Novel six-gene prognostic signature based on colon adenocarcinoma immune-related genes

Rui Zhou
Southern Medical University
ZhuoWei Gao
Shunde Hospital of Southern Medical Universit
YongLe Ju (richardju66@163.com)
Shunde Hospital of Southern Medical University

Research Article

Keywords: Colon adenocarcinoma, immunotherapy, cancer prognosis, prognostic prediction model, prognostic risk model, immune-related genes

Posted Date: May 10th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1601528/v1

Abstract

Colon adenocarcinoma (COAD) is one of the most common gastrointestinal tumors worldwide, and immunotherapy is one of the most promising emerging cancer treatments. Immune genes (IGs) play a key role in tumor immunotherapy. This study aimed to determine the prognostic value of IGs in patients with COAD and establish an immune-related gene signature. Differentially expressed RNAs (DEGs), immune-related genes (DEIRGs), and transcription factors (DETFs) were screened using The Cancer Genome Atlas (TCGA), Immunology Database and Analysis Portal, InnateDB database, and Cistrome database and we constructed a regulatory network between DEIRGs and DETFs. Using Weighted Gene Co-Expression Network Analysis, we conducted five co-expressed gene modules; six hub genes were obtained using univariate and multivariate regression analysis (CD1A, CD1B, FGF9, GRP, SERPINE1, and F2RL2), leading to the construction of a risk model. Patients from TCGA database were divided into high- and low-risk groups; GSE40967 datasets were used for validation. Survival analysis, somatic gene mutations, and tumor-infiltrating immune cells differed significantly between the high- and low-risk groups. This immune-related gene signature could play an important role in guiding treatment, predicting prognosis, and potentially, future clinical applications.

Introduction

Colorectal cancer (CRC) is one of the most common malignant tumors in the digestive system, accounting for about 1.2 million new cases and 600,000 deaths every year [1]. More than 2.2 million new [1]

colorectal cancer cases and 1.1 million deaths are projected in 2030 [2]. Colon adenocarcinoma (COAD) is the most common histopathological type of CRC. It often evolves into invasive cancer due to gene mutation and continuous accumulation of colonic adenomatous lesions [3]. Current treatments for patients with CRC include surgery, radiation, chemotherapy, targeted therapy and immunotherapy. The 5-year survival rate exceeds 90% for patients referred to curative surgery. However, most patients are in the mid to late

stage of the disease at diagnosis, in which case the 5-year survival rate decreases to approximately 10% [4]. Immunotherapy is becoming an important treatment in cancer control [5]; thus, understanding the tumor immune microenvironment is significant in exploring tumor-associated immune signature biomarkers [6]. It has been reported that polygenic prediction models possess better predictive ability for cancer prognosis than the single gene model. Therefore, it is important to explore the tumor immune microenvironment and establish an accurate polygenic prediction model.

We used recently established cancer databases [7], such as the TCGA database (https://portal.gdc.cancer.gov/) and IMMPORT (https://immport.niaid.nih.gov/home)Cistrome(http://www.cistrome.org/) to construct a prognostic

prediction model based on immune-related genes. This prognostic model can predict patient prognosis and provide promising therapeutic molecular targets for colorectal cancer.

Materials And Methods

Obtaining relevant data from network database

The gene expression clinical and somatic mutation data of COAD patients were obtained through the TCGA data portal (https://portal.gdc.cancer.gov/). Gene expression data and clinical information were downloaded from Gene Expression Omnibus (GEO) databases (https://www.ncbi.nlm.nih.gov/geo/). The immune signatures were obtained from the IMMPort (https://immport.niaid.nih.gov) and InnateDB databases (https://www.innatedb.ca/). The list of transcription factors (TFs) was obtained from the Cistrome Project (http://www.cistrome.org/). Furthermore, the Tumor Immune Estimation Resource (TIMER) web tool (http://timer.cistrome.org/) was used to obtain the genes in the tumor microenvironment. The c5.go.v7.4.symbols and c2.cp.kegg.v7.4.symbols datasets were downloaded from the MsigDB database (http://www.gsea-msigdb.org/gsea/index.jsp) on the GSEA website.

Obtaining differential genes

Data extraction and integration were conducted using Perl (v5.32.1). DEIGs, DEGs, and DETFs were analyzed using R version 4.1.1 and the relevant Bioconductor packages (e.g., limma v3.48.3, edgeR v3.34.1) [8] according to the screening criteria of |log2FC| >1 and false discovery rate (FDR) < 0.05, Plots were generated using the R package ggplot2 (v3.3.5), and heatmaps were drawn using the R package pheatmap (v1.0.12).

DEIRGs enriched and analyzed using GO and KEGG

GO and KEGG term enrichment analysis of screened genes was performed using the R package clusterProfiler. Finally, an FDR < 0.05 was set as the cut-off criteria to identify the outstanding GO terms and KEGG pathways visualized using bubble and circle diagrams.

Correlation analysis between DEIRGs and DETFs

Correlation analysis between DEIGs and DETFs was performed using the cor.test in R software (cor > 0.5, P < 0.001). Protein-protein interaction networks were generated using Cytoscape (version 3.8.2) [9].

Weighted Gene Co-expression Network Analysis of DEIRGs

Weighted Gene Co-expression Network Analysis (WGCNA) was performed by applying the R package “WGCNA” to DEIRGs to obtain a different module. The optimal module was used to generate the module network plots using the R “igraph” package. Subsequently, we determined the intersection of the module genes obtained from the GEO and TCGA databases.

Immune-related immune genes obtained from intersection genes

Clinical information and gene expression for the univariate analysis were analyzed using Cox regression, and the corresponding DEGs were screened out as prognostic immune-related genes for further study at $p < 0.05$ and |hazard ratio (HR)|≠1.

Prognostic model construction

Based on the expression of screened genes from the previous step, a risk model was built using multivariate Cox regression model analysis, calculated as follows: Risk score=Expgene1×coefgene1+Expgene2×coefgene2+…+Expgenen×coefgenen. Exp represents the expression level of the gene, and coef is the estimated regression coefficient of the gene derived from the multivariate Cox analysis.

Evaluation and analysis of risk model

The RNA sequencing data of COAD patients were obtained from the TCGA database and set as the training set, whereas the external validation cohorts were obtained from the GEO dataset (GSE40967). The patients were then separated into high- and low-risk groups using the mean risk score in the training set as the cut-off value. Using the files downloaded from the MsigDB database, GO, and KEGG enrichment analyses were performed for high-risk and low-risk groups using R software. K-M analysis, univariate and multivariate independent prognosis analyses were performed for the two groups using the survminer and survival R and packages, and the forest map was drawn. The ROC curve and time-dependent ROC-based areas under the curve (AUC) were plotted using the R package “timeROC,” We found a prediction model similar to ours from the literature, and compared the predictive accuracy of the prognostic models using a concordance index (C-index) and plotting ROC curves to compare the AUC.

Analyzing somatic mutation in high and low-risk groups

The tumor mutation burden (TMB) was calculated as mutations per megabase (mut/Mb). The tumor mutation of patients in the high and low-risk groups was analyzed using TMB data. The R package “maftools” was used to analyze and visualize the somatic mutation data.

Correlation analysis of immune cell infiltration

The proportion of infiltrating immune cells was calculated using the CIBERSORT algorithm [10], and the results were considered significant at $p < 0.05$. The Wilcoxon signed-rank test was used to analyze the differential abundance of infiltrating immune cells and immune cell function between the low- and high-risk groups, and box plots were created using the ggpubr package.

Validation of reliability of risk models in other databases

The TIMER (https://cistrome.shinyapps.io/timer/)[11]database was used to verify the difference in gene expression between tumor and normal samples. Immunohistochemistry was used to compare protein

expression between normal and tumor tissue using the Human Protein Atlas (HPA) (https://www..roteinatlas.rg/) [12] database.

Results

Data processing and identification of DEIGs, DEGs, and DETFs

The transcriptome RNA-sequencing data and clinical materials of 514 COAD patients were obtained from the TCGA database. The screening criterion for DEGs between tumors and normal samples was set up as the FDR < 0.05 and |log FC (fold change)| >1 (Figure 1A, B). A list of 2,660 immune genes was obtained from the Immunome database, downloaded from InnateDB and IMMPORT database, and a total of 649 DEIGs were obtained from the further screening (|log FC (fold change)| >1, FDR < 0.05). (Figure 1C, D). We obtained 318 TFs from the Cistrome program and 67 DETFs after screening (Figure 1E, F)

Functional analysis of DEIRGs

GO, KEGG enrichment analyses were conducted on the DEIGs. GO-BP analysis indicated that the DEIGs were mainly involved in B cell, immunoglobulin mediated immune response, complement activation, and regulation of immune system processes (Figure 2A,B). The KEGG pathway was mainly enriched in cytokine–cytokine receptor interaction, viral protein interaction with cytokine and cytokine receptor, chemokine signaling pathway, IL-17 signaling pathway, neuroactive ligand–receptor interaction, and the NF–kappa B signaling pathway (Figure 2 C, D).

GO

Building DEIG and DETF interaction network

The correlation between DEIGs and DETFs was obtained using the screening criteria cor > 0.5, P < 0.001, and the correlation network diagram was drawn using Cytoscape (Figure 3). The specific correlation results are shown in Table 1.

Weighted Gene Co-Expression Network analysis of DEIGs

WGCNA[13, 14]can identify gene modules with similar expression patterns by calculating the expression relationship between genes, analyzing the relationship between gene modules and phenotypes, and mapping the regulatory network between genes in the gene module and central genes (hubs). The WGCNA package in R was used to divide DEIGs into 5 modules (“MEgreen,” “MEblue,” “MEblue,” “MEyellow,” and “MEgrey”) (Figure 4). The optimal POWER value was 4. Prognostic models were built based on the minimal p-value. Therefore, genes in the green module were selected for subsequent analysis.

Obtaining immune genes related to the prognosis of COAD patients

After obtaining the “MEgreen” gene module, univariate regression analysis was performed on the clinical data of COAD patients in TCGA and GEO databases. Nine genes were screened as prognosis-related genes (CD1A, CD1B, FGF9, GRP, OXTR□SPHK1, BGN, SERPINE1, and F2RL2) using $p < 0.05$ as the criteria. The results of univariate analysis are shown in Figure 5A. The K-M curves of 9 genes in COAD patients are plotted using “survival” packages in R software (Figure 5B-J)

Building of survival prognosis model and performing a survival analysis

The 9 prognosis-related genes were included in multivariate Cox regression analysis. The inclusion criterion was $p < 0.05$, HR≠1. Six genes (*CD1A*, *CD1B*, *FGF9*, *GRP*, *SERPINE1*, and *F2RL2*) were incorporated into the risk core model. The relationship between genes and risk score is shown in Table 2. The composite risk scores of patients enrolled in the TCGA datasets were calculated. The COAD patients were separated into low- and high-risk groups using the mean risk score as the cut-off. The prognostic model was built using TCGA data and verified in the GEO datasets GSE40967. Further, we used the R software “Survival” and “timeROC” packages to draw two groups of K-M and ROC curves, respectively (Figure 6A-D), The “ComplexHeatmap” of R software was used to conduct a chi-square test between the low- and high-risk groups in the clinical traits, and draw the clinical correlation heat map ($p < 0.05$, Figure 6E). There were significant differences between the two groups in the T, N, M, and tumor stages. Univariate and multivariate Cox regression analyses revealed that age, T stage, and risk score were independent prognostic factors for COAD patients (Figure 7 A-B).

ssGSEA analysis for low-and high-risk groups

GO and KEGG enrichment analysis files (“c5.go.v7.4.smbols,” and “c2.cp.kegg.v7.4.symbols”) were downloaded from the GSEA database (http://www.gsea- msigdb.org/gsea/index.jsp).Subsequently, GO, and KEGG enrichment analyses were performed using the clusterProfiler and the org.Hs.eg.db packages in high- and low-risk groups. GO enrichment analysis of the high-risk group’s main enrichment was at the keratinocyte differentiation, skin development, collagen-containing extracellular matrix external encapsulating structure, and structural molecule activity. The most significantly enriched KEGG pathways were axon guidance, ECM receptor_interaction, focal adhesion, PPAR signaling pathway, and systemic lupus erythematosus. GO enrichment analysis of the low-risk group’s main enrichment was of activation of immune response, adaptive immune response, immune response regulating signaling pathway, immunoglobulin production, immunoglobulin complex. The KEGG pathway was enriched in allograft rejection, asthma, autoimmune thyroid disease, intestinal immune network for IgA production, and primary immunodeficiency (Figure 7C-F).

Comparison somatic mutation in high- and low-risk groups

Somatic mutation profiles of COAD patients downloaded from the TCGA database were analyzed and visualized using the “maftools” package [15]. A total of 388 patients exhibited mutation, of which 202 were high-risk, and a low-risk of 184 patients. The top 5 genes with the highest somatic mutation rate in

the high-risk group were *APC*, *TP53*, *TTN*, *KRAS*, and *SYNE1*. Missense mutations were the most common. The high-risk group achieved a higher mutation frequency than the low-risk group (Figure 8A-B).

Analyzing the immune infiltration of high- and low-risk groups

Next, we used CIBERSORT to estimate the proportions of 22 distinct immune cell types ($p < 0.05$) (Figure 8 E), The Wilcoxon signed-rank test was used to determine the immune infiltrating cells with differences between the high-risk groups. Additionally, the resting dendritic cells, follicular helper T cells (Tfh), were present in significantly higher fractions in low-risk patients than in high-risk patients ($p < 0.05$; Figure 8C). The immune cell function of the high-risk group was lower than that of the low-risk group in APC co-inhibition, APC co-expression, T cell function, and macrophage function (Figure 8D).

Comparison with the other models

We compared the four corresponding prognostic models: a 7-gene signature (Sun), 6-gene signature (Liang), 12-gene signature (Mia), and 7-gene signature (Chen) [16-19]. Similarly, taking the mean risk value of all samples as the division standard, the high- and low-risk groups were divided into high and low groups, and four models were used to calculate the risk value of the TCGA samples. The ROC and K-M curves for the four models are attached as Figure 9A-J. The AUCs of the four models at 5 years were 0.581, 0.521, 0.616, and 0.555, respectively, all significantly lower than our model (0.639). C-indexes were used to evaluate the prediction capability of the mixed-effect Cox model [20]. We then calculated the C-indexes of all models. The results showed that our model exhibited the highest C-index value (0.704; Figure 9K).

Verification of six IRGs in external databases

The TIMER online database was used to analyze the differential expression of six genes in this model in 17 types of tumors and adjacent tissues. It was seen that CD1A, CD1B, GRP, SERPINE1, and F2RL2 were highly expressed in tumor tissue. In contrast, FGF9 was highly expressed in normal colon tissue (Figure 11). An HPA database search was performed to verify the protein expression levels of CD1A, CD1B, SERPINE1, and FGF9 (Figure 10).

Discussion

Here, we downloaded the information on 514 COAD patients from the TCGA online database to identify DEGs using the limma R package (LogFC > 1, FDR< 0.05). Additionally, we identified the DEGs, DEIRGs, and DETFs from the IMMPORT, InnateDBt, and Cistrome databases (LogFC > 1, FDR< 0.05). Next, we structured the regulatory network between TFs and immune genes depicted by the Cytoscape software. The DETFs might facilitate DEIRGs (Fig. 3), and the function of our regulatory network was linked to immune regulation and complement activation. However, further studies are needed to elucidate this mechanism.

The differential immune genes were analyzed using WGCNA, and the module genes with the lowest p-value were obtained for further analysis. We presented univariate and multivariate analysis for module genes and identified 9 immune genes that were correlated with the prognosis of COAD patients (CD1A, CD1B, FGF9GRP, OXTR, SPHK1, BGN, SERPINE1, and F2RL2), The K-M curve of 9 immune genes are shown in Figure 5. Subsequently, we performed multivariate regression analysis ($p < 0.05$,$HR \neq 1$). Finally, CD1A, CD1B, FGF9, GRP, SERPINE1, and F2RL2CD1B were included in the risk model. CD1B is a lipid antigen-presenting molecule that can be inducibly expressed on monocytes and dendritic cells. As a marker, CD1B can be recognized by γδ T cells and plays an important role as effectors of tissue injury, infection, and cancer development [21]; it also regulates the differentiation and maturation of dendritic cells [22]. FGF9 plays a critical role in patients with colon cancers that developed resistance to EGFR-targeted therapy, and combination therapy with anti-EGFR inhibitor may reverse drug resistance [23]. GRP could serve as an independent prognostic factor for survival in patients with colon cancer [24]. SERPINE1 plays an essential role in remodeling the tumor microenvironment and tumor-infiltrating immune cells [25], and some noncoding RNAs can affect epithelial-mesenchymal transition (EMT) of colon cancer by regulating SERPINE1 [26-28].

A prognostic risk model was constructed for the six candidate genes (Table 2). We calculated the risk score for each COAD patient downloaded from the TCGA and GEO databases. The patients were assigned to low- and high-risk groups based on the mean risk Score using survival analyses of the high- and low-risk groups from the TCGA or GEO database; we found that the low-risk patient group patients had a better prognosis than the high-risk group patients ($p < 0.013$ and $p < 0.001$, respectively) (Figure 6C-D). Univariate and multivariate regression analysis showed that age, T stage, and risk score were independent prognostic factors for patients with COAD. In the analysis of clinical characteristics, there were significant differences in tumor T, N, and M stages between the two groups ($p < 0.001$) (Figure 6E). Moreover, GSEA analysis was performed for the high- and low-risk groups. The up-regulated molecular pathways in the high-risk group included axon guidance, ECM receptor interaction, focal adhesion, PPAR signaling pathways, and systemic lupus erythematosus (Figure 7E). Davide Pradella D et al. considered that UNC5B is an axon guidance regulator that controls developmental and tumor angiogenesis, and its expression was associated with tumor angiogenesis and poor prognosis [29]. Mitra et al. demonstrated that cell migration involves integrating ECM with the actin cytoskeleton through transmembrane receptors [30], and focal adhesion kinase (FAK) can be activated by the ECM that promotes focal adhesion maturation and formation of focal adhesion complexes [30-32]. Thus, the high-risk group may activate the FAK signaling pathway, leading to invasion in vitro and metastasis in vivo in colorectal cancer. In colon cancer, the absence of PPARα and PPARβ/δ expression could promote cancer growth and PPARγ could suppress tumorigenesis through the regulation of and interaction with β-catenin [33-35]. Systemic lupus erythematosus is associated with abnormal autoimmune reaction; however, the mechanisms have not been identified. Furthermore, colon cancer is one of the most common complications in gastrointestinal diseases [36]. Down-regulated molecular pathways in the low-risk group included allograft rejection, asthma, autoimmune thyroid disease, intestinal immune network for IgA production, and primary immunodeficiency (Figure 7 F). Asthma can be alleviated by reducing eosinophil

production; previous studies have proposed that the depletion of eosinophils due to asthma severely compromises antitumor immunity in syngeneic and genetic models of colorectal cancer. This association is possibly due to defective Th1 and CD8+ T cell responses [37]. Bacteria are a link to cancer [38], and the tumor microenvironment could be more favorable for IgA production by modulating the immune microenvironment and intestinal bacteria. Consequently, IgAs are widely used studies as biomarkers for early screening [39-41]. Recent reports have shown that patients with primary immunodeficiency tend to have a higher cancer incidence because of genomic instability due to defective DNA repair mechanisms [42, 43]. However, the association between autoimmune thyroid disease, allograft rejection, and colorectal cancer has not been found.

Our research also found that the high-risk group had a higher somatic gene mutation rate than the low-risk group (Figure 8A-B). Several relevant prediction models have been developed by various researchers. Sun et al. created a 7-gene signature based on the TCGA cohort, Liang et al. created a 6-gene signature based on the TCGA colon cancer cohort to guide COAD recurrence risk judgment, Miao et al. constructed a prognostic score model of a 12-gene signature, and Chen et al. established a 7-gene signature-associated risk model with good prognostic prediction performance. The ROC curve studies concluded that our risk model could better predict patient prognosis (Figure 9 A-J), and C-index analysis showed that our risk models outperformed these four models (Figure 9K). Therefore, the models developed here could guide clinical management and enhance precision prognostic medicine. However, the results of this study are mainly based on bioinformatics analysis and need to be further validated.

Conclusion

In summary, we created a 6-gene prognostic prediction model with good predictive capability in both the training and validation sets. Therefore, this model could help clinicians predict individual risks of COAD patients and in the development of personalized COAD treatment.

Declarations

Acknowledgments

The authors thank the staff members of TCGA, GEO, TIMER, IMMPort, InnateDB, Cistrome, MsigDB and

HPA databases.

Authors' contributions

Zhou Rui is responsible for collecting data and conducting preliminary analysis. Gao Zhuowei and Ju Yongle are responsible for guiding the use of statistical methods and the revision of the discussion part

Funding

The design, analysis, interpretation of data, and the writing of the manuscript were supported in part by the National Natural Science Foundation of China(81703914).

Availability of data and materials

The entire sequencing profle data and the clinical data of COAD patients in this study come from The Cancer Genome Atlas (TCGA, https://cancergenome.nih.gov/) database and Gene Expression Omnibus (GEO) databases (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE40967). The transcription factors (TFs) was obtained from the Cistrome Project (http://cistrome.org/CistromeCancer/CancerTarget/).The genes in the tumor microenvironment were obtained from Tumor Immune Estimation Resource (TIMER) web tool (http://timer.cistrome.org/), and the immune-related genes were obtained from IMMPort (https://www.immport.org/shared/genelists) and InnateDB databases (https://www.innatedb.ca/annotatedGenes.do?type=innatedb).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of of gastrointestinal surgery, Shunde Hospital of Southern Medical University,GuangDong, FoShan, China.[2]Medical department of tcm, Shunde Hospital of Guangzhou University of traditional Chinese Medicine, FoShan, China.*Department of of gastrointestinal surgery, Shunde Hospital of Southern Medical University,GuangDong, FoShan, China

References

1. Dekker, E., et al., *Colorectal cancer.* Lancet (London, England), 2019. **394**(10207): p. 1467-1480.
2. Arnold, M., et al., *Global patterns and trends in colorectal cancer incidence and mortality.* Gut, 2017. **66**(4): p. 683-691.
3. Mutch, M.G., *Molecular profiling and risk stratification of adenocarcinoma of the colon.* Journal of surgical oncology, 2007. **96**(8): p. 693-703.
4. Punt, C.J.A., M. Koopman, and L. Vermeulen, *From tumour heterogeneity to advances in precision treatment of colorectal cancer.* Nature reviews. Clinical oncology, 2017. **14**(4): p. 235-246.

5. Riley, R.S., et al., *Delivery technologies for cancer immunotherapy.* Nature reviews. Drug discovery, 2019. **18**(3): p. 175-196.
6. Koi, M. and J.M. Carethers, *The colorectal cancer immune microenvironment and approach to immunotherapies.* Future oncology (London, England), 2017. **13**(18): p. 1633-1647.
7. Chen, C., H. Huang, and C.H. Wu, *Protein bioinformatics databases and resources.* Methods in molecular biology (Clifton, N.J.), 2011. **694**.
8. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic acids research, 2015. **43**(7): p. e47.
9. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome research, 2003. **13**(11): p. 2498-2504.
10. Newman, A.M., et al., *Robust enumeration of cell subsets from tissue expression profiles.* Nature methods, 2015. **12**(5): p. 453-457.
11. Li, T., et al., *TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells.* Cancer research, 2017. **77**(21): p. e108-e110.
12. Pontén, F., et al., *The Human Protein Atlas as a proteomic resource for biomarker discovery.* Journal of internal medicine, 2011. **270**(5): p. 428-446.
13. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis.* BMC bioinformatics, 2008. **9**: p. 559.
14. Langfelder, P. and S. Horvath, *Fast R Functions for Robust Correlations and Hierarchical Clustering.* Journal of statistical software, 2012. **46**(11).
15. Mayakonda, A., et al., *Maftools: efficient and comprehensive analysis of somatic variants in cancer.* Genome research, 2018. **28**(11): p. 1747-1756.
16. Sun, Y.-L., et al., *A Prognostic Model Based on the Immune-related Genes in Colon Adenocarcinoma.* International journal of medical sciences, 2020. **17**(13): p. 1879-1896.
17. Liang, Y., Q. Su, and X. Wu, *Identification and Validation of a Novel Six-Gene Prognostic Signature of Stem Cell Characteristic in Colon Cancer.* Frontiers in oncology, 2020. **10**: p. 571655.
18. Miao, Y., et al., *Identification prognosis-associated immune genes in colon adenocarcinoma.* Bioscience reports, 2020. **40**(11).
19. Chen, S., et al., *Prediction and identification of immune genes related to the prognosis of patients with colon adenocarcinoma and its mechanisms.* World journal of surgical oncology, 2020. **18**(1): p. 146.
20. Harrell, F.E., et al., *Evaluating the yield of medical tests.* JAMA, 1982. **247**(18): p. 2543-2546.
21. Reijneveld, J.F., et al., *Human γδ T cells recognize CD1b by two distinct mechanisms.* Proceedings of the National Academy of Sciences of the United States of America, 2020. **117**(37): p. 22944-22952.
22. Chen, C.-A., et al., *Daphnoretin modulates differentiation and maturation of human dendritic cells through down-regulation of c-Jun N-terminal kinase.* International immunopharmacology, 2017. **51**: p. 25-30.

23. Mizukami, T., et al., *Significance of FGF9 gene in resistance to anti-EGFR therapies targeting colorectal cancer: A subset of colorectal cancer patients with FGF9 upregulation may be resistant to anti-EGFR therapies.* Molecular carcinogenesis, 2017. **56**(1): p. 106-117.
24. Li, C., et al., *Independent prognostic genes and mechanism investigation for colon cancer.* Biological research, 2018. **51**(1): p. 10.
25. Wang, S., et al., *SERPINE1 associated with remodeling of the tumor microenvironment in colon cancer progression: a novel therapeutic target.* BMC cancer, 2021. **21**(1): p. 767.
26. Yang, J., et al., *LncRNA CYTOR drives L-OHP resistance and facilitates the epithelial-mesenchymal transition of colon carcinoma cells via modulating miR-378a-5p/SERPINE1.* Cell cycle (Georgetown, Tex.), 2021. **20**(14): p. 1415-1430.
27. Wan, J., et al., *LINC00491 as a new molecular marker can promote the proliferation, migration and invasion of colon adenocarcinoma cells.* OncoTargets and therapy, 2019. **12**: p. 6471-6480.
28. Hu, B., et al., *MicroRNA-148a-3p Directly Targets SERPINE1 to Suppress EMT-Mediated Colon Adenocarcinoma Progression.* Cancer management and research, 2021. **13**: p. 6349-6362.
29. Pradella, D., et al., *A ligand-insensitive UNC5B splicing isoform regulates angiogenesis by promoting apoptosis.* Nature communications, 2021. **12**(1): p. 4872.
30. Mitra, S.K., D.A. Hanson, and D.D. Schlaepfer, *Focal adhesion kinase: in command and control of cell motility.* Nature reviews. Molecular cell biology, 2005. **6**(1): p. 56-68.
31. Webb, D.J., et al., *FAK-Src signalling through paxillin, ERK and MLCK regulates adhesion disassembly.* Nature cell biology, 2004. **6**(2): p. 154-161.
32. Bernabé-García, Á., et al., *Amniotic membrane promotes focal adhesion remodeling to stimulate cell migration.* Scientific reports, 2017. **7**(1): p. 15262.
33. Luo, Y., et al., *Intestinal PPARα Protects Against Colon Carcinogenesis via Regulation of Methyltransferases DNMT1 and PRMT6.* Gastroenterology, 2019. **157**(3).
34. Harman, F.S., et al., *Peroxisome proliferator-activated receptor-delta attenuates colon carcinogenesis.* Nature medicine, 2004. **10**(5): p. 481-483.
35. Andersen, C.B., et al., *Falcarindiol Purified From Carrots Leads to Elevated Levels of Lipid Droplets and Upregulation of Peroxisome Proliferator-Activated Receptor-γ Gene Expression in Cellular Models.* Frontiers in pharmacology, 2020. **11**: p. 565524.
36. Shi, J., et al., *Treatment of colon cancer in a patient with systemic lupus erythematosus: a case report.* BMC cancer, 2018. **18**(1): p. 961.
37. Arnold, I.C., et al., *The GM-CSF-IRF5 signaling axis in eosinophils promotes antitumor immunity through activation of type 1 T cell responses.* The Journal of experimental medicine, 2020. **217**(12).
38. Shirazi, M.S.R., et al., *Microbiome Dysbiosis and Predominant Bacterial Species as Human Cancer Biomarkers.* Journal of gastrointestinal cancer, 2020. **51**(3): p. 725-728.
39. Chen, M., et al., *Development of a panel of serum IgG and IgA autoantibodies for early diagnosis of colon cancer.* International journal of medical sciences, 2020. **17**(17): p. 2744-2750.

40. Bridges, K.M., et al., *Relating Stool Microbial Metabolite Levels, Inflammatory Markers and Dietary Behaviors to Screening Colonoscopy Findings in a Racially/Ethnically Diverse Patient Population.* Genes, 2018. **9**(3).
41. Kurt, M. and Z. Yumuk, *Diagnostic accuracy of Fusobacterium nucleatum IgA and IgG ELISA test in colorectal cancer.* Scientific reports, 2021. **11**(1): p. 1608.
42. Mayor, P.C., et al., *Cancer in primary immunodeficiency diseases: Cancer incidence in the United States Immune Deficiency Network Registry.* The Journal of allergy and clinical immunology, 2018. **141**(3): p. 1028-1035.
43. Xiao, Y., et al., *A Novel Four-Gene Signature Associated With Immune Checkpoint for Predicting Prognosis in Lower-Grade Glioma.* Frontiers in oncology, 2020. **10**: p. 605737.

Tables

Tables 1 and 2 are available in the Supplementary Files section

Figures

Figure 1

A) Heatmap of DEGs. B) Volcano plot of DEGs. C) Heatmap of DEIRGs. D) Volcano plot of DEIRGs. E) Heatmap of DETFs. F) Volcano plot of DETFs.

Figure 2

A) Bubble diagram of the GO enrichment analysis of DEIGs. B) Circle diagram of the GO enrichment analysis of DEIGs. C) Bubble diagram of the KEGG enrichment analysis of DEIGs. D) Circle diagram of the KEGG enrichment analysis of DEIGs.

Figure 3

Regulatory networks between DEIGs and DETFs

Green circle represents DEIGs, red triangles represent DETFs, and red lines represent positive regulation.

Figure 4

A) WGCNA identified 5 gene co-expression modules. B) Hierarchical clustering dendrogram of DEIGs. Different colors represent highly connected genes modules containing a cluster of functionally related genes.

Figure 5

A) Forest map of univariate regression analysis of prognosis-related immune genes. B-J) The K-M curve of prognosis-related immune genes.

Figure 6

A-B) K-M and ROC curves of high- and low-risk groups in GEO database. C-D) K-M and ROC curves of high- and low-risk groups in TCGA database. E) Clinical correlation heat map of high- and low-risk groups.

Figure 7

A-B) Univariate and multivariate regression analysis of high-risk and low-risk groups. C-F) GSEA analysis of high-risk and low-risk groups.

Figure 8

A-B) Somatic mutation in high- and low-risk groups. C) Differences in immune infiltrate cell distribution between high- and low-risk groups. D) Differences in immune cell function between high- and low-risk groups. E) Proportion of immune infiltrating cells in high- and low-risk groups.

Figure 9

A-B) 1, 3, and 5-year time-dependent ROC and K-M Chen signature curves. C-D) 1, 3, and 5-year time-dependent ROC and K-M Liang signature curves. E-F) 1, 3, and 5-year time-dependent ROC and K-M Miao signature curves. G-H) 1, 3, and 5-year time-dependent ROC and K-M Sun signature curves. K) Comparison of c-indices of five models.

Figure 10

Differential protein expression of six prognostic genes in colon cancer and normal tissue using HPA database (GRP and F2RL2 were not included). A-B) CD1A expression in colon cancer and normal tissue. C-D) CD1B expression in colon cancer and normal tissue. E-F) FGF9 expression in colon cancer and normal tissue. G-H) SERPINE1 expression in colon cancer and normal tissue.

Figure 11

Gene expression profiles of six genes in this model in the TIMER database

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1RegulatoryrelationshipbetweenDEIGsandDETFs.xls
- Table2DEIGsmultivariateCoxregressionanalysisandregressioncoefficient.xls
- NationalNaturalScienceFoundationofChina81703914.pdf
- NaturalScienceFoundationofGuangdongProvince2019A1515011275.pdf
- TCGAclinicaldata.xls
- TCGAdata.xlsx