly enriched in cyclin A/B1/B2-associated events during G2/M transition and APC-Cdc20 mediated degradation of Nek2A.

Hub genes

According to the node degree calculated by CytoHubba, 20 hub genes were screened out because they were all upregulated (Fig. 4B). The gene symbols and corresponding degree are shown in Table 5. Functional annotation of the 20 hub genes is shown in Fig. 4C-D. Heat map visualisation shows that the expression of these 20 hub genes in COAD tissues is higher than in normal tissues (Fig. 4E).

Survival based on hub gene expression

Because several hub genes were closely related to the cell cycle, we further analysed their survival curves using the GEPIA database. Our results show that overexpression of six hub genes influenced COAD prognosis, including *CCNB1*, *CCNA2*, *AURKA*, *NCAPG*, *DLGAP5*, and *CENPE*. Overexpression of the six genes was associated with favourable overall survival (OS) of colon cancer patients (Fig. 5A-5F). Additionally, *AURKA* and *CENPE* overexpression showed a favourable prognosis of disease-free survival (DFS) in COAD patients (Fig. 5G-5H).

Differential expression of hub genes

UALCAN was used to analyse mRNA expression of the identified hub genes. We found 5 hub genes were related to clinicopathological parameters, including *CDK1*, *CCNB1*, *CCNA2*, *MAD2L1*, and *DLGAP5*. Additionally, we observed that these five genes were significantly overexpressed in tumour tissues (Fig. 6A,6D,6G,6J,6M). Then, we analysed their mRNA expression under different clinicopathological parameters. Our results revealed that the mRNA expression of the five genes was significantly correlated

with the clinical stage, and that the highest mRNA expression appeared in the first tumour stage (Fig. 6B, 6E, 6H, 6K, 6N). Moreover, the mRNA expression of the five genes showed a significant correlation with lymph node metastasis, and the highest mRNA expression appeared at the N0 phase (Fig. 6C, 6F, 6I, 6L, 6O).

Moreover, we analysed the protein expressions of hub genes using IHC images from HPA. Our results showed that CDK1 staining was low in normal tissues and moderate in COAD tissues (Fig. 7A). CCNB1 and CCNA2 staining were moderate in normal colon tissues, whereas high staining was observed in COAD tissues (Fig. 7B-7C). DLGAP5 staining was not detected in normal tissues, while moderate staining was observed in COAD tissues (Fig. 7D). MAD2L1 was moderately stained in both tumour and normal tissues (Fig. 7E).

Alteration of hub genes

We also analysed alterations of the six prognostic hub genes *CCNB1*, *CCNA2*, *AURKA*, *NCAPG*, *DLGAP5*, and *CENPE* together with the five hub genes which were associated with clinicopathological parameters: *CDK1*, *CCNB1*, *CCNA2*, *MAD2L1*, and *DLGAP5*. Eight hub genes including *CDK1*, *CCNB1*, *CCNA2*, *AURKA*, *MAD2L1*, *NCAPG*, *DLGAP5*, and *CENPE* were detected by cBioPortal.

Altogether 378 samples of COAD were included, and our analysis revealed that the hub genes were altered in 42.86% of the 378 samples. *AURKA* (28%) was the most frequently altered gene of the eight hub genes (Fig. 8).

Discussion

Colon cancer was the fourth most commonly diagnosed malignant tumour worldwide in 2018, with increasing incidence in countries undergoing major developmental transition(30). Due to a lack of specific symptoms for early detection, patients are usually diagnosed at an advanced stage which leads to a poor prognosis(31). Therefore, it is crucial to uncover the underlying molecular mechanism and to explore key biomarkers for early colon cancer diagnosis.

In this study, we analysed three microarray datasets that included 127 tumours and 117 normal samples. A total of 436 DEGs were screened. Functional annotation showed that the DEGs were mainly enriched in biological processes associated with cell cycle phase transition, nuclear division, positive regulation of transferase activity, meiotic nuclear division, and DNA replication. These results suggest that these genes are closely related to the cell cycle. Many studies indicate that dysregulation of cell cycle progression is closely related to cancer progression(32, 33). Finetti et al. (34) found that several genes participate in regulating the cell cycle, like *CDK1* and *AURKA*. Moreover, their expression was correlated with breast cancer prognosis. In our colon cancer study, we obtained many DEGs involved in cell cycle progression, including *CCND1*, *BLM*, *BUB1*, *BUB1B*, *CCNA2*, *CCNB1*, *CDK1*, and *CDC20*. Some genes are closely related to the transformation of cancer. For example, *CCND1* belongs to the cyclin family whose members are characterised by dramatic periodicity in protein abundance throughout the cell cycle. Deregulation of

CCND1 is observed frequently in numerous human cancers, including pancreatic cancer, head and neck squamous cell carcinoma, breast cancer, and colorectal carcinoma(35, 36). Accumulation of *CCND1* in the nucleus causes uncontrolled cell cycle progression and acts as a tumour-initiating event.(37) Overexpression of cyclin D1 (T286A), an oncogenic mutant allele of *CCND1*, promotes stabilisation and overexpression of the DNA replication licensing factor, Cdt1, by inhibiting its proteolysis. This causes DNA re-replication and damage, and results in cellular aneuploidy, genomic instability, and further neoplastic growth(38). Cyclin dependent kinases (CDKs) are necessary functional partner kinases with cyclin D1. Thus, CDK inhibitors would be an effective drug for targeting malignant tumours(39). However, given the development of resistance and side effects of CDK inhibitors, further research is warranted(36).

Pathway analysis also revealed that DEGs were mainly enriched for terms associated with the cell cycle pathway. Cyclin A/B1/B2-associated events in the “G2/M transition” and “Regulation of TP53 Activity through Phosphorylation” pathways were closely related to tumorigenesis. Like the cyclin D1 mentioned above, cyclins A/B1/B2 are also cyclin members that bind to CDKs and regulate the cell cycle. Abundant evidence shows that G2/M phase arrest is closely related to the inhibition of tumour cell proliferation(40, 41). Additional studies focusing on cyclins aimed to identify novel therapeutic strategies for cancer treatment. Qiang(42) revealed that the microRNA MiR-219-5p downregulates *CCNA2* expression and induces G2/M phase arrest to inhibit tumour formation in oesophageal cancer. Xuezi et al.(43) found *CCNA2* is downregulated by the small molecule FH535 in colorectal cancer, which causes G2/M phase arrest and inhibits tumour proliferation. Downregulating *CCNA2* and *CCNB1* is a key mechanism for the anti-cancer activity of gossypin, a flavone extracted from *Hibiscus vitifolius*(44). Thus, inhibiting *CCNA2* and *CCNB1* may contribute to the development of novel anti-cancer drugs. The p53 signalling pathway significantly contributes to cell cycle regulation, suppression of tumour expression, metabolism, aging, development, and reproduction(45). Phosphorylation of p53 protein stabilizes the protein and extends its half-life, thus causing cell cycle arrest, apoptosis, and inhibited tumour cell proliferation(46). For instance, a study of natural polyphenols as anti-cancer agents revealed that polyphenols could induce apoptosis, which was achieved by stabilizing p53 protein through phosphorylation and showed remarkable effects in human gastric carcinoma cells(47). We also identified some pathways associated with metabolism, including triglyceride metabolism, carnitine metabolism, regulation of lipolysis in adipocytes, and Phase I - Functionalization of compounds. Among these pathways, we found that *FABP4*, which encodes fatty acid binding protein, was involved in fatty acid uptake, transport, metabolism, and is related to tumour metastasis. Kshipra et al.(48) observed that overexpression of *FABP4* plays a key role in aggressive metastasis of ovarian cancer via various metabolites and protein pathways. Likewise, *FABP4* has crucial effects on adipocyte-induced cholangiocarcinoma metastasis(49). Collectively, metabolic disorder is among the leading causes of tumour development. Thus, the study of tumour metabolism may provide new targets for tumour treatment.

The PPI network was built using STRING. Twenty hub genes were screened, which include *CDK1*, *CCNB1*, *CCNA2*, *AURKA*, *CDC20*, *AURKB*, *TPX2*, *BUB1*, *CDC45*, *MAD2L1*, *KIF2C*, *NCAPG*, *DLGAP5*, *FOXM1*, *CENPF*, *CENPE*, *BUB1B*, *TTK*, *ASPM*, and *KIF20A*. The functional annotation of these genes was most closely related to the cell cycle.

Survival analysis showed that higher mRNA expression of six hub genes was significantly related to longer OS in colon cancer patients, including *CCNB1*, *CCNA2*, *AURKA*, *NCAPG*, *DLGAP5*, and *CENPE*. Moreover, *AURKA* and *CENPE* exhibited favourable effects on both OS and DFS. Thus, the 6 hub genes showed clinical prognostic value and might provide potential therapeutic targets for colon cancer.

For early COAD diagnosis, we identified *CDK1*, *CCNB1*, *CCNA2*, *MAD2L1*, and *DLGAP5*, which were closely related to clinicopathological parameters. Our results showed that both the mRNA and protein expression of these five hub genes are higher in tumour tissue than in normal tissue, which indicates that the hub genes may be closely related to COAD progression. Previous studies observed that the expression of these genes was correlated with tumour size and stage(50–52). In our study, we found that mRNA expression of the five hub genes was significantly related to mild clinical pathological parameters, so these genes may play an important role in the early diagnosis of colon cancer. The relationship between these genes and COAD has not yet been fully determined, but our data indicate that the increased expression in early COAD stages may provide an indicator for early diagnosis.

We finally performed alteration analysis of eight hub genes which showed significant effects on survival analysis, including *CDK1*, *CCNB1*, *CCNA2*, *AURKA*, *MAD2L1*, *NCAPG*, *DLGAP5*, and *CENPE*. These result show that more than 40% of the patient tumours analysed had at least one hub gene alteration. *AURKA* was the most frequently altered (28%) of the 8 hub genes. The protein encoded by this gene is a cell cycle-regulated kinase that appears to be involved in spindle assembly, cytokinesis, centrosome maturation, and separation(53). Previous studies showed that *AURKA* is frequently upregulated and correlated with prognosis in several types of cancers, which may reveal an important role in human cancer(54, 55).

There were some limitations in this study. First, all the data analysed in our study was retrieved from online databases. Thus, further studies with larger sample sizes and biological experiments are required to validate our findings. Our future research will focus on experimental verification of these results. Second, we did not explore the underlying mechanisms of hub genes in COAD. Future studies should investigate the detailed mechanism between hub genes and COAD.

Conclusions

In conclusion, our study identified and analysed DEGs and 20 core genes associated with COAD, which might deepen the understanding of carcinogenesis and provide indicators for prognosis and early diagnosis of the disease.

Abbreviations

COAD
Colon adenocarcinoma
GEO

Gene Expression Omnibus
DEGs
differentially expressed genes
PPI
protein-protein interaction
GEPIA
Gene Expression Profiling Interactive Analysis
STRING
The Search Tool for the Retrieval of Interacting Genes
MCODE
Molecular Complex Detection
IHC
immunohistochemistry
TPM
transcripts per million
GO
gene ontology
KEGG
Kyoto Encyclopedia of Genes and Genomes
HPA
Human Protein Atlas
CDK
Cyclin dependent kinase

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets and analysis tools are all available online.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by grant from the Natural Science Foundation of Liaoning Province (Grant No.20170540997).

Authors' contributions

SX was responsible for drafting manuscripts, DL was responsible for database data collation, and was a contributor in writing the manuscript, MC was responsible for article immunohistochemical part and manuscript revision, Yao Z was responsible for interpreting and analyzing the results, Yu Z was responsible for making charts and data analysis, and HZ was responsible for topic selection, manuscript revision and funding acquisition. These authors contributed equally to this work. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394–424.
2. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut.* 2017;66(4):683–91.
3. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin.* 2019;69(5):363–85.
4. Issa IA, Nouredine M. Colorectal cancer screening: An updated review of the available options. *World J Gastroenterol.* 2017;23(28):5086–96.
5. Favoriti P, Carbone G, Greco M, Pirozzi F, Pirozzi RE, Corcione F. Worldwide burden of colorectal cancer: a review. *Updates Surg.* 2016;68(1):7–11.
6. Oncology VK. DEJNcp. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. 2008;5(10):588–99.
7. CC H, Cancers MD. WLJ. Bioinformatics Analysis for Circulating Cell-Free DNA in Cancer. 2019;11(6).
8. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 2013;41(Database issue):D991-5.
9. The Gene Expression Omnibus database. <https://www.ncbi.nlm.nih.gov/geo/>. Accessed 20 May 2020.
10. Uribe-Lewis S, Stark R, Carroll T, Dunning MJ, Bachman M, Ito Y, et al. 5-hydroxymethylcytosine marks promoters in colon that resist DNA hypermethylation in cancer. *Genome Biol.* 2015;16:69.
11. Vlachavas EI, Pilalis E, Papadodima O, Koczan D, Willis S, Klippel S, et al. Radiogenomic Analysis of F-18-Fluorodeoxyglucose Positron Emission Tomography and Gene Expression Data Elucidates the

- Epidemiological Complexity of Colorectal Cancer Landscape. *Comput Struct Biotechnol J*. 2019;17:177–85.
12. Sole X, Crous-Bou M, Cordero D, Olivares D, Guino E, Sanz-Pamplona R, et al. Discovery and validation of new potential biomarkers for early detection of colon cancer. *PLoS One*. 2014;9(9):e106748.
 13. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10(1):1523.
 14. Metascape. <http://metascape.org/>. Accessed 26 May 2020.
 15. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607-D13.
 16. The Search Tool for the Retrieval of Interacting Genes. <http://string-db.org/>. Accessed 6 June 2020.
 17. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
 18. GD B, bioinformatics HCJB. An automated method for finding molecular complexes in large protein interaction networks. 2003;4:2.
 19. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol*. 2014;8(Suppl 4):11.
 20. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009;25(8):1091–3.
 21. Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics*. 2013;29(5):661–3.
 22. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45(W1):W98–102.
 23. The Gene Expression Profiling Interactive Analysis. <http://gepia.cancer-pku.cn/>. Accessed 6 June 2020.
 24. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia*. 2017;19(8):649–58.
 25. UALCAN. <http://ualcan.path.uab.edu>. Accessed 6 June 2020.
 26. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347(6220):1260419.
 27. The Human Protein Atlas. <https://www.proteinatlas.org>. Accessed 6 June 2020.

28. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012;2(5):401–4.
29. cBioPortal. <http://cbioportal.org>. Accessed 6 June 2020.
30. Fidler MM, Soerjomataram I, Bray F. A global view on cancer incidence and national levels of the human development index. *Int J Cancer.* 2016;139(11):2436–46.
31. Brenner H, Altenhofen L, Stock C, Hoffmeister M. Prevention, early detection, and overdiagnosis of colorectal cancer within 10 years of screening colonoscopy in Germany. *Clin Gastroenterol Hepatol.* 2015;13(4):717–23.
32. Zhou Q, Zhang F, He Z, Zuo MZ. E2F2/5/8 Serve as Potential Prognostic Biomarkers and Targets for Human Ovarian Cancer. *Front Oncol.* 2019;9:161.
33. Sun J, Shi R, Zhao S, Li X, Lu S, Bu H, et al. Cell division cycle 45 promotes papillary thyroid cancer progression via regulating cell cycle. *Tumour Biol.* 2017;39(5):1010428317705342.
34. Finetti P, Cervera N, Charafe-Jauffret E, Chabannon C, Charpin C, Chaffanet M, et al. Sixteen-kinase gene expression identifies luminal breast cancers with poor prognosis. *Cancer Res.* 2008;68(3):767–76.
35. GA MMB,AB,MK,MR. F, A A, et al. The prognostic value of cyclin D1 expression in the survival of cancer patients: A meta-analysis. 2020;728:144283.
36. Qie S, Diehl JA. Cyclin D1, cancer progression, and opportunities in cancer treatment. *J Mol Med.* 2016;94(12):1313–26.
37. Alt JR. Phosphorylation-dependent regulation of cyclin D1 nuclear export and cyclin D1-dependent cellular transformation. *Genes Dev.* 2000;14(24):3102–14.
38. Aggarwal P, Lessie MD, Lin DI, Pontano L, Gladden AB, Nuskey B, et al. Nuclear accumulation of cyclin D1 during S phase inhibits Cul4-dependent Cdt1 proteolysis and triggers p53-dependent DNA rereplication. *Genes Dev.* 2007;21(22):2908–22.
39. MC C, M V-V, C A-A, drugs PRJEooi. Overview of cyclins D1 function in cancer and the CDK inhibitor landscape: past and present. 2014;23(3):295–304.
40. Sun J, Guo Y, Fu X, Wang Y, Liu Y, Huo B, et al. *Dendrobium candidum* inhibits MCF-7 cells proliferation by inducing cell cycle arrest at G2/M phase and regulating key biomarkers. *Onco Targets Ther.* 2016;9:21–30.
41. Lin Z, Xiong L, Lin Q. Knockdown of eIF3d inhibits cell proliferation through G2/M phase arrest in non-small cell lung cancer. *Med Oncol.* 2015;32(7):183.
42. Ma Q. MiR-219-5p suppresses cell proliferation and cell cycle progression in esophageal squamous cell carcinoma by targeting CCNA2. *Cell Mol Biol Lett.* 2019;24:4.
43. Tu X, Hong D, Jiang Y, Lou Z, Wang K, Jiang Y, et al. FH535 inhibits proliferation and migration of colorectal cancer cells by regulating CyclinA2 and Claudin1 gene expression. *Gene.* 2019;690:48–56.

44. Wang L, Wang X, Chen H, Zu X, Ma F, Liu K, et al. Gossypin inhibits gastric cancer growth by direct targeting of AURKA and RSK2. *Phytother Res.* 2019;33(3):640–50.
45. Kruiswijk F, Labuschagne CF, Vousden KH. p53 in survival, death and metabolic health: a lifeguard with a licence to kill. *Nat Rev Mol Cell Biol.* 2015;16(7):393–405.
46. S S, oncology SRJljo. Involvement of Bcl-2 family members, phosphatidylinositol 3'-kinase/AKT and mitochondrial p53 in curcumin (diferulolylmethane)-induced apoptosis in prostate cancer. 2007;30(4):905–18.
47. Lin HH, Huang HP, Huang CC, Chen JH, Wang CJ. Hibiscus polyphenol-rich extract induces apoptosis in human gastric carcinoma cells via p53 phosphorylation and p38 MAPK/FasL cascade pathway. *Mol Carcinog.* 2005;43(2):86–99.
48. Gharpure KM, Pradeep S, Sans M, Rupaimoole R, Ivan C, Wu SY, et al. FABP4 as a key determinant of metastatic potential of ovarian cancer. *Nat Commun.* 2018;9(1):2923.
49. Nie J, Zhang J, Wang L, Lu L, Yuan Q, An F, et al. Adipocytes promote cholangiocarcinoma metastasis through fatty acid binding protein 4. *J Exp Clin Cancer Res.* 2017;36(1):183.
50. Yang W-X, Pan Y-Y, You C-G. CDK1, CCNB1, CDC20, BUB1, MAD2L1, MCM3, BUB1B, MCM2, and RFC4 May Be Potential Therapeutic Targets for Hepatocellular Carcinoma Using Integrated Bioinformatic Analysis. *Biomed Res Int.* 2019;2019:1–16.
51. Branchi V, Garcia SA, Radhakrishnan P, Gyorffy B, Hissa B, Schneider M, et al. Prognostic value of DLGAP5 in colorectal cancer. *Int J Colorectal Dis.* 2019;34(8):1455–65.
52. Dong S, Huang F, Zhang H, Chen Q. Overexpression of BUB1B, CCNA2, CDC20, and CDK1 in tumor tissues predicts poor survival in pancreatic ductal adenocarcinoma. *Biosci Rep.* 2019;39(2).
53. Bolanos-Garcia VM. Aurora kinases. *Int J Biochem Cell Biol.* 2005;37(8):1572–7.
54. Guo M, Lu S, Huang H, Wang Y, Yang MQ, Yang Y, et al. Increased AURKA promotes cell proliferation and predicts poor prognosis in bladder cancer. *BMC Syst Biol.* 2018;12(Suppl 7):118.
55. Gao M, Kong W, Huang Z, Xie Z. Identification of Key Genes Related to Lung Squamous Cell Carcinoma Using Bioinformatics Analysis. *Int J Mol Sci.* 2020;21(8).

Tables

Table 1. Gene ontology (GO) annotation of DEGs in COAD.

GO	Category	Description	Count	%	Log10(P)	Log10(q)
GO:0044770	GO Biological Processes	cell cycle phase transition	45	10.32	-14.12	-10.06
GO:0000280	GO Biological Processes	nuclear division	36	8.26	-14.06	-10.06
GO:0051347	GO Biological Processes	positive regulation of transferase activity	37	8.49	-8.32	-5.38
GO:0140013	GO Biological Processes	meiotic nuclear division	17	3.90	-7.60	-4.94
GO:0006260	GO Biological Processes	DNA replication	19	4.36	-6.27	-3.79
GO:2000241	GO Biological Processes	regulation of reproductive process	14	3.21	-6.22	-3.75
GO:0015701	GO Biological Processes	bicarbonate transport	8	1.83	-6.06	-3.60
GO:0008285	GO Biological Processes	negative regulation of cell proliferation	34	7.80	-5.72	-3.29
GO:0048762	GO Biological Processes	mesenchymal cell differentiation	16	3.67	-5.61	-3.19
GO:0071407	GO Biological Processes	cellular response to organic cyclic compound	27	6.19	-5.60	-3.19
GO:0043408	GO Biological Processes	regulation of MAPK cascade	33	7.57	-5.52	-3.13
GO:0098771	GO Biological Processes	inorganic ion homeostasis	33	7.57	-5.46	-3.08
GO:1901699	GO Biological Processes	cellular response to nitrogen compound	30	6.88	-5.39	-3.02
GO:0000910	GO Biological Processes	cytokinesis	14	3.21	-5.33	-2.97
GO:0045177	GO Cellular Components	apical part of cell	34	7.80	-12.92	-9.27
GO:0005819	GO Cellular Components	spindle	27	6.19	-9.32	-6.23
GO:0005902	GO Cellular Components	microvillus	13	2.98	-8.20	-5.35

Abbreviations: DEGs differentially expressed genes; COAD Colon adenocarcinoma

GO	Category	Description	Count	%	Log10(P)	Log10(q)
GO:0016323	GO Cellular Components	basolateral plasma membrane	16	3.67	-5.41	-3.04
GO:0035173	GO Molecular Functions	histone kinase activity	6	1.38	-6.46	-3.97
GO:0016788	GO Molecular Functions	hydrolase activity, acting on ester bonds	33	7.57	-5.61	-3.19
Abbreviations: DEGs differentially expressed genes; COAD Colon adenocarcinoma						

Table 1. KEGG and Reactome annotation of DEGs in COAD.

GO	Category	Description	Count	%	Log10(P)	Log10(q)
hsa04110	KEGG Pathway	Cell cycle	16	3.67	-9.05	-6.08
hsa04964	KEGG Pathway	Proximal tubule bicarbonate reclamation	5	1.15	-4.32	-2.26
hsa04923	KEGG Pathway	Regulation of lipolysis in adipocytes	6	1.38	-3.38	-1.5
hsa04960	KEGG Pathway	Aldosterone-regulated sodium reabsorption	5	1.15	-3.3	-1.48
hsa04978	KEGG Pathway	Mineral absorption	5	1.15	-2.65	-1.03
R-HSA-1640170	Reactome Gene Sets	Cell Cycle	41	9.4	-10.9	-7.44
R-HSA-1475029	Reactome Gene Sets	Reversible hydration of carbon dioxide	5	1.15	-5.87	-3.47
R-HSA-382551	Reactome Gene Sets	Transport of small molecules	29	6.65	-4.18	-2.14
R-HSA-69273	Reactome Gene Sets	Cyclin A/B1/B2 associated events during G2/M transition	5	1.15	-4.13	-2.11
R-HSA-6804756	Reactome Gene Sets	Regulation of TP53 Activity through Phosphorylation	8	1.83	-3.58	-1.64
R-HSA-109582	Reactome Gene Sets	Haemostasis	24	5.5	-3.35	-1.5
R-HSA-8979227	Reactome Gene Sets	Triglyceride metabolism	5	1.15	-3.3	-1.48
R-HSA-418594	Reactome Gene Sets	G alpha (i) signalling events	17	3.9	-2.89	-1.16
R-HSA-420029	Reactome Gene Sets	Tight junction interactions	4	0.92	-2.76	-1.09
R-HSA-200425	Reactome Gene Sets	Carnitine metabolism	3	0.69	-2.74	-1.09
R-HSA-420092	Reactome Gene Sets	Glucagon-type ligand receptors	4	0.92	-2.55	-0.95
R-HSA-211945	Reactome Gene Sets	Phase I - Functionalization of compounds	7	1.61	-2.5	-0.91

Abbreviations: KEGG Kyoto Encyclopedia of Genes and Genomes; DEGs differentially expressed genes; COAD Colon adenocarcinoma

GO	Category	Description	Count	%	Log10(P)	Log10(q)
R-HSA-6785807	Reactome Gene Sets	Interleukin-4 and Interleukin-13 signalling	7	1.61	-2.46	-0.88
R-HSA-983189	Reactome Gene Sets	Kinesins	5	1.15	-2.38	-0.82
R-HSA-422085	Reactome Gene Sets	Synthesis, secretion, and deacylation of Ghrelin	3	0.69	-2.34	-0.8
Abbreviations: KEGG Kyoto Encyclopedia of Genes and Genomes; DEGs differentially expressed genes; COAD Colon adenocarcinoma						

Table 1. Functional annotation of the genes involved in the most significant module.

GO	Category	Description	Count	%	Log10(P)	Log10(q)
GO:0007059	GO Biological Processes	chromosome segregation	29	46.77	-37.06	-32.66
GO:0044770	GO Biological Processes	cell cycle phase transition	33	53.23	-35.09	-31.16
GO:0045787	GO Biological Processes	positive regulation of cell cycle	17	27.42	-15.9	-13.21
GO:0006260	GO Biological Processes	DNA replication	14	22.58	-14.21	-11.7
GO:0051321	GO Biological Processes	meiotic cell cycle	12	19.35	-11.78	-9.35
GO:0008608	GO Biological Processes	attachment of spindle microtubules to kinetochore	7	11.29	-11.68	-9.27
GO:0071103	GO Biological Processes	DNA conformation change	12	19.35	-10.45	-8.08
GO:0072331	GO Biological Processes	signal transduction by p53 class mediator	10	16.13	-8.81	-6.51
GO:0051347	GO Biological Processes	positive regulation of transferase activity	14	22.58	-8.75	-6.45
GO:0007062	GO Biological Processes	sister chromatid cohesion	6	9.68	-7.87	-5.62
GO:0032506	GO Biological Processes	cytokinetic process	5	8.06	-7.29	-5.1
GO:0044380	GO Biological Processes	protein localisation to cytoskeleton	5	8.06	-6.41	-4.27
GO:0005819	GO Cellular Components	spindle	23	37.1	-25.83	-22.66
GO:0030496	GO Cellular Components	midbody	14	22.58	-16.65	-13.9
GO:0005871	GO Cellular Components	kinesin complex	5	8.06	-6.53	-4.38
GO:0045171	GO Cellular Components	intercellular bridge	5	8.06	-5.97	-3.86
GO:0140097	GO Molecular Functions	catalytic activity, acting on DNA	7	11.29	-5.87	-3.78

Abbreviations: GO Gene ontology.

GO	Category	Description	Count	%	Log10(P)	Log10(q)
GO:0003682	GO Molecular Functions	chromatin binding	10	16.13	-5.86	-3.76
R-HSA-69273	Reactome Gene Sets	Cyclin A/B1/B2 associated events during G2/M transition	4	6.45	-6.32	-4.18
R-HSA-179409	Reactome Gene Sets	APC-Cdc20 mediated degradation of Nek2A	4	6.45	-6.32	-4.18
Abbreviations: GO Gene ontology.						

Table 1. Top 20 hub genes and corresponding degree

Gene symbol	Gene description	Score
CDK1	Cyclin Dependent Kinase 1	78
CCNB1	Cyclin B1	76
CCNA2	Cyclin A2	75
AURKA	Aurora Kinase A	75
CDC20	Cell Division Cycle 20	74
AURKB	Aurora Kinase B	72
TPX2	TPX2 Microtubule Nucleation Factor	71
BUB1	BUB1 Mitotic Checkpoint Serine/Threonine Kinase	70
CDC45	Cell Division Cycle 45	70
MAD2L1	Mitotic Arrest Deficient 2 Like 1	69
KIF2C	Kinesin Family Member 2C	69
NCAPG	Non-SMC Condensin I Complex Subunit G	69
DLGAP5	DLG Associated Protein 5	69
FOXM1	Forkhead Box M1	69
CENPF	Centromere Protein F	68
CENPE	Centromere Protein E	68
BUB1B	BUB1 Mitotic Checkpoint Serine/Threonine Kinase B	68
TTK	TTK Protein Kinase	68
ASPM	Abnormal Spindle Microtubule Assembly	68
KIF20A	Kinesin Family Member 20A	67

Figures

Figure 1

Venn diagram of DEGs from three datasets. (A) 267 downregulated DEGs. (B) 169 upregulated DEGs. Abbreviations: DEGs differentially expressed genes.

Figure 2

DEG and neighbouring gene enrichment analysis in COAD using Metascape. (A) Heatmap of GO enriched terms coloured by p-value. (B) Network of GO enriched terms coloured by p-value. Each node represents an enriched term. Dark colours indicate increased statistical significance. (C) Heatmap of KEGG and Reactome enriched terms coloured by p-value. (D) Network of KEGG and Reactome enriched terms coloured by p-value. Each node represents an enriched term. Darker colour indicates more statistical significance. Abbreviations: DEGs differentially expressed genes; COAD Colon adenocarcinoma; GO Gene ontology; KEGG Kyoto Encyclopedia of Genes and Genomes.

Figure 3

PPI network of DEGs, containing 369 nodes and 2708 edges. Red represents upregulated genes. Blue represents downregulated genes. Abbreviations: PPI protein-protein interaction; DEGs differentially expressed genes.

Figure 4

The most significant module gene network and hub genes analysis. (A) The most significant module in the PPI network contains 62 nodes and 1708 edges. (B) Network of 20 hub genes. Darker colours represent higher scores. (C) Biological process annotation of hub genes using ClueGO and CluePedia. $p < 0.01$ was considered statistically significant. (D) KEGG annotation of hub genes using ClueGO and CluePedia. $p < 0.01$ was considered statistically significant. (E) Heatmap of the top 20 hub genes was

constructed using the UALCAN database. Abbreviations: PPI protein-protein interaction; KEGG Kyoto Encyclopedia of Genes and Genomes.

Figure 5

Overall survival of the hub genes in COAD patients. (A-F) CCNB1, CCNA2, AURKA, NCAPG, DLGAP5, and CENPE showed a significant difference in overall survival (OS). High expression of the 6 genes indicated fa-vourable OS in COAD ($p<0.05$). (G-H) AURKA and CENPE showed statistically significant association with disease-free survival (DFS) and indicated fa-vourable disease free survival in COAD ($p<0.05$). Abbreviations: COAD Colon adenocarcinoma.

significantly related to nodal metastasis status, and the highest mRNA expression tended to appear at the N0 phase. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure 7

Protein expression analysis of the 5 hub genes were performed using the HPA database. Except for MAD2L1, the other 4 proteins showed a higher degree of staining in tumour tissue compared to normal tissues.

Figure 8

Alterations of the eight hub genes analysed by cBioPortal. (A) OncoPrint of genetic alterations in 378 COAD cases. (B) Alteration frequency of eight hub genes. Gene expression was altered in 42.86% of 378 cases. Abbreviations: COAD Colon adenocarcinoma.