Constructing a Tregs-associated signature to predict the prognosis of colorectal cancer patients
A STROBE-compliant retrospective study
Guoqiang Ping, BS*, Yichen Tian, BS*, Ziqiang Zhou, BS*,

Abstract
Colorectal cancer (CRC) ranks as the second leading cause of cancer-related mortality worldwide. Regulatory T cells (Tregs) are a key constituent of the immune cells in the tumor microenvironment (TME) and are significantly associated with patient outcomes. Our study aimed to construct a Treg-associated signature to predict the prognosis of CRC patients. The genes’ expression values and patients’ clinicopathological features were downloaded from TCGA and gene expression omnibus (GEO) databases. The single-cell RNA (scRNA) sequencing data of CRC were analyzed through the Deeply Integrated Human Single-Cell Omics database. WGCNA analysis was used to select Tregs-associated genes (TrAGs). The infiltrated levels of immune and stromal cells were accessed through the ESTIMATE algorithm. Cox regression analysis and the LASSO algorithm were implemented to construct prognostic models. Gene set enrichment analysis (GSEA) was performed to annotate enriched gene sets. Based on scRNA sequencing data, our study uncovered that more Tregs were significantly enriched in the TME of CRC. Then we identified 123 differentially expressed TrAGs which mainly participated in immune regulation. Given that CRC patients were reclassified into 2 subgroups with distinct overall survival based on 26 differentially expressed TrAGs with prognostic values, we subsequently constructed a signature for CRC. After training and validating in independent cohorts, we proved that this prognostic model can be well applied to predict the prognosis of CRC patients. Further analysis exhibited that more tumor-suppressing immune cells and higher immune checkpoint genes were enriched in CRC patients with high-risk scores. Moreover, immunohistochemistry analysis validated that the genes in the prognostic model were significantly elevated in CRC tissues. We were the first to construct a prognostic signature for CRC based on TrAGs and further revealed that the poor prognosis of patients was mainly attributed to the tumor-suppressing microenvironment and upregulated immune checkpoint genes in tumor tissues.

Abbreviations: ANTs = adjacent normal tissues, CRC = colorectal cancer, GEO = gene expression omnibus, GSEA = gene set enrichment analysis, LASSO = least absolute shrinkage and selection operator, ROC = receiver operating characteristic, scRNA = single-cell RNA, TME = tumor microenvironment, TrAGs = Tregs-associated genes, Tregs = regulatory T cells, WGCNA = weighted gene co-expression network analysis.

Keywords: colorectal cancer, GEO, prognosis, scRNA sequencing, TCGA, Tregs

1. Introduction
Colorectal cancer (CRC) ranks as the second leading cause of cancer-related death worldwide.[1-2] Despite advances in therapeutic methods, most CRC patients still suffer from unfavorable prognoses.[3,4] Therefore, novel therapeutic strategies are urgently needed to improve the clinical outcomes of CRC patients.

Tumor cells live in a complex microenvironment that consists of multiple types of cells, such as immune cells.[5] Foxp3 + regulatory T cells (Tregs) are a key constituent of the immune cells in the tumor microenvironment (TME) and are significantly associated with patients’ prognoses.[6] In normal conditions, Tregs maintain the immunologic self-tolerance and immune homeostasis by suppressing aberrant as well as excessive immune responses.[7] However, cancer cells often recruited Tregs in high frequencies to construct a suppressing TME through their antitumor immunoactivity.[8] With the development of single-cell RNA (scRNA) sequencing technology, it has been proved that immune cells are a highly heterogeneous population of cells and the communications between them are active.[9] For instance, tumor-infiltrating Tregs generally express higher levels of immune checkpoint genes, such as CTLA-4 and PD-1, to inhibit CD8+ T-cell activation.[10]
Given the close association between patients’ outcomes and Tregs, we aimed to construct a Tregs-associated signature to predict the prognosis of CRC patients.

In the present study, we uncovered that more Tregs infiltrated CRC tissues through scRNA sequencing analysis and that CRC patients can be reclassified into 2 subgroups with distinct overall survival based on Tregs-associated genes. Moreover, we established a prognostic signature for CRC and revealed that there were more tumor-suppressing immune cells and higher immune checkpoint genes in tumor tissues of CRC patients with high-risk scores.

2. Material and methods

2.1. The collection of data

The genes’ expression and patients’ clinicopathological information were collected from The Cancer Genome Atlas (TCGA, https://xenabrowser.net/datapages/) CRC cohort as well as Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) database. The “limma” R package was used to normalize genes’ expression values and calculate their fold changes. The scRNA sequencing data of CRC were analyzed in the Deeply Integrated human Single-Cell Omics database (https://www.immunesinglecell.org). The immunohistochemistry analysis of genes was detected through immunohistochemistry based on The Human Protein Atlas database (https://www.proteinatlas.org/).

2.2. WGCA analysis

The weighted gene co-expression network analysis (WGCNA) was conducted by using the R package “WGCNA.” According to Pearson’s correlation values, each gene was reclarified into a similarity matrix and was subsequently transformed into an adjacency matrix. The adjacency matrix was calculated by 
\[ a_{mn} = |c_{mn}|^\beta \]

among which the \( c_{mn} \) and \( a_{mn} \) mean Pearson’s correlation between paired genes and adjacency between paired genes, respectively. Parameter \( \beta \) was aimed to improve the correlation between different genes and when the power of \( \beta = 5 \), a topological overlap matrix was generated. To classify genes with similar expression patterns into distinct modules, a dynamic hybrid cutting method was applied by using a bottom-up algorithm whose module minimum size cutoff is 10.

2.3. The prognosis analysis of genes

The prognostic genes and clinicopathological features were selected through univariate and multivariate Cox regression analyses. The survival differences between the 2 groups were calculated through Kaplan–Meier curves with a log-rank test. The prognostic signature was constructed through the least absolute shrinkage and selection operator (LASSO) regression algorithm based on the “lasso” R package.

2.4. The function enrichment analysis of genes

The functional enrichment analysis, including Gene Ontology and Kyoto Encyclopedia of Genes and Genomes, were carried out in the DAVID database (https://david.ncifcrf.gov), and the results were visualized by the “ggplot2” R package. Gene Set Enrichment Analysis (GSEA) which is often used to determine significantly correlated gene sets was performed based on the MSigDB database (http://www.gsea-msigdb.org/gsea/msigdb/index.jsp). The number of permutations was set to 1000, and the permutation type was set to phenotype.

Figure 1. Analysis of the infiltration level of Tregs in CRC tissues through scRNA sequencing. (A) The identified immune cells in CRC tissues and ANTs. (B) The infiltration levels of various immune cells in CRC tissues and ANTs. (C) The hallmark genes of Tregs. * \( P < .05 \); ** \( P < .01 \). ANTs = adjacent normal tissues, CRC = colorectal cancer, Tregs = regulatory T cells.
2.5. Analysis of immune and stromal cells infiltration
The ESTIMATE algorithm was implemented to assess the infiltration levels of immune cells and stromal cells according to cell-specific marker genes. The infiltration analysis was performed by using the TIMER2 (http://timer.cistrome.org).

2.6. Statistical analysis
Statistical analysis was carried out using SPSS version 20.0 and GraphPad Prism Software 8.0. The student t test was used to compare the difference between the 2 groups. The Kaplan–Meier curves with a log-rank test and univariate Cox regression model were used to determine the survival difference. A P value of <.05 was considered significant.

3. Results
3.1. The upregulated infiltration level of Tregs in CRC tissues
To explore the infiltrated levels of immune cells in CRC tissues and adjacent normal tissues (ANTS), we first conducted scRNA sequencing and revealed 19 types of immune cells (Fig. 1A). Subsequently, we found more Tregs infiltrated CRC tissues compared to ANTs (Fig. 1B). As expected, the specific markers of Tregs, including FOXP3, IL2RA, TIGIT, and CD3D, were significantly upregulated in identified Tregs when compared to other cell lines (Fig. 1C).

3.2. Identification and analysis of Tregs-associated genes in CRC
The TCGA CRC cohort was applied to establish a co-expression network through the “WGCNA” analysis to identify genes associated with Tregs infiltration. First, $\beta = 4$ (scale-free $R^2 = 0.89$) was identified as the soft-thresholding power to construct the scale-free network (Fig. 2A). Then a hierarchical clustering tree was built by using dynamic hybrid cutting among which each branch represents genes with similar expression and each leaf was a single gene (Fig. 2B). Finally, 15 modules were constructed, among which the turquoise module was most correlated with Tregs ($R^2 = 0.5$, $P < .001$) (Fig. 2C). Therefore, these genes in the turquoise module were identified as Tregs-associated genes (TrAGs). Subsequently, we identified 3143 upregulated genes in CRC.
CRC (Fig. 2D) and overlapped them with TrAGs, and selected 123 significantly elevated TrAGs (Fig. 2E). Biological process analysis showed that these elevated TrAGs were mainly associated with inflammatory response, angiogenesis, and immune response (Fig. 2F). Kyoto Encyclopedia of Genes and Genomes analysis showed that these genes mainly participated in several pathways in cancer, such as PI3K-AKT, Rap1, MAPK, Ras, JAK-STAT, and TNF signaling pathways (Fig. 2G).

**Figure 3.** Dividing CRC patients into 2 novel clusters based on prognostic TrAGs. (A) Overlapping upregulated genes, TrAGs, and prognostic genes in CRC. (B) The forest plots of elevated prognostic TrAGs calculated by univariate Cox analysis. (C) The optimal number of clusters according to the consensus index. (D) The subclusters of CRC patients divided by consensus clustering analysis. (E) The Kaplan–Meier curves of CRC patients in cluster 1 and cluster 2. CRC = colorectal cancer, TrAGs = Tregs-associated genes.
3.3. Reclassifying CRC patients into 2 novel subgroups with distinct prognosis

To explore whether these TrAGs could reclassify CRC patients into novel clusters, we first identified 535 prognostic genes in CRC, overlapped them with differentially expressed TrAGs and selected 26 significantly elevated TrAGs with prognostic values (Fig. 3A and B). Then a consensus clustering analysis was implemented based on the “Consensus ClusterPlus” R package, and the lowest proportion of ambiguous clustering was identified as 2 (Fig. 3C). Based on the unsupervised clustering, CRC patients were eventually reclassified into 2 novel subgroups (Fig. 3D). The Kaplan–Meier curves exhibited that CRC patients in cluster 2 suffered worse overall survival than in cluster 1 (Fig. 3E).

3.4. Constructing a Tregs-associated prognostic signature for CRC patients

Considering that Tregs are significantly associated with the prognosis of patients, we intended to construct a signature for CRC based on the above 26 upregulated TrAGs with prognostic values. As shown in Figure 4A, after these genes were subjected to LASSO regression analysis, 7 (PGF, ADAM8, RGS16, TIMP1, SIX4, SLC39A13, and ELFN1) were selected to construct the prognostic model. The risk score was calculated by using the following formula: (PGF × 0.014) + (ADAM8 × 0.021) + (RGS16 × 0.027) + (TIMP1 × 0.057) + (SIX4 × 0.085) + (SLC39A13 × 0.21) + (ELFN1 × 0.27) (Fig. 4B). According to the median value of risk scores, CRC patients were reclassified into low- and high-risk score subgroups. The Kaplan–Meier curves analysis showed that the CRC patients in the high-risk score group underwent worse overall survival than that in the low-risk score group (Fig. 4C). The time-dependent receiver operating characteristic (ROC) analysis exhibited that the area under the ROC curve was 0.72, 0.70, and 0.75 for 1-, 3-, and 5-year survival (Fig. 4D). In addition, univariate and multivariate Cox regression analysis determined that the risk score was an independent prognostic factor for CRC patients (Fig. 4E and F).

3.5. Validating the utility of prognostic signature by an independent GEO dataset

To validate the utility of our prognostic signature, an independent GEO dataset (GSE17536) was used. After calculating the risk scores, 79 and 98 CRC patients were reclassified into low-risk and high-risk subgroups, respectively (Fig. 5A). As expected, CRC patients with high-risk scores underwent more deaths than those with low-risk scores (Fig. 5A). The KM curves analysis showed that the CRC patients with high-risk scores underwent worse overall survival than those with low-risk scores (Fig. 5B). The time-dependent ROC analysis showed that the area under the curve was 0.71, 0.67, and 0.73 for 1-, 3-, and 5-year survival (Fig. 5C). Besides, univariate and multivariate Cox regression analysis validated that the risk score was an independent prognostic factor for CRC patients (Fig. 5D and E). Taken together, the results proved that our prognostic model can be well applied to predict the overall survival of CRC patients.

3.6. The different tumor immune microenvironments between CRC patients in the high-risk score group and the low-risk score group

Subsequently, GSEA analysis was implemented to explore the difference between CRC patients in the high-risk and low-risk score groups. Intriguingly, immune-associated gene sets, such as HALLMARK_IL2_STAT5_SIGNALING, HALLMARK_INFLAMMATION_RESPONSE, HALLMARK_INTERFERON_ALPHA_RESPONSE, HALLMARK_TNFA_SIGNALING_VIA_NFKB, HALLMARK_COMPLEMENT, HALLMARK_INTERFERON_GAMMA_RESPONS
HALLMARK_IL6_JAK_STAT3_SIGNALING, HALLMARK_ALLOGRAFT_REJECTION, and HALLMARK_INFLAMMATORY_RESPONSE, were positively enriched in the low-risk score group (Fig. 6A). Then we further analyzed the infiltrated stroma and immune cell levels between 2 risk score groups through 3 algorithms (QUANTISEQ, XCELL, and CIBERSORT-ABS). As shown in Figure 6B to D, the infiltrated levels of Tregs, cancer-associated fibroblasts, tumor-associated macrophages, and eosinophils were significantly upregulated in the TME of CRC patients with high-risk scores. On contrary, the infiltrated levels of CD4+ T cell and CD8+ T cell were significantly downregulated in the TME of CRC patients with low-risk scores. In addition, most immune checkpoint genes, such as LAG3, PD-L1, and NRP1, were significantly upregulated in the high-risk group compared with the low-risk group (Fig. 6E).

3.7. Validating the expression of genes in the prognostic signature

To validate the expression of genes in the prognostic signature in CRC, the GSE9438 dataset was enrolled,[12] and the result showed that these 7 genes were significantly upregulated in CRC tissues compared to normal tissues (Fig. 7A). Moreover, the protein levels of SIX4, RGS16, SLC39A13, TIMP11, and ADAM8 were analyzed through immunohistochemistry. As shown in Figure 7B to F, the protein levels of SIX4, RGS16, SLC39A13, TIMP11, and ADAM8 were significantly elevated in CRC tissues.

4. Discussion

Tregs are part of the CD4+ T cell population and are generally identified by a combination of surface markers together with the transcription factor FOXP3.[13] The interaction between Tregs and cancer cells has been widely reported.[14] For example, CRC could educate γδ T cells into Tregs to promote tumor progression and metastasis.[15] Although mounting evidence revealed that the level of tumor-infiltrating Tregs was negatively correlated with patients’ survival, some studies reported that increased Tregs were associated with improved prognosis in specific tumors.[16,17] Salama et al demonstrated that FOXP3+ Treg density was higher in CRC tissue compared to the normal colonic mucosa, and intra-tumoral FOXP3+ Tregs show strong prognostic significance in CRC.[18] Based on scRNA sequencing data, we consistently validated that more Tregs infiltrated the TME of CRC tissues compared to that of adjacent normal tissues. However, there was no report about Treg-associated gene signatures to determine the prognosis of CRC patients. In the present study, we were the first to construct a prognostic signature for CRC based on TRAGs. Furthermore, we proved that our prognostic model can be well applied to predict the clinical outcomes of CRC patients after training and validating in 2 independent cohorts.

Intriguingly, we uncovered that the poor prognosis of CRC patients with high-risk scores was partly attributed to the immune-suppressing microenvironment, because of the significantly upregulated anti-immunity cells in CRC patients with high-risk scores. On contrary, only a few CD8+ T cells infiltrated the TME of CRC patients with low-risk scores. Accumulating evidence has proved that Tregs carry out anti-tumor immunity ability by suppressing T cell responses and activities of antigen-presenting cells.[8,19] For instance, Tregs can regulate the tumor-suppressing role of cytokine-induced killer cells in CRC through secreting IL-10 and TGF-β.[20] CAFs and TAMs are the most abundant nonmalignant cell types in the TME of cancers.[21] The crosstalk between Tregs and other immune cells has also been investigated.[22] For example, CD70-positive CAFs stimulate migration and significantly increase the frequency of naturally occurring Tregs.[23] Besides, Kos and colleagues demonstrated that TAMs promote the intra-tumoral conversion of conventional CD4 T cells into Tregs via PD-1 signaling.[24] Therefore, we hypothesized that targeting multiple immune cells simultaneously may improve the therapeutic effect of CRC.

Immune checkpoint genes are highly expressed on T cells, antigen-presenting cells, and cancer cells, and modulate the homeostasis of co-stimulatory and co-inhibitory signals.[25] In recent decades, immune checkpoint inhibitor treatment has revolutionized the
therapeutic paradigms of solid tumors. For example, inhibitors of programmed death-1 (PD-1)/PD-ligand 1 (PD-L1) have been approved by the American FDA to manage lung cancer. In this study, we discovered that immune checkpoint genes, such as PD-1, were also significantly upregulated in CRC patients with high-risk scores, which suggested that these patients may be more suitable for immune checkpoint inhibitor treatment.

We have to admit that several limitations exist in the present study. First, scRNA sequencing data revealed that several immune cells differentially infiltrated CRC tissues compared to normal tissues. It may improve the prognostic value of our signature to enroll other immune cell-associated genes, such as mast cells. Second, the expression levels of genes in our signature ought to be validated by more clinical tissue samples. Third, our hypothesis should be further verified before clinical application that CRC patients in the high-risk score group are more suitable for immune checkpoint inhibitor treatment.

5. Conclusion

In conclusion, we were the first to construct a prognostic signature for CRC patients based on Treg-associated genes, according to which patients were well reclassified into 2 subclusters with different clinical outcomes. Besides, our study further revealed that the poor prognosis of CRC with high-risk scores was mainly attributed to the anti-tumor immunity microenvironment and elevated immune checkpoint genes in tumor tissues.

Acknowledgments

We acknowledge the TCGA database for providing their platforms and contributors for uploading their meaningful datasets.

Author contributions

Conception and design of the research: ZZQ; acquisition, analysis, and interpretation of data: PGQ; data collection: TYC; drafting the manuscript: ZZQ; All authors read and approved the final manuscript.

Conceptualization: Ziqiang Zhou.
Data curation: Guoqiang Ping, Yichen Tian.
Formal analysis: Guoqiang Ping.
Funding acquisition: Yichen Tian.
Software: Yichen Tian.
Validation: Guoqiang Ping.
Visualization: Yichen Tian.
Writing – original draft: Ziqiang Zhou.
Writing – review & editing: Ziqiang Zhou.

References
[1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA. 2020;70:7–30.
[2] Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. CA. 2016;66:115–32.
[3] Dhillon S. Regorafenib: a review in metastatic colorectal cancer. Drugs. 2018;78:1133–44.
[4] Vardy JL, Dhillon HM, Pond GR, et al. Cognitive function in patients with colorectal cancer who do and do not receive chemotherapy: a prospective, longitudinal, controlled study. J Clin Oncol. 2015;33:4085–92.
[5] Wu T, Dai Y. Tumor microenvironment and therapeutic response. Cancer Lett. 2017;387:61–8.
[6] Revenko A, Carnevalli Ls, Sinclair C, et al. Direct targeting of FOXP3 in Tregs with AZD8701, a novel antisense oligonucleotide to relieve immunosuppression in cancer. J ImmunoTher Cancer. 2022;10:e003892.
[7] Zimmer N, Trzeciak ER, Graefen B, et al. GARP as a therapeutic target for the modulation of regulatory T cells in cancer and autoimmunity. Front Immunol. 2022;13:928450.

[8] Zhu J, Liu JQ, Shi M, et al. IL-27 gene therapy induces depletion of Tregs and enhances the efficacy of cancer immunotherapy. JCI Insight. 2018;3.

[9] Chen X, Song E. Turning foes to friends: targeting cancer-associated fibroblasts. Nat Rev Drug Discov. 2019;18:99–115.

[10] Vignali DA, Collison LW, Workman CJ. How regulatory T cells work. Nat Rev Immunol. 2008;8:523–32.

[11] Chen MS, Lo YH, Chen X, et al. Growth factor-independent 1 is a tumor suppressor gene in colorectal cancer. Mol Cancer Res. 2019;17:697–708.

[12] Hong Y, Downey T, Eu KW, et al. A “metastasis-prone” signature for early-stage mismatch-repair proficient sporadic colorectal cancer patients and its implications for possible therapeutics. Clin Exp Metastasis. 2010;27:83–90.

[13] Winerdal ME, Krantz D, Hartana CA, et al. Urinary bladder cancer tregs suppress MMP2 and potentially regulate invasiveness. Cancer Immunol Res. 2018;6:528–38.

[14] Obata-Ninomiya K, de Jesus Carrion S, Hu A, et al. Emerging role for thymic stromal lymphopoietin-responsive regulatory T cells in colorectal cancer progression in humans and mice. Sci Transl Med. 2022;14:eabi6960.

[15] Zhan Y, Zheng L, Liu J, et al. PLA2G4A promotes right-sided colorectal cancer progression by inducing CD39+γδ Treg polarization. JCI Insight. 2021;6.

[16] Shang B, Liu Y, Jiang SJ, et al. Prognostic value of tumor-infiltrating FOXP3+ regulatory T cells in cancers: a systematic review and meta-analysis. Sci Rep. 2015;5:15179.

[17] Winerdal ME, Marits P, Winerdal M, et al. FOXP3 and survival in urinary bladder cancer. BJU Int. 2011;108:1672–8.

[18] Salama P, Phillips M, Grieu F, et al. Tumor-infiltrating FOXP3+ T regulatory cells show strong prognostic significance in colorectal cancer. J Clin Oncol. 2009;27:186–92.

[19] Wang H, Franco F, Ho PC. Metabolic regulation of tregs in cancer: opportunities for immunotherapy. Trends Cancer. 2017;3:583–92.

[20] Zhong W, Fang C, Liu H, et al. LAP+CD4+T cells regulate the anti-tumor role of CIK cells in colorectal cancer through IL-10 and TGF-β. Am J Transl Res. 2022;14:3716–28.

[21] Denton AE, Roberts EW, Fearon DT. Stromal cells in the tumor micro-environment. Adv Exp Med Biol. 2018;1060:99–114.

[22] Huang H, Wang Z, Zhang Y, et al. Mesothelial cell-derived antigen-presenting cancer-associated fibroblasts induce expansion of regulatory T cells in pancreatic cancer. Cancer Cell. 2022;40:656–673.e7.

[23] Jacobs J, Deschoolemeester V, Zwaenepoel K, et al. Unveiling a CD70-positive subset of cancer-associated fibroblasts marked by pro-migratory activity and driving regulatory T cell accumulation. Oncoimmunology. 2018;7:e1440167.

[24] Kos K, Salvagno C, Wellenstein MD, et al. Tumor-associated macrophages promote intratumoral conversion of conventional CD4(+) T cells into regulatory T cells via PD-1 signalling. Oncoimmunology. 2022;11:2063225.

[25] Zhao X, Subramanian S. Intrinsic resistance of solid tumors to immune checkpoint blockade therapy. Cancer Res. 2017;77:817–22.

[26] Long L, Zhao C, Ozarina M, et al. Targeting immune checkpoints in lung cancer: current landscape and future prospects. Clin Drug Investig. 2019;39:341–53.