

RESEARCH

Open Access



Tumor microenvironment characterization in colorectal cancer to identify prognostic and immunotherapy genes signature

Xian-wen Guo^{1†}, Rong-e Lei^{2†}, Qing-nan Zhou¹, Guo Zhang¹, Bang-li Hu^{3*} and Yun-xiao Liang^{1*}

Abstract

Background The tumor microenvironment (TME) plays a crucial role in tumorigenesis, progression, and therapeutic response in many cancers. This study aimed to comprehensively investigate the role of TME in colorectal cancer (CRC) by generating a TMEscore based on gene expression.

Methods The TME patterns of CRC datasets were investigated, and the TMEscores were calculated. An unsupervised clustering method was used to divide samples into clusters. The associations between TMEscores and clinical features, prognosis, immune score, gene mutations, and immune checkpoint inhibitors were analyzed. A TME signature was constructed using the TMEscore-related genes. The results were validated using external and clinical cohorts.

Results The TME pattern landscape for CRC was examined using 960 samples, and then the TMEscore pattern of CRC datasets was evaluated. Two TMEscore clusters were identified, and the high TMEscore cluster was associated with early-stage CRC and better prognosis in patients with CRC when compared with the low TMEscore clusters. The high TMEscore cluster indicated elevated tumor cell scores and tumor gene mutation burden, and decreased tumor purity, when compared with the low TMEscore cluster. Patients with high TMEscore were more likely to respond to immune checkpoint therapy than those with low TMEscore. A TME signature was constructed using the TMEscore-related genes superimposing the results of two machine learning methods (LASSO and XGBoost algorithms), and a TMEscore-related four-gene signature was established, which had a high predictive value for discriminating patients from different TMEscore clusters. The prognostic value of the TMEscore was validated in two independent cohorts, and the expression of TME signature genes was verified in four external cohorts and clinical samples.

Conclusion Our study provides a comprehensive description of TME characteristics in CRC and demonstrates that the TMEscore is a reliable prognostic biomarker and predictive indicator for patients with CRC undergoing immunotherapy.

Keywords Tumor microenvironment, Colorectal cancer, Prognosis, Immunotherapy

[†]Xian-wen Guo and Rong-e Lei contributed equally to the study.

*Correspondence:

Bang-li Hu
hubangli@gxmu.edu.cn
Yun-xiao Liang
petterry_wy@163.com

¹Department of Gastroenterology, The People's Hospital of Guangxi Zhuang Autonomous Region, No.6 Tao-Yuan Road, Nanning 530021, Guangxi, China

²Department of Gastroenterology, First Affiliated Hospital of Guangxi Medical University, Nanning 530021, Guangxi, China

³Department of Research, Guangxi Medical University Cancer Hospital, No.71 Hedi Road, Nanning 530021, Guangxi, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Colorectal cancer (CRC) is the third most common cancer and the fourth leading cause of cancer-related deaths, according to a recent report by GLOBOCAN 2020 [1]. Numerous advancements have been made over the past few decades with regard to CRC treatment, and the five-year survival rate has increased significantly [2]. Recently, immunotherapy, such as immune-checkpoint blockade programmed cell death 1 (PD-1) and programmed cell death 1 ligand 1 (PD-L1), has further improved the prognosis of patients with CRC, especially those with late-stage cancer [3, 4]. However, owing to the heterogeneity of CRC, only a subset of patients is suitable for immunotherapy [5, 6]; therefore, it is essential to investigate the pathogenesis of CRC and screen patients that may be appropriate for immunotherapy.

Tumor components are very complex; besides tumor cells, there are numerous immune, stromal, and inflammatory cells in the tumor microenvironment (TME) that modulate tumor development and other biological functions [7]. Immune cells in the TME are associated with the pathogenesis, development, therapeutic response, and prognosis of various cancers. Immune cell abundance varies greatly in tumor tissues, and different types of immune cells display distinct functions that exert pro- and anti-tumor effects at different stages [8]. A dynamic balance exists between the pro- and anti-tumor effects within the TME, profoundly influencing the prognosis of patients with cancer [9].

Recently, the abundance of immune and other cells in the TME has been assessed quantitatively using computational methods. This has been termed the TMEscore, which has great value in identification of patients suitable for precision therapy or immunotherapy in several cancers, such as gastric [10], bladder [11] and prostate cancer [12]. However, the comprehensive landscape of the TME cells in CRC has not yet been characterized, and the utility of the TMEscore for CRC remains to be elucidated. Therefore, the aim of the present study was to estimate the TME patterns of CRC using a meta-cohort with larger tumor samples. Additionally, we intended to systematically analyze the TMEscore with genomic characteristics and CRC clinical out, which could provide key biomarkers for predicting responsiveness to immunotherapy, thereby further improving precision immunotherapy for CRC.

Materials and methods

CRC dataset acquisition and preprocessing

CRC datasets were systematically searched from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>), and only datasets with sufficient clinical information were collected. In total, five datasets were selected (GSE103479, GSE29621,

GSE72970, GSE39582, and GSE41258) with survival information regarding samples from patients with CRC. Four datasets (GSE20916, GSE21815, GSE3629, and GSE89287) with tumor tissues and the corresponding control tissues were also downloaded. We downloaded The Cancer Genome Atlas (TCGA)-COADREAD dataset (615 samples) from the TCGA database (GDC hub: <https://gdc.xenahubs.net>), which included 449 colon cancer and 166 rectal cancer samples, and we also downloaded the corresponding clinical information. The raw data from the GEO datasets were preprocessed using the RMA algorithm for background adjustment before analysis. The raw data from the TCGA dataset were transformed into transcripts per kilobase million values prior to analysis. The “sva” R package was used to process the RNA expression data from both the GEO and TCGA cohorts with the aim of reducing the batch effect.

Quantification of tumor infiltrating cells, immune scores, and tumor purity

The CIBERSORT algorithm [13] was used to quantify the proportions of immune infiltrating cells in CRC samples, which allowed for sensitive and specific discrimination of 22 human immune cell phenotypes by calculating gene expression. The immune score, stromal score, Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE) score, and tumor purity of CRC samples were calculated using the ESTIMATE algorithm [14]. This algorithm provides researchers with scores for tumor purity, the level of stromal cells present, and the infiltration level of immune cells in tumor tissues based on gene expression.

Consensus clustering analysis

Unsupervised clustering methods for dataset analysis (TME pattern and TMEscore) were used to classify samples and determine the optimal number of clusters for further analysis. The procedures were undertaken using the ConsensusClusterPlus package [15] in R software, and were repeated 1,000 times, to ensure classification stability. The tool implements the consensus clustering method and provides visualizations, including item tracking, item-consensus, and cluster-consensus plots. The results are in the form of cumulative distribution function curves, which are used to determine the number of clusters, with k values ranging from 2 to 6. Furthermore, a consensus heatmap of the clusters was determined and visualized.

Differentially expressed gene analysis

For the GEO datasets, the differentially expressed genes (DEGs) among different groups were screened using the “limma” package, which adopts an empirical Bayesian approach to estimate gene-expression changes using

moderated *t* tests. For the TCGA dataset, DEGs were screened using the “edgeR” package, which is specifically used to screen DEGs in TCGA datasets. The DEGs were determined based on the $|\log_{2}FC| > 0.5$ and $P < 0.01$ criteria.

Functional and pathway enrichment analysis

Gene annotation enrichment analysis was performed on DEGs from TMEscore-related genes using the clusterProfiler R package [16]. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) terms of the genes were identified with a cutoff of $P < 0.01$ and a false discovery rate < 0.05 . Gene set variation analysis (GSVA) was used to screen significantly enriched pathways between the two clusters using the Molecular Signatures Database (MSigDB) version 7.4 [17]. Significant pathway terms were set at a P -value < 0.05 .

TMEscore calculation and establishment of the TME signature

The “TMEscore” package [10, 18] was used to calculate the TMEscore and establish TME clusters for the gene datasets, which estimates the TMEscore based on gene expression. This provides functionality for calculating the TMEscore using principal component analysis or *z*-score methods. The results generated four types of data: TMEscoreA, TMEscoreB, TMEscore, and TME_binary. TME_binary divides the samples into high or low clusters based on the TMEscore value.

Reverse transcription-polymerase chain reaction (RT-PCR) assay using CRC clinical samples

Gene expression in clinical CRC samples was tested using RT-PCR. Forty tumor tissues and their corresponding adjacent tissues were collected from our hospital between February 2020 and August 2021. Total RNA from the tissues was isolated using TRIzol reagent (Invitrogen, Waltham, MA, USA), according to the manufacturer’s instructions. The primers used for RT-PCR are as follows: CXCL10 FORWARD: CTC TCT CTA GAA CTG TAC GCT G, REVERSE: ATT CAG ACA TCT CTT CTC ACC C; IDO1 FORWARD: CTG CCT GAT CTC ATA GAG TCT G, REVERSE: TTG TGG TCT GTG AGA TGA TCA A; MAB21L2 FORWARD: CTA TCT CTC AGC GCG TAA GAT C, REVERSE: CAT CTT GAC CAC ATC CCG ATA G; LZTS2 FORWARD: TCC TCC TCC TCC TCT TCC TCC TC, REVERSE: GCA GGC TGG ACA GTG AGT TTCG. RT-PCR was performed using the SYBR[®] Premix Ex Taq kit (Takara, Dalian, China). The relative expression of each gene was calculated using the $2^{-\Delta\Delta CT}$ method. This study was approved by the hospital’s ethics committee.

Statistical analysis

Statistical significance for normally distributed variables was tested using the unpaired Student’s *t* test when comparing two groups; otherwise the Mann–Whitney U test was used. The Kruskal–Wallis test and one-way Analysis of Variance were used as nonparametric and parametric methods, respectively, for the comparison of more than two groups. The Chi-square test was used to compare categorical variables. Kaplan–Meier and log-rank (Mantel–Cox) tests were used to determine the statistical significance of patient survival. Two machine learning methods, LASSO regression and XGBoost, were used to screen significant genes using the “glmnet” and “xgboost” packages, respectively. Receiver Operating Characteristic (ROC) curves were used to calculate the area under the curve (AUC) and evaluate the predicted value of the signature. All statistical analyses were performed using R language. P -values were two-sided, and $P < 0.05$ were considered to indicate statistical significance.

Results

Construction and analysis of TME landscape patterns in CRC

The TME pattern of CRC was evaluated by calculating the immune cell infiltrating fraction from four CRC datasets (TCGA-CORDREAD, GSE103479, GSE29621, and GSE72970) with a total of 960 CRC tissue samples, using the CIBERSORT algorithm. After removing the batch effect for the datasets, each immune cell infiltrating fraction was combined into a large meta-cohort to construct the landscape of the TME pattern. Subsequently, an unsupervised hierarchical clustering method using the ConsensusClusterPlus package was used to determine the optimal cluster number for the TME pattern, and two robust clusters (clusters I and II) of the meta-cohort were identified (Fig. 1A and B). Next, the correlation of each type of immune cell was analyzed, and the TME cell network was visualized as a landscape of CRC tumor–immune cell interactions and CRC cell lineages (Fig. 1C and F). Based on the two clusters of TME cell patterns, we found that the patients with CRC in cluster II had better prognosis than those in cluster I (log-rank test, $P = 0.001$; HR = 2.404; Fig. 1D). We also found that the infiltrating fractions of plasma, T, CD8, and resting dendritic cells were considerably higher in cluster I than in cluster II. In addition, the infiltrating fraction of Macrophages M0, T cells, follicular helper T cells, and resting CD4 memory cells were significantly higher in cluster II than in cluster I (Fig. 1E).

Identification of TME clusters in TCGA dataset

To characterize the clinical significance of the TME pattern between the two clusters, we focused on the TCGA-CORDREAD dataset, which contains comprehensive

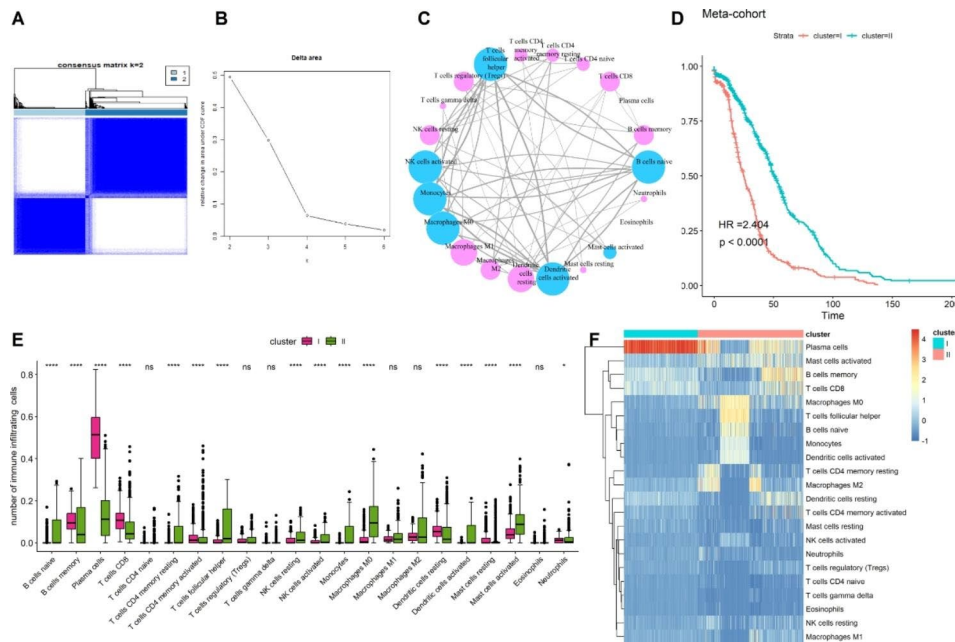


Fig. 1 Landscape of TME patterns in four CRC cohorts (TCGA-CORDREAD, GSE103479, GSE29621 and GSE72970) with 960 samples. **(A)** Heatmap of the consensus clustering matrix for $k=2$; **(B)** CDF curves of the consensus score from $k=2$ to 6. **(C)** Cellular interaction of the TME cell types. Cell cluster I: blue; cell cluster II: pink. The size of each cell represents survival of each TME cells type, the smaller log-rank test P value, the larger size of node. The lines connecting TME cells represent cellular interactions. The thickness of the line represents the strength of correlation estimated by Spearman correlation analysis. The more correlation, the more thickness of the lines. **(D)** Kaplan–Meier curves for overall survival (OS) of 960 CRC patients from four CRC cohorts between two TME clusters. **(E)** Comparison of 22 immune cells infiltrating fraction between TME cluster I and cluster II. **(F)** Unsupervised clustering of 22 immune cells in TME for 960 patients from four CRC cohorts

clinical information for most patients. Here, we calculated the immune cell infiltrating fraction for the TCGA dataset, and the TMEscore of the immune cells was calculated using the “TMEscore” package. Subsequently, unsupervised hierarchical cluster analysis was performed to screen the optimal clusters for the TMEscore. Two distinct clusters were identified for the TMEscore (TME-cluster I and TME-cluster II). TME-cluster II was associated with a higher immune score than TME-cluster I (Kruskal–Wallis, $P<0.001$; Fig. 2A). Survival analysis revealed that patients in TME-cluster II exhibited better survival rates than those in TME-cluster I (log-rank test, $P=0.001$; HR=1.838; Fig. 2B). The clinical characteristics analysis results indicated that patients in TME-cluster II exhibited increased MSI-H status, T1+T2 stage, N0 stage, M0 stage, and I+II tumor stage than those in TME-cluster I (Fig. 2C–F); however, there was no significant difference in sex or tumor location. The results indicate that patients in TME-cluster II displayed high TMEscores and were associated with early-stage CRC. Subsequently, we explored the pathways involving the two TME clusters using the GSVA method, and observed immunodeficiency, antigen processing and presentation, and natural killer cell-mediated cytotoxicity pathways underlying the TME clusters (Fig. 2G).

Construction of the TME signature and functional annotation

To further determine the underlying biological characteristics of TME clusters, we screened DEGs between TME-cluster I and TME-cluster II of the TCGA dataset using the “edgeR” package. Here, 1953 DEGs were identified based on the ($|\logFC| > 0.5$ and P value < 0.01) criteria, with 1208 upregulated and 745 downregulated genes (Fig. 3A). The DEGs were then incorporated into the unsupervised hierarchical cluster analysis, and two clusters, TME-H and TME-L, were identified. The samples in the two signatures were significantly consistent with the clustering results of TME-clusters I and II (χ^2 tests, $P<0.05$). Survival analysis showed that patients with TME-H signature displayed a better prognosis than those with TME-L signature (log-rank test, $P=0.001$; Fig. 3B). The results also revealed that the TME-H signature exhibited an elevated immune score (Fig. 3C).

GO enrichment and KEGG analyses of the signature genes were conducted using the R package clusterProfiler. We observed that upregulated genes were involved in the biological processes of leukocyte-mediated immunity, T-cell activation, and lymphocyte-mediated immunity. The pathways are involved in cytokine–cytokine receptor interaction and viral protein interaction with cytokines and cytokine receptors. The downregulated genes were involved in the phospholipase C-activating

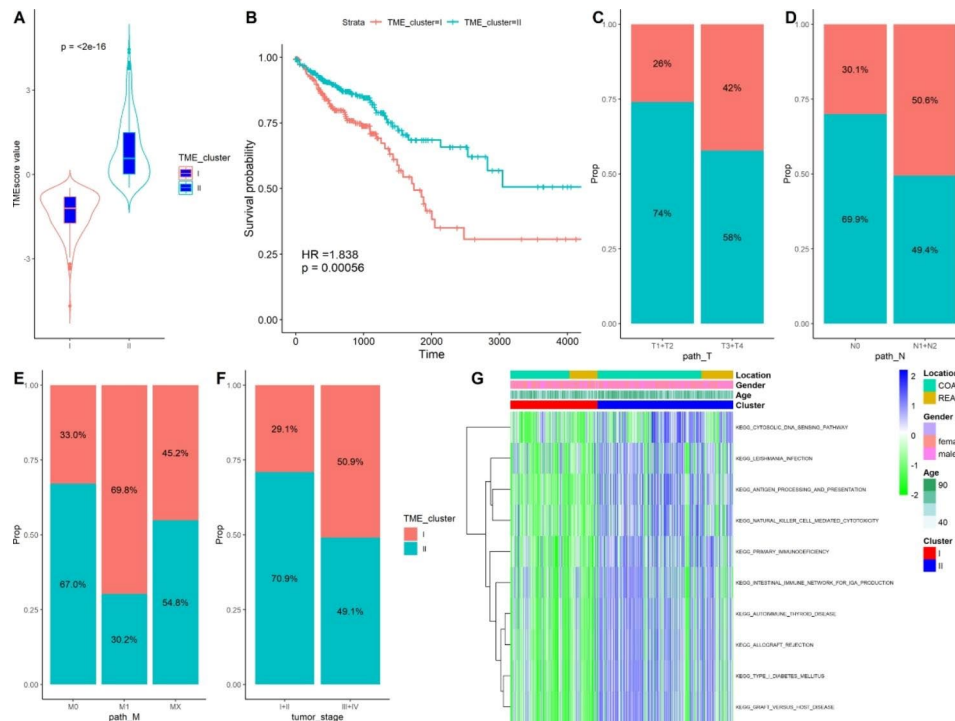


Fig. 2 Identification of TME clusters in TCGA dataset. (A) Comparison of TME scores between TME-cluster I and TME-cluster II. (B) Kaplan–Meier curves for overall survival of CRC patients between two TME clusters. Comparison of T stage (C), N stage (D), M stage (E), Tumor stage (F) between TME-cluster I and TME-cluster II. (G) GSVA method revealed the pathways that the TME-clusters involved

G protein-coupled receptor signaling pathway and positive regulation of cytosolic calcium ion concentration involved in the phospholipase C-activating G protein-coupled signaling pathway. The calcium and Wnt signaling pathways were found to be involved (Fig. 3D–G). Subsequently, we tested the robustness of the TME signature using DEGs in the GSE39582 and GSE41258 datasets. The results verified that the TME signatures constructed by DEGs in the GSE39582 and GSE41258 datasets were consistent to the TCGA-COADREAD dataset, and patients with CRC with the TME-H signature displayed a better prognosis (Fig. 3H and I). Finally, the correlations between TME signature scores and TME patterns were determined, and the TME-H signature was associated with elevated immune scores in the TCGA-COADREAD, GSE39582, and GSE41258 datasets (Fig. 3J and K), suggesting the reliability of the TME signature.

Association of TME signature with the immune environment and immune checkpoint inhibitors

To determine the association of the TME signature with immune cell scores and tumor purity in CRC, we used the ESTIMATE algorithm to quantify the composition of CRC. The results indicated that TME-H had higher immune, stromal, and ESTIMATE scores, and lower tumor purity, than the TME-L signature ($P < 0.05$, Fig. 4A), demonstrating high abundances of immune and

stromal cells and low tumor purity in patients with the TME-H signature. Next, we examined the association between the TME signature and immune cell infiltration fraction in CRC. As Fig. 4B illustrates, follicular helper T cells and M1 macrophages were more abundant in the TME-H cluster, whereas other immune cell infiltrating fractions showed little difference between the TME-H and TME-L signatures ($P > 0.05$). Finally, we examined the associations between the TME signatures with known immune checkpoint inhibitors, including CTLA4, PD1 (PDCD1), CD80, CD86, PD-L1, PD-L2 (PDCD1LG2), and (CD274). As shown in Fig. 4C, the expression levels of the six immune checkpoint inhibitors were increased in the TME-H when compared with those in the TME-L signature. The findings demonstrate that the TME signature is critically associated with the immune environment and may be used to screen patients with CRC who are suitable for immunotherapy.

Association of TME signature with gene mutations and tumor mutation burden

Gene mutation and Tumor Mutation Burden (TMB) are closely related to the pathogenesis of various cancers, including CRC [19, 20]. Therefore, we explored the association of the TME signature with CRC gene mutations and TMB in the TCGA dataset using the “maftools” package. We first examined the whole gene mutation

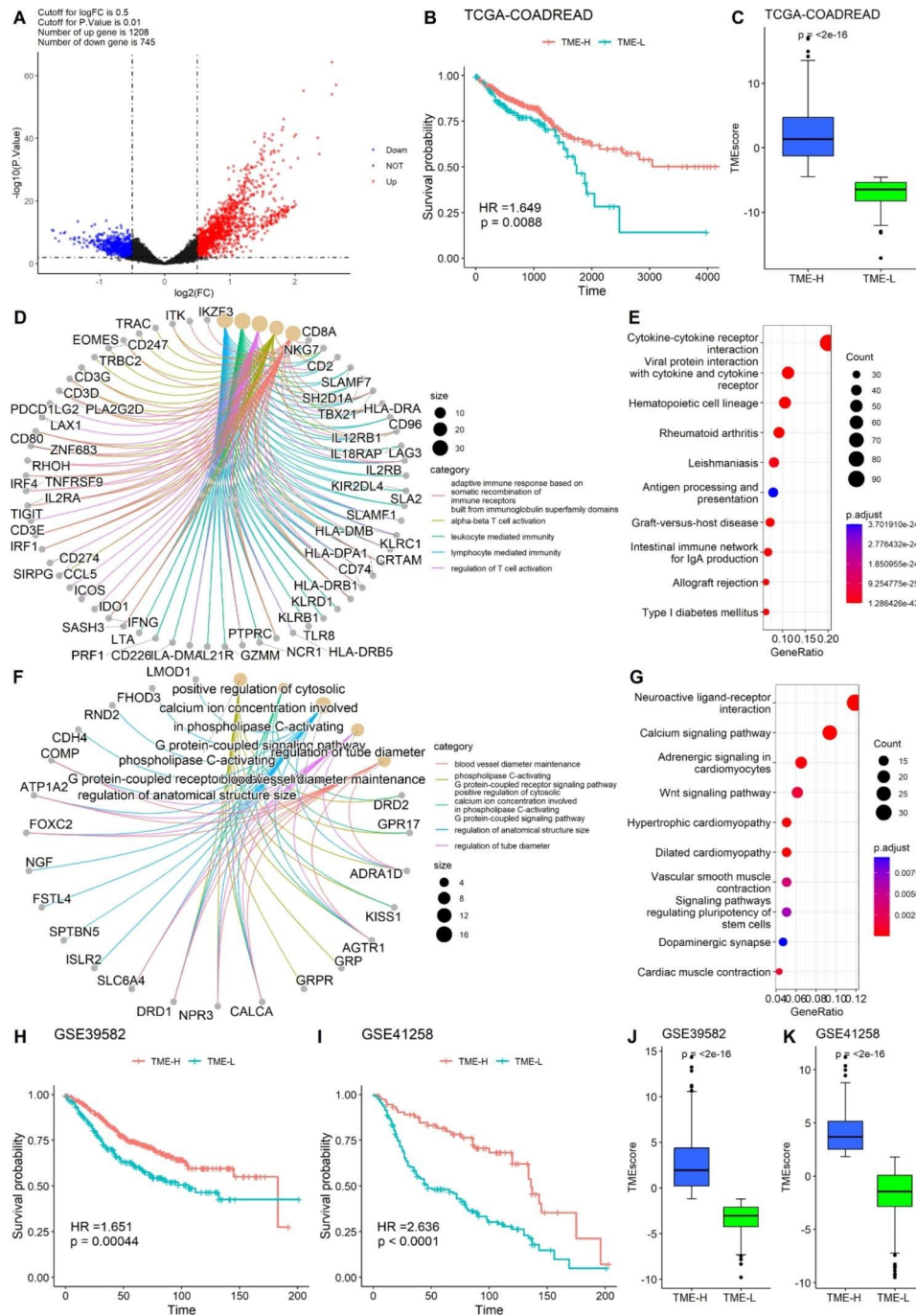


Fig. 3 Construction of the TME signature and functional annotation. **(A)** Volcano plot for the differential expressed genes (DEGs) between TME-cluster I and TME-cluster II. **(B)** Kaplan–Meier curves for overall survival of CRC patients between TME-H and TME-L. **(C)** Comparison of immune scores between TME-H and TME-L signature. **(D–E)** GO and KEGG analysis for the upregulated genes. **(F–G)** GO and KEGG analysis for the downregulated genes. Validated survival impact of TME signature using the DEGs in **(H)** GSE39582 dataset and **(I)** GSE41258 dataset. Validated immune scores of TME signatures using the DEGs in **(J)** GSE39582 dataset and **(K)** GSE41258 dataset

between the TME-H and TME-L signatures and found that the whole mutated frequency was higher in TME-H than in TME-L (99.71% vs. 96.97%). However, the top 10 mutated genes of the two signatures were similar, namely *APC*, *TP53*, *TTN*, *KRAS*, *PIK3CA*, *PTEN*, *ATM*, *SYNE1*, *MUC16*, and *SMAD4*, demonstrating that the two

signatures harbor similar top mutated genes (Fig. 5A and B). Next, we screened the genes with significant differential mutation frequency between the two signatures and revealed that *ITPR3*, *DSCAML1*, *EP400*, *RNF43*, *TRPM3*, *RYR1*, *TEP1*, *GRIK2*, *MCF2*, and *DNAH5* showed higher mutation frequencies in the TME-H signature than in

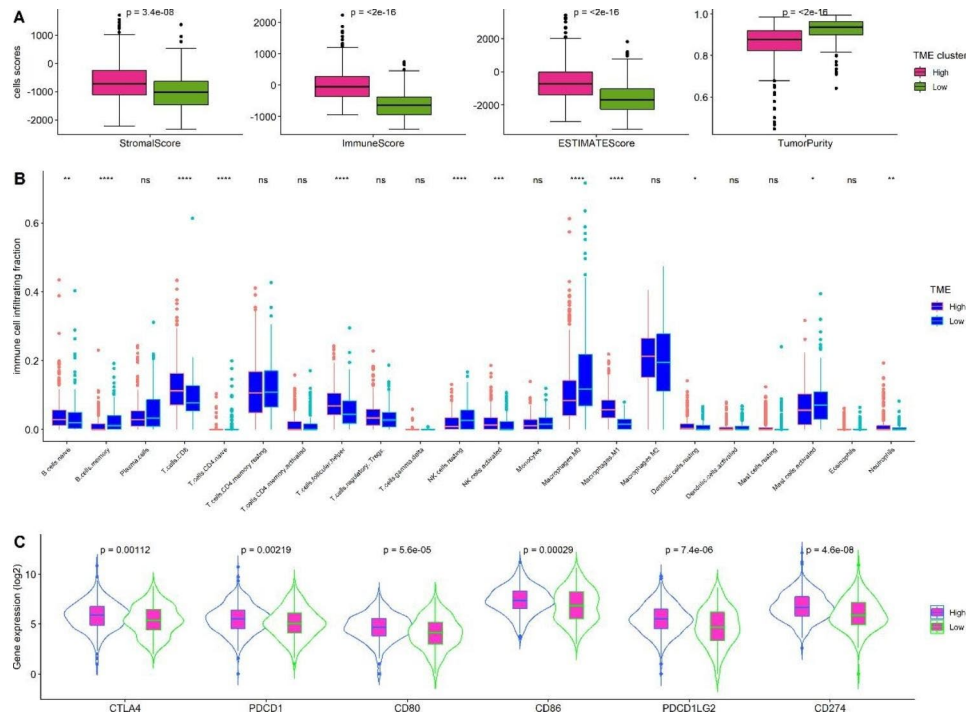


Fig. 4 TME signature was associated with (A) immune cells scores and tumor purity; (B) tumor immune cells fraction; (C) immune checkpoint inhibitors

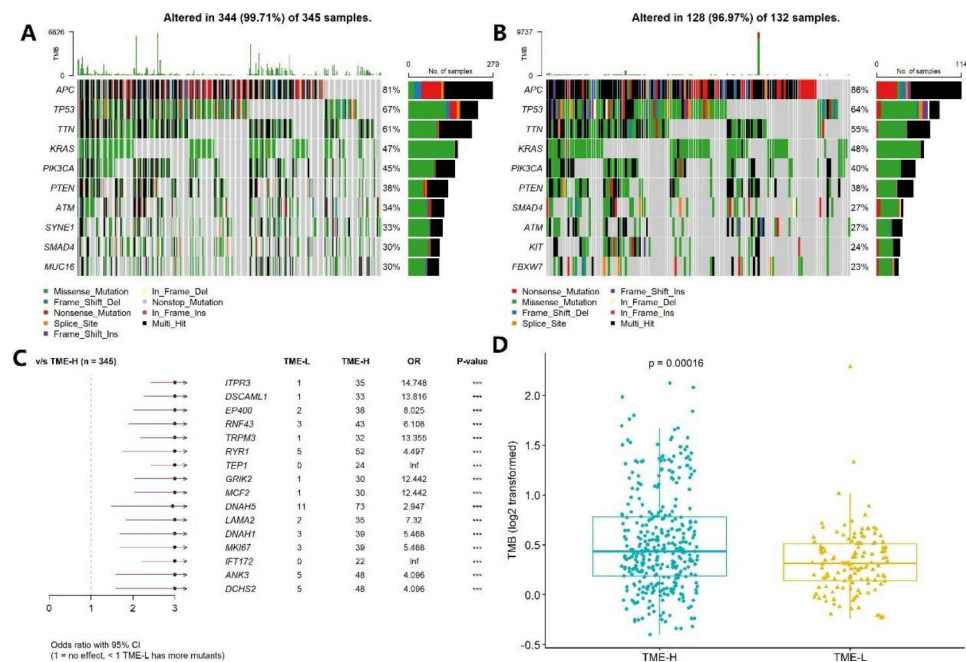


Fig. 5 TME signature was associated with gene mutation and TMB. (A-B) Top 10 mutated genes in the TME-H and TME-L signature. (C) Comparison of gene mutation between TME-H and TME-L clusters. (D) Comparison of tumor mutation burden between TME-H and TME-L signature

the TME-L signature. However, we did not observe a significant difference in the top 10 mutated genes between the two signatures (Fig. 5C). In addition, we observed that the TMB value was much higher in the TME-H signature than in the TME-L signature ($P < 0.001$; Fig. 5D). The findings indicate that the TME-H signature harbored

higher gene mutations and TMB than the TME-L signature.

Establishment of a predictive model for the TME signature
To establish a predictive model for the TME signature, the CRC samples were randomly divided into two sets

(training and test sets) at a 7:3 ratio as previous studies did [21, 22], with 430 and 185 samples in each set, respectively. Two machine learning algorithms (LASSO and XGBoost) were employed, by incorporating TME-score DEGs into the model, and the key genes related to the TME signature were screened. Five-fold cross validation was performed to search the best hyper-parameters, and the “nround” was 25 in the present analysis, and the min logloss and min logloss index were 2.54 and 43, respectively. The LASSO and XGBoost algorithms identified 59 and 6 genes associated with the TME signature, respectively. The ROC curve showed that both models constructed using LASSO and XGBoost algorithms had good predictive value in differentiating TME-H from the TME-L signature (both AUC values > 0.90; Fig. 6A and B). We then superimposed the genes from the two algorithms and observed that four genes (*CXCL10*, *LZTS2*, *IDO1*, and *MAB21L2*) were identified by the two machine learning algorithms (Fig. 6C). Subsequently, we applied a multivariate logistic regression analysis to construct a predictive model using the four genes. As shown in Fig. 6D, the optimal cutoff value of the model for discriminating the two clusters was 0.552, which means that patients with scores < 0.552 were grouped in the TME-H signature; otherwise, they were grouped in the TME-L signature. The ROC curve demonstrated that the predictive value of the model reached 0.982, with a sensitivity and specificity of 90.1% and 95.4%, respectively, for the entire dataset. Finally, we tested the expression of the four

genes in the training and test sets between the TME-H and TME-L signatures. The results showed that *CXCL10* and *IDO1* expression levels increased, while *LZTS2* levels decreased in the TME-H signature compared with that in TME-L signature; however, there was no significant difference between the two signatures regarding *MAB21L2* expression (Fig. 6E and F).

Validation of TME signature-related genes in independent cohorts and CRC clinical samples

The expression of four TME-related genes (*CXCL10*, *LZTS2*, *IDO1*, and *MAB21L2*) that predict the TME signature of CRC was validated in four GEO datasets: GSE20916 (145 samples), GSE21815 (141 samples), GSE3629 (121 samples), and GSE89287 (71 samples), which contained normal control and tumor samples. Furthermore, 40 clinical CRC tissues and corresponding adjacent tumor tissues were collected, and the expression of the four genes was determined using RT-PCR assay. As shown in Fig. 7, the four GEO datasets and our clinical samples confirmed that the expression levels of *CXCL10*, *IDO1*, and *MAB21L2* were significantly higher in tumor tissues than in control tissues ($P < 0.05$). Moreover, the expression level of *LZTS2* decreased in tumor tissues from the GSE20916 and GSE3629 datasets, but appeared increased in samples from GSE21815 and GSE89287 datasets and our clinical samples ($P < 0.05$). Collectively, the results indicated that the levels of the three TME signature genes (*CXCL10*, *IDO1*, and *MAB21L2*) were all

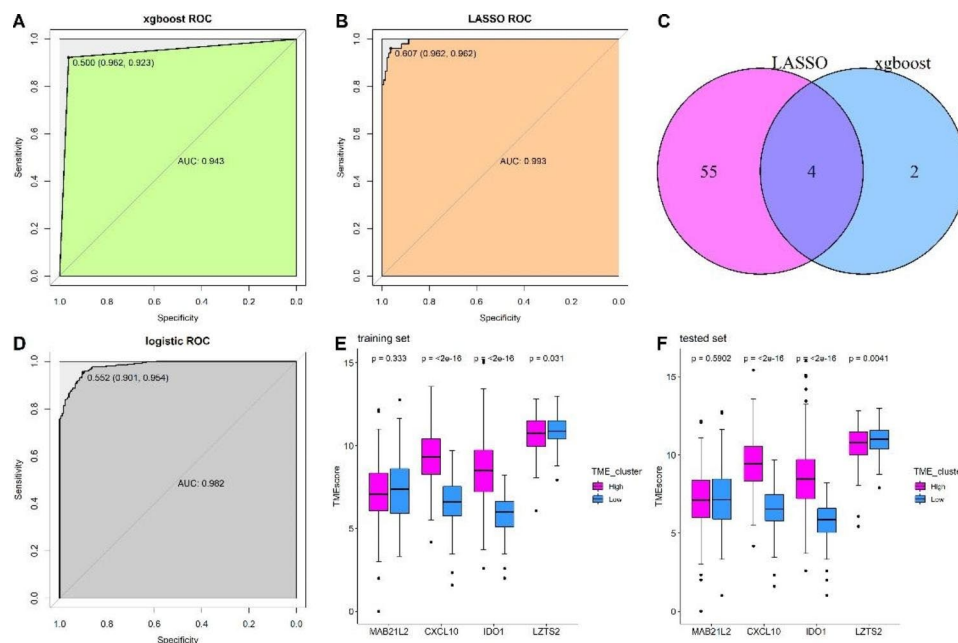


Fig. 6 Establishment of predictive model for TME signature. **(A)** xgboost algorithms identified a model predicting the TME signature; **(B)** LASSO algorithms identified a model predicting the TME signature. **(C)** Venn diagram identified overlapped genes between xgboost and LASSO algorithms. **(D)** Multivariate logistic regression model revealed the predictive value for the TME signature. **(E)** Expression of the four overlap genes in training dataset between TME-H and TME-L signature. **(F)** Expression of the four overlap genes in tested dataset between TME-H and TME-L signature

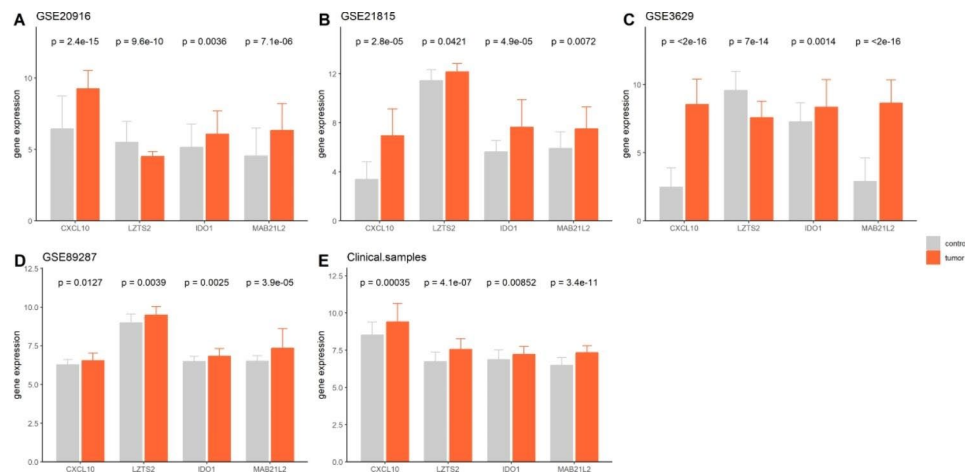


Fig. 7 Validation of four genes in datasets and clinical samples. Comparison of CXCL10, LZTS2, IDO1 and MAB21L2 between CRC and control tissues in (A) GSE20916 dataset; (B) GSE21815 dataset; (C) GSE3629 dataset; (D) GSE89287 dataset; (E) clinical tissues

increased in CRC tissues when compared with the control tissues, but the expression of *LZTS2* in CRC tissues remained need to further validate.

Discussion

Increasing evidence indicates that the TME profoundly influences tumorigenesis, progression, and therapeutic response; however, the utility of the TME in clinical settings is hampered by the inaccurate combination and uncertain interactions of the TME [23, 24]. Over the past few decades, flow cytometry or immunohistochemistry assays have mostly been used to determine various cell types in clinical tissues; however, these methods are subject to cumbersome procedures and low feedback efficiency, which limits the application of the TME in the clinical setting. With the development of sequencing techniques and bioinformatics, many deconvolution methods are available for predicting cell types and proportion information in various tissues [25, 26]. TMEscore, a parameter designed to quantitatively evaluate the composition of the TME using computational algorithms, was recently introduced as a biomarker for predicting the prognosis of patients with cancers and to guide more effective immunotherapy strategies. However, a comprehensive characterization of the TME in CRC has not yet been performed. The present study demonstrated that TMEscore and its signature are promising tools [10, 18].

Herein, we performed a systematic analysis to explore the value of the TMEscore with the clinical outcome in patients with CRC based on several gene datasets. We first depicted the TME cell landscape using a large meta-cohort, divided the cells into two clusters, and observed that the clusters were associated with the survival of patients with CRC. We then focused on the TCGA dataset and calculated its TMEscore, and the cluster analysis revealed its significant prognostic value

and clinical significance, with a high TMEscore indicating better patient prognosis. Next, we developed a TMEscore signature using TMEscore-related genes that were divided into TME-H and TME-L signatures. We then determined the prognostic value and association of the TMEscore signature with tumor immune scores, tumor purity, immunotherapy response, and genomic mutations, which exhibit high prognostic value and are closely associated with the above indicators. Moreover, we screened genes that represented the TMEscore signature using two machine learning algorithms and validated the predictive value of the genes using independent cohorts. Overall, the comprehensive estimation of the cellular, molecular, and genetic factors associated with TME characterization has shed light on the mechanism of tumor response to immunotherapy and may guide the identification of patients who are suitable for immunotherapy.

Consistent with previous studies regarding other types of cancers, such as gastric and prostate cancer [10, 12], patients with CRC with a high TMEscore showed a better prognosis than those with a low TMEscore. We noted that the DEGs between the TMEscore clusters were involved in immune cell processes, such as T cell activation and differentiation, and that the TMEscore clusters were related to the immune cell score and tumor purity. High TMEscores indicated more immune cells and fewer tumor cells, indicating that the TMEscore clusters were critical biomarkers for immune cells and may serve as a predictor of immunotherapy response. Currently, anti-tumor immunity through immune checkpoint inhibitors, specifically anti PD-1/PD-L1 interaction, is a new line of treatment for patients with CRC, especially for those at later stages of cancer [27]. However, only a minority of patients exhibit response to the immune checkpoint blockade, and studies have found that PD-1/PD-L1 expression and mutation load are not efficient

biomarkers for predicting the benefits of immune checkpoint blockade [28, 29]. Therefore, the establishment of reliable predictive biomarkers for checkpoint immunotherapy is important for maximizing therapeutic benefits [30, 31]. In the present study, the top 10 mutated genes (*APC*, *TP53*, *KRAS*, *PIK3CA*, *ATM*, *PTEN*, *TTN*, *SYNE1*, *MUC16* and *SMAD4*), which have well documented in the pathogenesis of CRC [32–37]. Regarding the ten genes with a significant differential mutation frequency between the two signatures, mutation of *EP400*[38], *RNF43*[39], *TRPM3*[40], *TEP1*[41], *MCF2*[42] have been reported to associated with pathogenesis or prognosis of CRC, and the expression of *DSCAML1*[43], *RYR1*[44] and *DNAH5*[45] was related to CRC development. However, although *ITPR3*[46] and *GRIK2*[47] were reported to associated with cancers development, but the no evidence reported their roles in CRC. These results revealed that these TMEscore signatures also exhibited high gene mutation, and thus potentially influence the tumor microenvironment. Moreover, some genes mutation was not reported in CRC, suggesting further study is warrant to explore their role in CRC. Furthermore, we observed that the TMEscore clusters were associated with the expression of six immune checkpoint inhibitors and that genomic mutation and TMB were greatly increased in the TME-H cluster, demonstrating that patients with high TMEscores may benefit from immunotherapy.

Compared with previous studies that investigated the value of TMEscore in other cancers [10, 12, 25], our study constructed a four-gene TME signature using TMEscore-related genes (*CXCL10*, *LZTS2*, *IDO1*, and *MAB21L2*), making it easy for clinicians to assess patients who might belong to the TME-H or TME-L signature. The role of *CXCL10* and *IDO1* in CRC has been described in previous studies; for example, *CXCL10* was found to be related to immune infiltration [48], and *IDO1*(+) Paneth cells promote the immune escape of CRC [49]. *LZTS2* protein reduces the level of nuclear β -catenin in CRC cells (SW480 cells) [50], while *MAB21L2* is reduced in CRC and is associated with the Wnt pathway [51]. Therefore, our results suggest that further studies are required to explore their roles in that aspect. However, their roles in predicting immunotherapy have not yet been reported. Notably, *LZTS2* expression trend from GSE20916 and GSE3629 datasets was contrary to those from the other two datasets and our clinical samples. We speculated that the disparity is due to sample sources, so that the tissue stage, gene mutation status, and detection methods could influence gene expression. Hence, more studies are warranted to validate the gene expression by considering other factors related to gene expression.

In the present study, we first combined the TCGA dataset and other GEO datasets to construct a larger cohort, which could increase the reliable of the results

due to the larger samples. In addition, we also employed other public datasets and clinical samples to validate the results, this strategy could guarantee the robustness of the results. Our study showed that the TMEscore is a reliable prognostic biomarker and predictive indicator in patients with CRC, and the gene expression results were verified using independent cohorts and clinical samples. Nevertheless, our results still require further validation in a prospective cohort of patients with CRC undergoing immunotherapy, and their gene expression profiles should be tested to confirm the results of the TMEscore signature. In addition, the number of immune cells in the tumors was determined by computational algorithms, the exact number of which remained to be determined using flow cytometry or immunohistochemistry. Furthermore, the treatment response of CRC is affected by numerous factors, and our study only included a few, and more clinical factors should be incorporated into predictive models to further improve the accuracy of the signature.

Conclusion

The present study provides a comprehensive description of TME characteristics in CRC. It also demonstrates that the TME score is a potential reliable prognostic biomarker and predictive indicator for patients with CRC undergoing immunotherapy. Furthermore, our results provide a novel strategy for the precise treatment of patients with CRC.

Acknowledgements

We thank Dr. Jianming Zeng (University of Macau), and all the members of his bioinformatics team, biotrainee, for generously sharing their experience and codes.

Authors' contributions

Study concept and design: GXW and LYX; Collection and assembly of data: GXW and HBL; Performed the experiment: ZG, HBL and GXW; Data analysis and interpretation: ZQN, HBL, LRE and GXW; Manuscript writing and review: All authors. All authors have read and approved the manuscript in its current state.

Funding

This study was partially supported by research funding from the National Natural Science Foundation of China (No. 82060104; No. 81860417). Scientific Research Project of Guangxi Health Commission (No. Z20190717).

Data Availability

The data used to support the findings of this study are available from GEO databases (www.ncbi.nlm.nih.gov/geo/), with the access number as GSE103479, GSE29621, GSE72970, GSE39582, GSE41258, GSE20916, GSE21815, GSE3629, and GSE89287, respectively. The TCGA-COADREAD datasets was downloaded from Xena database (<https://xenabrowser.net/datapages/>). All data used in this study is available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

All the procedures were carried out in accordance with institutional guidelines. The study is approved by the Ethics Committee of People's Hospital of Guangxi Zhuang Autonomous Region, written informed consent was

obtained from each patient. All the procedures were carried out in accordance with Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 2 February 2023 / Accepted: 8 August 2023

Published online: 18 August 2023

References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and Mortality Worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–49.
- Carlsen L, Huntington KE, El-Deiry WS. Immunotherapy for Colorectal Cancer: mechanisms and predictive biomarkers. *Cancers (Basel)*. 2022;14(4):1028.
- Han Y, Liu D, Li L. PD-1/PD-L1 pathway: current researches in cancer. *Am J Cancer Res*. 2020;10(3):727–42.
- Wang Z, Aguilar EG, Luna JI, Dunai C, Khuat LT, Le CT, Mirsoian A, Minnar CM, Stoffel KM, Sturgill IR, et al. Paradoxical effects of obesity on T cell function during tumor progression and PD-1 checkpoint blockade. *Nat Med*. 2019;25(1):141–51.
- Golshani G, Zhang Y. Advances in immunotherapy for colorectal cancer: a review. *Th Adv Gastroenterol*. 2020;13:1756284820917527.
- Plundrich D, Chikhladze S, Fichtner-Feigl S, Feuerstein R, Briquez PS. Molecular Mechanisms of Tumor Immunomodulation in the Microenvironment of Colorectal Cancer. *Int J Mol Sci* 2022, 23(5).
- Chen L, Jiang X, Zhang Q, Li Q, Zhang X, Zhang M, Yu Q, Gao D. How to overcome tumor resistance to anti-PD-1/PD-L1 therapy by immunotherapy modifying the tumor microenvironment in MSS CRC. *Clin Immunol* 2022;108962.
- Zhu PF, Wang MX, Chen ZL, Yang L. Targeting the Tumor Microenvironment: A literature review of the Novel Anti-Tumor mechanism of statins. *Front Oncol*. 2021;11:761107.
- Ozga AJ, Chow MT, Luster AD. Chemokines and the immune response to cancer. *Immunity*. 2021;54(5):859–74.
- Zeng D, Li M, Zhou R, Zhang J, Sun H, Shi M, Bin J, Liao Y, Rao J, Liao W. Tumor Microenvironment characterization in gastric Cancer identifies prognostic and immunotherapeutically relevant Gene Signatures. *Cancer Immunol Res*. 2019;7(5):737–50.
- Zhang H, Song J, Dong J, Liu Z, Lin L, Wang B, Ma Q, Ma L. Tumor Microenvironment Analysis Identified Subtypes Associated with the prognosis and the Tumor response to immunotherapy in bladder Cancer. *Front Genet*. 2021;12:551605.
- Lu G, Cai W, Wang X, Huang B, Zhao Y, Shao Y, Wang D. Identifying prognostic signatures in the microenvironment of prostate cancer. *Transl Androl Urol*. 2021;10(11):4206–18.
- Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–7.
- Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612.
- Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. 2010;26(12):1572–3.
- Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *The Innovation* 2021.
- Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417–25.
- Zeng D, Wu J, Luo H, Li Y, Xiao J, Peng J, Ye Z, Zhou R, Yu Y, Wang G et al. Tumor microenvironment evaluation promotes precise checkpoint immunotherapy of advanced gastric cancer. *J Immunother Cancer* 2021, 9(8).
- Schrock AB, Ouyang C, Sandhu J, Sokol E, Jin D, Ross JS, Miller VA, Lim D, Amanam I, Chao J, et al. Tumor mutational burden is predictive of response to immune checkpoint inhibitors in MSI-high metastatic colorectal cancer. *Ann Oncol*. 2019;30(7):1096–103.
- Sahin IH, Akce M, Alese O, Shaib W, Lesinski GB, El-Rayes B, Wu C. Immune checkpoint inhibitors for the treatment of MSI-H/MMR-D colorectal cancer and a perspective on resistance mechanisms. *Br J Cancer*. 2019;121(10):809–18.
- Chen Y, Li H. Prognostic and predictive models for left- and right- colorectal Cancer patients: a Bioinformatics Analysis based on ferroptosis-related genes. *Front Oncol*. 2022;12:833834.
- Wang Z, Wang Y, Yang T, Xing H, Wang Y, Gao L, Guo X, Xing B, Wang Y, Ma W. Machine learning revealed stemness features and a novel stemness-based classification with appealing implications in discriminating the prognosis, immunotherapy and temozolomide responses of 906 glioblastoma patients. *Brief Bioinform* 2021, 22(5).
- Chen F, Qi X, Qian M, Dai Y, Sun Y. Tackling the tumor microenvironment: what challenge does it pose to anticancer therapies? *Protein Cell*. 2014;5(11):816–26.
- Hu H, Chen Y, Tan S, Wu S, Huang Y, Fu S, Luo F, He J. The Research Progress of Antiangiogenic Therapy, Immune Therapy and Tumor Microenvironment. *Front Immunol*. 2022;13:802846.
- Peng L, Hayatullah G, Zhou H, Chang S, Liu L, Qiu H, Duan X, Han L. Tumor microenvironment characterization in cervical cancer identifies prognostic relevant gene signatures. *PLoS ONE*. 2021;16(4):e0249374.
- Chiu YJ, Hsieh YH, Huang YH. Improved cell composition deconvolution method of bulk gene expression profiles to quantify subsets of immune cells. *BMC Med Genomics*. 2019;12(Suppl 8):169.
- Yaghoubi N, Soltani A, Ghazvini K, Hassanian SM, Hashemy SI. PD-1/ PD-L1 blockade as a novel treatment for colorectal cancer. *Biomed Pharmacother*. 2019;110:312–8.
- Fuchs CS, Doi T, Jang RW, Muro K, Satoh T, Machado M, Sun W, Jalal SI, Shah MA, Metges JP, et al. Safety and Efficacy of Pembrolizumab Monotherapy in patients with previously treated Advanced gastric and gastroesophageal Junction Cancer: phase 2 clinical KEYNOTE-059 trial. *JAMA Oncol*. 2018;4(5):e180013.
- Roh W, Chen PL, Reuben A, Spencer CN, Prieto PA, Miller JP, Gopalakrishnan V, Wang F, Cooper ZA, Reddy SM et al. Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci Transl Med* 2017, 9(379).
- Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskova S, Berent-Maoz B, Pang J, Chmielowski B, Cherry G, et al. Genomic and transcriptomic features of response to Anti-PD-1 therapy in metastatic melanoma. *Cell*. 2016;165(1):35–44.
- Nishino M, Ramaiya NH, Hatabu H, Hodi FS. Monitoring immune-checkpoint blockade: response evaluation and biomarker development. *Nat Rev Clin Oncol*. 2017;14(11):655–68.
- Sell T, Klotz C, Fischer MM, Astaburuaga-Garcia R, Krug S, Drost J, Clevers H, Sers C, Morkel M, Bluthgen N. Oncogenic signaling is coupled to colorectal cancer cell differentiation state. *J Cell Biol* 2023, 222(6).
- Marbut VMG, Erlina L, Lalisang TJM. Genomic landscape of pathogenic mutation of APC, KRAS, TP53, PIK3CA, and MLH1 in Indonesian colorectal cancer. *PLoS ONE*. 2022;17(6):e0267090.
- Peng L, Li Y, Gu H, Xiang L, Xiong Y, Wang R, Zhou H, Wang J. Mucin 4 mutation is associated with tumor mutation burden and promotes antitumor immunity in colon cancer patients. *Aging*. 2021;13(6):9043–55.
- Ge W, Hu H, Cai W, Xu J, Hu W, Weng X, Qin X, Huang Y, Han W, Hu Y, et al. High-risk stage III colon cancer patients identified by a novel five-gene mutational signature are characterized by upregulation of IL-23A and gut bacterial translocation of the tumor microenvironment. *Int J Cancer*. 2020;146(7):2027–35.
- Karczmarski J, Goryca K, Pachlewski J, Dabrowska M, Pysiak K, Paziewska A, Kulecka M, Lenarcik M, Mroz A, Mikula M, et al. Mutation profiling of Premalignant Colorectal Neoplasia. *Gastroenterol Res Pract*. 2019;2019:2542640.
- Oh JH, Jang SJ, Kim J, Sohn I, Lee JY, Cho EJ, Chun SM, Sung CO. Spontaneous mutations in the single TTN gene represent high tumor mutation burden. *NPJ Genom Med*. 2020;5:33.
- Hoshi D, Kita E, Maru Y, Kogashi H, Nakamura Y, Tatsumi Y, Shimozato O, Nakamura K, Sudo K, Tsujimoto A, et al. Derivation of pancreatic acinar cell

- carcinoma cell line HS-1 as a patient-derived tumor organoid. *Cancer Sci.* 2023;114(3):1165–79.
39. Quintanilha JCF, Graf RP, Oxnard GR. BRAF V600E and RNF43 co-mutations predict patient outcomes with targeted therapies in real-world cases of Colorectal Cancer. *Oncologist.* 2023;28(3):e171–4.
 40. Ishaque N, Abba ML, Hauser C, Patil N, Paramasivam N, Huebschmann D, Leupold JH, Balasubramanian GP, Kleinheinz K, Toprak UH, et al. Whole genome sequencing puts forward hypotheses on metastasis evolution and therapy in colorectal cancer. *Nat Commun.* 2018;9(1):4782.
 41. Dicuonzo G, Angeletti S, Garcia-Foncillas J, Brugarolas A, Okrouzhnov Y, Santini D, Tonini G, Lorino G, De Cesaris M, Baldi A. Colorectal carcinomas and PTEN/MMAC1 gene mutations. *Clin Cancer Res.* 2001;7(12):4049–53.
 42. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, et al. The consensus coding sequences of human breast and colorectal cancers. *Science.* 2006;314(5797):268–74.
 43. Wu B, Yang J, Qin Z, Yang H, Shao J, Shang Y. Prognosis prediction of stage IV colorectal cancer patients by mRNA transcriptional profile. *Cancer Med.* 2022;11(24):4900–12.
 44. Yin H, Harrison TA, Thomas SS, Sather CL, Koehne AL, Malen RC, Reedy AM, Wurscher MA, Hsu L, Phipps AI, et al. T cell-inflamed gene expression profile is associated with favorable disease-specific survival in non-hypermutated microsatellite-stable colorectal cancer patients. *Cancer Med.* 2023;12(6):6583–93.
 45. Ward TM, Cauley CE, Stafford CE, Goldstone RN, Bordeianou LG, Kunitake H, Berger DL, Ricciardi R. Tumour genotypes account for survival differences in right- and left-sided colon cancers. *Colorectal Dis.* 2022;24(5):601–10.
 46. Wu Y, Liu Z, Tang D, Liu H, Luo S, Stinchcombe TE, Glass C, Su L, Lin L, Christiani DC, et al. Potentially functional variants of HBEGF and ITPR3 in GnRH signaling pathway genes predict survival of non-small cell lung cancer patients. *Transl Res.* 2021;233:92–103.
 47. Wu CS, Lu YJ, Li HP, Hsueh C, Lu CY, Leu YW, Liu HP, Lin KH, Hui-Ming Huang T, Chang YS. Glutamate receptor, ionotropic, kainate 2 silencing by DNA hypermethylation possesses tumor suppressor function in gastric cancer. *Int J Cancer.* 2010;126(11):2542–52.
 48. Wang P, Li X, Gao Y, Guo Q, Ning S, Zhang Y, Shang S, Wang J, Wang Y, Zhi H, et al. LnCeVar: a comprehensive database of genomic variations that disturb ceRNA network regulation. *Nucleic Acids Res.* 2020;48(D1):D111–7.
 49. Pflugler S, Svinka J, Scharf I, Crncec I, Filipits M, Charoentong P, Tschurtschenthaler M, Kenner L, Awad M, Stift J, et al. IDO1(+) paneth cells promote immune escape of colorectal cancer. *Commun Biol.* 2020;3(1):252.
 50. Thyssen G, Li TH, Lehmann L, Zhuo M, Sharma M, Sun Z. LZTS2 is a novel beta-catenin-interacting protein and regulates the nuclear export of beta-catenin. *Mol Cell Biol.* 2006;26(23):8857–67.
 51. Segditsas S, Sieber O, Deheragoda M, East P, Rowan A, Jeffery R, Nye E, Clark S, Spencer-Dene B, Stamp G, et al. Putative direct and indirect wnt targets identified through consistent gene expression changes in APC-mutant intestinal adenomas from humans and mice. *Hum Mol Genet.* 2008;17(24):3864–75.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.