DOI: 10.53469/jcmp.2024.06(12).25

Molecular Mechanisms and Clinical Predictive Value of Colorectal Cancer Susceptibility Genes: WGCNA Analysis and Mendelian Randomization Study

Dan Liu, Qinlang Liu, Lei Song, Linmei Sun*

The Second Ward of Anorectal Disease Hospital, Xi'an Hospital of Traditional Chinese Medicine, Xi'an 710016, China
**Correspondence Author*

Abstract: ***Background and objective:*** ***Colorectal cancer (CRC) is a highly heterogeneous disease, making treatment and prognosis prediction challenging. Early diagnosis of CRC and identification of gene expressions associated with its onset are crucial for prognosis, especially before clinical symptoms appear. This study aims to explore potential key genes involved in CRC and evaluate their clinical application in predicting the disease.*** ***Methods:*** ***This study utilizes differential expression analysis and Weighted Gene Co-expression Network Analysis (WGCNA) to identify novel susceptibility modules and key genes associated with colorectal cancer (CRC). Through KEGG and GO analyses, we aim to investigate the potential functions of these key genes. Subsequently, we will construct a Nomogram model and assess its diagnostic value for CRC using ROC curves. Based on genome-wide association studies, a Mendelian randomization analysis will be conducted to determine the causal relationship between these key genes and CRC. Finally, we will explore the association between these key genes, which are causally linked to CRC risk factors, and immune cell infiltration.*** ***Results:*** ***A gene co-expression network was constructed using WGCNA, from which key modules related to colorectal cancer (CRC) were identified, along with 963 overlapping key genes derived from WGCNA. GO and KEGG pathway enrichment analyses revealed that these genes are involved in the biosynthesis of ribonucleoprotein complexes, rRNA metabolic processes, chromatin organization-regulated signaling pathways, as well as cell cycle, DNA replication, and ribosome-related pathways. Using Cytoscape software, we identified the top five highly expressed genes: CDC2, CCNB1, CCNA2, TOP2A, and CCNB2. We then developed a Nomogram model, which effectively predicts the risk of CRC. The performance of this model in CRC diagnosis was further validated through ROC curve analysis, showing promising diagnostic accuracy. Finally, we focused on CDC2 and observed a causal relationship between CDC2 and immune cell infiltration in CDAD. Through inverse variance-weighted analysis, we found that CDC2 significantly increased the risk of CDAD, with an OR of 1.0005 (95% CI = 1.0001-1.001, P = 0.01).*** ***Conclusion:*** ***We successfully identified the core genes associated with colorectal cancer (CRC). This finding provides important insights for further research into early diagnostic methods for CRC, while also contributing to the understanding of the molecular mechanisms underlying CRC risk genes.***

Keywords: CRC, WGCNA, CDC2, Mendelian randomization analysis**.**

1. Introduction

Colorectal cancer (CRC) is considered one of the most aggressive cancers globally, causing over 50,000 deaths annually [1]. The disease is characterized by insidious onset, rapid progression, and a high propensity for chemotherapy resistance [2,3], placing a heavy burden on both society and healthcare systems. Advanced CRC patients often face the risk of recurrence and metastasis, with a 5-year survival rate of less than 10% [4-6]. In contrast, early-stage CRC patients who undergo surgical treatment experience a significant increase in 5-year survival rates, reaching up to 90% [7]. Therefore, early diagnosis and identification of gene expressions associated with CRC are crucial for both prognosis and treatment [8,9].

The etiology and pathogenesis of CRC are complex and multifactorial, and their full mechanisms remain unclear, potentially involving genetic, environmental, and lifestyle factors. With ongoing advancements in science and technology, bioinformatics and Mendelian randomization (MR) analysis have become essential tools in CRC research, providing valuable insights at different levels of disease investigation. In this study, we utilized multiple bioinformatics software and databases to identify pathways associated with the disease, including Weighted Gene Co-expression Network Analysis (WGCNA), Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis, and Gene Set Enrichment Analysis (GSEA). WGCNA is a bioinformatics method used to explore gene expression data, aiming to uncover gene-gene interactions and their relationships with biological traits. This approach is widely applied in biological research, especially for identifying gene networks linked to complex diseases and biological processes. It can be used with various types of high-dimensional data, including proteomics, genetic marker data, gene expression profiles, and metabolomics [10]. Additionally, WGCNA can assist in identifying potential therapeutic targets and candidate biomarkers. Therefore, our study aims to employ these methods to identify CRC-related genes, novel biomarkers, and potential disease mechanisms.

By conducting an in-depth analysis of CRC-related data, we identified differentially expressed genes (DEGs). We then used WGCNA to pinpoint gene modules most strongly associated with CRC, narrowing down the pool of candidate genes. Subsequently, a protein-protein interaction (PPI) network was constructed using Cytoscape software, and the top 5 hub genes were selected based on degree centrality scores using the CytoHubba plugin. These genes—CDC2, CCNB1, CCNA2, CCNB2, and TOP2A—contribute to the diagnostic model for CRC and shed light on the potential

mechanisms underlying CRC risk genes.

Mendelian randomization (MR) is a biostatistical method used to investigate causal relationships, named after Austrian geneticist Gregor Mendel, one of the founders of genetics. Recently, MR has gained widespread popularity as a reliable method, utilizing single nucleotide polymorphisms (SNPs) as instrumental variables (IVs) to assess causal relationships between exposures and outcomes [11,12]. MR employs genetic variations closely related to exposure factors as IVs, allowing the inference of causal relationships between exposures and outcomes. In our study, MR was used to explore the causal relationship between CDC2 and CRC. Additionally, bioinformatics analyses, such as organ-specific localization and immune cell infiltration analysis, provided molecular foundations and guidance for personalized drug therapy, aiming for precise, individualized, and optimized treatment. The research workflow is illustrated in Figure 1.

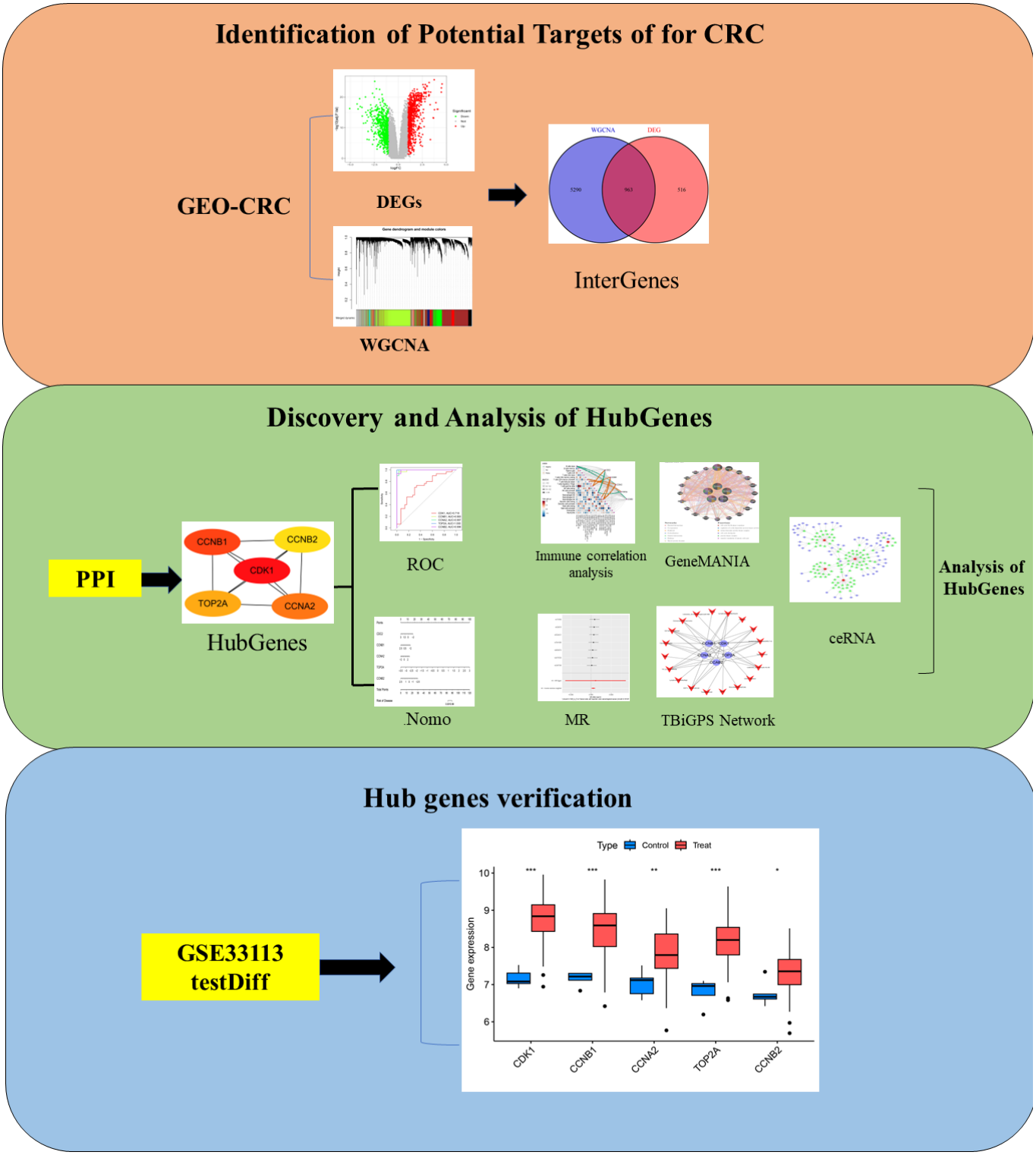

Figure 1: Work flow chart

2. Materials and Methods

2.1 Data source

We used the GSE74602 dataset, which includes 30 normal samples and 30 colorectal tumor samples, as the training set, and the GSE17536 dataset as the validation set. All data were retrieved from the Gene Expression Omnibus (GEO) database.

2.2 Identification of Potential Targets of for CRC

2.2.1 Differentially expressed genes identification

First, we used R software (version 4.4.1) to read the GSE74602 dataset and performed preprocessing, including batch correction and normalization. Next, we employed the 'limma' package to conduct differential expression analysis (DEG) screening. After performing statistical analysis on the expression levels, we generated volcano plots and DEG expression heatmaps using the 'pheatmap' and 'ggplot2' packages in R.

2.2.2 Weighted gene co-expression network analysis

WGCNA is a powerful tool that helps identify modules and key genes within gene co-expression networks, as well as their relationships with biological traits. It can also be used to identify candidate biomarkers [13]. In this study, we used the 'WGCNA' package in R to construct a colorectal cancer (CRC) gene co-expression network. Finally, we evaluated the correlation between different modules and the pathogenesis of CRC, selecting the module most strongly associated with CRC as the core module of the WGCNA analysis.

2.2.3 Screening of candidate pivotal genes and Go/KEGG analysis

The intersecting genes between WGCNA and DEGs were selected as candidate core genes related to the pathogenesis of CRC. We also utilized KEGG, a comprehensive database resource for gene functional analysis. Next, we performed Gene Ontology (GO) and KEGG enrichment analyses using the 'clusterProfiler' package in R to gain deeper insights into the potential mechanisms of disease progression and pathogenesis.

2.3 Discovery and Analysis of HubGenes

2.3.1 Protein–protein interaction network hub gene

STRING and Cytoscape were used to predict and visualize molecular interactions and the protein-protein interaction (PPI) network. Subsequently, the degree centrality algorithm in Cytoscape was applied to rank the key genes within the PPI network.

2.3.2 Nomogram model construction

The 'rms' package was used to construct a Nomogram model for predicting the risk of colorectal cancer (CRC). To evaluate the performance of the Nomogram model, we calculated Harrell's Concordance Index, a metric used to assess predictive accuracy. Next, the 'ROC' package was used to construct receiver operating characteristic (ROC) curves to validate the diagnostic effectiveness of the candidate biomarkers. The area under the ROC curve (AUC) was used to represent accuracy, with an AUC value between 0.7 and 1.0 generally considered indicative of good accuracy.

2.3.3 Immune cell analysis

To investigate the function of immune cells in CRC, we used the CIBERSORT analysis method to assess the immune cell infiltration levels of 22 immune cell types in colorectal cancer (CRC) [14].

2.3.4 Mendelian randomization

In this study, we employed a two-sample Mendelian randomization (MR) approach to explore the causal relationship between hub genes and CRC risk, with SNPs defined as instrumental variables (IVs). The hub gene data were obtained from publicly available genome-wide association study (GWAS) datasets. We selected CDC2, the gene with the highest degree centrality, and colon cancer as a representative disease of CRC lesions for the MR analysis. Data for CDC2 can be found at [https://gwas.mrcieu.ac.uk/datasets/?trait__icontains=CDC2], and data for colon cancer are available at [https://gwas.mrcieu.ac.uk/datasets/?gwas_id__icontains=&year__iexact=&trait__icontains=colon+cancer&consortium__icontains=]. MR analysis was conducted using the 'TwoSampleMR' package, and inverse variance weighting (IVW) was applied to evaluate the relationship between the central gene levels and the risk of CRC. Additional sensitivity analyses were performed using MR-Egger [15,16].

3. Results

3.1 Identification of Potential Targets of for CRC

3.1.1 DEGs screening

CRC datasets (GSE74602) were obtained from the GEO database, and differentially expressed genes (DEGs) in CRC were identified. We identified 1,480 genes (811 upregulated and 668 downregulated) as shown in Figure 2-A.

3.1.2 Construction of WGCNA network and identification of neuropathic pain-related module

To determine whether potential gene modules are associated with CRC, we performed WGCNA analysis on all candidate genes from the CRC dataset (GSE24982) (Figure 2-B). We identified 11 distinct modules (Figure 2-C). Finally, we extracted the genes from the MEbrown module, which had the highest correlation coefficient and the lowest P-value in the GSE24982 dataset, for subsequent analysis.

3.1.3 Go/KEGG analyses

We screened for shared genes between the WGCNA-derived module genes and the DEGs. A total of 963 overlapping genes were selected as candidate HubGenes, which may play an important role in the occurrence and progression of CRC

(Figure 2-D). We conducted GO and KEGG analyses to further explore the potential functions of these 440 overlapping genes (Figure 2-E, 2-F). GO enrichment analysis indicated that the shared genes primarily affect biological functions such as ribonucleoprotein complex biogenesis, ribosome biogenesis, rRNA processing, and rRNA metabolic processes. KEGG enrichment analysis showed that these genes mainly influence pathways such as the cell cycle, DNA replication, purine metabolism, ribosome, and ribosome biogenesis in eukaryotes.

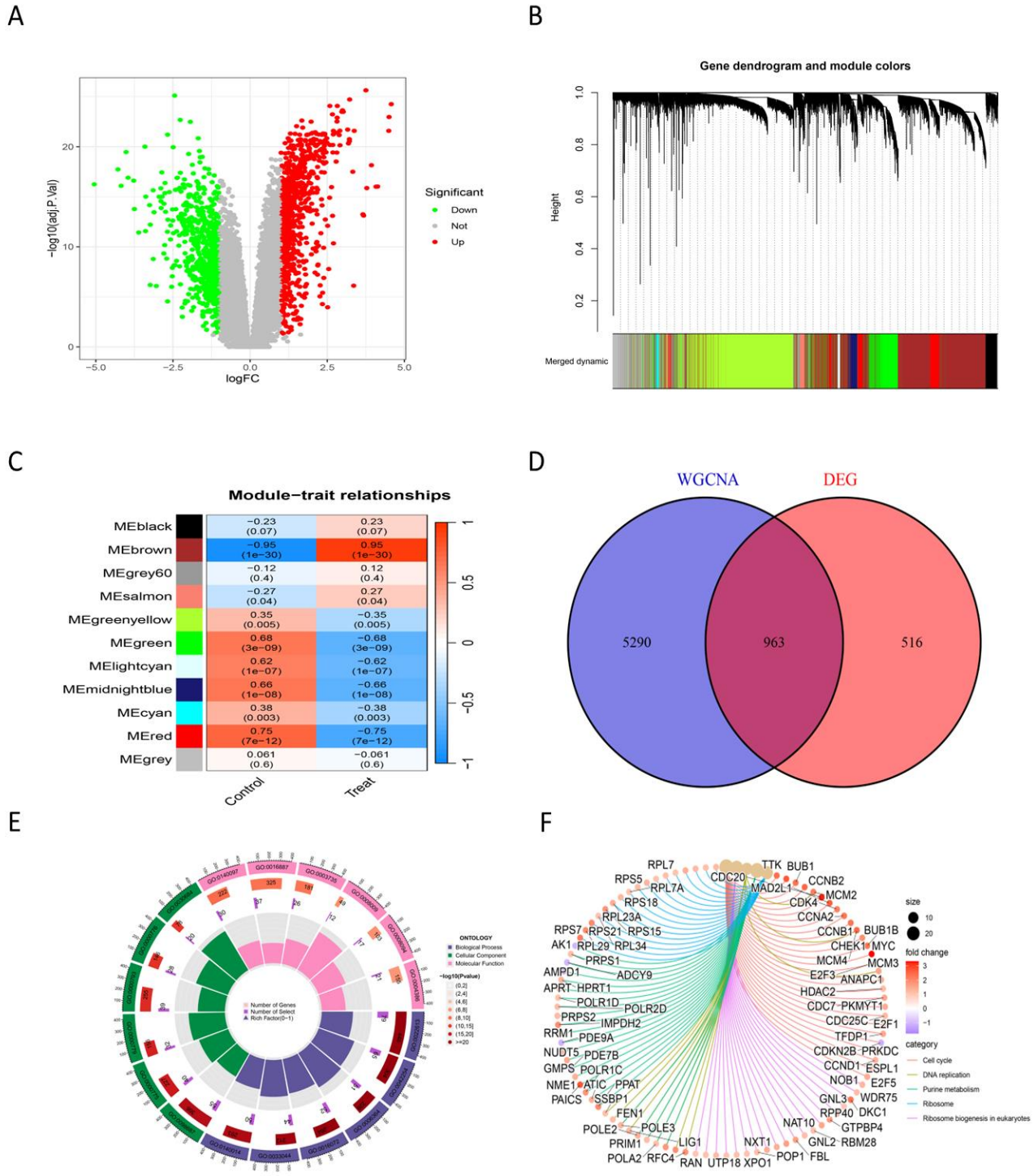

Figure 2: 2-A Volcano map of differentially expressed genes, red for up-regulated genes and green for down-regulated genes; 2-B Clustering of dendrograms of all genes in the GSE24982 dataset based on the topological overlap matrix (1-TOM). Each branch in the clustering tree represents a gene, and co-expression modules are constructed in different colors; 2-C Clustered gene modules in the GSE24982 dataset and module-trait heatmap of CRC. Each module contains the corresponding correlation coefficient and P value; 2-D Venn diagram showing 963 overlapping candidateHubGenes; 2-E GO enrichment analysis. 2-F KEGG pathway enrichment analysis.

3.2 Discovery and Analysis of HubGenes

3.2.1 PPI network analysis for hub genes

The STRING online tool was used to construct a PPI network with an interaction score greater than 0.9 (Figure S1-A). Subsequently, Cytoscape software was employed to visualize the top 5 upregulated genes (Figure 3-A). The five core genes identified were CDC2, CCNB1, CCNA2, TOP2A, and CCNB2. The color intensity represents the score, with darker colors indicating higher scores. According to the gene expression difference volcano plot (Figure 3-B), all these HubGenes were upregulated. We performed differential expression analysis of these five HubGenes in the validation cohort GSE33113. Consistent with our predictions, the mRNA expression levels of the five hub genes were significantly upregulated in IBD samples compared to control samples (Figure 3-C).

3.2.2 Construction of nomogram model for CRC risk prediction

A nomogram model was constructed using the five core genes CDK1 (CDC2), CCNB1, CCNA2, TOP2A, and CCNB2 to predict CRC risk (Figure 3-D). The model demonstrated good performance in CRC prediction. Subsequently, ROC curves for the five core genes were calculated to assess diagnostic efficacy. The nomogram's AUC for distinguishing CRC control and experimental samples showed high accuracy (Figure 3-E). The AUC values for the five core genes were as follows: CDK1 (CDC2) AUC = 0.718, CCNB1 AUC = 0.993, CCNA2 AUC = 0.997, TOP2A AUC = 1.000, and CCNB2 AUC = 0.996. All five core genes were upregulated (Figure 3-D). To further understand the disease risk, we also constructed a line chart analysis for the five HubGenes (Figure 3-C). Among the five hub genes, the most accurate prediction of disease incidence was for TOP2A, with a prediction rate of 99%.

3.2.3 Protein-Protein Interaction (PPI) of HubGenes Using GeneMANIA

GeneMANIA was used to predict functionally similar genes to the HubGenes. The identified HubGenes were input into the GeneMANIA database (https://genemania.org/) [17], and 20 similar hub genes were obtained (Figure 3-F). A PPI network was constructed, with the HubGenes in the inner circle and the predicted genes in the outer circle. The functions of these genes were primarily associated with cell cycle G2/M phase transition, regulation of cyclin-dependent protein kinase activity, serine/threonine protein kinase complex, cell cycle checkpoint, protein kinase complex, and negative regulation of mitotic cell cycle. These findings align with previous studies on inflammatory bowel disease (IBD) [18-21]. Dysregulation of cyclin-dependent protein kinases (CDKs) plays a critical role in the onset and progression of cancers, particularly in colorectal cancer (CRC). Cell cycle checkpoints, which are key control points in the cell division process, ensure that essential events and DNA repair are completed before a cell progresses to the next stage of the cell cycle. In cancers like CRC, disruption of these checkpoints can lead to abnormal cell proliferation and the development of cancer cells.

3.2.4 Construction of the Key Target-Organ Tissue Network

The BioGPS database is an online resource platform for gene annotation and querying gene expression across tissues and cells. The five HubGenes were imported into BioGPS, with human as the selected species. The top 10 organs or tissues with the highest gene expression for each HubGene were queried via the Interactive Image feature. The resulting data was used to construct the "HubGenes-Organ Tissue Network" using Cytoscape (Figure 3-G). The five HubGenes primarily affect Leukemia lymphoblastic (MOLT-4), Cardiac Myocytes, Leukemia promyelocytic-HL-60, Lymphoma Burkitt's (Daudi), Colorectal adenocarcinoma, Thymus, and Bronchial Epithelial Cells. Although the BioGPSHubGenes information did not include the spleen, expression was noted in immune cells closely related to the spleen, such as CD105+_Endothelial, 721_B_lymphoblasts, CD71+_EarlyErythroid, CD34+, and CD33+_Myeloid. This validates the immune regulatory role of the HubGenes from an immunological perspective and provides a biological basis for immunotherapy in CRC. Interestingly, these five HubGenes are directly localized to colorectal cancer tissues.

3.2.5 Construction of lncRNA-miRNA-mRNA ceRNA Network

Based on the miRDB database (http://mirdb.org/), miRTarBase database (http://mirtarbase.mbc.nctu.edu.tw/php/index.php), and TargetScan (https://www.targetscan.org/vert_80/), we identified the miRNA target genes of the HubGenes. A total of four HubGenes were found to target 111 miRNAs. Subsequently, using the SpongeScan database, we identified 64 lncRNAs that interact with these miRNAs, resulting in the construction of an lncRNA-miRNA-mRNA network. This network was imported into Cytoscape 3.10.2.0 for visualization of the ceRNA network (Figure 3-H). This diagram visually reflects the competition between LncRNAs and HubGenes for binding miRNAs.

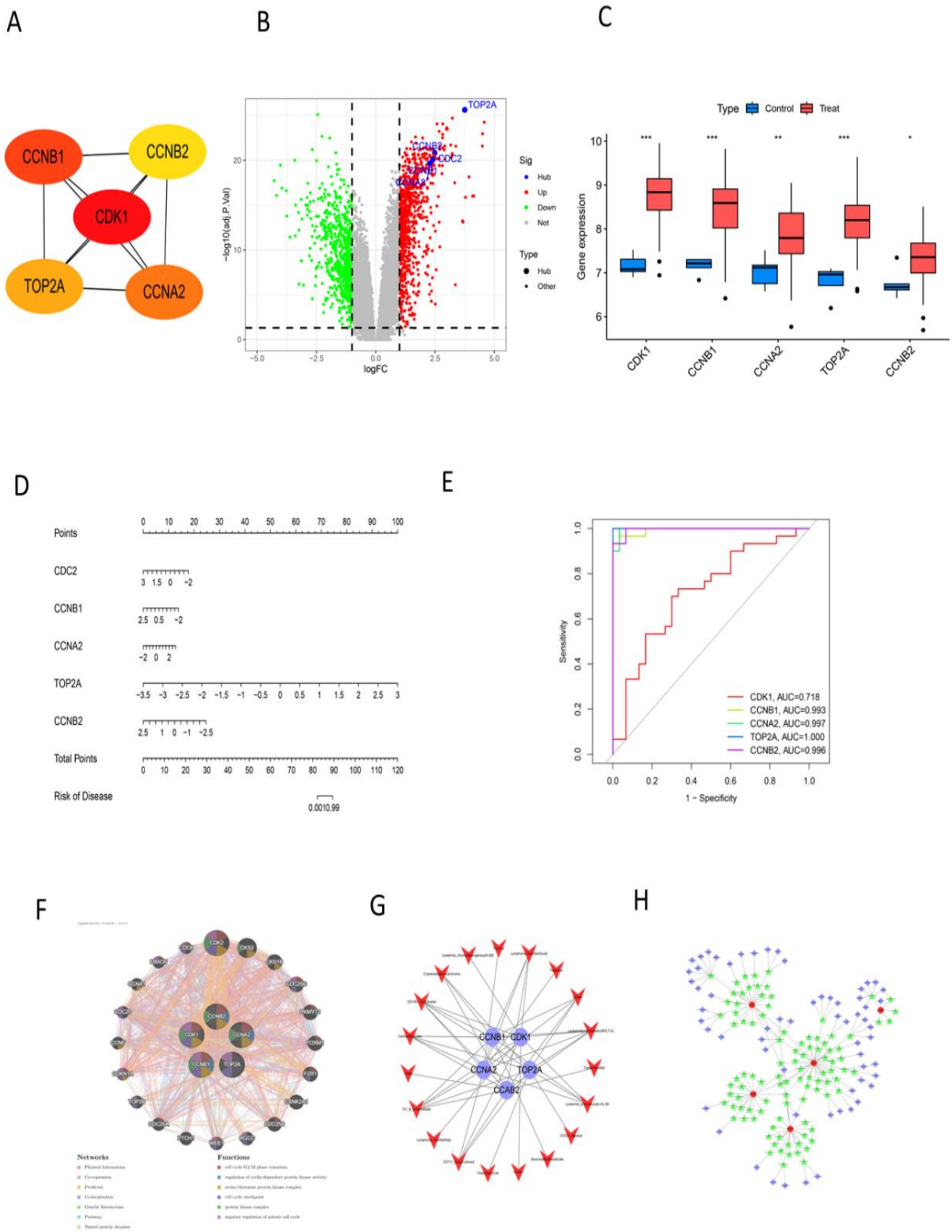

Figure 3: 3-A The core genes of the interactive network are obtained through the degree algorithm; 3-B Hub gene volcano map, red is up-regulated genes, green is down-regulated genes; 3-C HubGenes differential analysis; 3-D Hub gene Nomogram model; 3-E ROC curve, evaluate the diagnostic effect of the nomogram model and each hub gene; 3-F GeneMANIA; 3-G BioGPSNetWork, Note: The blue circle represents HubGenes, and the red "V" represents organ tissue. 3-H ceRNANetwaork, Note: The red circle represents HubGenes, the green "V" represents miRNA, and the blue diamond represents lncRNA.

4. Assessment of Immune Cell Infiltration in CRC

We used the CIBERSORT algorithm to calculate the abundance of various immune cell infiltrates in CRC patients (Figure 4-A). The violin plot showing differences in immune cell infiltration is displayed in Figure 4-B. Compared to the normal control samples, the experimental group exhibited significantly higher proportions of B cells naive, plasma cells, T cells CD4 memory resting, macrophages M0, M1, M2, mast cells activated, and neutrophils (P<0.001). Through immune cell correlation analysis, we assessed the relationship between immune cells and HubGenes. As shown in Figure 4-D, the upper right corner clearly indicates that CDC2 is positively correlated with T cells gamma delta and negatively correlated with monocytes; CCNB1 is positively correlated with both T cells gamma delta and macrophages M2; CCNA2 is positively correlated with macrophages M2 and T cells CD4 memory activated; TOP2A is negatively correlated with monocytes and positively correlated with T cells CD4 memory activated; CCNB2 is negatively correlated with B cells naive. Additionally, immune cell PCA analysis (Figure 4-C) visually reflects the ability to distinguish between normal and experimental groups based on immune cell content. These results demonstrate the direct relationship between immune cell infiltration and the key regulatory factor CDC2 in cell cycle regulation in colorectal cancer.

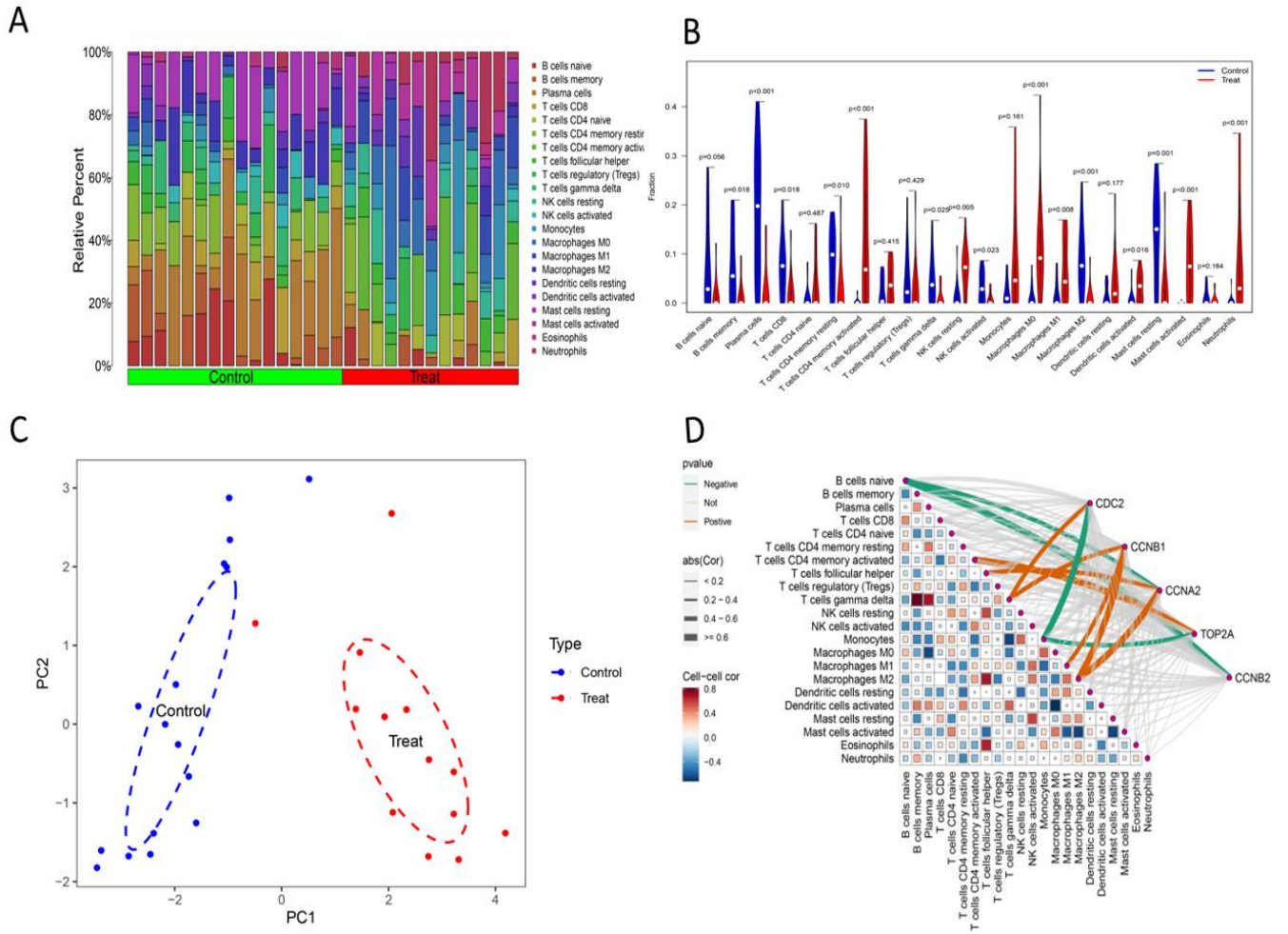

Figure 4: 4-A Abundance of immune cell infiltration; 4-B immune cell differential analysis; 4-C immune cell PCA analysis; 4-D immune cell correlation; in the lower left corner of the graph, red represents positive correlation, blue represents negative correlation, the red line in the upper right corner represents positive correlation, the green line represents negative correlation, and the thickness of the line represents the absolute value of the correlation coefficient.

5. CDC2 was Causally Associated with the Risk of Colon Cancer

The SNP characteristics of CDC2 and colon cancer are shown in the supplementary table Table S3. MRresult. None of the SNPs are weak instrumental variables. The causal effects of each genetic variation on colon cancer are illustrated in Figure 5-A and 5-B. We evaluated the causal relationship between CDC2 levels and colon cancer. Using the IVW method, we found that CDC2 is associated with the risk of colon cancer, with an odds ratio (OR) of 1.001 (95% CI = 1.0001–1.001, P = 0.01). The MR-Egger method showed no significant statistical significance [OR = 1.001, 95% CI = 0.993–1.009, P = 0.78]. The funnel plot of the causal effect appears roughly symmetrical (Figure 5-C), and the intercept of the MR-Egger regression showed no evidence of horizontal pleiotropy (P = 0.96), further suggesting that pleiotropy does not bias the causal effect. As shown in Figure 5-D, after removing each SNP, we systematically re-conducted MR analysis on the remaining SNPs. The results remained consistent, indicating that the causal relationship calculated from all SNPs is significant. This also suggests that there are no dominant SNPs affecting the relationship between CDC2 levels and

colon cancer, validating the previous MR results.

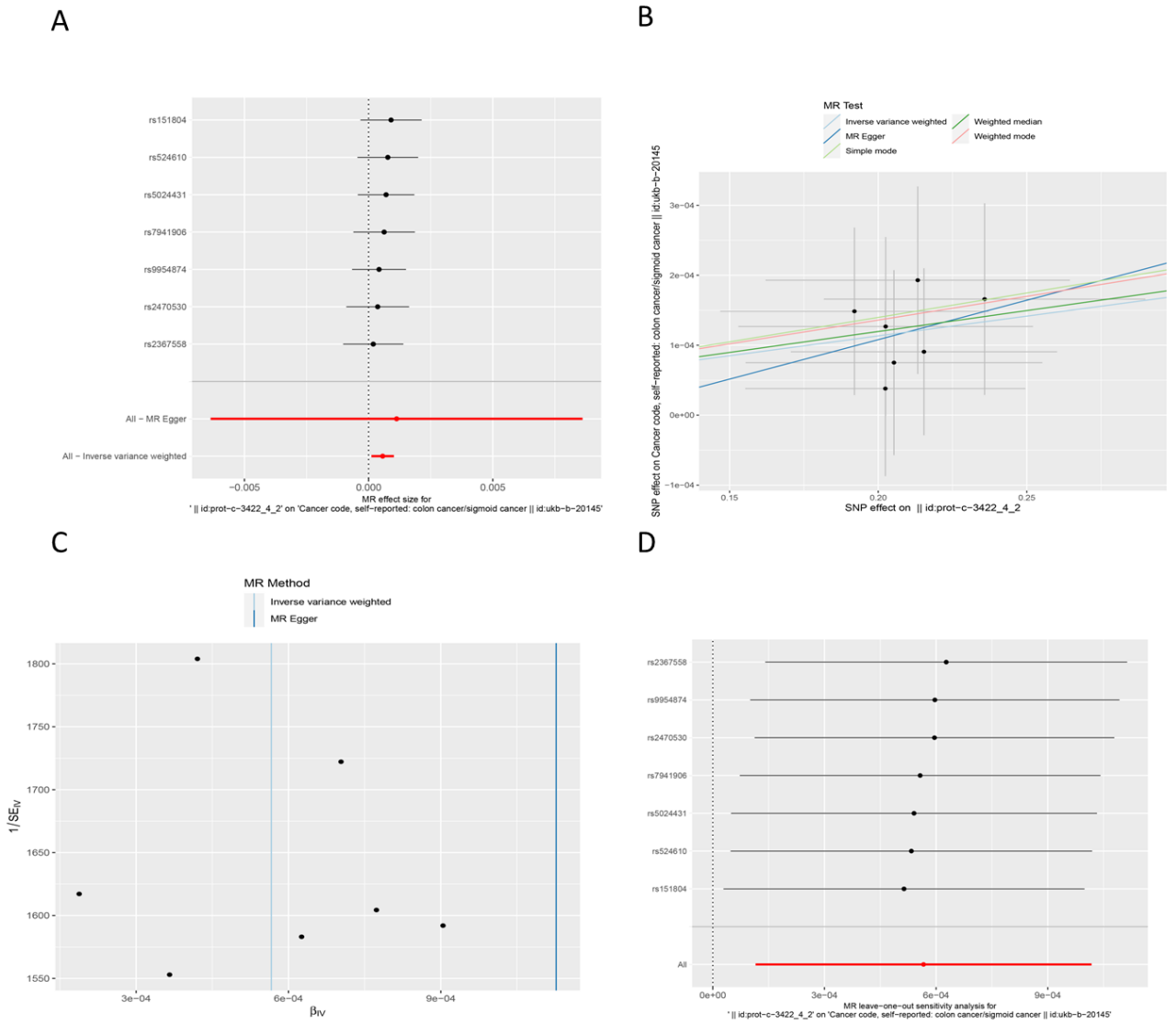

Figure 4: 4-A Forest plot shows the causal effect of each SNP on colon cancer risk; 4-B Scatter plot shows the causal effect of CDC2 on colon cancer risk; 4-C Funnel plot shows the overall heterogeneity of MR estimates of the effect of CDC2 on colon cancer. 4-D Visualization of the causal effect of CDC2 on colon cancer risk when one SNP is removed.

6. Discussion

Colorectal cancer (CRC) is a malignant tumor whose treatment and prognosis are largely stage-dependent. Early diagnosis and intervention can significantly enhance patient survival rates. With advancements in molecular biology, CRC-related genes have emerged as key areas of research [22]. By quantifying gene expression across thousands of genes and obtaining genome-wide expression data, researchers can unravel complex gene regulatory networks, uncover molecular characteristics of CRC, and identify potential diagnostic and therapeutic targets. These findings contribute to improved therapeutic strategies, offering better prognostic outcomes and increased survival opportunities for CRC patients.

Diagnostic Biomarkers and Current Limitations

Diagnostic biomarkers are molecules or substances that aid in disease diagnosis, progression tracking, treatment planning, and efficacy evaluation. Currently, endoscopic pathology is the gold standard for diagnosing CRC. However, this method is invasive, costly, and demands high expertise from endoscopists, often leading to poor patient compliance [23]. Considering CRC's high mortality rate and low early detection rates, identifying sensitive early diagnostic markers and reliable prognostic indicators is an urgent need.

Serum carcinoembryonic antigen (CEA) is widely used due to its simplicity, rapidity, and minimally invasive nature. However, its sensitivity and specificity are suboptimal [24,25]. Most researchers agree that a single tumor marker cannot meet clinical requirements, and combined testing strategies are essential to improve diagnostic accuracy [26-28]. In our study, Weighted Gene Co-expression Network Analysis (WGCNA) was used to identify hub genes associated with CRC.

Key Findings from Hub Gene Analysis

Our nomogram model exhibited strong predictive performance for CRC. The ROC curves for five hub

genes—CDC2 (CDK1), CCNB1, CCNA2, TOP2A, and CCNB2—demonstrated their diagnostic effectiveness, with high area under the curve (AUC) values, effectively distinguishing CRC cases from controls (Figure 3-C).

Protein-protein interaction (PPI) network analysis underscored the critical roles of these hub genes in CRC pathogenesis. CDC2, also known as CDK1, is a cyclin-dependent kinase pivotal in cell cycle regulation. Abnormal CDC2 activity contributes to unregulated proliferation, a hallmark of cancer. Clinical applications of CDK1 inhibitors have shown promise in disrupting cancer cell cycles [29-31]. CCNA2, encoding cyclin A, and CCNB1/CCNB2, encoding cyclin B, form active complexes with CDC2, orchestrating DNA replication and mitosis during the cell cycle. Overexpression of these cyclins correlates with heightened malignancy, metastasis, and poor prognosis in CRC [32-35].

TOP2A, encoding DNA topoisomerase IIα, is critical for DNA replication and repair. Its overexpression disrupts genomic stability and is associated with increased CRC malignancy and drug sensitivity to topoisomerase inhibitors [36].

Pathway Enrichment Analysis

Gene Ontology (GO) analysis revealed enrichment in ribosomal biogenesis and RNA metabolic processes, suggesting that ribosomal RNA abnormalities contribute to CRC pathogenesis. Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis highlighted significant enrichment in cell cycle and DNA replication pathways. These pathways regulate cell growth, DNA synthesis, and mitosis, which are frequently dysregulated in CRC [37-41].

Abnormal activation of pathways like Wnt, p53, and RAS-RAF-MEK-ERK further contributes to genomic instability and uncontrolled proliferation in CRC. Our results emphasize the cell cycle and DNA replication pathways' critical roles in CRC initiation, progression, and metastasis.

Novel Insights from Mendelian Randomization Analysis

Using Cytoscape software, CDC2 emerged as the most significant hub gene in the PPI network. This is the first study employing two-sample Mendelian randomization (MR) to explore the causal relationship between CDC2 levels and CRC risk using genome-wide association study (GWAS) data. Our MR analysis revealed a potential causal link between serum CDC2 levels and increased CRC risk. Rigorous controls for confounders and SNP-related biases, coupled with MR-Egger regression testing, ensured robust findings.

CeRNA Network Analysis

In CRC, competitive endogenous RNA (ceRNA) networks play a pivotal role in tumor progression. Certain long non-coding RNAs (lncRNAs) and messenger RNAs (mRNAs) regulate each other's expression by competitively binding to microRNAs (miRNAs). Our ceRNA network analysis identified 168 nodes and 175 edges, underscoring the hub genes' strong associations with CRC.

Focusing on miRNAs related to CDC2, 22 key candidates were identified, including hsa-miR-590-3p, hsa-miR-182-5p, and hsa-miR-802, which are implicated in CRC cell proliferation, invasion, and apoptosis [42-50]. Further investigation into these miRNAs may uncover new biomarkers and therapeutic targets.

Limitations and Future Directions

While our study provides meaningful insights, limitations exist. First, only one dataset was used due to limited microarray data in the CRC field, restricting generalizability. Second, our findings rely on bioinformatics analyses, necessitating further experimental validation of core gene mechanisms. Prospective studies integrating larger datasets and experimental approaches will strengthen these findings.

In summary, our study identifies critical hub genes and molecular pathways involved in CRC, providing valuable insights into potential diagnostic and therapeutic strategies. Further validation of these targets could significantly advance CRC management and patient outcomes.

7. Conclusion

This study identified key genes associated with colorectal cancer (CRC) through differential expression analysis and weighted gene co-expression network analysis (WGCNA), highlighting their potential applications in CRC diagnosis and prognosis. By constructing a gene co-expression network, we pinpointed 963 key genes. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses demonstrated the involvement of these genes in critical biological processes such as ribonucleoprotein complex biogenesis, rRNA metabolism, the cell cycle, and DNA replication. Specifically, five highly expressed genes—CDC2, CCNB1, CCNA2, TOP2A, and CCNB2—were identified through Cytoscape analysis as potentially pivotal in CRC development and progression.

Additionally, a nomogram model was constructed to predict CRC risk, with its diagnostic efficacy validated through receiver operating characteristic (ROC) curves, demonstrating high accuracy. Among the identified genes, CDC2 emerged as a crucial factor associated with immune cell infiltration. Using Mendelian randomization studies, we established a causal relationship between CDC2 expression and immune cell infiltration in CRC, revealing that elevated CDC2 levels significantly increase the risk of CRC-related immune dysfunction. These findings provide novel insights and theoretical support for the early diagnosis of CRC and its immunotherapy strategies.

In conclusion, this study proposes new biomarkers and molecular targets for the early detection and treatment of CRC, with a particular focus on CDC2. The results underscore its significant role in CRC immune infiltration and pathological progression, offering a theoretical basis for personalized treatment approaches and prognosis evaluation in CRC management.

Funding

This work was supported by the Science and Technology Department of Shaanxi Province (2023-YBSF-500); Xi 'an health committee topic (2023yb35).

References

[1] Haggar, Fatima A, and Robin P Boushey. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors[J]. Clinics in colon and rectal surgery vol. 22,4 (2009): 191-7. doi:10.1055/s-0029-1242458.

[2] He J, Pei L, Jiang H, et al. Chemoresistance of colorectal cancer to 5-fluorouracil is associated with silencing of the BNIP3 gene through aberrant methylation[J]. Journal of Cancer vol. 8,7 1187-1196. 9 Apr. 2017, doi:10.7150/jca.18171.

[3] Marin JJ, Sanchez de Medina F, Castaño B, et al. Chemoprevention, chemotherapy, and chemoresistance in colorectal cancer[J]. Drug metabolism reviews vol. 44,2 (2012): 148-72. doi:10.3109/03602532.2011.638303.

[4] Bertheloot D, Latz E, Franklin BS. Necroptosis, pyroptosis and apoptosis: an intricate game of cell death[J]. Cellular & molecular immunology vol. 18,5 (2021): 1106-1121. doi:10.1038/s41423-020-00630-3.

[5] Bhandari A, Woodhouse M, Gupta S. et al. Colorectal cancer is a leading cause of cancer incidence and mortality among adults younger than 50 years in the USA: a SEER-based analysis with comparison to other young-onset cancers[J]. Journal of investigative medicine: the official publication of the American Federation for Clinical Research vol. 65,2 (2017): 311-315. doi:10.1136/jim-2016-000229.

[6] Benson AB 3rd, Venook AP, Cederquist L, et al. Colon Cancer, Version 1.2017, NCCN Clinical Practice Guidelines in Oncology[J]. Journal of the National Comprehensive Cancer Network: JNCCN vol. 15,3 (2017): 370-398. doi:10.6004/jnccn.2017.0036.

[7] Barclay J, Clark AK, Ganju P et al. Role of the cysteine protease cathepsin S in neuropathic hyperalgesia[J]. Pain. 2007;130(3):225-234. doi:10.1016/j.pain.2006.11.017.

[8] von Schack D, Agostino MJ, Murray BS, et al.Dynamic changes in the microRNA expression profile reveal multiple regulatory mechanisms in the spinal nerve ligation model of neuropathic pain[J]. PloS one vol. 6,3 e17670. 14 Mar. 2011, doi:10.1371/journal.pone.0017670.

[9] Nangraj AS, Selvaraj G, Kaliamurthi S et al. Integrated PPI- and WGCNA-Retrieval of Hub Gene Signatures Shared Between Barrett's Esophagus and Esophageal Adenocarcinoma[J]. Frontiers in pharmacology vol. 11 881. 31 Jul. 2020, doi:10.3389/fphar.2020.00881

[10] Benson AB 3rd, Venook AP, Cederquist L et al. Colon Cancer, Version 1.2017, NCCN Clinical Practice Guidelines in Oncology[J]. Journal of the National Comprehensive Cancer Network: JNCCN vol. 15,3 (2017): 370-398. doi:10.6004/jnccn.2017.0036

[11] Jansen H, Samani NJ, Schunkert H. Mendelian randomization studies in coronary artery disease[J]. European heart journal vol. 35,29 (2014): 1917-24. doi:10.1093/eurheartj/ehu208.

[12] Emdin CA, Khera AV, Kathiresan S. Mendelian Randomization. JAMA vol. 318,19 (2017): 1925-1926. doi:10.1001/jama.2017.17219

[13] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis [J]. BMC bioinformatics vol. 9 559. 29 Dec. 2008, doi:10.1186/1471-2105-9-559.

[14] Chen B, Khodadoust MS, Liu CL, et al. Profiling Tumor Infiltrating Immune Cells with CIBERSORT[J]. Methods in molecular biology (Clifton, N.J.) vol. 1711 (2018): 243-259. doi:10.1007/978-1-4939-7493-1_12

[15] Emdin CA, Khera AV, Kathiresan S. Mendelian Randomization. JAMA vol. 318,19 (2017): 1925-1926. doi:10.1001/jama.2017.17219.

[16] Dudbridge, Frank. Polygenic Mendelian Randomization[J]. Cold Spring Harbor perspectives in medicine vol. 11,2 a039586. 1 Feb. 2021, doi:10.1101/cshperspect.a039586

[17] Franz M, Rodriguez H, Lopes C, et al. GeneMANIA update 2018[J]. Nucleic acids research vol. 46, W1 (2018): W60-W64. doi:10.1093/nar/gky311.

[18] Thoma OM, Neurath MF, Waldner MJ. et al. Cyclin-Dependent Kinase Inhibitors and Their Therapeutic Potential in Colorectal Cancer Treatment [J]. Frontiers in pharmacology vol. 12 757120. 21 Dec. 2021, doi:10.3389/fphar.2021.757120.

[19] Vu T, Datta A, Banister C et al. Serine-threonine Kinase Receptor-Associated Protein is a Critical Mediator of APC Mutation-Induced Intestinal Tumorigenesis Through a Feed-Forward Mechanism[J]. Gastroenterology vol. 162,1 (2022): 193-208. doi:10.1053/j.gastro.2021.09.010.

[20] Solier S, Zhang YW, Ballestrero A, et al. DNA damage response pathways and cell cycle checkpoints in colorectal cancer: current concepts and future perspectives for targeted treatment[J]. Current cancer drug targets vol. 12,4 (2012): 356-71. doi:10.2174/156800912800190901.

[21] Lu PH, Chen MB, Ji C, et al. Aqueous Oldenlandiadiffusa extracts inhibits colorectal cancer cells via activating AMP-activated protein kinase signalings [J]. Oncotarget vol. 7,29 (2016): 45889-45900. doi:10.18632/oncotarget.9969.

[22] Malki A, ElRuz RA, Gupta I, et al. Molecular Mechanisms of Colon Cancer Progression and Metastasis: Recent Insights and Advancements[J]. International journal of molecular sciences vol. 22,1 130. 24 Dec. 2020, doi:10.3390/ijms22010130.

[23] Ladabaum U, Dominitz JA, Kahi C. Strategies for Colorectal Cancer Screening[J]. Gastroenterology vol. 158,2 (2020): 418-432. doi:10.1053/j.gastro. 2019.06. 043.

[24] Siregar, Gontar Alamsyah, and Henry Sibarani. Comparison of Carcinoembryonic Antigen Levels Among Degree of Differentiation and Colorectal Cancer's Location in Medan[J]. Open access Macedonian journal of medical sciences vol. 7,20 3447-3450. 14 Oct. 2019, doi:10.3889/oamjms.2019. 442.

[25] McKeown E, Nelson DW, Johnson EK, et al. Current approaches and challenges for monitoring treatment response in colon and rectal cancer[J]. Journal of Cancer vol. 5,1 31-43. 1 Jan. 2014, doi:10.7150/jca.7987.

[26] Scherman P, Syk I, Holmberg E, et al. Influence of primary tumour and patient factors on survival in patients undergoing curative resection and treatment for liver metastases from colorectal cancer[J]. BJS open vol. 4,1 (2020): 118-132. doi:10.1002/bjs5.50237.

[27] Li X, Guo D, Chu L, et al. Potential diagnostic value of combining inflammatory cell ratios with carcinoembryonic antigen for colorectal cancer[J]. Cancer Management and Research. 2019; Volume 11: 9631–9640. doi: 10.2147/CMAR.S222756.

[28] He CZ, Zhang KH, Li Q, et al. Combined use of AFP, CEA, CA125 and CAl9-9 improves the sensitivity for the diagnosis of gastric cancer[J]. BMC Gastroenterology. 2013;13(1):p. 87. doi: 10.1186/1471-230X-13-87.

[29] Wu CX, Wang XQ, Chok SH,et al. Blocking CDK1/PDK1/β-Catenin signaling by CDK1 inhibitor RO3306 increased the efficacy of sorafenib treatment by targeting cancer stem cells in a preclinical model of hepatocellular carcinoma[J]. Theranosticsvol. 8,14 3737-3750. 13 Jun. 2018, doi:10.7150/thno.25487.

[30] Kvedaraite E, Ginhoux F. Human dendritic cells in cancer. Sci Immunol. 2022;7(70):eabm9409. doi:10.1126/sciimmunol.abm9409.

[31] Huang J, Chen P, Liu K, et al. CDK1/2/5 inhibition overcomes IFNG-mediated adaptive immune resistance in pancreatic cancer[J]. Gut. 2021;70(5):890-899. doi:10.1136/gutjnl-2019-320441

[32] Yuan J, Li X, Zhang G, et al. USP39 mediates p21-dependent proliferation and neoplasia of colon cancer cells by regulating the p53/p21/CDC2/cyclin B1 axis[J]. Mol Carcinog. 2021;60(4):265-278. doi:10.1002/mc.23290.

[33] Wu WJ, Hu KS, Wang DS, et al. CDC20 overexpression predicts a poor prognosis for patients with colorectal cancer[J]. Transl Med. 2013;11:142. Published 2013 Jun 10. doi:10.1186/1479-5876-11-142.

[34] Chigbrow M, Nelson M. Inhibition of mitotic cyclin B and cdc2 kinase activity by selenomethionine in synchronized colon cancer cells[J]. Anticancer Drugs. 2001;12(1):43-50. doi:10.1097/00001813-200101000-00006

[35] Liu KC, Shih TY, Kuo CL, et al. Sulforaphane Induces Cell Death Through G2/M Phase Arrest and Triggers Apoptosis in HCT 116 Human Colon Cancer Cells[J]. Am J Chin Med. 2016;44(6):1289-1310. doi:10.1142/S0192415X16500725.

[36] Zhang R, Xu J, Zhao J, et al.. Proliferation and invasion of colon cancer cells are suppressed by knockdown of TOP2A[J]. Cell Biochem. 2018;119(9):7256-7263. doi:10.1002/jcb.26916.

[37] Zhao H, Ming T, Tang S, et al. Wnt signaling in colorectal cancer: pathogenic role and therapeutic target[J]. Molecular cancer vol. 21,1 144. 14 Jul. 2022, doi:10.1186/s12943-022-01616-7.

[38] Rong Z, Luo Z, Fu Z,et al. The novel circSLC6A6/miR-1265/C2CD4A axis promotes colorectal cancer growth by suppressing p53 signaling pathway[J]. Journal of experimental & clinical cancer research: CR vol. 40,1 324. 16 Oct. 2021, doi:10.1186/s13046-021-02126-y.

[39] Jafari M, Laraqui A, Baba W, et al. Prevalence and patterns of mutations in RAS/RAF/MEK/ERK/MAPK signaling pathway in colorectal cancer in North Africa [J]. BMC cancer vol. 22,1 1142. 7 Nov. 2022, doi:10.1186/s12885-022-10235-w.

[40] Zhu H, Swami U, Preet R et al. Harnessing DNA Replication Stress for Novel Cancer Therapy[J]. Genes vol. 11,9 990. 25 Aug. 2020, doi:10.3390/genes11090990.

[41] Huang TT, Lampert EJ, Coots C et al. Targeting the PI3K pathway and DNA damage response as a therapeutic strategy in ovarian cancer [J]. Cancer treatment reviews vol. 86 (2020): 102021. doi:10.1016/j.ctrv.2020.102021.

[42] Feng ZY, Xu XH, Cen DZ, et al. miR-590-3p promotes colon cancer cell proliferation via Wnt/β-catenin signaling pathway by inhibiting WIF1 and DKK1[J]. European review for medical and pharmacological sciences vol. 21,21 (2017): 4844-4852.

[43] Wu K, Zhao Z, Xiao Y, et al. Roles of mitochondrial transcription factor A and microRNA-590-3p in the development of colon cancer[J]. Molecular medicine reports vol. 14,6 (2016): 5475-5480. doi:10.3892/mmr.2016.5955.

[44] Wu H, Zhong W, Zhang R, et al. G-quadruplex-enhanced circular single-stranded DNA (G4-CSSD) adsorption of miRNA to inhibit colon cancer progression[J]. Cancer medicine vol. 12,8 (2023): 9774-9787. doi:10.1002/cam4.5721.

[45] Uhr K, Prager-van der Smissen WJC, Heine AAJ,et al. MicroRNAs as possible indicators of drug sensitivity in breast cancer cell lines[J]. PloS one vol. 14,5 e0216400. 7 May. 2019, doi:10.1371/journal.pone.0216400.

[46] Gao T, Zou M, Shen T, et al. Dysfunction of miR-802 in tumors[J]. Journal of clinical laboratory analysis vol. 35,11 (2021): e23989. doi:10.1002/jcla.23989.

[47] Ge W, Goga A, He Y, et al. miR-802 Suppresses Acinar-to-Ductal Reprogramming During Early Pancreatitis and Pancreatic Carcinogenesis[J]. Gastroenterology vol. 162,1 (2022): 269-284. doi:10.1053/j.gastro.2021.09.029.

[48] Jing X, Xie M, Ding K, et al. Exosome-transmitted miR-769-5p confers cisplatin resistance and progression in gastric cancer by targeting CASP9 and promoting the ubiquitination degradation of p53[J]. Clinical and translational medicine vol. 12,5 (2022): e780. doi:10.1002/ctm2.780.

[49] Luan PB, Jia XZ, Yao J. MiR-769-5p functions as an oncogene by down-regulating RYBP expression in gastric cancer[J]. European review for medical and pharmacological sciences vol. 24,12 (2020): 6699-6706. doi:10.26355/eurrev_202006_21657.

[50] Lee, Daniel. miR-769-5p is associated with prostate cancer recurrence and modulates proliferation and apoptosis of cancer cells[J]. Experimental and therapeutic medicine vol. 21,4 (2021): 335. doi:10.3892/etm.2021.9766.