Establishment and Validation of a Genetic Label Associated With M2 Macrophage Infiltration to Predict Survival in Colon Cancer Patients and Assist Immunotherapy

Boyang Xu

Department of Gastroenterology, the First Affiliated Hospital of China Medical University, Shenyang, China

Ziqi Peng

Department of Breast Surgery, the First Affiliated Hospital of China Medical University, Shenyang, China

Guanyu Yan

Department of Gastroenterology, the First Affiliated Hospital of China Medical University, Shenyang, China

Ningning Wang

Department of Gastroenterology, the First Affiliated Hospital of China Medical University, Shenyang, China

Moye Chen

Department of Gastroenterology, the First Affiliated Hospital of China Medical University, Shenyang, China

Xue Yao

Department of Surgical Oncology, the First Hospital of China Medical University, Shenyang, China

Mingjun Sun (✉ smjzgykdx@163.com)

Department of Gastroenterology, the First Affiliated Hospital of China Medical University, Shenyang, China

Yue An

Department of Endoscopy, the First Hospital of China Medical University, Shenyang, China

Research Article

Keywords: Colon cancer, genetic label, M2 macrophage infiltration, patients, assist immunotherapy

Posted Date: July 30th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-712255/v1>

Abstract

Background: Colon cancer is a kind of malignant tumor with high morbidity and mortality. Researchers have tried to interpret it from different perspectives and divide it into different subtypes in order to achieve individualized treatment. With the rise of immunotherapy, its value in the field of tumor has initially emerged. Based on the above background, from the perspective of immune infiltration, this study classified colon cancer according to the infiltration of M2 macrophages in patients with colon cancer and further explored it.

Methods: Cibersort was used to analyze the level of immune cell infiltration in colon cancer patients in the TCGA database. WGCNA, Consensus Clustering analysis, Lasso analysis, and univariate KM analysis were used to screen and verify the hub genes associated with M2 macrophages. PCA was used to establish the M2 macrophage-related score—M2I Score. The correlation between M2I Score and somatic cell variation and microsatellite instability were analysed. Furthermore the correlation between M2 macrophage score and differences in immunotherapy sensitivity was also explored.

Results: M2 macrophage infiltration was associated with poor prognosis. Four hub genes (ANKS4B, CTSD, TIMP1, and ZNF703) were selected as the progression-related genes associated with M2 macrophages. A stable and accurate M2I Score for M2 macrophages used in COAD was constructed based on four hub genes. M2I Score was positively correlated with tumor mutation load (TMB). The M2I Score of MSI-H group was higher than that of MSI-L group and MSS group. Combine with the TCIA database, we concluded that patients with a high M2I Score were more sensitive to PD-1 inhibitors and PD-1 inhibitors combined with CTLA-4 inhibitors. The low rating group may have better efficacy without immune checkpoint inhibitors or with CTLA4 inhibitors alone.

Conclusion: Four prognostic hub genes associated with M2 macrophages were screened to establish the M2I Score and divided the patients into two subgroups: high M2I Score group and low M2I Score group. TMB, microsatellite instability and sensitivity to immunotherapy were higher in the high-rated group. PD-1 inhibitors or PD-1 combined with CTLA-4 inhibitors are preferred for patients in the high-rated group who are more sensitive to immunotherapy.

Introduction

According to the latest global cancer data released by the International Agency for Research on Cancer (IARC) of the World Health Organization (WHO) in 2020, colon cancer accounted for 10% of the new cases of all malignant tumors, ranking the third among all malignant tumors, and its deaths reached nearly 940,000 cases, accounting for 9.4% of all cancer deaths¹. Despite continued research and improvement in treatment regimens for colon cancer, the five-year survival rate for patients with colon cancer remains low². In recent years, immunotherapy has attracted a great deal of attention for its success in treating previously difficult solid tumors, such as melanoma and lung cancer^{3,4}. In colon cancer, immunotherapy, particularly immunocheckpoint inhibitor therapy, has shown promising results in

patients with mismatched repair defects (DMMR) or high levels of microsatellite instability (MSI-H). And it was approved by regulators in 2017 to treat those tumors with severe mutations^{5,6}.

Macrophages are one of the most abundant white blood cells in the colon and play an important role in a variety of intestinal diseases such as inflammatory bowel disease and bowel cancer^{7,8}. Tumor associated macrophages (TAM) have different polarization directions, among which M1 macrophages are called pro-inflammatory macrophages, while M2 macrophages are called anti-inflammatory macrophages⁹. In tumors, M1 macrophages can effectively clear tumor cells by presenting antigen to T cells, activating specific immune response, regulating and promoting the immune response of Th1 cells¹⁰. In contrast, M2 macrophages inhibit the proliferation and activation of T cells by secreting immunosuppressive factors, cytokines and growth factors, regulate and promote Th2 immune response, promote the growth of tumor cells, participate in tumor angiogenesis, and promote tumor invasion and metastasis¹¹. Both M1 and M2 macrophages exist in the tumor microenvironment (TME), but with the progress of the tumor, M2 macrophages gradually increase and become the macrophage type with a higher proportion, leading to the poor prognosis of the patients¹². Due to the importance of macrophages in tumors, tumor therapy strategies targeting macrophages have received considerable attention¹³. In solid tumors such as gastric cancer and melanoma, high levels of M2 macrophage infiltration are associated with higher expression levels of immune checkpoints such as PD-1 and PD-L1, suggesting that macrophages can be used as a potential therapeutic target for tumors^{14,15}. In tumor therapy, the use of anti-M2 macrophages combined with immune checkpoint inhibitors improves the therapeutic effect, which also provides a new idea for the treatment of tumors¹⁶.

With the development of high-throughput sequencing technology, cancer patients are classified into different subtypes according to the expression of specific genes, and individualized treatment regimens based on the characteristics of different subtypes provide a new direction to improve the prognosis of patients¹⁷. Based on the above background, we believe that immune infiltration is a good starting point. From the perspective of immune cell infiltration, this study applied Cibersort algorithm to analyze the infiltration of M2 macrophages in patients with colon cancer, and divided colon cancer into two subtypes by screening Hub genes related to M2 macrophages. Finally, M2 macrophage infiltration score (M2I Score) was constructed to accurately predict the prognosis of patients and their sensitivity to immunotherapy, and to provide some reference for the clinical individualized medication diagnosis and treatment.

Results

Data downloading and collection

A total of 432 patients with colon cancer from the TCGA dataset were included in this study. The validation set was from the GEO database. The data set GSE39582 included 566 samples of colon

cancer tissue and 19 samples of paracancerous normal tissue. The cancer tissue samples were selected for further analysis. TCGA and GEO related patient information is shown in Table S1.

Immunoinfiltration analysis and screening of M2 macrophage related genes in COAD using WGCNA

We used Cibersort to analyze the immune cell infiltration in patients with TCGA-COAD, and obtained the M2 macrophage score of each sample. Survival analysis showed that the infiltration level of M2 macrophages had an impact on the survival time of patients (Fig. 1A). Patients with higher M2 macrophage scores had poor survival($p = 7.163e-03$). To further dig out the genes associated with M2 macrophage infiltration in colon cancer, we performed WGCNA analysis on the data. We performed a WGCNA analysis on the data. We found that when we set the soft threshold to 7, it accords with the scale-free property of biological network, so we used $\beta = 7$ to construct the weighted network. Then, we carried out average linkage hierarchical clustering, identified the modules based on TOM's differences, dynamic tree pruning and merging processing, and obtained a total of 20 meaningful modules with different colors (Fig. 1B). Then, all the modules analyzed in WGCNA were correlated with M2 macrophage scoring results, and we found that the tan module had the strongest correlation with M2 macrophages ($R = 0.62$) and $P = 6.1e-13$ was statistically significant (Fig. 1C). We enriched and analyzed 110 genes in the tan module, and found that the biological functions of this group of genes were related to immunity (Fig. 1D;figure 1E). So far, we have found a group of stable genes related to M2 macrophages in intestinal cancer, and their biological functions are correlated with tumor immunity to a certain extent.

Screening of hub genes based on cluster analysis

In order to further explore the characteristics of genes in the TAN module, we performed K-value based consistent clustering based on the expression of 110 genes involved in the tan module. According to the cumulative distribution function (CDF), $k = 2$ was selected as the optimal parameter, and TCGA-COAD patients were divided into two groups, named ClusterA and ClusterB (Fig. 2A). Further, through survival analysis, it was found that the OS of patients in ClusterA and ClusterB had significant differences($P = 0.003$), as shown in Fig. 2B. Thus, we infer that M2 macrophage-related genes in the TAN module can affect the OS of TNBC patients through immune-related pathways.

In order to screen hub genes in the module, univariate Cox analysis was conducted on the genes in the TAN module to screen the genes that were related to the survival of patients. When the P value in univariate Cox analysis was less than 0.05, 5 candidate genes could be screened (Table 1). Furthermore, through 1000 LASSO analyses, four hub genes (Fig. 2C) were finally screened out, namely: ANKS4B, CTSD, TIMP1 and ZNF703, which are the most stable prognostic-related genes associated with M2 macrophages in colon cancer.

Table 1
Results of the univariate Cox regression analysis
between gene expression and OS.

ID	HR	HR.95L	HR.95H	pvalue
ANKS4B	0.745	0.575	0.965	0.026
TIMP1	1.421	1.102	1.831	0.007
CTSD	1.479	1.062	2.061	0.021
NFKB2	1.437	1.022	2.019	0.037
ZNF703	0.773	0.604	0.988	0.040

In order to verify the prognostic value of the four hub genes, according to the expression of these four genes, we applied the consistency cluster analysis based on the K value, and selected $k = 2$ as the optimal parameter. TCGA-COAD patients were divided into two groups, named ClusterA and ClusterB (Fig. 2D). Survival analysis showed that the OS of ClusterB was significantly lower than that of ClusterA, and the difference was statistically significant ($P = 0.004$) (Fig. 2E).

Validation of screened hub genes

In order to verify the stability of the consistency clustering method based on K value for the classification of TCGA-COAD patients, another classification method, namely PCA analysis method, was selected for verification again. The results are shown in Fig. 3A. Meanwhile, ssGSEA results showed that the macrophage score of ClusterB was higher than that of ClusterA, and the difference was statistically significant ($P < 0.001$; Fig. 3B). The expression of the four hub genes in different clusters is shown in Fig. 3C. At the same time, we chose another dataset, GSE39582, to prove that the change of the dataset would not affect our conclusion. In GSE39582, cluster analysis was conducted according to the expression of hub genes, and the results showed that these 4 genes could well divide patients with colon cancer into two clusters (Fig. 3D). The survival analysis between the two groups showed statistically significant differences in OS (Fig. 3E). Since then, we have obtained consistent results in different datasets and different classification methods, proving the stability and accuracy of the above key gene screening.

Construction of M2I Score

In order to construct M2 macrophage score in TCGA-COAD patients, PC1 values of genes in gene signature A and B were calculated by principal component analysis (PCA) and the sum of PC1 (SPC1a and SPC1b) of gene signature A and B were calculated. Subsequently, the difference between SPC1A and SPC1B was used as M2 macrophage score – M2I Score. Patients in the TCGA cohort were divided into two groups according to the M2I Score by using the optimal cutoff value obtained by X-tile software. Figure 4A shows the distribution of patients with high and low scores. Patients with high scores were mainly from ClusterB while those with low scores were mainly from ClusterA. Simultaneously, survival

analysis showed that the survival of patients in the high-rated group was significantly worse than that in the low-rated group ($p < 0.001$) (Fig. 4B). Meanwhile, according to Wilcoxon test, immune checkpoints (CD274 and CTLA4) and M2 macrophage marker genes (CCl2, CCR2, CD163, CD40, CSF1R, MRC1 and PDGFB) were significantly overexpressed in the high M2I Score group (Fig. 4C). At this point, We have an accurate M2 macrophage score – M2I Score, which reflects the level of immune checkpoints and M2 macrophage marker genes.

Correlation between M2I Score and somatic variation

There is substantial evidence that a higher tumor mutation load (TMB) represents a better patient response to immunotherapies such as immun checkpoint inhibitor therapy¹⁸. Considering the important clinical significance of TMB, we decided to explore the relationship between M2I Score and TMB. For this reason, we first analyzed the differences in TMB scores between groups with high and low M2I Scores, and the results showed that TMB was significantly higher in the group with high scores than in the group with low scores ($p = 1e-06$; Fig. 5A). Meanwhile, Spearman correlation analysis showed that M2I Score was positively correlated with TMB ($R = 0.17$, $p = 0.0016$; Fig. 5B). In addition, we also analyzed the differences of somatic cell variation driver genes in the high and low M2I Score group of colon cancer. The top 20 driver genes with the highest frequency of change were selected using the R package "Maftools" for analysis, and the mutation frequency of 16 genes in the high-rated group was higher than that in the low-rated group (Fig. 5C; Fig. 5D). These results suggest that patients in the highly rated group may have a better response to immunotherapy.

The role of M2I Score in predicting the benefit of immunotherapy

In colon cancer, higher microsatellite instability often represents patients' ability to obtain better immunotherapeutic effects⁶. In order to further explore the relationship between M2I Score and microsatellite instability, relevant data from the TCIA database were used for analysis. According to the TCIA database, the MSI of patients with TCGA-COAD is divided into three levels: 1. MSS – Microsatellite stabilization; 2. MSI-L – low instability of microsatellites; 3. MSI-H – Microsatellite is highly unstable. The proportion of patients with three kinds of MSI levels in the high and low M2I Score group was calculated. The results showed that the proportion of MSI-H in the high M2I Score group was 43% higher than that in the low M2I Score group (11%), and the chi-square test showed that the difference was statistically significant (figure. 6A). Meanwhile, patients were grouped according to the MSI level, and the differences of M2I Score among different groups were compared. The results showed that the M2I Score of patients in the MSI-H group was higher than that of patients in the MSI-L and MSS groups ($p = 1.1e-07$ and $p = 2.3e-09$, respectively) (Fig. 6B). This suggests that patients with a high M2I Score are more likely to benefit from immunotherapy.

Based on the above results, we analyzed the difference of efficacy between PD-1 inhibitor and CTLA4 inhibitor in patients with different rating groups according to the sensitivity data of immunotherapy in TCIA database. The results showed that patients in the M2I high-level group, previously associated with

somatic variation and microsatellite instability analysis, who might be more sensitive to immunotherapy were more sensitive to PD-1 inhibitors ($p = 0.022$, Fig. 7A) and PD-1 inhibitors in combination with CTLA4 inhibitors ($p = 0.0015$, Fig. 7B). For the group with low sensitivity to immunotherapy, we did not use immune checkpoint inhibitors ($p = 0.00048$, Fig. 7C) or CTLA4 inhibitors alone ($p = 0.012$, Fig. 7D), which may achieve better efficacy. Taken together, these data suggest that M2I Score may be associated with immunotherapeutic response and may have implications for the selection of immunosuppressive agents in clinical treatment.

Discussion

As one of the most common malignancies, colon cancer has been the third most common cancer in terms of new cases and the second most common cancer in terms of deaths in 2020. Therefore, developing more effective treatments for colon cancer has been an urgent problem for researchers¹⁹. As the most abundant immune cells in colon, macrophages play an important role in the interaction between tumor cells and tumor microenvironment. As one of the carcinogenic differentiation types, M2 macrophages have attracted extensive attention as a potential therapeutic target¹². In this study, we developed a method to quantify M2 macrophage infiltration in CRC. Our results indicate that M2 macrophage infiltration score can accurately predict patient survival and has potential guiding significance for the selection of immunotherapy drugs.

More and more studies have shown that M2 macrophages are involved in the occurrence and development of colon cancer, and promote its invasion and metastasis, thus leading to poor prognosis of patients. Therefore, it is very important to construct a score that can evaluate the infiltration degree of M2 macrophages in patients with colon cancer^{20,21}. In this study, immune cell infiltration analysis was performed on TCGA-COAD samples according to Cibersort algorithm, and WGCNA was used to identify modules associated with M2 macrophage infiltration. Consistency clustering based on the genes within the module could divide the patients into two clusters with different survival conditions. Subsequently, univariate and Lasso Cox regression analyses were performed to screen the most robust prognostic biomarkers to establish the M2-macrophage-related genetic signature. Finally, four hub genes were obtained: ANKS4B, CTSD, TIMP1 and ZNF703. ANKS4B is expressed in intestinal cells and is distinct in the distal part of the brushlike microvilli. Studies have shown that this gene plays an important role in the assembly of brushlike microvilli²². Destruction or malformation of the brush border is associated with a number of intestinal diseases, including infections of attached and disappearing microorganisms and microvillous inclusion diseases^{23,24}. In addition, the loss of brush-imbic proteins involved in cell polarity in colon cancer is important for tumor development²⁵, suggesting that ANKS4B may influence the occurrence and development of colon cancer. Proteins encoded by CTSD are involved in a variety of biological processes, such as proteolysis in highly acidic lysosomal environments during development and maintenance of tissue homeostasis²⁶. In tumor research, CTSD secreted by tumor cells into the extracellular space plays an important role in the invasion and metastasis of breast cancer and ovarian cancer²⁷. In colon cancer, activation of the Wnt/ β -catenin signaling pathway can lead to increased levels

of endogenous CTSD, and thus enhance the proliferation and invasiveness of colon cancer cells²⁸. The proteins encoded by TIMP1 may inhibit the activity of matrix metalloproteinases (MMPs) and regulate the balance of matrix remodeling during extracellular matrix degradation²⁹. In colon cancer, TIMP1 has higher expression levels in tumor tissues and lymph node metastases than in normal tissues, and affects the prognosis of patients through the FAK-PI3K/ AKT and MAPK pathways³⁰. ZNF703 plays an important role in the occurrence and development of head and neck squamous cell carcinoma, non-small cell lung cancer and other tumors^{31,32}. In colon cancer, the inhibition of ZNF703 expression can inhibit the proliferation and migration of CRC cells, which is considered as a potential therapeutic target for metastatic colon cancer³³.

Based on these four hub genes, CRC patients were divided into two groups, Cluster A and Cluster B, by consistent clustering method, and the results showed that the prognosis of the two groups was significantly different. At the same time, ssGSEA was used to analyze the immune cell infiltration of patients, and the results showed that there were also significant differences in the infiltration of macrophages between the two groups. Thus, we concluded that these 4 hub genes were related to M2 macrophages in CRC patients, and could affect the OS of patients. In addition, the external verification set from the GEO database also verified the accuracy of the genetic signature constituted by the four hub genes in predicting the prognosis of patients.

Considering the individual differences in immune cell infiltration, it is important to quantify the infiltration pattern of M2 macrophages in individual tumors. Individual-based models based on tumor subtype specific biomarkers have been well used in breast cancer and other cancers to improve the accuracy of patient prognosis prediction³⁴. In this study, we used the above hub gene as a potential "subtype biomarker" and established an M2I score to quantify M2 macrophage infiltration in each sample. Survival analysis showed that higher M2I scores were associated with poorer survival. Ju et al. have shown that tumor-associated macrophages can induce tumor cells to express PD-L1 through IL-6 and TNF- α signaling¹⁴. In colon cancer, macrophages are present in higher concentrations in patients with DMR-MSI-H, which represents good immunotherapeutic sensitivity^{35,36}. Based on the above conclusions, we decided to further explore the association between M2I score and immunotherapy sensitivity. Numerous studies have shown a clear association between gene mutations and patients' responsiveness to immunotherapy such as immune checkpoint inhibitor therapy³⁷. In this study, we found that TMB was significantly increased in patients with higher M2I scores, and the correlation between M2I score and TMB was 0.17. In addition, the mutation frequency of several genes was also different between the high and low M2I rating groups. Further analysis of the relationship between M2I Score and microsatellite instability also showed that higher M2I Score also represented higher MSI level. Therefore, we preliminarily concluded that M2I Score may be associated with immunotherapy sensitivity. Finally, according to the relevant data in the TCIA database, the sensitivity differences of patients with low and low M2I ratings to PD-1 and CTLA-4 inhibitors were analyzed. We found that patients with high M2I Score had higher sensitivity to PD-1 inhibitors and PD-1 inhibitors combined with CTLA4 inhibitors. For patients in the low-rated group, no immunocheckpoint inhibitors or CTLA4 inhibitors alone may yield better results. According to Fiegle et al.,

dual inhibition of CTLA-4 and PD-L1 resulted in tumor growth arrest and complete blocking of liver metastasis, whereas inhibition of CTLA-4 and PD-L1 alone only modestly reduced metastatic spread of colon cancer cells³⁸. In this context, we conclude that dual immuncheckpoint suppressive therapy for CTLA-4 and PD-L1 may be the preferred immunotherapy for patients with high M2I scores.

In general, 4 hub genes associated with M2 macrophages in colon cancer were screened out in this study, and they could affect patients' OS through immune-related pathways. Based on these four genes, we constructed an M2I Score which can predict patients' survival. In addition, we hypothesized that these four genes may be involved in the sensitivity of colon cancer patients to immune checkpoint inhibitors. There are still some limitations in this paper, which are mainly reflected in the fact that this study is based on the exploration of public database, and experimental verification is still needed, and the specific regulatory mechanism of characteristic genes on colon cancer still needs to be explored.

Methods

Data collection

The transcriptome data (RNA-seq; Fragments Per Kilobase Million [FPKM] value) and related clinical information of patients with colon cancer (TCGA-COAD) were downloaded from the TCGA database (<https://portal.gdc.cancer.gov/>). The R-package "LIMMA" was used to standardize transcriptome data and delete transcriptome data from multiple samples from the same patient. Chip data of colon cancer patients numbered GSE39582 in the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) was selected as the verification set. The validation set was selected according to the following criteria: (a) with mRNA expression data (b) with complete patient prognosis information. Data from the GEO database were used to verify the conclusions of the TCGA database analysis.

Analysis of M2 macrophage infiltration and identification of related genes

The Cibersort algorithm was used to analyze the levels of 22 kinds of immune cell infiltration in TCGA patients with colon cancer. CIBERSORT (<https://cibersort.stanford.edu/>), which is based on the principle of the linear support vector regression on immune cell subtype of deconvolution of the expression of matrix a tool that can be used to estimate the immune cell infiltration. Combined with the overall survival time of patients, the effect of M2 macrophage infiltration level on survival was analyzed using the R pack "Survival".

To define the genes associated with M2 macrophages in colon cancer, we used the WGCNA algorithm. Weighted gene co-expression network analysis (WGCNA) is a method to construct gene co-expression network based on gene expression data. First, we selected the first 5000 genes after the mean absolute deviation (MAD) sequencing, and used the R package of WGCNA to construct the co-expression network of mRNA expression of the above genes. We then chose a soft threshold parameter β to construct a proximity matrix that matches our gene distribution to a connection-based scale-free network. Then, the adjacency is transformed into topological overlap matrix (TOM), and the linkage hierarchical clustering is

performed according to the average of different measures based on TOM. In the end, we chose a gene-tree with a minimum (genome) of 30 and a module tree with a cut line of 0.25, combined with several modules to produce more rigorous results.

After the WGCNA analysis, a correlation analysis was conducted between the M2 macrophage score and the co-expression modules enriched by WGCNA, and a gene co-expression module related to M2 macrophages was obtained. Subsequently, we enriched and analyzed the genes in the co-expression module through the clusterprofiler package to explore the biological functions of the genes in the module.

Screening of hub genes

Consensus Clustering is an algorithm that can be used to identify the members and number of clusters in a dataset. In this study, Consensus ClusterPlus package was used to conduct conformance cluster analysis on the selected genes in the co-expression modules and K-M survival analysis was used to judge the difference of survival time among different clusters. Univariate Cox regression analysis was used to screen the prognostic genes in the co-expression module, followed by 1000 times of Lasso analysis to select the most stable genes as hub genes. Finally, conformance cluster analysis was performed on the finally screened hub genes again to determine the survival differences between different clusters.

Validation of selected hub genes

First, we verify the reliability of the clustering analysis method. PCA analysis was used to verify the results of the previous concordant analysis, so as to prove that the hub gene screened out by us can well divide colon cancer patients into two categories. Subsequently, ssGSEA was used to calculate the scores of immune cell infiltration in TCGA-COAD patients, and the differences of macrophage scores among different clusters were analyzed. Then, in order to verify the reliability of the data set, we selected GSE39582 data set from the GEO database to verify the grouping based on hub genes and the prognostic correlation. So far, the stability of the results has been verified in different clustering algorithms and different datasets.

Generation of M2 macrophage score and the difference between high and low rating groups

According to the previous typing results, the gene was classified as gene signature A when the gene expression decreased with the increase of typing value, while the gene expression increased with the increase of typing value, and the gene was classified as gene signature B (Removing the use of Bruta). PCA algorithm was used to calculate the M2I Score of each sample. The calculation formula is as follows:

$$M2I\ Score = \sum PC1A - \sum PC1B$$

The optimal truncation value was calculated according to the survival status of the patients, and the patients were divided into the high rating group and the low rating group, and the differences in survival status between the high rating group and the low rating group were analyzed. Meanwhile, the differences

in immune checkpoints and M2 macrophage marker gene expression between the different rating groups were analyzed by Wilcoxon test.

Relationship between M2 macrophage score and somatic variation

The mutation data of TCGA-COAD patients were downloaded and Perl was used to count the number of non-synonymous mutations. The total number of somatic gene coding errors, base substitutions, gene insertion or deletion errors detected per million bases was defined as the tumor mutation load (TMB). The difference of TMB between the high and low rating groups was calculated, and the P value ($p < 0.05$) was considered statistically significant. Spearman correlation coefficient was used to analyze the relationship between M2I Score and TMB. Then, the R package "maftools" was used to identify the COAD driver genes, and the situation of the top 20 genes with the highest mutation frequency in the high rating group and the low rating group was further analyzed.

Differences in microsatellite instability and immunotherapy sensitivity between high and low rating groups

Data on microsatellite instability (MSI) and immunotherapy sensitivity of TCGA-COAD were downloaded from The Cancer Immunome Atlas database (<https://tcia.at/>). The Cancer Immunome Atlas database (TCIA) was developed and maintained by the Institute of Bioinformatics (ICBI). The database can query data on gene expression of specific immune-related gene sets, cell composition of immune infiltrates, and tumor heterogeneity. We analyzed the difference of M2I Score among patients with different levels of MSI (MSS, MSI-L, MSI-H) and the proportion of different levels of MSI between groups with high and low M2I Scores. Subsequently, according to the sensitivity scores of TCGA-COAD patients to PD-L1 and CTLA4 inhibitors in the TCIA database, we analyzed the differences in sensitivity to immunotherapy between the groups with high and low ratings.

Conclusions

M2 macrophage infiltration is associated with poor prognosis in colon cancer. Four prognostic hub genes associated with M2 macrophages in colon cancer were screened out as follows: ANK4B, CTSD, TIMP1 and ZNF703, and the stability of these results was verified by different clustering methods and GSE39582 dataset. The M2I Score was constructed and the patients with colon cancer were divided into two subgroups: high M2 Score group and low M2 Score group. The correlation between M2I Score and somatic cell variation and microsatellite instability was analyzed, respectively. The results showed that in the high-rated group, TMB was higher, microsatellite instability was stronger, and immunotherapy was more sensitive. Combined with the above results and the TCIA database, we concluded that patients in the high-rated group who were more sensitive to immunotherapy were given priority to use PD-1 inhibitors or PD-1 combined with CTLA-4 inhibitors, while patients in the low-rated group were given priority to use no immune checkpoint inhibitors or CTLA4 inhibitors alone.

Declarations

Acknowledgements

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article material.

Author contributions statement

YA and MS conceived the study. BX conducted all statistical analyses. ZP and YA reviewed relevant literature and drafted the manuscript. NW, GY, MC and XY provided guidance to the study. All authors read and approved the final manuscript.

Funding

This work was supported by grants from the National Key R&D Program of China (No. 2016YFC1303601) and Fund for Liaoning Province Science and Technology Plan Project: Study on Pathogenicity and Antibiotic Resistance of HofE-positive *Helicobacter pylori* Strains (NO.20180550049).

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians* **70** (2020).
2. Siegel R, Desantis C, Jemal A. Colorectal cancer statistics, 2014. *CA Cancer J Clin.* 2014; 64: 104-17.
3. Gandhi, L. *et al.* Pembrolizumab plus Chemotherapy in Metastatic Non–Small-Cell Lung Cancer. *New England Journal of Medicine*, NEJMoa1801005 (2018).
4. Amm, E. *et al.* Adjuvant Pembrolizumab versus Placebo in Resected Stage III Melanoma. *The New England journal of medicine* **378** (2018).
5. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science (New York, N.Y.)* **357**, 409 (2017).
6. Ganesh, K. *et al.* Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nature Reviews Gastroenterology & Hepatology* (2019).
7. Lee & S., H. Quantitative analysis of total macrophage content in adult mouse tissues. Immunochemical studies with monoclonal antibody F4/80. *Journal of Experimental Medicine* **161**, 475-489 (1985).
8. Platt, A. M., Bain, C. C., Bordon, Y., Sester, D. P. & Mowat, A. M. An Independent Subset of TLR Expressing CCR2-Dependent Macrophages Promotes Colonic Inflammation. *Journal of Immunology*

- 184**, 6843-6854 (2010).
9. B, A. M. A. *et al.* The chemokine system in diverse forms of macrophage activation and polarization - ScienceDirect. *Trends in Immunology* **25**, 677-686 (2004).
 10. Sica & Antonio. Macrophage plasticity and polarization: in vivo veritas. *Journal of Clinical Investigation* (2012).
 11. Biswas, S. K. *et al.* A distinct and unique transcriptional program expressed by tumor-associated macrophages (defective NF-kappaB and enhanced IRF-3/STAT1 activation). *Blood* **107**, 2112 (2006).
 12. Badawi, M. A., Abouelfadl, D. M., El-Sharkawy, S. L., El-Aal, W. & Abbas, N. F. Tumor-Associated Macrophage (TAM) and Angiogenesis in Human Colon Carcinoma. *Open Access Macedonian Journal of Medical Sciences* **3** (2015).
 13. Denardo, D. G. & Ruffell, B. Macrophages as regulators of tumour immunity and immunotherapy. *Nature Reviews Immunology* **19**, 1 (2019).
 14. Xj, A., Hz, B., Zz, A., Miao, C. & Qiang, W. D. Tumor-associated macrophages induce PD-L1 expression in gastric cancer cells through IL-6 and TNF- α signaling - ScienceDirect. *Experimental Cell Research* **396** (2020).
 15. Kim, Y. J., Chong, H. W., Mi, W. L., Choi, J. H. & Lee, W. J. Correlation Between Tumor-Associated Macrophage and Immune Checkpoint Molecule Expression and Its Prognostic Significance in Cutaneous Melanoma. *Journal of Clinical Medicine* **9**, 2500 (2020).
 16. Gordon, S. R., Maute, R. L., Dulken, B. W., Hutter, G. & Weissman, I. L. PD-1 expression by tumor-associated macrophages inhibits phagocytosis and tumor immunity. *Nature* **545**, 495-499 (2017).
 17. Cook, D. P. & Vanderhyden, B. C. Ovarian cancer and the evolution of subtype classifications using transcriptional profiling†. *Biology of Reproduction*, 3 (2019).
 18. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nature Genetics* (2019).
 19. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA Cancer J Clin* **70**, 7-30, doi:10.3322/caac.21590 (2020).
 20. Jingqin *et al.* M2 macrophage-derived exosomes promote cell migration and invasion in colon cancer. *Cancer Research* (2018).
 21. Vinnakota *et al.* M2-like macrophages induce colon cancer cell invasion via matrix metalloproteinases.
 22. Graves, M. J., Matoo, S., Choi, M. S., Storad, Z. A. & Crawley, S. W. A cryptic sequence targets the adhesion complex scaffold ANKS4B to apical microvilli to promote enterocyte brush border assembly. *Journal of Biological Chemistry* **295**, jbc.RA120.013790 (2020).
 23. Vallance, B. A., Chan, C., Robertson, M. L. & Finlay, B. B. Enteropathogenic and Enterohemorrhagic Escherichia coli Infections: Emerging Themes in Pathogenesis and Prevention. *Canadian journal of gastroenterology = Journal canadien de gastroenterologie* **16**, 771-778 (2002).

24. Wilson, W., Scott, R. B., Pinto, A. & Robertson, M. A. Intractable diarrhea in a newborn infant: microvillous inclusion disease. *Canadian Journal of Gastroenterology* **15**, 61-64 (2016).
25. Rocco, M. *et al.* Brush border myosin Ia has tumor suppressor activity in the intestine. *Proceedings of the National Academy of Sciences of the United States of America* **109** (2012).
26. Benes, P., Vetvicka, V. & Fusek, M. Cathepsin D—many functions of one aspartic protease. *Critical Reviews in Oncology/hematology* (2008).
27. Pranjol, Z. I. & Whatmore, J. L. *Cathepsin D in the Tumor Microenvironment of Breast and Ovarian Cancers*. (Tumor Microenvironment, 2020).
28. Basu, S. *et al.* Increased expression of cathepsin D is required for L1-mediated colon cancer progression. *Oncotarget* **10** (2019).
29. Batra, J., Robinson, J., Soares, A. S., Fields, A. P. & Radisky, E. S. Matrix metalloproteinase-10 (MMP-10) interaction with tissue inhibitors of metalloproteinases TIMP-1 and TIMP-2: binding studies and crystal structure. *Journal of Biological Chemistry* **287**, 15935-15946 (2012).
30. Song, G., Xu, S., Zhang, H., Wang, Y. & Wang, X. TIMP1 is a prognostic marker for the progression and metastasis of colon cancer through FAK-PI3K/AKT and MAPK pathway. *Journal of Experimental & Clinical Cancer Research* **35**, 148 (2016).
31. Orhan, C., Bakr, B., Dalay, N. & Buyru, N. ZNF703 is an important player in head and neck cancer. *Clinical Otolaryngology* **44** (2019).
32. Baykara, O., Da Lay, N., Kaynak, K. & Buyru, N. ZNF703 Overexpression may act as an oncogene in non-small cell lung cancer. *Cancer Med* **5**, 2873-2878 (2016).
33. Ma, F. *et al.* ZNF703 promotes tumor cell proliferation and invasion and predicts poor prognosis in patients with colorectal cancer. *Oncology Reports* **32**, 1071 (2014).
34. Callari, M., Cappelletti, V., D 'Aiuto, F., Musella, V. & Bianchini, G. Subtype-Specific Metagene-Based Prediction of Outcome after Neoadjuvant and Adjuvant Treatment in Breast Cancer. *Clinical Cancer Research An Official Journal of the American Association for Cancer Research* **22**, 337 (2016).
35. Hu, W., Yang, Y., Qi, L., Chen, J. & Zheng, S. Subtyping of microsatellite instability-high colorectal cancer. *Cell Communication and Signaling* **17** (2019).
36. Narayanan, S., Kawaguchi, T., Peng, X., Qi, Q. & Takabe, K. Tumor Infiltrating Lymphocytes and Macrophages Improve Survival in Microsatellite Unstable Colorectal Cancer. *Scientific Reports* **9** (2019).
37. Pan, D., Hu, A. Y., Antonia, S. J. & Li, C. Y. A Gene Mutation Signature Predicting Immunotherapy Benefits in Non-Small Cell Lung Cancer Patients. *Journal of Thoracic Oncology* **16** (2020).
38. Fiegler, E. *et al.* Dual CTLA-4 and PD-L1 Blockade Inhibits Tumor Growth and Liver Metastasis in a Highly Aggressive Orthotopic Mouse Model of Colon Cancer. *Neoplasia* **21**, 932-944 (2019).

Figures

Figure 1

Analysis and screening of M2 macrophage related genes A. Survival analysis of patients with different levels of M2 macrophage infiltration B. Analysis of gene distribution in WGCNA network C. Correlation analysis of Tan module and M2 macrophages D. GO analysis of genes in the Tan module E. KEGG analysis of genes in the TAN module

Figure 2

Screening of key genes based on cluster analysis A. Consistency cluster analysis based on genes within the TAN module B. Survival analysis of patients between subgroups differentiated by genes within the TAN module C. Lasso analysis and screening of hub gene D. Consistency cluster analysis based on four hub genes E. Survival analysis of patients between subgroups differentiated by hub gene

Figure 3

Validation of screened key genes A. PCA diagram of the TCGA queue B. Comparison of ssGSEA scores among different subgroups C. Distribution of hub gene among different subgroups D. Consistent clustering analysis based on four hub genes in GSE39582 E. Survival analysis of patients in different subgroups in the GEO cohort

Figure 4

Construction of the M2I score A. Sankey plot of survival outcomes in the distribution set of M2I scores in different subgroups B. Survival difference between groups with high and low M2I rating C. Immun checkpoint related genes and M2 macrophage related genes were expressed in the subgroups with high and low M2I scores.

Figure 5

Correlation between M2I Score and somatic variation A. TMB difference between high and low M2I score subgroups B. Correlation analysis between M2I score and mutation load C-D. Oncoprint of the top 20 genes with the highest mutation frequency in the subgroup of high and low M2I scores

Figure 6

Relationship between M2I score and MSI A. The proportion of different MSI levels in the subgroups with high and low M2I scores B. Differences in M2I scores among groups with different MSI levels

Figure 7

Role of M2I score in predicting immunotherapy benefits A-D. Sensitivity of patients with high and low M2I score subgroups to four treatments (A.Use of PD-1 inhibitor alone B. Use of PD-1 inhibitor in combination with CTLA-4 inhibitor C.Do not use immune checkpoint inhibitors D. CTLA-4 inhibitor alone)

Supplementary Files

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

- [additionalfile1tableS1.docx](#)