TCGA database analysis of the tumor mutation burden and its clinical significance in colon cancer

Junjie Chen[1][#], Anwaier Apizi[2][#], Lin Wang[2][#], Guanting Wu[1], Zheng Zhu[1], Huihui Yao[1], Guoliang Chen[1], Xinyu Shi[1], Bo Shi[1], Qingliang Tai[1], Chenglong Shen[3], Guoqiang Zhou[3], Lingzhi Wu[4], Songbing He[1]

[1]Department of General Surgery, The First Affiliated Hospital of Soochow University, Suzhou, China; [2]Department of Gastrointestinal Tumors, The First Affiliated Hospital of Xinjiang Medical University, Urumqi, China; [3]Department of Gastrointestinal Surgery, Changshu No. 2 Hospital, Changshu, China; [4]Department of Oncology, The First Affiliated Hospital of Soochow University, Suzhou, China

Contributions: (I) Conception and design: J Chen, S He; (II) Administrative support: G Wu, H Yao, G Chen, X Shi, Q Tai; (III) Provision of study materials or patients: X Shi, B Shi, G Chen, C Shen, G Zhou, L Wu, S He; (IV) Collection and assembly of data: J Chen, A Apizi, L Wang, G Wu, H Yao; (V) Data analysis and interpretation: J Chen, A Apizi, L Wang, Z Zhu, G Zhou, L Wu, S He; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Songbing He, MD, PhD. Department of General Surgery, The First Affiliated Hospital of Soochow University, Suzhou, China. Email: hesongbing1979@suda.edu.cn; Lingzhi Wu, MD, PhD. Department of Oncology, First Affiliated Hospital of Soochow University, Suzhou, China. Email: wulingzhi@suda.edu.cn.

Background: Colon cancer is one of the most common malignant tumors, with high rates of incidence and death. The tumor mutational burden (TMB), which is characterized by microsatellite instability, has been becoming a powerful predictor which can show tumor behavior and response to immunotherapy.

Methods: In this study, we analyzed 437 mutation data of colon cancer samples obtained from The Cancer Genome Atlas (TCGA) and divided patients into low- and high-TMB groups according to the TMB value. Then we identified differentially-expressed genes (DEGs), conducted immune cell infiltration and survival analyses between groups.

Results: The higher TMB of the patients with colon cancer predicts a poorer prognosis. Functional analysis was performed to assess the prognostic value of the top 30 core genes. The CIBER-SORT algorithm was used to investigate the correlation between the immune cells and TMB subtypes. An immune prognosis model was constructed to screen out immune genes related to prognosis, and the tumor immunity assessment resource (TIMER) was then used to determine the correlation between gene expression and the abundance of tumor-infiltrating immune cell subsets in colon cancer. We observed that APC, TP53, TTN, KRAS, MUC16, SYNE1, PIK3CA have higher somatic mutations. DEGs enrichment analysis showed that they are involved in the regulation of neuroactive ligand-receptor interaction, the Cyclic adenosine monophosphate (cAMP) signaling pathway, the calcium signaling pathway, and pantothenate and Coenzyme A (CoA) biosynthesis. The difference in the abundance of various white blood cell subtypes showed that Cluster of Differentiation 8 (CD8) T cells (P=0.008), activated CD4 memory T cells (P=0.019), M1 macrophages (P=0.002), follicular helper T cells (P=0.034), activated Natural killer (NK cell) cells (P=0.017) increased remarkably, while M0 macrophages significantly reduced (P=0.025). The two immune model genes showed that secretin (SCT) was negatively correlated with survival, while Guanylate cyclase activator 2A (GUCA2A) was positively correlated.

Conclusions: This study conducted a systematically comprehensive analysis of the prediction and clinical significance of TMB in colon cancer in identification, monitoring, and prognosis of colon cancer, and providing reference information for immunotherapy.

Keywords: Tumor mutation burden (TMB); colon cancer; prognosis; TCGA

Introduction

According to the global malignant tumor statistics from the International Agency for Research on Cancer (IAFC) in 2020, there were 1,148,515 newly diagnosed cases of colon cancer, with 576,858 deaths (1). With the development of medicine and improving treatment technology, colon cancer therapy has made significant progress by means of surgery, chemotherapy, radiotherapy, and targeted treatment; yet, the prognosis of colon cancer remains unsatisfactory (2,3). Studies have shown that 20–50% of colon cancer patients have tumor transfer or recurrence after surgery and the 5-year survival rate is only 57% (4). In recent years, the progress of colon cancer treatment has been slow, although immunotherapy has shown huge clinical application potential in the combined treatment of colon cancer (5). Although a series of past research found that TMB can be as an effective molecular diagnosis of colon cancer biomarkers (6,7), but most of them focus on the study of microsatellite instability and the combination of the TMB diagnostic value, even if some scholars research the tumor microenvironment and the relationship between immune cell infiltration (8), but still lack of correlation with tumor mutation load research. However, there are no effective biomarkers to evaluate the effects of colon cancer after immunotherapy.

It is an emerging biomarker that the tumor mutational burden (TMB) referring to the total number of mutations (per megyl, per 1 million bases) in tumor tissues, and it is increasingly applied in predicting tumor immunotherapy efficacy (9). The TMB is a novel cancer feature, which is related to microsatellite instability, deoxyribonucleic acid (DNA) replication defects, and reactions to programmed cell death protein 1 (PD-1) and Programmed cell death 1 ligand 1 (PD-L1) blocking immunotherapy (10). Among them, microsatellite instability is caused by the functional defects of DNA mismatch repair (MMR) in tumor tissues, resulting in pro-immune response, and the emergence of more highly expressed $CD8^+$ T cells, activated NK cells and M1 macrophages. Most $CD8^+$ T cells have a high level of PD1 expression, and PD1 can be identified as a clinical target of inhibitors, so as to produce good immunotherapy effect (11,12). The increase in TMB is attributed to both endogenous and exogenous factors (13). Studies have shown that cancers with a higher TMB will have a better reaction to immunotherapy (14). Therefore, TMB can be used as a useful biomarker for predicting survival and the effectiveness of immunotherapy (15).

Tumor tissue structure is complex. In addition to tumor cells, there are stromal cells, inflammatory cells, vascular, extracellular matrix and so on which together constitute the tumor microenvironment (16). In addition, only APC and KRAS mutations were more common in tumor tissues of high-purity. Pathways associated with focal adhesion, ECM-receptor interaction and Calcium signaling pathway were significantly increased in tumor tissues of low-purity. Immunotherapy related markers (PD-1, PD-L1, TIM-3, LAG-3, CTLA-4) were highly expressed in colon cancer tissues of low-purity, and tumor purity was negatively correlated with M1 macrophages, M2 macrophages, and neutrophils (17,18).

In colon cancer, ICI and TMB are closely related to its diagnosis, clinical sensitivity and prognosis (19,20). If it has the higher ICI score, the higher proportion of $CD8^+$ T cells, plasma cells, memory resting $CD4^+$ T cells, eosinophils, monocytes and dendritic cells, and the better prognosis they will be (19). Macrophages and neutrophils increased and overall survival rate decreased in colon cancer patients with low ICI score (19). Immuno-checkpoint molecules (PDCD1, CTLA-4, LAG3, CD274, IDO1, HAVCR2 and TECH) were overexpressed significantly in subgroups with low ICI score (19). In addition, patients with low TMB and high ICI score had the best overall survival (19).

Since The Cancer Genome Atlas (TCGA, https://cancergenome.nih.gov) established, the production of large-scale tumor genomic data sets and comprehensive biological information analysis has become possible (21). The present study extracted the gene expression profile data of colon cancer from TCGA, and explored the potential utility of the TMB in immunotherapy and individualized medication.

We present the following article in accordance with the REMARK reporting checklist (available at https://dx.doi.org/10.21037/jgo-21-661).

Methods

Data Collection and assembly

TCGA database (https://portal.gdc.cancer.gov/) (21) is the world's largest cancer gene information database, which has the application of genomic analysis techniques. TCGA has not only developed a large-scale genome sequence, but it also contains rich samples and more than 30 kinds of cancers. Most importantly for our purposes is that TCGA includes very detailed prognosis information. In this study, the gene expression profiles and related clinical data in colon cancer were collected from the TCGA-GDC (Genomic Data Commons) database (up to January 11, 2020), and the PERL language script was used to handle raw data. Finally, the clinical information (including ID (Identity document) number, survival time, survival status, age, gender, clinical installment, Tumor staging, lymph node transfer state, distant transfer state) of 398 tumor tissue samples and 39 adjacent normal tissue samples was obtained. The data of sample mutation was acquired, analyzed, and visualized using 'maftools' tool (Version 2.8.05, https://github.com/PoisonAlien/maftools) in the R package (R version 4.0.0 (2020-04-24), https://mirrors.tuna.tsinghua.edu.cn/CRAN/). This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

TMB value estimation

Generally, the TMB, the total number of mutations (per megabase), represents the mutation density of tumor genes which means the average number of mutations in the tumor genome that includes the total number of gene coding errors, gene insertions, base substitutions or deletion errors (15). The 38Mb was as usual extracted based on the length of human exons, so the TMB of each sample was estimated to be equal to the total mutation frequency. Through dividing the total number of mutations by the size of the target coding region, the TMB (per megabase) was calculated.

To a certain extent, it can be understood that the greater the total number of codes that are changed in a unit area, the higher the tumor mutation load, and thus, the more tumor-related carcinogenic mutations may be associated with each other. This in turn leads to more prominent personalities of each tumor with greater differences, thereby making it is easier to be recognized by the immune system and become the target of tumor immunity. In theory, the higher the TMB, the more effective immunotherapy the patient receives (15).

Relationship between TMB value and overall survival

Colon cancer patients were divided into 2 subtypes, which are low and high TMB subtype, according to the median TMB value (3.657) of all samples. Also, the R package (R version 4.0.0 (2020-04-24), https://mirrors.tuna.tsinghua.edu.cn/CRAN/) was used to perform Kaplan-Meier survival analysis on the relationship between two subtypes and survival, and subsequently to explore the predictive value of TMB for the prognosis of colon cancer.

Correlation between TMB value and clinicopathological features

The R package was used to analyze the relationship between TMB value and clinicopathological characteristics, including age, gender, tumor grade, stage grading and TNM (tumor, regional lymph node, metastasis) stage.

Relationship between TMB and differentially-expressed genes

Firstly, we used the Wilcox rank sum test to screen the differentially-expressed genes (DEGs) between the high and low TMB subgroups (22). Next, we used the 'limma' (Version 3.48.3, http://bioinf.wehi.edu.au/limma) R software package to calculate, filter and export all DEGs that satisfied the following conditions: False Discovery Rate (FDR) <0.05 and |log2 FC|>0.5. Lastly, we used "pheatmap" (Version 1.16.0, https://bioconductor.org/packages/heatmaps/) package to draw a heat map showing the different gene results.

DEGs Functional analysis

We used R package 'clusterProfiler (Version 4.0.5, https://yulab-smu.top/biomedical-knowledge-mining-book/), org.Hs.eg.db (Version 3.13.0, https://bioconductor.org/packages/org.Hs.eg.db/) plot, ggplot2 (Version 3.0.4, https://yulab-smu.top/treedata-book/)' to perform Gene Ontology (GO) pathway enrichment analysis of DEGS (23) and Kobas-Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. The protein-protein interaction (PPI) network of DEGs was constructed using the STRING database (24), and Cytoscape software (Version

3.6.1, https://cytoscape.org/) was employed to visualize the number of core gene nodes in the PPI network (25).

Survival analysis of core genes

The survival package in R (R version 4.0.0 (2020-04-24), https://mirrors.tuna.tsinghua.edu.cn/CRAN/) was used to assess the prognostic value of the top 30 core genes in colon cancer.

CIBER-SORT analysis

The CIBER-SORT analysis platform (https://cibersort.stanford.edu/) (26) was used to quantify the relative levels of different immune cell types in complex gene expression mixtures. CIBER-SORT's deconvolution of gene expression data can provide valuable information about the composition of immune cells in the sample.

Statistical analysis

We use the 'survival' package of the R language to perform survival analyses of immune-related model genes (27). The overall survival (OS) rate was tested using the Kaplan-Meier and log-rank tests. Differences in subgroups were analyzed using the Wilcox test or Kruskal test. The P value <0.05 was considered statistically significant.

Construction of immune prognosis model based on TMB

The immune-associated genes (IAGs) were screened from the differential genes between the high and low TMB subtypes. Also, using the 'survival' R package, we performed single-factor and multivariate Cox regression analysis of the IAGs and survival data of colon cancer patients, and then screened out the immune genes related to prognostic IAGs (pIAGs) (P<0.05). The risk score of all colon cancer patients was calculated according to the following prognostic model formula: $\text{Risk Score} = \sum_{i=1}^{x} \text{Exp}i \times \text{coef}i$, where coef is the regression coefficient of a gene, and Exp is the amount of expression of a gene in the sample. Furthermore, the patients were divided into high- and low-risk groups based on the risk scores, and the cutoff value was determined by the median risk score value. Finally, the 'survival' software package was used to draw Kaplan-Meier survival curves and construct ROC (receiver operating characteristic curve) curves to verify the reliability of the model. The results were visualized using the 'survivalROC' (version 1.0, https://cran.r-project.org/web/packages/survivalROC/index.html) software package.

Correlation analysis between prognosis-related immune gene copy number variation (CNV) and immune cells

In this study, we investigated the relationship between gene expression and immune cell infiltration using the tumor immunity evaluation resources (TIMER) database (28), which is a useful resource for comprehensive analysis of tumor-infiltrating immune cells. The TIMER database was also used to determine the correlation of infiltration abundance between tumor-infiltrating immune cell subsets (B cells, cluster of differentiation (CD) 4+ T cells, CD8+ T cells, macrophages, neutrophils, and dendritic cells) in colon cancer, as well as secretin (SCT) and guanylate cyclase activator 2A (GUCA2A) expression.

Results

Mutations in colon cancer

Firstly, we used the PERL language script to process the original number for each colon cancer sample collected from TCGA database, in order to perform mutation statistical analysis. Next, to understand the mutational factors related to the occurrence of colon cancer mutations, we used the R package 'maftools' tool to obtain, analyze, and visualize these factors. The results showed that missense mutations, SNP, and C>T mutations were much more common, and the highest mutation frequency was 8,541 (*Figure 1A*). A waterfall chart indicated the integration status of somatic mutations in colon cancer based on a MutSigCV algorithm, and the results showed that APC, TP53, TTN, KRAS, MUC16, SYNE1, PIK3CA, and PIK3CA had higher somatic mutations (P<0.001, *Figure 1B*).

TMB and clinical relevance

The samples were divided into high or low TMB groups according to the median value (3.657) of all samples' TMB. Also, to assess the potential correlation between TMB value of colon cancer and prognosis, we used the R package to group analyze the high and low TMB values by Kaplan-Meier survival analysis. The results reveled that TMB (*Figure 2A*, P=0.008) was negatively correlated with prognosis that elevated TMB resulted in poorer prognosis.
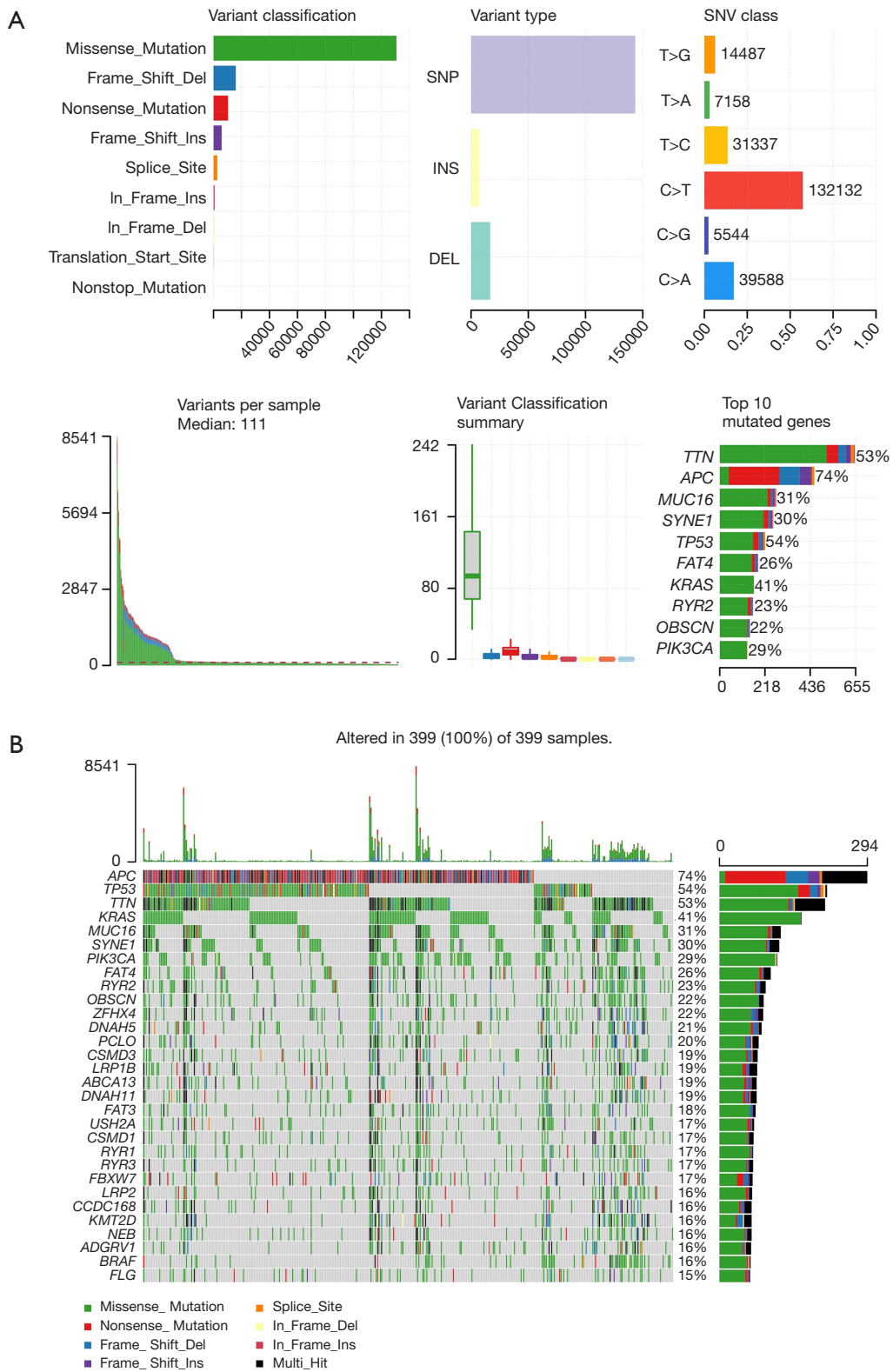
Figure 1 The Cancer Genome Atlas (TCGA) colon cancer mutation cohort. (A) Overview of cohort mutations in TCGA colon cancer. (B) Top 30 mutated genes waterfall chart in the TCGA colon cancer cohort.
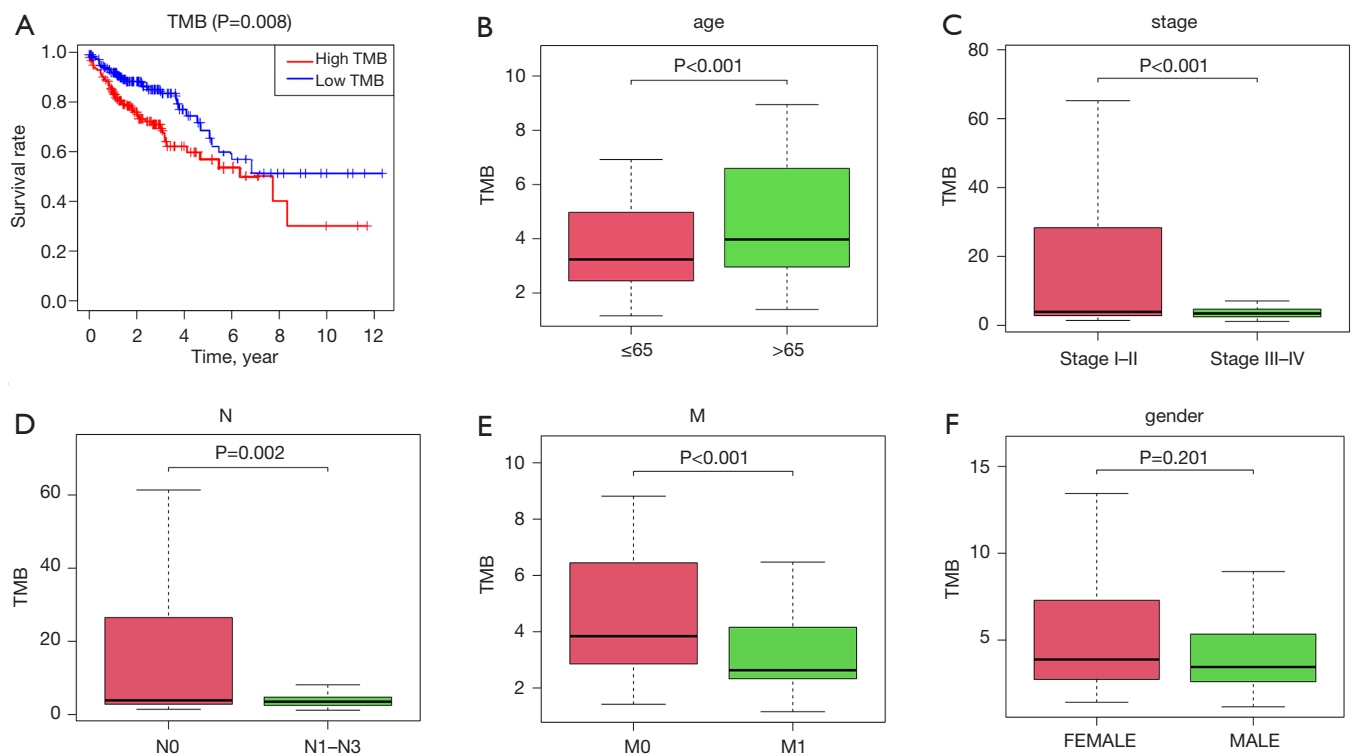
Figure 2 The tumor mutational burden (TMB) correlation analysis. (A) Overall survival Kaplan-Meier curves of the high and low TMB groups. (B) Wilcox test of patients stratified by age. (C) Wilcox test of patients stratified by stage. (D) Wilcox test of patients stratified by lymph node metastasis N. (E) Wilcox test of patients stratified by distant metastasis M. (F) Wilcox test of patients stratified by gender.

The correlation between colon cancer TMB and patient age, gender, tumor grade (tumor cell differentiation), and TNM stage showed that colon cancer TMB was related to age (*Figure 2B*, P<0.001), tumor grade (*Figure 2C*, P<0.001), with or without lymph node metastasis N (*Figure 2D*, P<0.001), and with or without distant metastasis M (*Figure 2E*, P<0.001); however, was unrelated to gender (*Figure 2F*, P=0.201). Therefore, it can be concluded that the TMB of young colon cancer patients is higher than that of elderly patients, the TMB value of early colon cancer patients is higher, and the TMB value of patients without lymph nodes and distant metastasis is higher.

Functional enrichment analysis of DEGs

The gene expression profiles between the high and low TMB value groups were analyzed using the 'limma'(Version 3.48.3, http://bioinf.wehi.edu.au/limma)software package in the R language, and a total of 164/236 DEGs were identified. Compared with normal samples, 121 genes were

down-regulated and 43 genes were up-regulated in colon cancer samples. The heatmap (*Figure 3*) shows the DEGs of different samples between the groups of the low and the high TMB value.

The GO enrichment analysis results of the 164 selected DEGs (*Figure 4A*) showed that the top 10 biological processes (BP) were multicellular organismal homeostasis, keratinization, cornification, positive regulation of ion transport, digestion, embryonic skeletal system development, digestive system process, peptide cross-linking, epithelial structure maintenance, and maintenance of gastrointestinal epithelium. Also, the top 10 cytology components (CC) were synaptic membrane, postsynaptic membrane, neuron-to-neuron synapse, axon terminus, neuron projection terminus, integral component of postsynaptic membrane, intrinsic component of postsynaptic membrane, integral component of postsynaptic density membrane, intrinsic component of postsynaptic density membrane, and cornified envelope. Moreover, the top 10 molecular functions (MF) were DNA-binding transcription activation activity, Ribonucleic Acid (RNA) polymerase
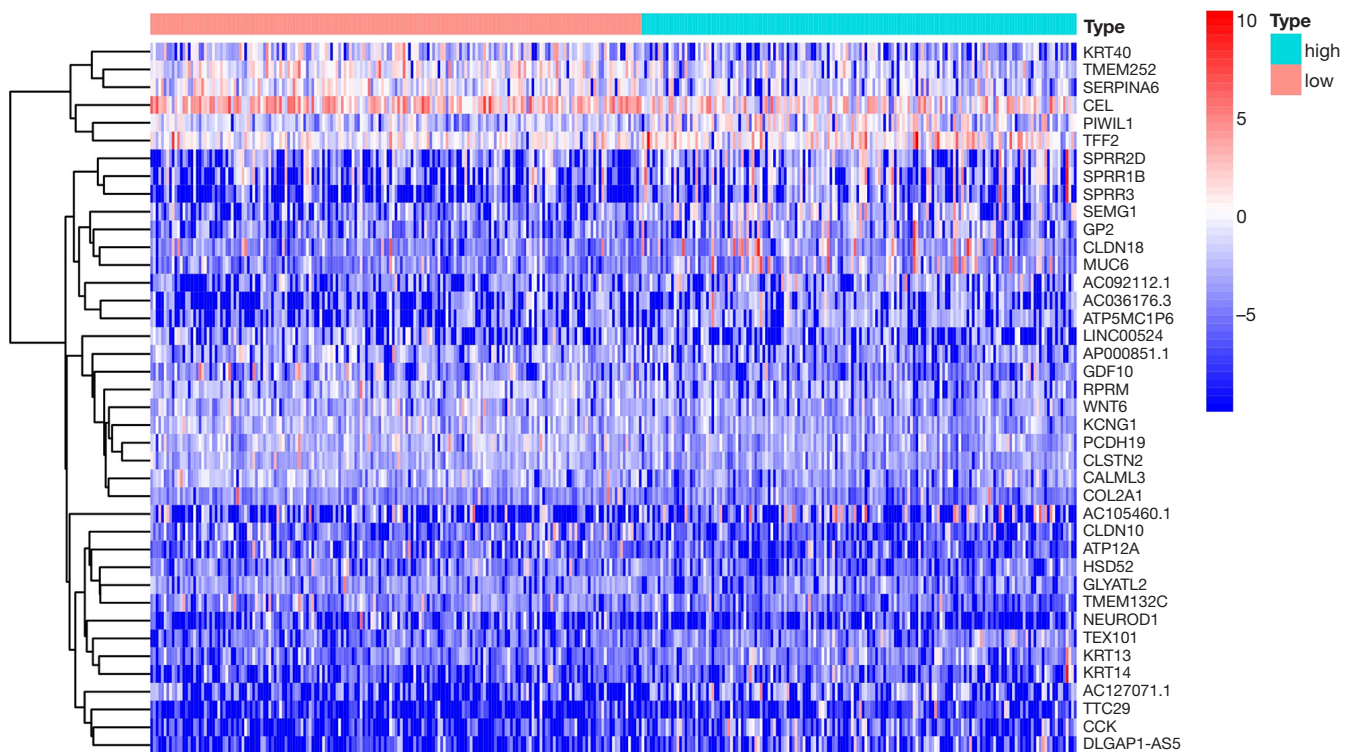
2250

Chen et al. TCGA analysis of TMB and its clinical significance in CC

Figure 3 Hierarchical clustering heatmap of differentially-expressed genes (DEGs) between the high and low TMB groups. Red: The higher expressed genes; Green: The lower expressed genes; White: genes with the same expression level.

II-specific, endopeptidase inhibitor activity, G protein-coupled amine receptor activity, activated transcription factor binding, serine-type endopeptidase inhibitor activity, peptidoglycan muralytic activity, tropomyosin binding, adrenergic receptor activity, catecholamine binding, and ionotropic glutamate receptor binding. KEGG enrichment analysis results (*Figure 4B*) indicated that the DEGs were mainly involved in neuroactive ligand-receptor interaction, the Cyclic adenosine monophosphate (cAMP) signaling pathway, the calcium signaling pathway, as well as pantothenate and coenzyme A (CoA) biosynthesis.

Association between TMB and the tumor immune microenvironment

For the combined analysis of the TMB and immune infiltration, CIBER-SORT was used to calculate the immune cell infiltration in each sample; the ratio of 22 immune cells was obtained from all samples, and the result is displayed as a bar graph (*Figure 5*). *Figure 6* displays the violin chart comparing the difference in immune cell infiltration of the high and low TMB values between the

two groups. The results showed that the top 3 in the low TMB group were M0 macrophages, resting CD4 memory T cells, and M2 macrophages, while the top 3 in the high TMB group were M0 macrophages, resting CD4 memory T cells, and CD8 T cells. The difference in the abundance of various white blood cell subtypes between the high and low TMB groups showed that the immune cells in the high TMB group that were significantly increased were CD8 T cells (P=0.008), activated CD4 memory T cells (P=0.019), M1 macrophages (P=0.002), follicular helper T cells (P=0.034), activated NK cells (P=0.017), while the M0 macrophages were significantly reduced (P=0.025). The difference in tumor-infiltrating immune cells between the two groups suggested that they might have important potential clinical implications.

Construction of the prognostic model based on TMB immune gene

In order to jointly analyze the TMB and immune infiltration, we took the intersection of the differential genes between the high and low TMB subgroups and
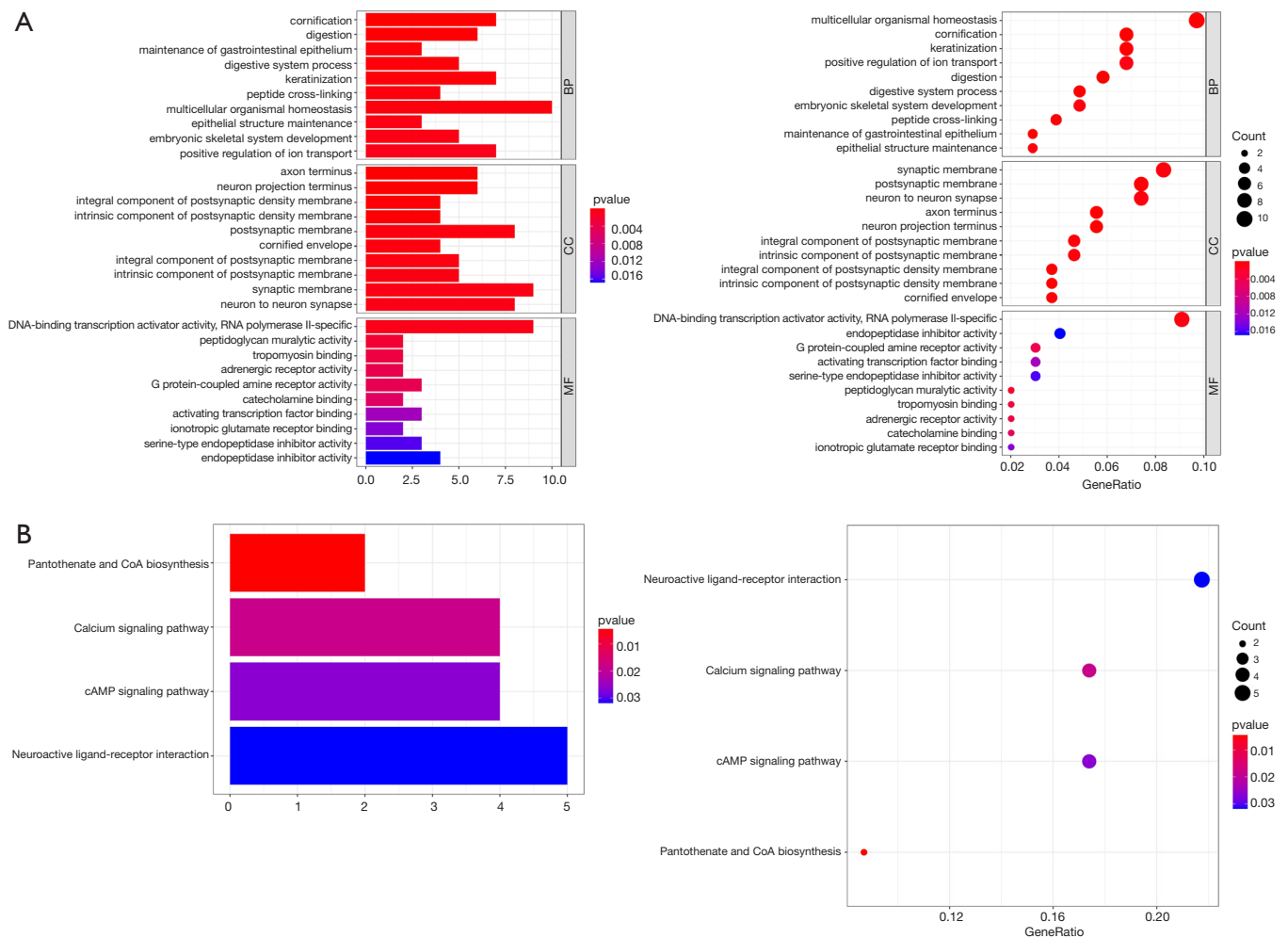
Figure 4 Functional enrichment analysis of differentially-expressed genes (DEGs). (A) Functional analysis of the top 10 enriched biological processes (BPs), cellular component (CC), and molecular function (MF) of GO analysis. (B) KEGG enrichment diseases analysis.

the 1,781 immune genes downloaded from the ImmPort database, and obtained 30 differentially-expressed immune genes (or immune-associated genes, IAGs) (*Figure 7*). We then performed univariate and multivariate Cox regression analysis on the 30 immune genes identified, and two pIAGs (P<0.05) were identified: SCT (risk factor code: 0.038791) and GUCA2A (risk factor code: –0.010505). Next, we established a TMB-based immune prognosis model for colon cancer patients according to the formula: Risk value = (0.038791 × SCT expression) + (–0.010505 × GUCA2A), and calculated the patient's risk value.

We then divided the patients into low- or high-risk groups according to the median value of risk. Kaplan-Meier survival analysis (*Figure 8A*) showed that high-risk scores were significantly associated with poor survival outcomes

(P=0.01). We used the R package to draw the survival and ROC curves, and found that the overall survival of low-risk patients was better than that of the high-risk. We performed ROC curve verification (*Figure 8B*), and the area under the curve (AUC) was 0.064, indicating that the model has good prediction accuracy. This demonstrated that our model has a particular sensitivity and specificity for predicting the prognosis of colon cancer patients.

In addition, we performed survival analysis of these two immune model genes (*Figure 8C,8D*), and the results showed that the SCT and GUCA2A gene expression levels were significantly correlated with patient survival; SCT expression was negatively correlated with survival, while GUCA2A expression was positively correlated with survival.
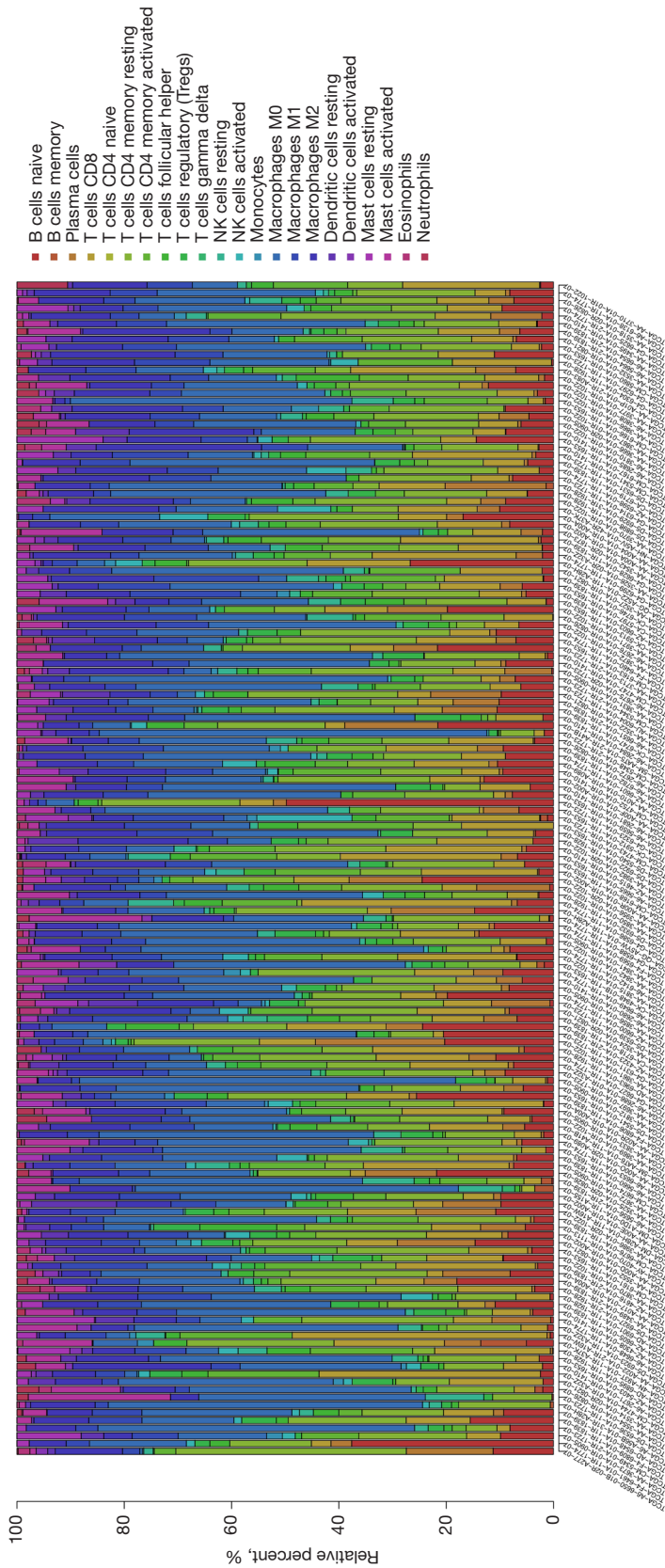
2252

Chen et al. TCGA analysis of TMB and its clinical significance in CC

Figure 5 The average proportion of different tumor-infiltrating immune cells between the low and high TMB groups. Red: high-TMB; Green: low-TMB.

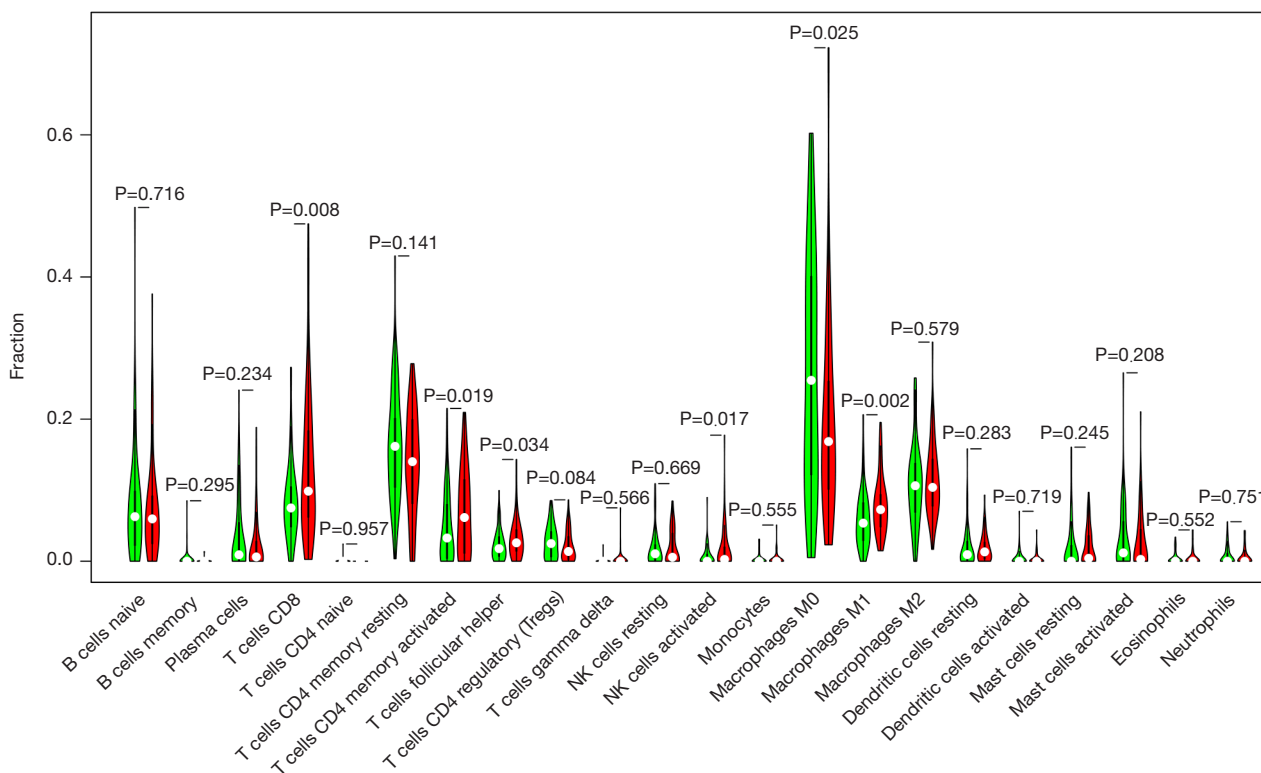**Figure 6** Differential analysis of tumor-infiltrating immune cells between the high and low TMB groups. Red: high-TMB; Green: low-TMB.
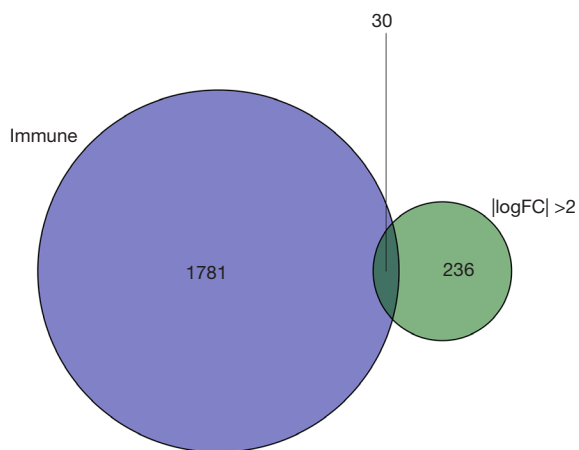

Figure 7 Thirty differentially expressed immune genes obtained from 1,781 immune genes.

Correlation analysis between prognosis-related immune gene CNV and immune cells

In order to more intuitively analyze the correlation between immune genes and immune cell infiltration richness,

we drew two pictures based on the data identified in the TIMER database (*Figure 9A,9B*). As shown in *Figure 9A*, a copy number variation (CNV) in the SCT gene signified that the CD8[+] T cell content was different from the other groups (P<0.05). As shown in *Figure 9B*, mutations that occurred during GUCA2A gene replication indicated that the content of CD8[+] T cells and B cells was significantly different from the other groups (P<0.001). Also, when one copy number variation was missing or obtained, the content of CD8[+] T cells was much more significantly different from the other groups (P<0.001). In other words, the gene expression levels of SCT and GUCA2A were significantly related to the content of immune cells in patients; with abnormal gene expression, the content of immune cells also increased.

Discussion

Tumor occurrence is a gradual process and is the result of gene mutations caused by the interaction of multiple immune cells in the tumor microenvironment (29). Among
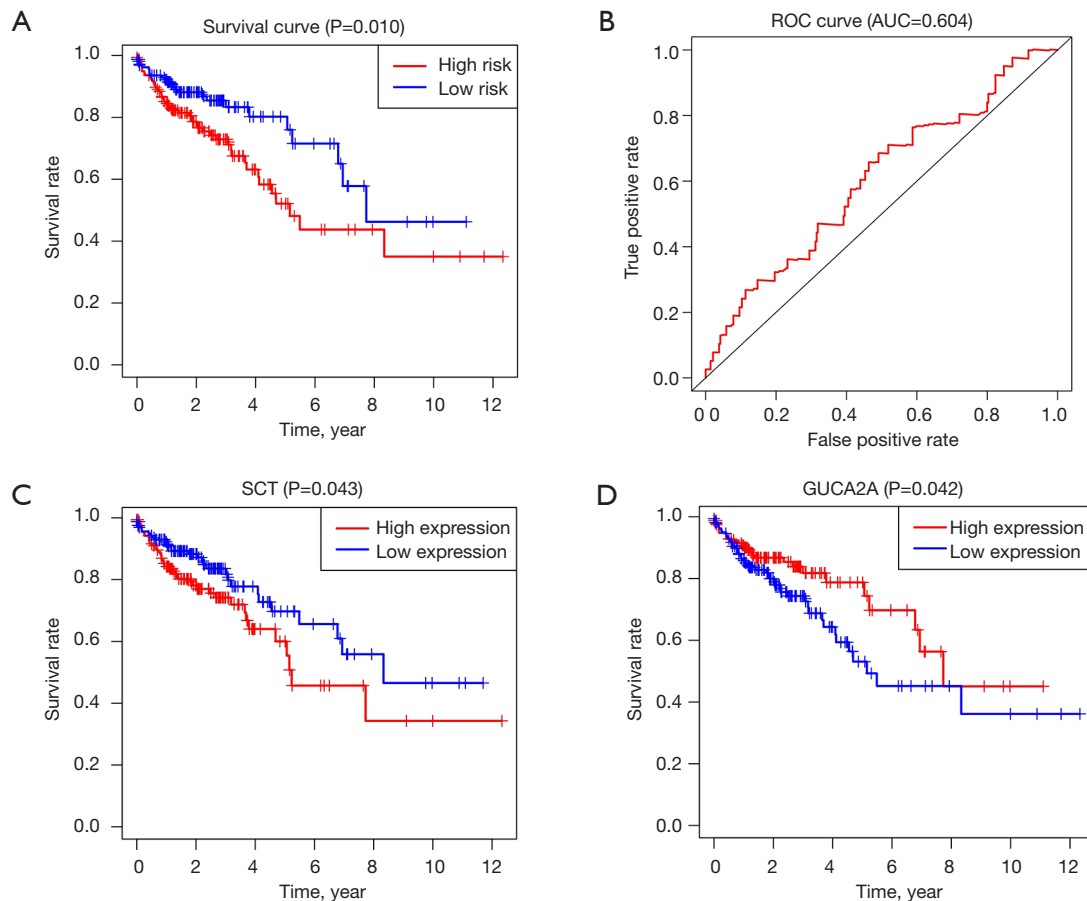
J Gastrointest Oncol 2021;12(5):2244-2259 | https://dx.doi.org/10.21037/jgo-21-661

Figure 8 Survival analysis and receiver operating characteristic curve (ROC) curve (A) survival analysis of low risk and high risk TMB groups. (B) ROC curve verifies the accuracy of prediction. (C) Survival analysis for secretin (SCT) in colon cancer. (D) Survival analysis for guanylate cyclase activator 2A (GUCA2A) in colon cancer.

these mutations, missense mutations are the most common type (30). Understanding the relationship between TMB and highly immunogenic tumors may have a good auxiliary effect on the evaluation of tumor immunotherapy effects and provide a theoretical basis for the prognosis of patients in clinical practice. Clinical trials (31) have shown that appropriate enhancement of the tumor immune response can lead to a long-term clinical response and patient benefit. Studies (32) have also shown that monoclonal antibodies that block the interaction between PD-1 and PD-L1 by binding to ligands or receptors have demonstrated significant clinical efficacy in patients with colorectal cancer. However, there is currently no reliable biomarker to evaluate the effect of immunotherapy for colon cancer.

In this study, we analyzed the mutations in colon cancer samples. These results showed that missense mutations, SNP, and C>T mutations are the most common mutations

in colon cancer. Former studies (33-35) have also confirmed that missense mutations are the most common type of colon cancer, and SNPs located in regulatory elements can affect the expression of coding genes through remote regulation, leading to cancer. We found that the four genes with the highest mutation frequency were APC, TP53, TTN, and KRAS, which is consistent with the results of previous studies (36). Some researchers (37) have confirmed that APC mutations cause intestinal tumors by activating Wnt signaling in epithelial cells. In addition, previous studies (38) have shown that TP53 mutation affects its protein structure, folding, and stability, and also affects its DNA binding ability and physiological activity, thus promoting tumor formation. Moreover, some scholars have suggested that TTN and TP53 double mutations may take part in tumorigenesis by influencing downstream pathways through the participation of other co-expressed genes on
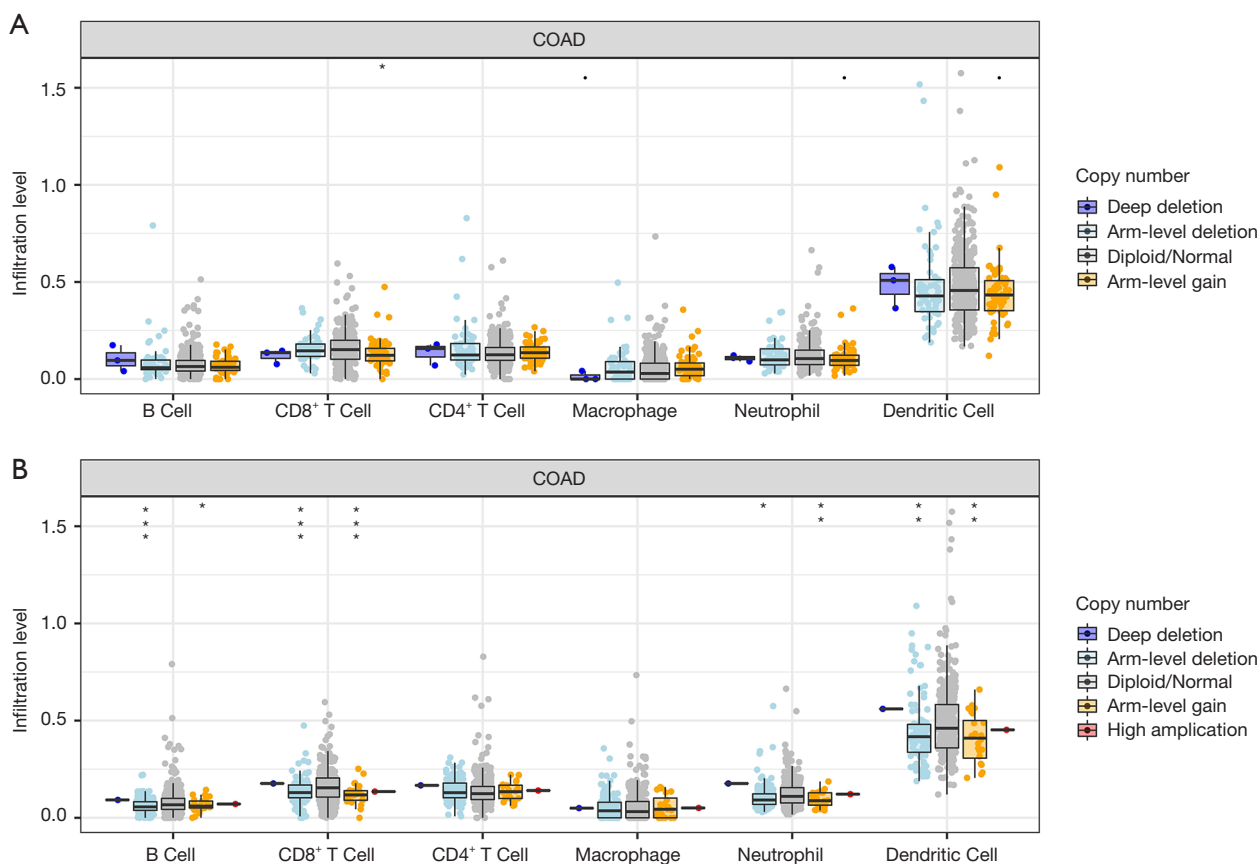
Figure 9 The correlation with immune cell infiltration richness (A) The correlation between secretin (SCT) and immune cell infiltration richness. (B) The correlation between guanylate cyclase activator 2A (GUCA2A) and immune cell infiltration richness. (*, P<0.05; **, P<0.01; ***, P<0.001).

the signaling network (39,40). Similarly, when KRAS is mutated, the downstream signaling pathway (mitogen-activated protein kinase, MAPK) is activated, leading to cell proliferation and tumor progression (41,42).

We also analyzed the clinical significance of the TMB in colon cancer. These results showed that the TMB was negatively correlated with prognosis that the higher TMB the patients are, the poorer prognosis. Lots of studies (43) have confirmed that in patients with colorectal cancer who have undergone radical surgery supplemented with fluoropyrimidine and oxaliplatin chemotherapy, those with higher a TMB are associated with better prognosis. Researchers have also confirmed that the response to PD-1/PD-L1 blockers is more significant in patients with a high TMB (44). Some scholars have also predicted the number of mutation-related neoantigens, which seems to be proportional to the actual number of mutation-related neoantigens, and tumors with a large number of

actual mutation-related neoantigens are more likely to stimulate the immune system to respond to the tumor (45). Therefore, it is predicted that patients with a high TMB may have a poor prognosis without drug intervention; however, the prognosis may improve after drug intervention. Our analysis also demonstrated that the TMB of young colon cancer patients is higher than that of elderly patients, the TMB value of early colon cancer patients is higher, and the TMB value of patients without lymph nodes and distant metastasis is also higher, which was quite different from the results of previous studies (46). This may require extensive experimental data to verify. TMB is a determinant of immune-mediated patient survival, and identifies candidate immune regulatory mechanisms related to immune cold tumors in breast cancer (47). Therefore, TMB may be considered to be an independent predictor that measure response to various cancer immunotherapy including colon cancer (44,48,49).

In addition, we analyzed the potential biological functions of DEGs related to TMB. The function of TMB-related DEGs was mainly related to neuroactive ligand-receptor interaction, the cAMP signaling pathway, the calcium signaling pathway, as well as pantothenate and CoA biosynthesis. The calcium signaling pathway can activate certain RAS guanine nucleotide exchange factor (GEF) or RAS GTP as activating protein (GAP) to promote or inhibit the activation of RAS and RAS-dependent signals, leading to tumorigenesis (50). The cAMP signaling pathway interacts with other intracellular signaling pathways, including cytokines and the Ras-Raf-Erk pathway, which also leads to tumorigenesis (51). The extracellular matrix is maintained in a highly dynamic balance, and runs through various cellular biological behaviors. Imbalanced dynamics of the extracellular matrix will lead to cancer and other diseases (52). Therefore, it is predicted that the extracellular matrix interacts with various signal pathways to cause the occurrence of tumors.

The status of the immune microenvironment was reflected by analyzing the relationship between TMB, related immune genes CNV, and infiltrating immune cells. In this study, compared with the low TMB group, the CD8 T cells and M1 macrophages were abundant in the high TMB group, and the CD4 memory T cells were activated, while M0 macrophages were significantly reduced. At the same time, the gene expression levels of SCT and GUCA2A were significantly related to the content of immune cells in the patients. We observed that with abnormal gene expression, the content of immune cells also increased. These two results suggested that patients with a higher infiltration level of CD8 T cells, M1 macrophages, and resting CD4 memory T cells, as well as lower levels of M0 macrophages will have a better immunotherapy effect. These findings evidenced that CD4 T cells, CD8 T cells, and macrophages may be the major participants of antitumor immunity in high TMB colon cancers patients. This is consistent with the research results of Deng *et al.* (19), and Zhou *et al.* (20).

Finally, we established a TMB immune gene prognosis model to predict the prognosis of colon cancer patients. This model showed the expression levels of SCT and GUCA2A were significantly correlated with patient survival; SCT expression was negatively correlated with survival, while GUCA2A expression was positively correlated. This demonstrates that by reducing the expression level of SCT and increasing the expression level of GUCA2A, the survival time of colon cancer patients can be prolonged.

In this study, we found that colon cancer is affected by genes, signaling pathways, immune cells and tumor microenvironment. Missense mutation, SNP and C>T mutation were the most common mutations in colon cancer. The functions of TMB related DEGs mainly involve Neuroactive ligand-receptor interaction and Calcium signaling Pathway, cAMP Signaling Pathway, Pantothenate and CoA biosynthesis, etc. In addition, CD8T cells, macrophages M1 and CD4+ T cells were more abundant in tumor microenvironment with high TMB, while the M0 abundance of macrophages was significantly reduced, and high TMB colon cancer has a better immunotherapy effect.

Although the results of this study were obtained through precise statistical processing, there are certain limitations in this study that should be taken into consideration when interpreting the results. For example, this study is an analysis of the overall patients in the database, and does not include a separate patient cohort study. Therefore, in order to further improve the results of this study, more large-scale *in vivo* or *in vitro* experiments are needed.

Conclusions

In conclusion, this study conducted a systematically comprehensive analysis of the prediction and clinical significance of TMB in colon cancer in identification, monitoring, prognosis of colon cancer, and providing reference information for immunotherapy.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the REMARK reporting checklist. Available at https://dx.doi.org/10.21037/jgo-21-661

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://dx.doi.org/10.21037/jgo-21-661). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Institutional ethical approval and informed consent were waived.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.
2. Benson AB, Venook AP, Al-Hawary MM, et al. Colon Cancer, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw 2021;19:329-59.
3. Mody K, Baldeo C, Bekaii-Saab T. Antiangiogenic Therapy in Colorectal Cancer. Cancer J 2018;24:165-70.
4. Peck J. Presentation, diagnosis and treatment of colorectal cancer. Nurs Stand 2018. [Epub ahead of print].
5. Stein A, Folprecht G. Immunotherapy of Colon Cancer. Oncol Res Treat 2018;41:282-5.
6. Xiao J, Li W, Huang Y, et al. A next-generation sequencing-based strategy combining microsatellite instability and tumor mutation burden for comprehensive molecular diagnosis of advanced colorectal cancer. BMC Cancer 2021;21:282.
7. Schrock AB, Ouyang C, Sandhu J, et al. Tumor mutational burden is predictive of response to immune checkpoint inhibitors in MSI-high metastatic colorectal cancer. Ann Oncol 2019;30:1096-103.
8. Picard E, Verschoor CP, Ma GW, et al. Relationships Between Immune Landscapes, Genetic Subtypes and Responses to Immunotherapy in Colorectal Cancer. Front Immunol 2020;11:369.
9. Genetic Tumor Markers Collaboration Group, Tumor Biomarker Committee, China Anti-cancer Association, Molecular Pathology Collaboration Group, Tumor Pathology Committee, China Anti-Cancer Association. Chinese expert consensus on tumor mutational burden testing and clinical application. Chinese Journal of Oncology Prevention and Treatment 2020;12:485-94.
10. Hatakeyama K, Nagashima T, Urakami K, et al. Tumor mutational burden analysis of 2,000 Japanese cancer genomes using whole exome and targeted gene panel sequencing. Biomed Res 2018;39:159-67.
11. Bao X, Zhang H, Wu W, et al. Analysis of the molecular nature associated with microsatellite status in colon cancer identifies clinical implications for immunotherapy. J Immunother Cancer 2020;8:e001437.
12. Llosa NJ, Cruise M, Tam A, et al. The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints. Cancer Discov 2015;5:43-51.
13. Phillips DH. Mutational spectra and mutational signatures: Insights into cancer aetiology and mechanisms of DNA damage and repair. DNA Repair (Amst) 2018;71:6-11.
14. Carbone DP, Reck M, Paz-Ares L, et al. First-Line Nivolumab in Stage IV or Recurrent Non-Small-Cell Lung Cancer. N Engl J Med 2017;376:2415-26.
15. Klebanov N, Artomov M, Goggins WB, et al. Burden of unique and low prevalence somatic mutations correlates with cancer survival. Sci Rep 2019;9:4848.
16. Tang H, Qiao J, Fu YX. Immunotherapy and tumor microenvironment. Cancer Lett 2016;370:85-90.
17. Mao Y, Feng Q, Zheng P, et al. Low tumor purity is associated with poor prognosis, heavy mutation burden, and intense immune phenotype in colon cancer. Cancer Manag Res 2018;10:3569-77.
18. Cen S, Liu K, Zheng Y, et al. BRAF Mutation as a Potential Therapeutic Target for Checkpoint Inhibitors: A Comprehensive Analysis of Immune Microenvironment in BRAF Mutated Colon Cancer. Front Cell Dev Biol 2021;9:705060.
19. Deng D, Luo X, Zhang S, et al. Immune cell infiltration-associated signature in colon cancer and its prognostic implications. Aging (Albany NY) 2021;13:19696-709.
20. Zhou Z, Xie X, Wang X, et al. Correlations Between Tumor Mutation Burden and Immunocyte Infiltration and Their Prognostic Value in Colon Cancer. Front Genet 2021;12:623424.
21. Mayakonda A, Koeffler HP. Maftools: efficient analysis, visualization and summarization of MAF files from large-scale cohort based cancer studies. BioRxiv 2016. DOI:

2258

Chen et al. TCGA analysis of TMB and its clinical significance in CC

10.1101/052662.

22. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.

23. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015;43:D447-52.

24. Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 2017;45:D362-8.

25. Bauer-Mehren A. Integration of genomic information with biological networks using Cytoscape. Methods Mol Biol 2013;1021:37-61.

26. Chen B, Khodadoust MS, Liu CL, et al. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. Methods Mol Biol 2018;1711:243-59.

27. Amezquita RA, Lun ATL, Becht E, et al. Orchestrating single-cell analysis with Bioconductor. Nat Methods 2020;17:137-45.

28. Li T, Fan J, Wang B, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. Cancer Res 2017;77:e108-10.

29. Hsu YC, Chang YH, Chang GC, et al. Tumor mutation burden and recurrent tumors in hereditary lung cancer. Cancer Med 2019;8:2179-87.

30. Brown SD, Warren RL, Gibb EA, et al. Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. Genome Res 2014;24:743-50.

31. Angell HK, Bruni D, Barrett JC, et al. The Immunoscore: Colon Cancer and Beyond. Clin Cancer Res 2020;26:332-9.

32. Gordon SR, Maute RL, Dulken BW, et al. PD-1 expression by tumour-associated macrophages inhibits phagocytosis and tumour immunity. Nature 2017;545:495-9.

33. Korphaisarn K, Pongpaibul A, Roothumnong E, et al. High Frequency of KRAS Codon 146 and FBXW7 Mutations in Thai Patients with Stage II-III Colon Cancer. Asian Pac J Cancer Prev 2019;20:2319-26.

34. Cong Z, Li Q, Yang Y, et al. The SNP of rs6854845 suppresses transcription via the DNA looping structure alteration of super-enhancer in colon cells. Biochem Biophys Res Commun 2019;514:734-41.

35. Zhang X, Cowper-Sal lari R, Bailey SD, et al. Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.3 prostate cancer risk locus. Genome Res 2012;22:1437-46.

36. Wolff RK, Hoffman MD, Wolff EC, et al. Mutation analysis of adenomas and carcinomas of the colon: Early and late drivers. Genes Chromosomes Cancer 2018;57:366-76.

37. Nakayama M, Oshima M. Mutant p53 in colon cancer. J Mol Cell Biol 2019;11:267-76.

38. Barbosa K, Li S, Adams PD, et al. The role of TP53 in acute myeloid leukemia: Challenges and opportunities. Genes Chromosomes Cancer 2019;58:875-88.

39. Wang X, Duanmu J, Fu X, et al. Analyzing and validating the prognostic value and mechanism of colon cancer immune microenvironment. J Transl Med 2020;18:324.

40. Cheng X, Yin H, Fu J, et al. Aggregate analysis based on TCGA: TTN missense mutation correlates with favorable prognosis in lung squamous cell carcinoma. J Cancer Res Clin Oncol 2019;145:1027-35.

41. Fu X, Wang X, Duanmu J, et al. KRAS mutations are negatively correlated with immunity in colon cancer. Aging (Albany NY) 2020;13:750-68.

42. Arrington AK, Heinrich EL, Lee W, et al. Prognostic and predictive roles of KRAS mutation in colorectal cancer. Int J Mol Sci 2012;13:12153-68.

43. Lee DW, Han SW, Bae JM, et al. Tumor Mutation Burden and Prognosis in Patients with Colorectal Cancer Treated with Adjuvant Fluoropyrimidine and Oxaliplatin. Clin Cancer Res 2019;25:6141-7.

44. Goodman AM, Kato S, Bazhenova L, et al. Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. Mol Cancer Ther 2017;16:2598-608.

45. Le DT, Uram JN, Wang H, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N Engl J Med 2015;372:2509-20.

46. Chalmers ZR, Connelly CF, Fabrizio D, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome Med 2017;9:34.

47. Thomas A, Routh ED, Pullikuth A, et al. Tumor mutational burden is a determinant of immune-mediated survival in breast cancer. Oncoimmunology 2018;7:e1490854.

48. Segal NH, Parsons DW, Peggs KS, et al. Epitope landscape in breast and colorectal cancer. Cancer Res 2008;68:889-92.

49. Devarakonda S, Rotolo F, Tsao MS, et al. Tumor Mutation Burden as a Biomarker in Resected Non-Small-Cell Lung Cancer. J Clin Oncol 2018;36:2995-3006.

50. Pierro C, Cook SJ, Foets TC, et al. Oncogenic

K-Ras suppresses IP$_3$-dependent Ca^{2+} release through remodelling of the isoform composition of IP$_3$Rs and ER luminal Ca^{2+} levels in colorectal cancer cell lines. J Cell Sci 2014;127:1607-19.

51. Sapio L, Gallo M, Illiano M, et al. The Natural cAMP Elevating Compound Forskolin in Cancer Therapy: Is It

Time? J Cell Physiol 2017;232:922-7.

52. Walker C, Mojares E, Del Río Hernández A. Role of Extracellular Matrix in Development and Cancer Progression. Int J Mol Sci 2018;19:3028.

(English Language Editor: A. Kassem)