Two immune-enhanced molecular subtypes differ in inflammation, immune checkpoints, mutations and outcome in Stage I-II colonic carcinoma.

Dongxin Tang

First affiliated Hospital of Guiyang College of Traditional Chinese Medicine

Wei Huang

Chengdu University of Traditional Chinese Medicine

Zhu Yang

First Affiliated Hospital of Guiyang College of Traditional Chinese Medicine

Xin Wu

Zhejiang Chinese Medical University

Xianan Sang

Zhejiang Chinese Medical University

Kuilong Wang

Zhejiang Chinese Medical University

Gang Cao

Zhejiang Chinese Medical University

Min Hao (✉ keyanprof@163.com)

Zhejiang Chinese Medical University <https://orcid.org/0000-0003-1007-2254>

Primary research

Keywords: Colonic carcinoma, TCGA, molecular subtype, immune microenvironments, immunotherapy

Posted Date: February 12th, 2020

DOI: <https://doi.org/10.21203/rs.2.23306/v1>

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

[Read Full License](#)

Abstract

Background: The purpose of this paper is to investigate the immune function of the tumor microenvironment and its clinical correlation with colonic carcinoma. **Methods :** Immune genes were downloaded from the The Cancer Genome Atlas (TCGA) database. Five subtypes are obtained by cluster screening based on immune gene expression data. **Results:** The C3 and C4 subtypes show stronger immune activity. In addition, the C4 subtype has the largest number of gene mutations and the worst prognosis. Most of the immune signatures are upregulated in the C4 subtype, while most of the immune infiltration-related cells are upregulated in the C3 and C4 subtypes. **Conclusions:** The different immune microenvironments between these subtypes may provide new ideas for immunotherapy strategies in colon carcinoma.

Background

Colonic carcinoma is one of the most common malignant tumors in the world. The incidence and mortality of colonic carcinoma rank third among all malignant tumors, and the incidence rate is on the rise (Xu et al., 2017). Most of the cases are found in the middle and late stages. Approximately 1/4 of colonic carcinoma patients have metastasis at the time of diagnosis, and the 5-year survival rate is less than 10%. The vast majority of patients with colonic carcinoma suffer from recurrence within two years after surgery [2]. Although there are many studies on colonic carcinoma, its pathogenesis is not yet completely understood [3]. The main treatment of colonic carcinoma is surgery, although treatment also includes chemotherapy, radiotherapy, and biotherapy. Nevertheless, the effects of these treatments are often unsatisfactory [4]. Hematogenous metastasis and lymphatic metastasis often occur in the late stage of colonic carcinoma. Postoperative recurrence and metastasis are the common causes of death [5]. At present, there is still a lack of effective new treatments. With the innovation and improvement in basic science, immunotherapy is regarded as the most likely way to cure tumors [6]. Tumor immunotherapy is a therapeutic method to control and remove tumors by restarting and maintaining tumor-immune circulation and restoring the normal anti-tumor immune response [7]. Compared with conventional tumor therapy, tumor immunotherapy mainly targets the immune system or tumor microenvironment rather than the tumor cells themselves. Tumor immunotherapy can also induce the synergistic anti-tumor effect in combined therapy [8].

Normally, after a tumor develops, the anti-tumor effect is initially performed by the innate immune system. Effector cells include NK cells, neutrophils, macrophages and so on. Subsequently, the adaptive immune response is initiated, including humoral immunity and cellular immunity, which produces antibodies and effector T cells to play an anti-tumor role [9]. As the tumor grows, tumor cells successfully evade the attack of the host immune system through immunosuppressive mechanisms. Different tumors can inhibit the effective identification and killing of tumor cells by the immune system through different mechanisms to produce immune tolerance and even promote the occurrence and development of tumors [10]. At present, a variety of immunotherapeutic drugs have been developed for this mechanism, including monoclonal antibody immune checkpoint inhibitors, therapeutic antibodies, cancer vaccines, cell

therapies and small molecular inhibitors [11–13]. In recent years, there has been substantial good news about tumor immunotherapy. At present, tumor immunotherapy has shown strong anti-tumor activity in the treatment of a variety of solid tumors, such as melanoma, non-small cell lung cancer, kidney cancer and prostate cancer. Several tumor immunotherapeutic drugs have been approved for clinical use by the Food and Drug Administration (FDA) in the United States [14–15].

However, the complexity of the tumor immune microenvironment not only increases the difficulty of tumor immunotherapy but also affects the efficiency of immunotherapy [16]. How to optimize these immunotherapy methods, screen the dominant population, or carry out immunotherapy for a specific group of people is a current urgent question to be answered by precision medicine [17]. There are many evidence-based medical studies proving that a comprehensive understanding of various immunomodulatory molecules in the tumor-related microenvironment, according to the effective immune molecule classification to guide treatment, is an important task for precision immunotherapy. At present, the expression patterns of immune checkpoint molecules in colonic carcinoma and the potential clinical correlations are unclear. Therefore, there is an urgent need for in-depth studies of the overall immune status and immune gene expression information of patients, according to clinical information and immune characteristics of patient samples, for more accurate classification to guide the clinical application of immunotherapy.

In summary, the purpose of this study is to explore the overall immune characteristics and clinical correlation of patients with colonic carcinoma. We have identified five molecular subtypes of colonic carcinoma based on the expression data of immune genes in the TCGA database. Among these subtypes, the C3 and C4 subtypes show stronger immune activity. Nevertheless, these subtypes are significantly different in clinical features, immune scores, gene mutations, prognosis and immune checkpoints, and metabolic pathways involved in regulation. These experimental results have been further verified in the GSE39582 dataset of the GEO database.

Methods

Data download

We download clinical follow-up information from the RNA-Seq dataset of colonic carcinoma from the GDC. From previous studies, we collect 13 types of immune metagenes, six types of immune cells corresponding to each sample of colonic carcinoma downloaded from Timer (<https://cistrome.shinyapps.io/timer/>), and immune genes downloaded from the ImmPort database (<https://immpart.niaid.nih.gov>). We use the R package estimate to calculate the immune score and matrix score for each sample.

Data preprocessing

First, we convert FPKM to TPM by R script. Next, we extract the immune gene set using the expression level. The samples with an expression level greater than 0 are screened out. More than 30% of the genes

are included in this study as immune genes. We match the expression profile samples and clinical follow-up samples, and we select both samples. We analyze all the Stage combinations, finding that the Stage I + II results are the best, which are related to the prognosis. Therefore, we further select the samples of Stage I and Stage II for this study. We then extract the immune gene set with the expression level from the expression profile. The samples with an expression level greater than 0 are screened out, and more than 30% of the genes are included in this study as immune genes. Finally, 1206 genes and 251 samples are included, as shown in Supplement Table 1.

Table 1
Relationship between five subtypes and clinical features (chi-square test)

	C1	C2	C3	C4	C5	P value
Age						0.076
> 60	34	65	29	40	24	
<=60	18	23	8	5	5	
Gender						0.4313
Male	26	37	15	26	13	
Female	26	51	22	19	16	
Stage T						0.675
T1	2	2	1	2	2	
T2	12	26	7	9	11	
T3	35	56	27	30	16	
T4	3	3	2	4	0	
Stage N						None
N0	52	88	37	45	29	
Stage M						None
M0	47	76	34	41	27	
Stage						0.238
StageI	14	29	8	11	13	

Molecular subtype screening based on immune genes

The expression profile of immune genes is used for consensus clustering. We use R software package ConsensusClusterPlus to screen molecular subtypes, which is the same method chosen by Zhang et al

[18]. We use the Euclidean distance method to calculate the similarity distance between samples and use K-means to cluster. We use a resampling scheme to sample 80% of the samples and resample 100 times. The optimal clustering number is determined by the cumulative distribution function (CDF). The clustering significance between these subtypes is further analyzed by using the R software package sigclust.

Relationship between subtypes and clinical features

Different clinical features are closely related to the development of disease. By analyzing the relationship between subtypes and clinical features, we can more clearly observe the relationship between subtypes and disease development. We extract Age, Gender, T, N, M, and Stage information from the clinical follow-up data of patients, and observe the relationship between subtypes and Age, Gender, T, N, M, and Stage.

Relationship between subtypes and immunity

Previous studies have reported the key gene sets involved in the immune process. We collect 13 types of immune metagenes to analyze the relationship between these metagenes and subtypes, finding that the immune components of tumor tissue are closely related to the tumor prognosis. We analyze the relationship among matrix, immune scores and molecular subtypes. It is found that the score of immune cell infiltration directly reflects the degree of immune infiltration in tumor tissues and is also closely related to the occurrence and development of tumors. We also analyze the relationship between the scores of immune cell infiltration in subtypes and use variance analysis to evaluate the differences of the above scores in different subtypes.

Relationship between subtypes and survival prognosis

We extract data from the sample follow-up information and use K-M to analyze the prognostic differences of different subtypes.

Results

Screening of molecular subtypes based on immune genes

We use the R software package ConsensusClusterPlus for consensus clustering based on immune gene expression profiles to screen molecular subtypes. We use the Euclidean distance method to calculate the similarity distance between samples and use K-means to cluster. We use a resampling scheme to sample 80% of the samples and resample 100 times. The optimal clustering number is determined by CDF as shown in Fig. 1A, from which we can see that the clustering results are more stable when the Cluster is 5. Further observation of the CDF delta area curve is shown in Fig. 1B, from which we can find that there is a stable clustering result when the Cluster is selected as 5. We eventually choose $k = 5$ and obtain five molecular subtypes. Furthermore, the clustering significance between the five subtypes is analyzed by using the R software package sigclust. It is found that the classification between C1 and C2 is not

significant ($p = 1$). However, C1-vs-C4 and C3-vs-C5 have a significantly different expression distribution, $p < 0.05$ (as shown in Supplement Table 2).

Table 2
Statistics of the
number of genes
corresponding to each
module

Modules	Genes
blue	112
brown	101
grey	155
turquoise	115

Cluster analysis of the expression profiles of the five molecular subtypes

According to the consensus clustering results, we select the stable clustering results of $k = 5$, as shown in Fig. 2A. It can be seen that 251 tumor samples have been assigned to these five categories. Furthermore, we use the expression profiles of 1206 immune gene set for subtype difference analysis and use the Kolmogorov-Smirnov test to screen for genes with high expression in one subgroup relative to the other subtype samples. Using $FDR < 0.05$ as the threshold, 80 genes with significantly high expression in the C1 subtype, 79 genes in the C2 subtype, 153 genes in the C3 subtype, 148 genes in the C4 subtype and 73 genes in the C5 subtype are screened out. We then select the expression profiles of the first 100 genes in the most significantly high expression of each subtype for PCA (if the genes with significantly high expression are fewer than 100, all the differentially expressed genes are taken). The first two principal components are obtained to draw a scatter plot, as shown in Fig. 2C, from which we can see that the five subtypes can be clearly separated. The expression profiles of these genes are further used to make heat maps, as shown in Fig. 2D. It can be seen that each subtype has a clear boundary in the expression profile of these genes and has obvious expression patterns as well.

Relationship between the five subtypes and clinical features

We analyze the relationship between the five subtypes and Age, Gender, T, N, M and Stage, as shown in Table 1. It can be seen that there is no significant relationship between the five subtypes and these clinical features. We further analyze the relationship between these five subtypes and CMS subtypes, and count the CMS ratios in the different subtypes, as shown in Fig. 3A. The highest proportion in the C1, C2, C3, C4 and C5 subtypes is CMS2, CMS3, CMS4, CMS1, and CMS2, respectively. We analyze the enrichment of each CMS class in the different subtypes. CMS2 in the C1 subtype, CMS3 in the C2

subtype, CMS4 in the C3 subtype, and CMS1 in the C4 subtype are significantly enriched. $FDR < 0.001$. These are shown in Fig. 3B. This result suggests that there is a strong correlation between the C1, C2, C3, C4 and CMS subtypes in the five molecular types. Previous studies have shown that microsatellite unstable subsets (CMS1) with good prognosis are characterized by the overexpression of specific genes in cytotoxic lymphocytes. The Mesenchymal subset (CMS4) with poor prognosis expresses the markers of cells derived from lymphocytes and monocytes, and the mesenchymal subsets also show the characteristics of angiogenesis, inflammation and immunosuppression. Pathological examination shows that the characteristics of mesenchymal subtypes are high-density fibroblasts. This mesenchymal subtype may produce chemokines and cytokines that are conducive to tumor-related inflammation and support angiogenesis, resulting in a poor prognosis. The classical subtype (CMS2) and metabolic subtype (CMS3) with intermediate prognosis exhibit low immune and inflammatory characteristics.

Relationship between the five subtypes and immunity

To analyze the relationship between these five subtypes and immunity, we collect 13 types of immune metagenes, tumor immune component (matrix, immunity, and tumor purity) scores, six types of immune cell infiltration scores, and MCP counts of 10 immune-related cells from previous studies. We analyze the relationship between the immune-related scores of the three groups and the five subtypes. Most of the 13 types of immune metagenes have high expression in C3 and C4, as shown in Fig. 4A. The immune score of the C4 group is significantly higher than that of the other groups. The matrix score of group C3 is significantly higher than that of other groups, as shown in Fig. 4B. In the ten kinds of immune-related cells, Endothelial cells and Fibroblasts in the C3 group have higher expression than those in the other groups, while T cells and Cytotoxic lymphocytes in the C4 group have higher expression than those in other groups, as shown in Fig. 4C. Among the six types of immune infiltration scores, the B_cell and CD8_cell scores are significantly higher in the C4 group than in the other groups. The CD4_Tcell, Neutrophil, Dendritic and Macrophage scores are significantly higher in the C3 and C4 groups than in the C1, C2 and C5 groups, as shown in Fig. 4D. In summary, most of the immune signatures are upregulated in the C4 subtype compared to the C1, C2 and C5 subtypes. Most of the immune infiltration-related cells are upregulated in the C3 and C4 subtypes compared to the C1, C2 and C5 subtypes. This result suggests that the immune microenvironment is enhanced in the C3 and C4 subtypes, as shown in Fig. 5.

Prognosis difference analysis of the five subtypes

To observe the relationship between five subtype samples and prognosis, we use Kaplan-Meier to analyze the prognostic differences of the five type samples, as shown in Fig. 6A. It can be seen that there are significant differences in prognosis among the five types of samples. Among the five types of samples, the prognosis of the C4 samples is the worst.

Compared with the other categories, the prognosis of the C4 subtype is significantly worse than that of the C1 and C3 samples, as shown in Fig. 4B. This result suggests that there are two opposite clinical prognostic outcomes in the immune-enhanced C3 and C4 subtypes in colonic carcinoma.

Relationship between the five subtypes and mutations

Gene mutations of the APC, K-RAS and p53 genes are considered to be essential in the development of colonic carcinoma. BRAF is also a common mutation in colorectal cancer species. Therefore, we analyze the relationship between TP53, KRAS, BRAF and APC mutations in these five types of samples. First, the mutation data for TP53, KRAS, BRAF and APC are extracted from the SNP data processed by mutect of TCGA. The proportions of the TP53, KRAS, BRAF, and APC mutation groups and nonmutation groups for the five subtypes are analyzed, as shown in Fig. 7A-D. There are significant differences in the proportion of mutations in these five genes among the five types of samples. In TP53, the mutation rate in the C2 type samples is significantly lower than that in the C1 and C3 samples. In KRAS, the mutation proportions of C3 and C4 are significantly lower than that of the other types. In BRAF, the mutation proportion of C4 is significantly higher than that of the other five types. There is no mutation in C5. In APC, the C4 mutation proportion is significantly lower than that of the other five types. We further analyze the distribution of the number of all mutant genes in the five types of samples, as shown in Fig. 7E. There are significant differences in the frequency of gene mutations among the five types of samples. Among the five types of samples, the frequency of gene mutations in the C4 type samples is significantly higher than that of the other groups, $p < 0.0001$.

Relationship between the five subtypes and the expression of eight immune checkpoint genes

We further analyze the relationship between the expression of eight immune checkpoint genes in these five subtypes, as shown in Fig. 8A. The expression of PDCD1LG2, CTLA4, CD86 and CD80 in C3 and C4 subtypes is significantly higher than those in other types. The expression of PDCD1, CD274 and CD267 in C4 is higher than those of other groups. Further statistics of the gene expression distribution of the eight immune checkpoints, as shown in Fig. 8B, reveal that there are significant differences in the expression distribution in the five types of samples.

Mining of immune-enhanced subtype-related modules by WGCNA analysis

To further explore the prognostic markers related to the immune microenvironment of colonic carcinoma, we collect the expression profile data for the five subtypes of immune-related genes. A total of 483 samples are obtained. We further use the Pearson correlation coefficient to calculate the distance between each transcript. The R software package WGCNA is used to construct a weight co-expression network. We select the soft threshold of 4 to screen the co-expression module. The results show that the co-expression network conforms to the scale-free network. That is, the logarithmic log (k) of the node with connection degree k is negatively correlated with the logarithmic log(P(k)) of the probability of the node, and the correlation coefficient is greater than 0.8. To ensure that the network is scale-free, we choose $\beta = 4$ (Fig. 9A, B). The next step is to convert the expression matrix into an adjacency matrix and then

transform the adjacency matrix into the topological matrix. Based on TOM, we use the average-linkage hierarchical clustering method to cluster genes, according to the standard of a mixed dynamic shear tree, and set the minimum number of genes in each gene (lncRNA) network module to 30. After using the dynamic shearing method to determine the gene module, we calculate the eigenvector value (eigengenes) of each module in turn. Then, the cluster analysis of the module is carried out, and the modules with closer proximity are merged into new modules, setting height = 0.25, deepSplit = 2, and minModuleSize = 30. A total of 4 modules are obtained, as shown in Fig. 9C. It should be noted that the gray module is a collection of genes that cannot be clustered into other modules. The transcript statistics for each module are shown in Table 2, from which we can find that 328 transcripts are assigned to the three co-expression modules. We calculate the correlation between the eigenvectors of the four modules and the five subtypes respectively, as shown in Fig. 9D. It can be seen that blue module is positively correlated with C3, the brown module is positively correlated with C4, and the turquoise module is positively correlated with C3 and C4. The number of transcripts in the three modules is 112, 101 and 115 genes, respectively.

We further analyze the functions of the genes in the blue and brown modules that are most relevant to C3 and C4. We use the R software package clusterProfiler for KEGG enrichment analysis and select a significance of $FDR < 0.05$. In the two modules, the blue module is enriched into 72 pathways, among which the most significant 20 pathways are shown in Fig. 10B. The brown module is enriched into 79 pathways, among which the most significant 20 pathways are shown in Fig. 10A. For all results, please refer to Supplement Table 3. We analyze the relationship between the paths in which these two modules are enriched, as shown in Fig. 10. It can be seen that the two modules are completely enriched to 103 pathways. There are 18 (17.5%) KEGG pathways enriched between the two modules, with fewer intersections, as shown in Fig. 10C. This result suggests that the C3 and C4 samples have different pathway anomalies and may have different regulatory mechanisms.

External validation

We select the genes in the gene co-expression module (blue and brown modules) that are closely related to the C3 and C4 subtypes to calculate the correlation between the genes and the modules. A total of 30 genes with correlation coefficients greater than 0.75 are selected, from which the expression profile is extracted as the training set. The classification model is established by using the support vector machine (SVM), and the samples are classified. The classification accuracy is 100%. To further validate these five subtypes, we download GSE39582 standardized data from the GEO database, containing a total of 585 samples. The expression profiles of genes as well as Stage I and II samples are extracted from the blue and brown modules. The samples are classified by the model. It is predicted that there are 80 C1 samples, 72 C2 samples, 86 C3 samples, 57 C4 samples and 14 C5 samples, as shown in Supplement Table 4. First, we analyze the expression distribution of 13 immune metagenes in each subtype of these five samples, as shown in Fig. 11A. The high expression of most metagenes in C3 and C4 is consistent with the validation set. Further analysis of sample immune scores is shown in Fig. 11B. It can be observed that the immune score of the C4 group is significantly higher than that of other groups, and the matrix score of the C3 group is significantly higher than that of other groups, which is consistent with the

training set. Ultimately, we analyze the prognostic differences of the five subtypes, as shown in Fig. 11E. It can be seen that there is no significant difference in prognosis among the five subtypes, but the prognosis of C4 is worse than that of other subtypes. Further analysis of the prognostic differences between the five types of samples and recurrence shown in Fig. 11F reveals that there is no significant difference in prognosis. The prognosis of C4 is worse than that of the other types. No significant difference in prognosis may be due to the short median follow-up time (Median OS = 58.5 months, RFS = 53 months).

Discussion

There are many publications studying the mechanism of cancer based on gene expression data. For example, Yang Yong et al [19] used systematic bioinformatics methods to explore the DNA methylation regulatory role in colonic carcinoma samples in the TCGA database. Liu Rong et al. [20] used a gene weighted co-expression network analysis technique to extract the co-expressed gene network from mRNA expression and identified a total of 11 co-regulatory modules in the sample data set of 461 patients with colonic carcinoma. Most of these studies are based on genome-wide research. Such research is comprehensive, but often ignores the characteristics of immune-related genes [21]. Our research focuses on immune genes, which helps obtain additional details about immunity.

In this study, we run consensus clustering of immune gene expression profiles in the TCGA database and obtain five subtypes after screening. The subsequent principal component analysis (PCA) has proven that the five molecular subtypes can be clearly distinguished. It is found that the C3 and C4 subtypes show stronger immune activity among the five subtypes. These subtypes are different in terms of clinical features, immune scores, gene mutations, prognosis, immune checkpoints, and metabolic pathways involved in regulation, which are validated in external GEO datasets. These findings may provide some new insights into immunotherapy for colonic carcinoma.

First, we compare the relationship between the five subtypes and clinical features, in which C1, C2, C3, and C4 are strongly correlated with the common molecular typing (CMS) of colonic carcinoma. Subtypes C1 and C2 are closely related to CMS2 (classic type) and CMS3 (metabolic type), respectively. These subtypes often show low immunity and inflammatory features. The tumors lacking infiltration of lymphocytes and immunomodulatory factors are usually PDL1 negative. It is suggested that the immunogenicity is poor [22–23], which may be the reason for the low immune activity of the C1 and C2 subtypes. In the following immune scores, the scores of the C1 and C2 subtypes are lower in 13 types of immune metagenes, tumor immune components (matrix, immunity, and tumor purity) scores, six types of immune cell infiltration scores and ten kinds of immune-related cells MCP count scores. The C3 subtype is closely related to the CMS4 type (mesenchymal type). For the characteristics of the CMS4 type, tumor matrix infiltrating epithelial cells, tumor-associated fibroblasts (CAFs) and innate immune cells produce more inflammatory factors [24], which may be the reason for the strong immune activity of the C3 subtype. Most of the 13 types of immune metagene are highly expressed in the C3 subtype. The matrix score of the C3 subtype is significantly higher than that of other subtypes. Among the 10 types of

immune-related cells, in Endothelial cells, Fibroblasts and six types of immune infiltration scores, CD4_Tcell, Neutrophil, Dendritic and Macrophage in the C3 subtypes are significantly higher than those in the C1, C2 and C5 subtypes. The C4 subtype is closely related to CMS1 (immunophenotype). CMS1 is characterized by hypermutation and hypermethylation. The immune cells significantly infiltrate the tumor microenvironment, especially CD8 + T cells, CD4 + T cells and NK cells [25], which may be the reason for the stronger immune activity of the C4 subtype. In the subsequent immune scores, it can be observed that most of the 13 types of immune metagenes are highly expressed in the C4 subtype. The immune score of the C4 subtype is significantly higher than that of other groups. Among the 10 immune-related cells, the scores of B_cell, CD8_cell, CD4_Tcell, Neutrophil, Dendritic, and Macrophage, T cells, Cytotoxic lymphocytes and six kinds of immune infiltration scores is significantly higher in the C4 group than in the C1, C2, and C5 groups. Overall, most of the immune signatures are upregulated in the C4 subtype, and most of the cells associated with immune infiltration are upregulated in the C3 and C4 subtypes, suggesting an enhanced immune microenvironment in the C3 and C4 subtypes. The relationship between the gene expression of 8 immune checkpoints in the five subtypes is analyzed. The results show that the expression of PDCD1LG2, CTLA4, CD86 and CD80 in C3 and C4 subtypes is significantly higher than that in other subtypes, and the expression of PDCD1, CD274 and CD267 in C4 is higher than that in other groups. The high expression of immune checkpoints may be the cause of stimulating the body to produce stronger immune activity. C1, C2 and C5 subtypes have lower immune activity expression, while the prognosis of these three subtypes is significantly different and not the worst, which is found in the prognostic analysis afterwards. It is worth mentioning that although C4 has a high expression of immune activity, its prognosis is the worst, which suggests that the level of immune activity in colonic carcinoma cannot directly indicate the quality of prognosis.

In addition, the mutation characteristics of five subtypes also reflect the characteristics of different subtypes. Microsatellite instability (MSI) of tumors mainly refers to CMS1. The subtype is mainly presented with high mutations in MSH6, RNF43, ATM, TGFBR2, BRAF and PTEN. CMS1 is closely related to the C4 subtype. According to the relationship between five subtypes and mutations, the proportion of the C4 mutation in BRAF is significantly higher than that in other subtypes. Further analysis of the number of all mutant genes in the five subtypes show that C4 type samples are significantly higher than in other groups, which is consistent with the characteristics of CMS1 mutations. Tumors with chromosomal instability (CIN) mainly include CMS2-4. The main mutations that exist in this type are mutations of APC, KRAS, TP53, SMAD4 and PIK3CA, which are closely related to the C1, C2 and C3 subtypes [26]. According to the relationship between the five subtypes and mutations, the proportion of C4 mutations in KRAS and APC is lower than that in other types. Ultimately, we obtain a blue module, brown module and turquoise module, which are closely related to C3 and C4 subtypes through WGCNA analysis. We apply the external dataset GSE39582 as the test set and use the SVM method to verify our previous experimental results to illustrate the reliability of the method and data.

Conclusions

In summary, we identify five molecular subtypes of colonic carcinoma based on TCGA immune gene expression data. These five subtypes are significantly different in terms of clinical features, immune scores, gene mutations, prognosis, and the expression of checkpoint genes. The external dataset validates our experimental results. The discovery of the immune microenvironments may provide new ideas for immunotherapy strategies for colonic carcinoma.

Declarations

Acknowledgments

Thanks to *American Journal Experts* for polishing this article

Author Contributions

Dongxin Tang, Wei Huang: Conceptualization, Writing, Data curation and Formal analysis; Zhu Yang, Xin Wu and Xianan Sang: Methodology and Supervision;

Kuilong Wang: Writing - original draft; Gang Cao and Min Hao: review & editing, and financial support

Availability of data and materials

The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that there are no conflicts of interest

References

1. Xu Z, Jiang H, Zhu Y, Wang H, Jiang J, Chen L, Xu W, Hu T, Cho CH. Cryptotanshinone induces ROS-dependent autophagy in multidrug-resistant colon cancer cells. *Chem. Biol. Interact.* 2017 Aug 01, 273.
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, He J. *Cancer Statistics in China, 2015*. CA: Cancer Journal for Clinicians, 2016(66), 115-132.
3. Yu C, Hong H, Zhang S, Zong Y, Ma J, Lu A, Sun J, Zheng M. Identification of key genes and pathways involved in microsatellite instability in colorectal cancer. *Mol Med Rep*, 2019 Mar, 19(3).
4. Duan X, Chan C, Han W, Guo N, Weichselbaum RR, Lin W. Immunostimulatory nanomedicines synergize with checkpoint blockade immunotherapy to eradicate colorectal tumors. *Nat Commun*, 2019,

04(23);10(1).

5. Shuchun Li, Junjun Ma, Xizhou Hong, Minhua Zheng, Shigenori Goto, Rishu Takimoto, Takashi Kamigaki, Lu Zang. Significant clinical response of advanced colorectal cancer to combination therapy involving capecitabine and adoptive cell transfer therapy: a case report. *Translational Cancer Research*, 2019, 8(2), 693-698.
6. Fanipakdel A, Seilianian Toussi M, Rezazadeh F, Mohamadian Roshan N, Javadinia SA. Overexpression of cancer-testis antigen melanoma-associated antigen A1 in lung cancer: A novel biomarker for prognosis, and a possible target for immunotherapy. *J. Cell. Physiol.* 2019, Jul, 234(7).
7. Wang BC, Zhang ZJ, Fu C, Wang C. Efficacy and safety of anti-PD-1/PD-L1 agents vs chemotherapy in patients with gastric or gastroesophageal junction cancer: a systematic review and meta-analysis. *Medicine (Baltimore)* 2019 Nov, 98(47).
8. Silina K, Kroeger D. Editorial: Immune Outposts on the Inflammatory Frontier: Tertiary Lymphoid Structures as Targets for Immunotherapy of Cancer and Autoimmunity. *Front Immunol*, 2019, 10.
9. Ahn S, Li J, Sun C, Gao K, Hirabayashi K, Li H, Savoldo B, Liu R, Dotti G. Cancer Immunotherapy with T Cells Carrying Bispecific Receptors That Mimic Antibodies. *Cancer Immunol Res*, 2019 May, 7(5).
10. Coosemans A, Vankerckhoven A, Baert T, Boon L, Ruts H, Riva M, Blagden S, Delforge M, Concin N, Mirza MR, Ledermann JA, du Bois A, Vergote I. Combining conventional therapy with immunotherapy: A risky business? *Eur. J. Cancer*, 2019, May;113.
11. Boursi B, Werner TJ, Gholami S, Margalit O, Baruch E, Markel G, Eshet Y, Houshmand S, Shacham-Shmueli E, Mitchell TC, Mamtani R, Alavi A, Yang YX. Physiologic colonic fluorine-18-fluorodeoxyglucose uptake may predict response to immunotherapy in patients with metastatic melanoma. *Melanoma Res.* 2019 Jun; 29 (3):318-321.
12. Khalili N, Keshavarz-Fathi M, Shahkarami S, Hirbod-Mobarakeh A, Rezaei N. Passive-specific immunotherapy with monoclonal antibodies for prostate cancer: A systematic review. *J Oncol Pharm Pract*, 2019 Jun;25(4).
13. Poletajew S, Krajewski W, Adamowicz J, Radziszewski P. A systematic review of preventive and therapeutic options for symptoms of cystitis in patients with bladder cancer receiving intravesical bacillus Calmette-Guérin immunotherapy. *Anticancer Drugs*, 2019 Jun;30(5).
14. Najafi M, Farhood B, Mortezaee K. Contribution of regulatory T cells to cancer: A review. *J. Cell. Physiol.* 2019 Jun;234(6).
15. Sadreddini S, Baradaran B, Aghebati-Maleki A, Sadreddini S, Shanehbandi D, Fotouhi A, Aghebati-Maleki L. Immune checkpoint blockade opens a new way to cancer immunotherapy. *J. Cell. Physiol.* 2019, Jun;234(6).
16. Yang L, Sun J, Liu Q, Zhu R, Yang Q, Hua J, Zheng L, Li K, Wang S, Li A. Synergetic Functional Nanocomposites Enhance Immunotherapy in Solid Tumors by Remodeling the Immunoenvironment. *Adv Sci (Weinh)*, 2019, Apr 17, 6(8).
17. Aline Talhouk, Heather Derocher, Pascal Schmidt, Samuel Leung, Katy Milne, C. Blake Gilks, Michael S. Anglesio, Brad H. Nelson and Jessica N. McAlpine. Molecular Subtype Not Immune Response Drives Outcomes in Endometrial Carcinoma. *Clinical Cancer Research*, 2019, 25(8), 2537-2548.
18. Zhang S, Wang Y, Gu Y, Zhu J, Ci C, Guo Z, Chen C, Wei Y, Lv W, Liu H, Zhang D, Zhang Y. Specific

- breast cancer prognosis-subtype distinctions based on DNA methylation patterns. *Mol Oncol*, 2018 06;12(7).
19. Yang Y, Chu FH, Xu WR, Sun JQ, Sun X, Ma XM, Yu MW, Yang GW, Wang XM. Identification of regulatory role of DNA methylation in colon cancer gene expression via systematic bioinformatics analysis. *Medicine (Baltimore)*, 2017, Nov;96(47)
20. Liu R, Zhang W, Liu ZQ, Zhou HH. Associating transcriptional modules with colon cancer survival through weighted gene co-expression network analysis. *BMC Genomics* 2017, 05 09;18(1).
21. Kumar P, Khadirnaikar S, Shukla SK. PILAR1, a novel prognostic lncRNA, reveals the presence of a unique subtype of lung adenocarcinoma patients with KEAP1 mutations. *Gene*, 2019 Apr 05;691.
22. Trinh A, Ládach C, Dawson HE, Ten Hoorn S, Kuppen PJK, Reimers MS, Koopman M, Punt CJA, Lugli A, Vermeulen L, Zlobec I. Tumour budding is associated with the mesenchymal colon cancer subtype and RAS/RAF mutations: a study of 1320 colorectal cancers with Consensus Molecular Subgroup (CMS) data. *Br. J. Cancer* 2018 11;119(10).
23. Barresi V, Cinnirella G, Valenti G, Spampinato G, Musso N, Castorina S, Condorelli DF. Gene expression profiles in genome instability-based classes of colorectal cancer. *BMC Cancer* 2018 Dec 18;18(1).
24. Smeets D, Miller IS, O'Connor DP, Das S, Moran B, Boeckx B, et al. Copy number load predicts outcome of metastatic colorectal cancer patients receiving bevacizumab combination therapy. *Nat Commun*, 2018, 10 05;9(1).
25. Cohen R, Heran M, Pudlarz T, Hilmi M, Tournigand C, André T, Rousseau B. Going beyond microsatellite instability for immunotherapy in metastatic colorectal cancer: Consensus molecular subtypes and tumor mutational burden. *Bull Cancer*, 2019, Feb;106(2).
26. Dienstmann R, Vermeulen L, Guinney J, Kopetz S, Tejpar S, Tabernero J. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat. Rev. Cancer*, 2017 02;17(2).

Supplemental Tables

Supplement Table 1: The number of genes and samples included in this study

Supplement Table 2: Clustering significance between pairs of the five subtypes

Supplement Table 3: Functional enrichment analysis of genes in the blue and brown modules most relevant to C3 and C4

Supplement Table 4: Functional enrichment analysis of genes in the blue and brown modules most relevant to C3 and C4

Figures

Figure 1

A: CDF curve; B: CDF delta area curve, Delta area curve of consensus clustering represents the relative variation of the area under the cumulative distribution function (CDF) curve of each category number k compared with $k-1$. The horizontal axis represents the category number k , and the vertical axis represents the relative change in the area under the CDF curve.

Figure 2

A: Sample clustering heat map when consensus k is 5; B: Intersection Venn diagram for significantly highly expressed genes of the various subtypes; C: PCA analysis of the expression profile of the top 100 highly expressed genes. Scatter plot of the first two principal components; D: Heat map of significantly highly expressed genes of the top 100 subtypes

Figure 3

A: Distribution of the CMS subtypes in the five subtypes. The upper half represents the proportion of CMS in the different subtypes. The lower half represents the number of CMS in the different subtypes. There are five subtypes horizontally. B: Enrichment significance of the CMS subtypes in the different subtypes. The value in the figure represents significance-log10 (FDR).

Figure 4

A: The relationship between the expression of 13 kinds of immune metagenes and the five subtypes; B: Relationship between the matrix scores, immune scores, tumor purity scores and five subtypes; C: The relationship between ten kinds of immune-related cells scores and the five subtypes; D: The relationship between the scores of six types of immune cell infiltration and five subtypes

Figure 5

Expression distribution of immune metagenes and immune infiltration scores in five subtypes; variance analysis of various score differences.

Figure 6

AKM curve of relationship prognosis of the five subtypes B: KM curve of prognostic difference between C4 and the other five subtypes

Figure 7

A-D: Proportion of TP53, KRAS, BRAF, and APC mutations and wild types in the five types of samples; Differences in the mutation proportion distribution by chi-square test; C: Distribution of gene mutation numbers in the five types of samples; Variance analysis, $p < 0.0001$.

Figure 8

The expression of eight immune checkpoint genes in the five subtypes, as tested by variance analysis.

Figure 9

A, B: Analysis of network topology for various soft-thresholding powers Figure C: Gene dendrogram and module colors Figure D: Correlation between each module and the five subtypes

Figure 11

A: The relationship between the expression of 13 kinds of immune metagenes and the five subtypes; B: Relationship between the matrix scores, immune scores, tumor purity scores and five subtypes; C: The relationship between ten kinds of immune-related cells scores and the five subtypes; D: The relationship between the scores of six types of immune cell infiltration and five subtypes. E: Difference in overall survival prognosis of the five subtypes in the verification set. F: Differences in recurrence prognosis of the five subtypes in the verification set

Supplementary Files

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

- [SupplementTable1.xlsx](#)
- [SupplementTable3.xlsx](#)
- [SupplementTable2.xlsx](#)
- [SupplementTable4.xlsx](#)