

# Identification and validation of prognostic models and survival predictions of T cell-inflamed related genes in patients with colorectal cancer

X. Shen<sup>1</sup>, Y. Zhang<sup>2</sup>, Y. Dong<sup>2</sup>, J. Qian<sup>1\*</sup>

<sup>1</sup>Department of Gastroenterology, Nantong First People's Hospital, Nantong, Jiangsu 226014, China

<sup>2</sup>Department of Gastroenterology, Nanjing First Hospital, Nanjing Medical University, Nanjing, Jiangsu 210000, China

## ABSTRACT

**Background:** The anti-tumor immune reaction has a critical part in the progression together with survival of colorectal cancer (CRC). The expression of T cell-inflamed gene as a potential indicator that predicts the reaction of immunogenic cancer types to checkpoint inhibitor immunotherapy, but its prognostic value in CRC is unclear.

**Materials and Methods:** Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses identified enrichment analysis of T cell-inflamed associated genes. The prognostic model is built and the prognostic genes screened using the LASSO Cox regression model. Relationship among risk score, disease subtypes and immune infiltration was constructed. Besides, risk score was combined with clinicopathological features to further improve prognostic models and survival predictions. **Results:** We identified 14 T cell-inflamed associated genes were differentially expressed in CRC tissues. GO along with KEGG analyses revealed that T cell-inflamed associated genes were implicated in the regulation of immune cells. Through LASSO Cox regression, a 5-gene prognostic model was constructed as well as classified all CRC patients into a low- or high-risk group or different disease subtypes (C1, C2 and C3). In the C3 group the OS rate of CRC patients was longer compared to the C1 group. Additionally, the prognosis Decision Curve Analysis (DCA) map displayed that the risk score was increased compared to the extreme curve and higher than other indicators, showing the strongest survival prediction ability. **Conclusion:** These results indicated that T cell-inflamed associated gene-based risk score reflected immune microenvironment features as well as predicted prognosis in CRC.

## ► Original article

### \*Corresponding author:

Junbo Qian

E-mail: [ntybobo@163.com](mailto:ntybobo@163.com)

Received: November 2023

Final revised: December 2023

Accepted: January 2024

*Int. J. Radiat. Res.*, July 2024;  
22(3): 595-602

DOI: [10.61186/ijrr.22.3.595](https://doi.org/10.61186/ijrr.22.3.595)

**Keywords:** Colorectal cancer, T cell, prognosis, immune.

## INTRODUCTION

Colorectal cancer (CRC) is one of the most prevalent forms of cancer along with high morbidity as well as death rate in the world<sup>(1)</sup>. As reported, about 1.3 million new cases of CRC along with 600,000 CRC-associated deaths occur every year<sup>(2)</sup>. The 5-year survival time of CRC patients was over 90% upon diagnosis at early stages<sup>(3)</sup>. Nevertheless, CRC is often diagnosed with late stage due to lacking adequate diagnostic methods<sup>(4)</sup>. In spite of the improvements in diagnosis together with therapy, the long-time survival for CRC patients remains low<sup>(5)</sup>. Thus, it is urgently required to understand the molecular mechanisms underlying CRC as well as identify novel biomarkers utilized for the early detection together with prognosis assessment of CRC.

Immunotherapy along with immune checkpoint inhibitors (ICIs) belongs to currently common for cancers, CRC included<sup>(6,7)</sup>. A crucial feature in the immune system is to differentiate "foreign" cells and normal cells in the body, attacking foreign cells whereas reserving normal cells<sup>(8)</sup>. A positive immune

response is discovered before therapy in patients sensitive to immunotherapy, featured by infiltration of antigen-specific T cells<sup>(9)</sup>. This phenomenon is identified as the T cell-inflamed tumor microenvironment (TME), which can be adopted for prediction of responding together with non-responding tumors<sup>(10)</sup>. The TME belongs to the environment surrounding the tumor, composed of peripheral blood vessels, immune cells, fibroblasts, along with extracellular matrix<sup>(11)</sup>. Tumors are strongly linked to the surrounding microenvironment as well as constantly interact with each other<sup>(12)</sup>. In particular, tumors can affect the microenvironment via extracellular signals, accelerating tumor angiogenesis, together with stimulating peripheral immune tolerance. In turn, cancer cells growth is influenced by immune cells in the microenvironment<sup>(13)</sup>.

In recent years, a signature known as the T cell-inflamed gene expression profile (GEP) has been established to suggest responsiveness to PD-1 blocking treatment<sup>(14)</sup>. The T cell-inflamed GEP is associated with chemokine expression, cytolytic

activity, as well as adaptive immune resistance<sup>(15)</sup>. In pan-cancer clinical datasets, it has revealed good predictive performance against overall survival (OS) following PD-1 therapy. Additionally, the prognostic effect of T cell-inflamed GEP on CRC<sup>(16)</sup>.

In this research we aimed to identify as well as validate prognostic models related to T cell inflammation. Our research provided a new target for the diagnosis of CRC, and also provides a new direction in clinical immunotherapy in the future, so as to help CRC patients with more accurate diagnosis and effective treatment, and improve the survival rate of patients.

## MATERIALS AND METHODS

### Differential expression analysis

To verify the gene expression between CRC tissues together with normal tissues, the limma package was implemented. The criteria for differentially expressed T cell-inflamed related were  $|\log_2(\text{fold change})| > 1$  together with the false discovery rate (FDR) less than 0.05.

### Functional enrichment analysis

Using edgeR R package for high-risk together with low-risk group identification of genes between  $|\log_2(\text{fold change})| > 1$  together with FDR less than 0.05, Functional annotations based on GO together with KEGG were also made by means of the 'clusterProfiler' R package (adjusted p values < 0.05). Metascape (<https://metascape.org>) was implemented to further certify the function enrichment of these genes.  $P < 0.05$  was significance.

### Protein-protein interaction (PPI) network

The PPI information in T cell-inflamed-related genes was identified by means of STRING website (<http://www.string-db.org/>).

### Construction of T cell-inflamed-related prognostic signature

Using the LASSO prognostic model, patients with CRC were divided into high- and low-risk groups, or C1, C2, and C3 groups, and the OS was compared using the Kaplan-Meier methodology. Additionally, the survival ROC was calculated to verify the predictive accuracy. To investigate the relationship between clinical variables and risk score, both univariate and multivariate Cox regression models were used. A nomogram containing calibration maps was implemented to assess the clinical outcome of CRC patients using an RMS R package.

### Comparison of immune cell infiltration

The normalized matrix data of gene expression levels were examined using the R software's limma package in order to investigate the T cell-inflamed-related genes in further detail between low- and high

-risk immunological score colorectal cancer cases. Furthermore, genes associated with immunity were obtained from the ImmPort database (<https://www.immport.org/>), and these genes were used to determine immune-related T cell-inflamatory genes in order to create a predictive risk model. To compute the microenvironment score, the "ESTIMATE" software was used.

### Statistical analysis

All statistical analyses were implemented by means of R software (version 3.6.3, <http://www.R-project.org>). We used the t-test for comparing the variables. Kaplan-Meier curves were applied to assess the survival curves, and a log-rank test was carried out to evaluate them. Differences were significant when  $P < 0.05$ .

## RESULTS

### Expression difference and enrichment analysis of T cell-inflamed related genes

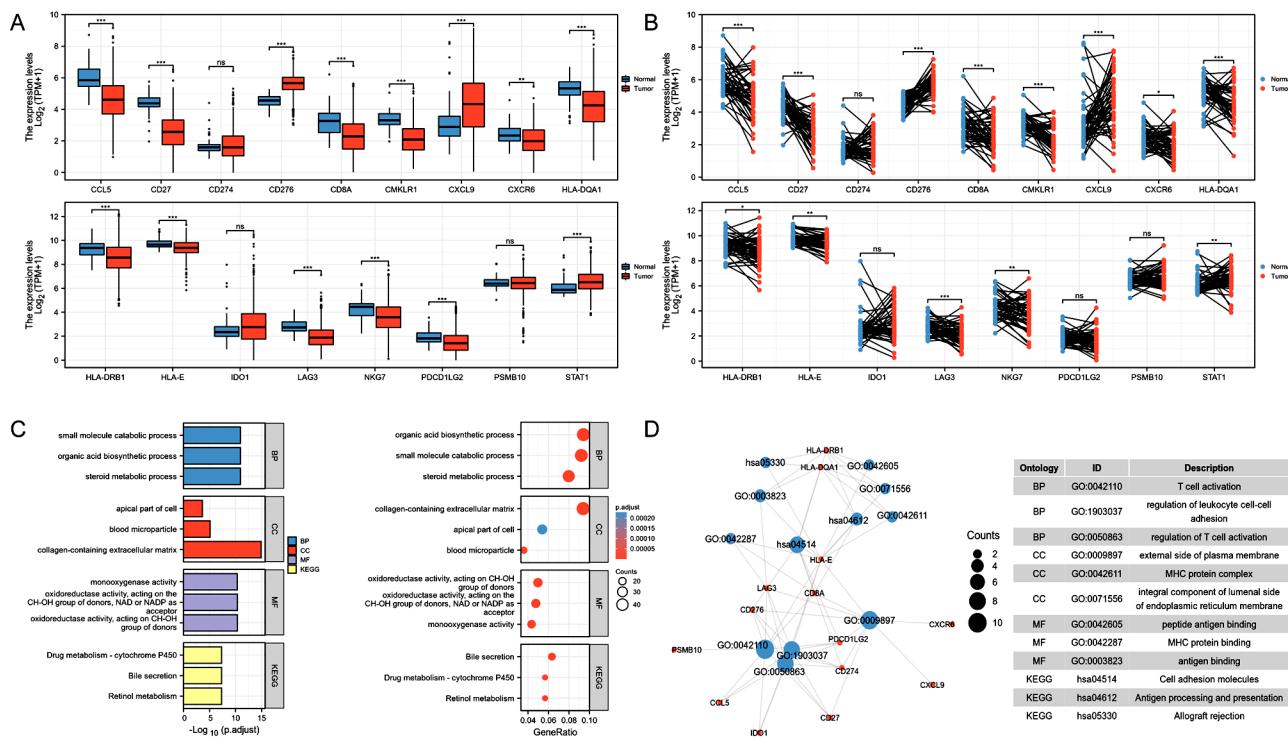
Expression difference in transcription levels of 17 T cell-inflamed related genes were analyzed in unpaired samples. In comparison with normal tissues, 14 genes containing CCL5, CD27, CD8A, CMKLR1, CXCR6, HLA-DQA1, HLA-DRB1, HLA-E, LAG3, NKG7 and PDCD1LG2 were down-regulated while CD276, CXCL9 and STAT1 were up-regulated in tumor tissues (figure 1A). At the same time, expression differences in transcription levels of these 17 genes were analyzed in paired samples. The outcomes displayed that, CCL5, CD27, CD8A, CMKLR1, CXCL9, CXCR6, HLA-DQA1, HLA-E, LAG3 and STAT1 were significantly down-regulated while CD276, HLA-DRB1 and NKG7 were high-expressed in tumor tissues (figure 1B). To assess the biological functions of these T cell-inflamed related genes, GO enrichment together with KEGG pathway analyses were implemented. As displayed in figure 1C, these T cell-inflamed related genes were mainly enriched in organic acid biosynthetic process, collagen-containing extracellular process, steroid metabolic process together with small molecule catabolic process. Besides, scatter plot of GO together with KEGG enrichment analyses exhibited that T cell-inflamed related genes were majorly implicated in T cell activation, cell adhesion molecules, regulation of T cell activation, external side of plasma membrane and regulation of leukocyte cell-cell adhesion (figure 1D).

### Enrichment analysis of T cell-inflamed linked genes based on Metascape

A significant heat map was constructed based on biological items from different datasets. As displayed in figure 2A, T cell-inflamed related genes were mainly enriched in hallmark interferon gamma response, regulation of T cell activation and positive

regulation of immune response. Besides, a significant heat map was constructed based on DisGeNET (<https://www.disgenet.org/>) database. It was found that T cell-inflamed related genes were mainly enriched in immunosuppression, hepatitis C, chronic and infection (figure 2B). Furthermore, Metascape was used to analyze the GO data set, and discovered that T cell-inflamed related genes were mainly involved in immune system process, regulation of

biological process along with positive regulation of biological process (figure 2C). In addition, the PPI network was built. It was observed that T cell-inflamed related genes were majorly implicated in hallmark interferon gamma response, regulation of T cell activation and positive regulation of immune response (figure 2D - 2E). The two most important subnetworks in the PPI network were exhibited in figure 2F.



**Figure 1.** Expression difference and enrichment analysis of T cell-inflamed related genes. **(A)** Unpaired samples revealed differences in transcription levels of 17 genes. **(B)** Differences in gene transcription levels in paired samples. **(C)** Bar charts and bubble charts showed the GO/KEGG enrichment analysis results of T cell-inflamed gene sets. **(D)** Scatter diagram of GO/KEGG enrichment analysis.

### The Lasso regression model is used to screen the prognostic genes and construct the prognostic model

A prognostic risk evaluation model on the basis of 5 T cell-inflamed related genes (CD276, CXCL9, CXCR6, HLA-DQA1 and LAG3) was constructed using the LASSO model. The cvfit together with lambda curve were displayed in figures 3A and 3B. Then, an X-tile diagram indicated the optimal cutoff value for the risk score. Based on this cutoff value, The CRC patients were separated into a high-risk group together with a low-risk group. Prognostic curves and scatter plots show the risk score of each case with CRC as well as survival status (figure 3C). Besides, the death cases were majorly presented in the high-risk group. Additionally, the heat map of candidate mRNAs demonstrated that CD276, CXCL9, CXCR6, HLA-DQA1 and LAG3 were high-expressed in the group of high risk. Three different disease subtypes were constructed based on consistent clustering (figure 3D-3E). Kaplan-Meier survival analysis

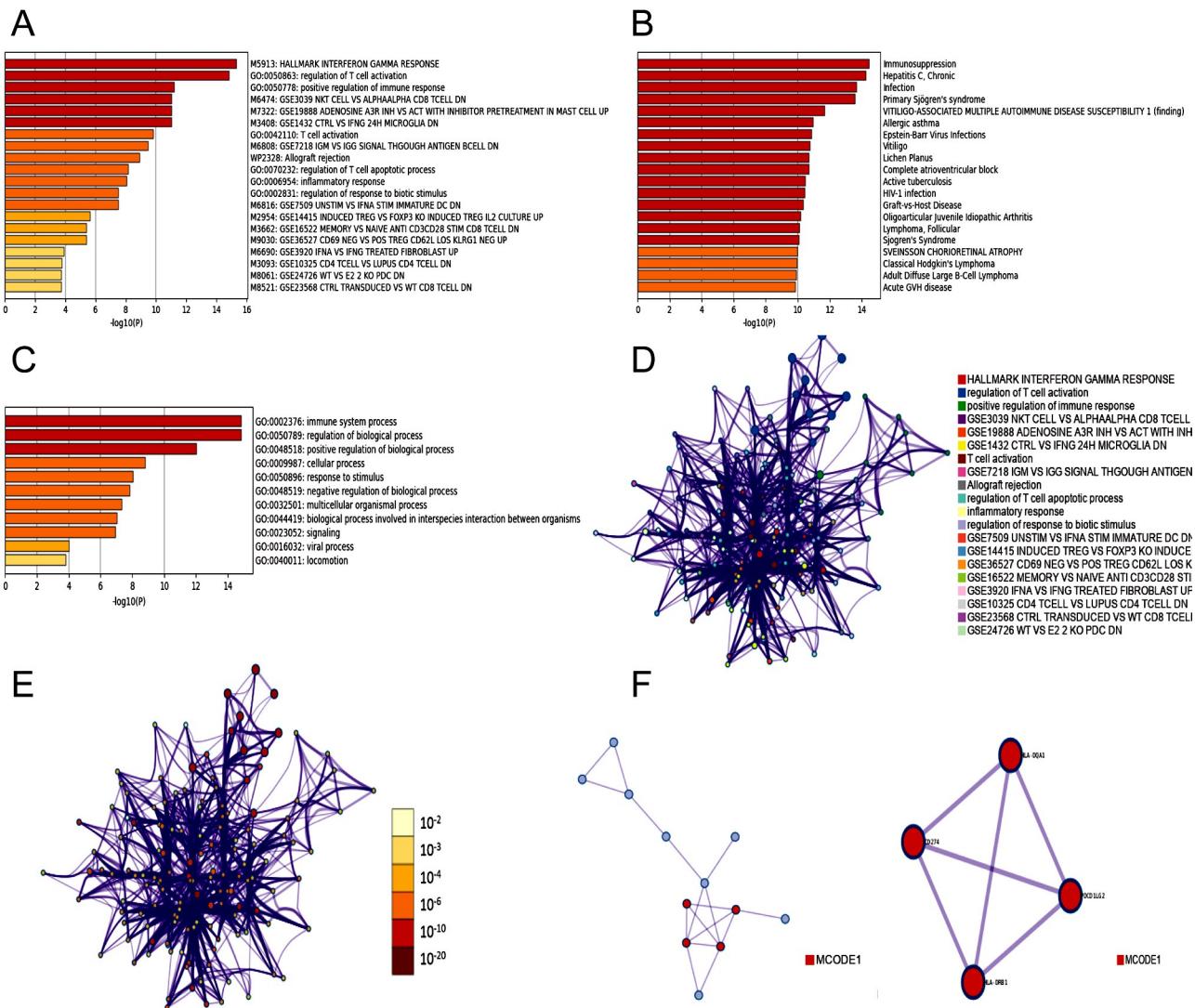
presented that the OS time of CRC cases in the C3 group was better compared to the C1 group (figure 3F). Besides, compared to C2 and C3, the risk score was elevated in patients with C1 (figure 3G). The AUC values of CRC in predicting the OS time was 0.68 for C1, 0.67 for C2 and 0.62 for C3 (figure 3H).

### Relationship among risk score, disease subtypes and immune infiltration

It was investigated how the infiltration levels of twenty-four different kinds of immune cells varied. High-risk patients showed higher NK CD56bright cells together with NK cells while the low-risk group showed higher aDC, B cells, eosinophils, T cells, T helper cells, Tcm, Tgd, Th1 cells, Th17 cells and Th2 cells (figure 4A). Meanwhile, the illness subtype was used to categorize the variations in infiltration levels in 24 different types of immune cells. As displayed in Figure 4B, compared to C1 and C3, C2 obtained a higher immune score, which insinuated an immunologically “hot” phenotype. The correlation

between risk score and infiltration scores of twenty-four kinds of immune cells was positively correlated (figure 4C). Patients with high scores of T helper cells, Tgd and NK CD56bright cells had shorter survival time compared to low scores, while patients with high scores of TH2 cells, aDC, B cells, TH17 cells along with NK cells harbored longer survival time

(figure 4D). The correlation between risk score and matrix score, immune score and EASIMATE score was constructed based on EASIMATE algorithm, and was found to be positively correlated (figure 4E). In contrast to high risk group, CRC patients possessing low risk scores had higher immune scores and EASIMATE scores (figure 4F).

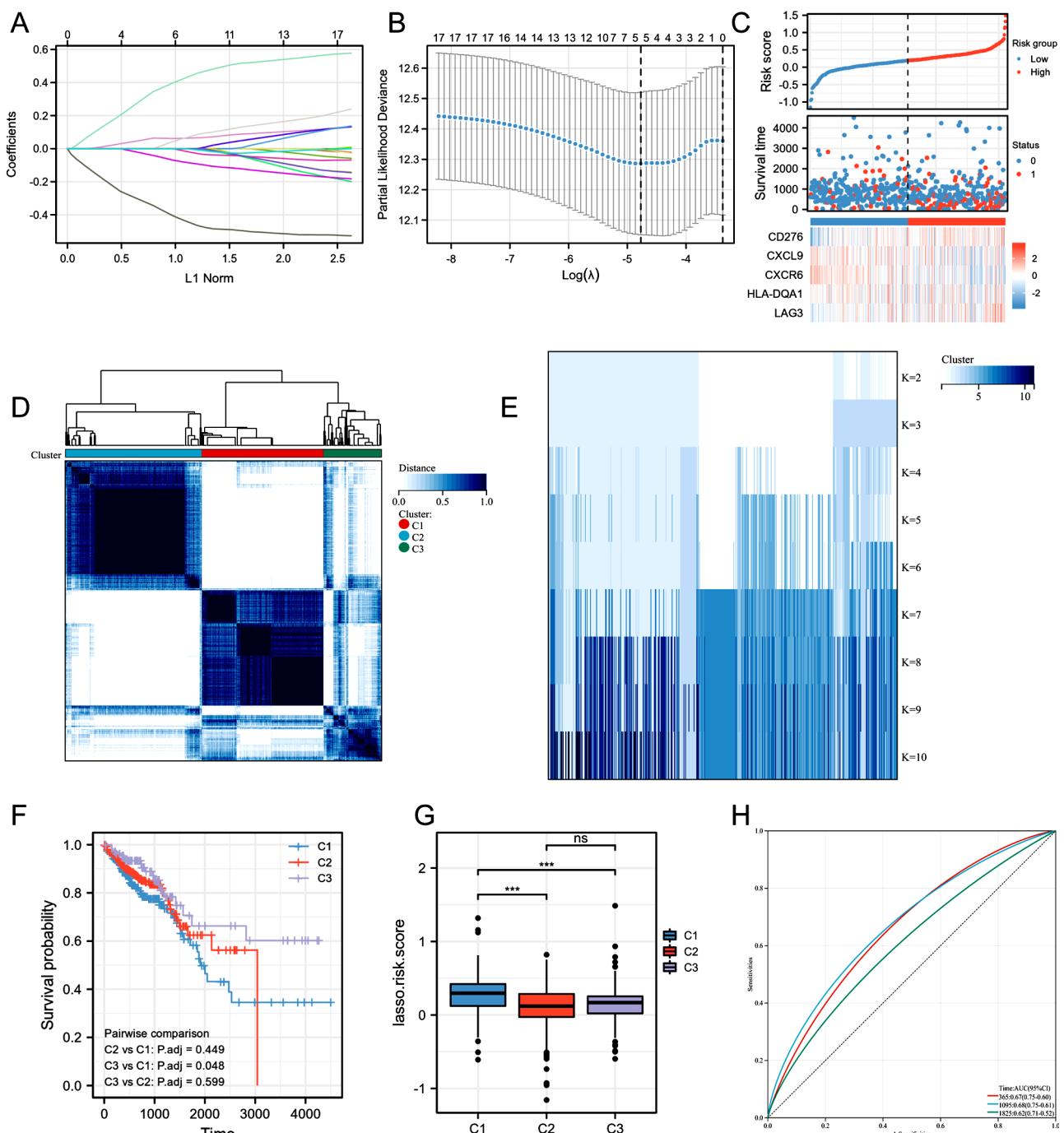


**Figure 2.** Enrichment analysis of T cell-inflamed related gene set based on Metascape. **(A)** A significant heat map was constructed based on biological entries from different data sets. **(B)** A significant heat map was established based on DisGeNET database. **(C)** A significant heat map was built based on GO data set. **(D)** Protein-protein interaction network was constructed based on protein interaction data, and color genes with P-values in protein-protein interaction networks. **(E)** Coloring genes with P-values in protein-protein interaction networks. **(F)** Two of the most important subnetworks in the protein-protein interaction network.

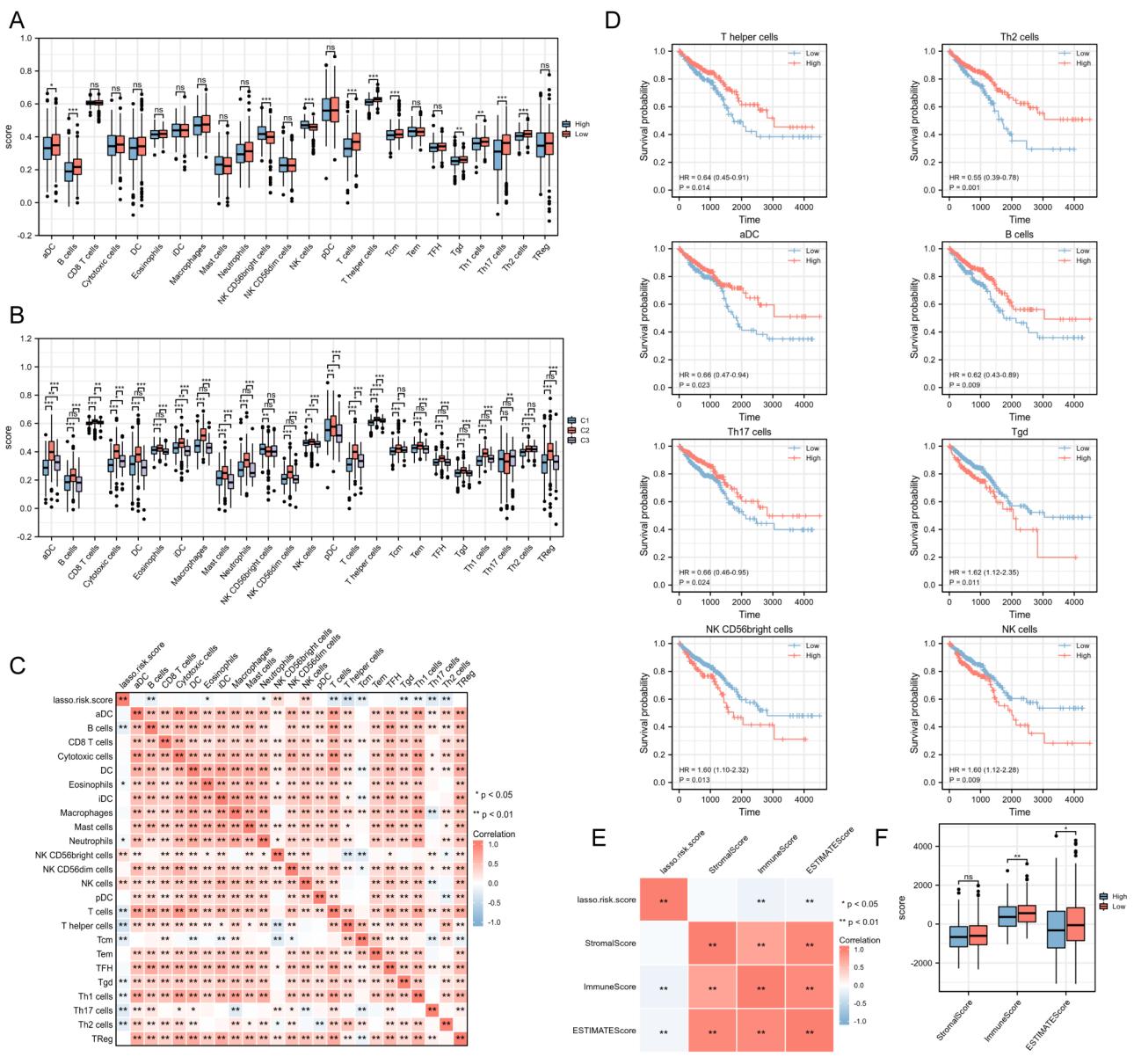
### Risk score is combined with clinicopathological features to further improve prognostic models and survival predictions

As revealed in figure 5A, CRC patients with low risk scores harbored longer survival time. The impact of five potential biomarkers and disease subtypes on patient outcomes was displayed in figure 5B.

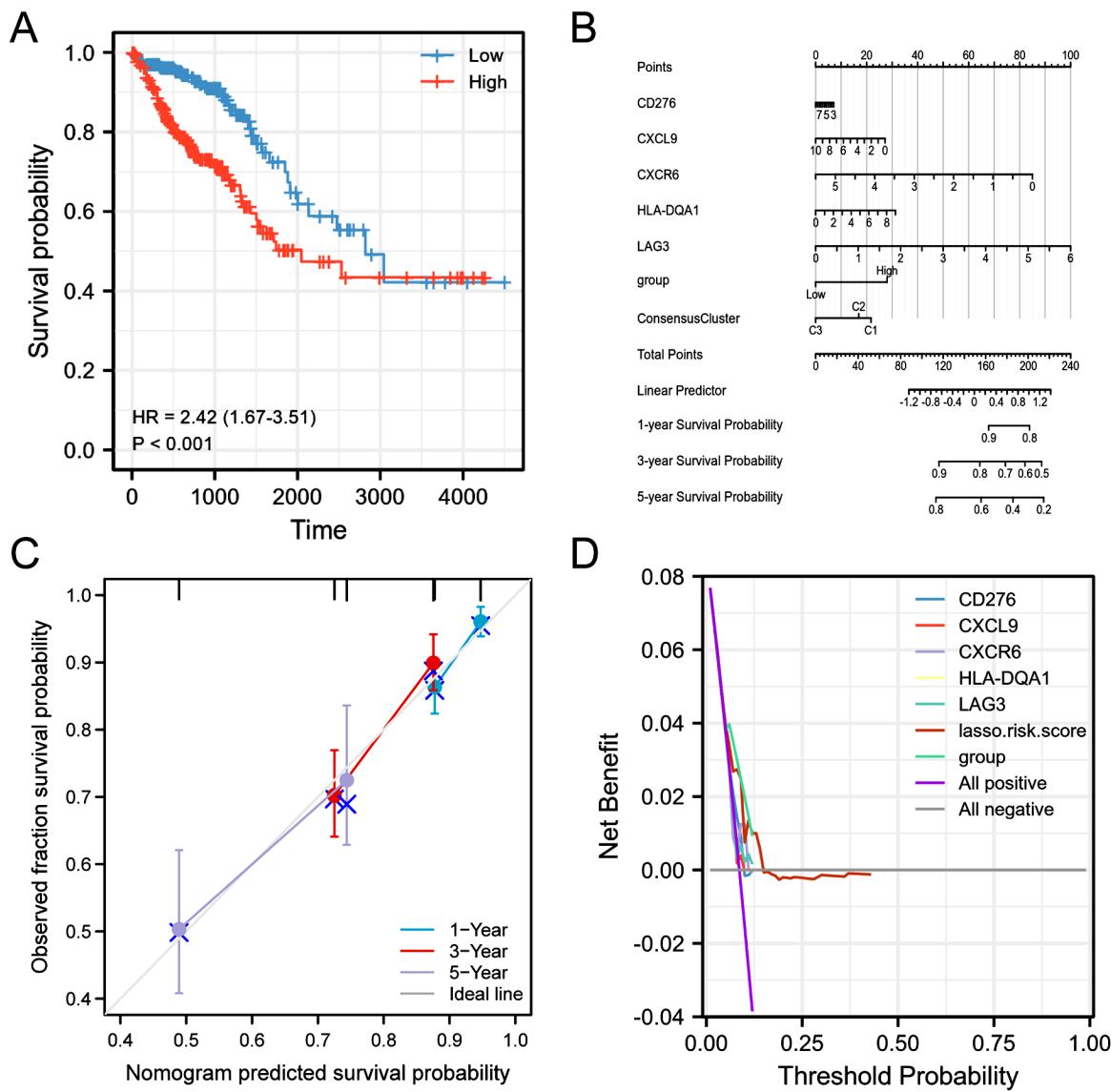
Calibration curve evaluated the predictive veracity of T cell-inflamed correlation prognostic model (figure 5C). Additionally, the prognosis DCA map revealed that the risk score was higher than the extreme curve and higher than other indicators, showing the strongest survival prediction ability (figure 5D).



**Figure 3.** The Lasso regression model is used to screen the prognostic genes and construct the prognostic model. **(A)** Variable trajectory diagram. **(B)** Lasso coefficient screening results. **(C)** T cell-inflamed prognostic model and risk factor map. An X-tile diagram indicated the optimal cutoff value for the risk score. **(D)** Three different disease subtypes were constructed based on consistent clustering. **(E)** Sample clustering consistency. **(F)** Survival curves of three disease subtypes were drawn based on TCGA database. **(G)** Risk scores were compared in the groups of three disease subtypes. **(H)** ROC curve verified the stability of the prognostic model.



**Figure 4.** Relationship among risk score, disease subtypes and immune infiltration. **(A)** The difference between the infiltration levels of 24 kinds of immune cells was explored in groups with high or low risk score. **(B)** Differences in infiltration levels in 24 types of immune cells grouped by disease subtype. **(C)** Heat maps of correlation between risk score and infiltration scores of 24 kinds of immune cells. **(D)** Survival curve was constructed based on different immune cell infiltration scores and patient prognosis data in TCGA. **(E)** Correlation between risk score and matrix score, immune score and EASIMATE score based on EASIMATE algorithm. **(F)** The difference between matrix score, immune score and EASIMATE score was compared in groups according to high or low risk score.



**Figure 5.** Risk score is combined with clinicopathological features to further improve prognostic models and survival predictions. **(A)** Survival curve based on high and low risk score groups of CRC patients. **(B)** The tabulated table showed the impact of five potential biomarkers and disease subtypes on patient outcomes. **(C)** Calibration curve evaluated the predictive accuracy of T cell-inflamed correlation prognostic model. **(D)** Prognostic DCA chart indicated that the risk score was higher than the extreme curve. And higher than other indicators, showing the strongest survival prediction ability.

## DISCUSSION

Our study firstly explored the expression difference and enrichment assessment of T cell-inflamed associated genes. It was observed that 14 T cell-inflamed related genes were significantly highly expressed or down-regulated in CRC tissues. Besides, GO enrichment together with KEGG pathway analyses was implemented to display the biological potentials of these T cell-inflamed linked genes in CRC. The results indicated that these T cell-inflamed linked genes were mostly implicated in T cell activation, modulation of leukocyte cell-cell adhesion, cell adhesion molecules, and external side of plasma membrane and regulation of T cell activation. Previous studies have revealed a fundamental association between T-cell-inflamed profiles and

tumor immune infiltration, cancer progression, and treatment response. T-cell-inflamed profiles have the potential to serve as biomarkers for cancer diagnosis, prognosis, and treatment response (17).

In addition, the prognosis of T cell-inflamed associated genes was analyzed by LASSO regression. It was found that a large proportion of the death cases were presented in the group of high risk. Additionally, the heat map displayed that CD276, CXCL9, CXCR6, HLA-DQA1 along with LAG3 were high-expressed in the group of high risk. It has been reported that overexpression of CD276 predicts poor outcome in CRC patients, and can promote CRC angiogenesis (18,19). CXCL9 expression is related to good prognosis for CRC patients (20). CXCR6 is linked to unfavorable prognosis of CRC (21). CRC patients harboring expression of HLA-DQA1 have good

survival (22). Tumour-infiltrating lymphocytes with high LAG-3 expression are found in CRC (23). Besides, 3 different disease subtypes were constructed based on consistent clustering. According to Kaplan-Meier survival analysis, the C3 group's OS time for CRC cases was superior to that of the C1 group.

To further identify whether these specific T cell-inflamed related genes were utilized for prognostic factors, the prognostic prediction model was built. In groups with high or low risk scores, the variations in the infiltration levels of 24 different types of immune cells were investigated. In the meantime, illness subtype was used to categorize the variations in infiltration levels in 24 different types of immune cells. Compared to C1 and C3, C2 obtained a higher immune score, which insinuated an immunologically "hot" phenotype. There was a positive association found between the risk score and the infiltration scores in 24 different types of immune cells. Accordance to the EASIMATE algorithm, a positive correlation was discovered between the risk score and matrix score, immune score and EASIMATE score. Additionally, to enhance prognostic models and survival forecasts, risk score was paired with clinicopathological variables. We found that the risk score had the highest survival prediction capacity, being greater than other indicators and increased when compared to the extreme curve. These studies provided a promising avenue for future study by demonstrating for the first time the prognostic significance of the T cell-inflamed related gene for CRC patients.

Nonetheless, this study has limitations. First, the outcomes are short of in vitro together with in vivo experiments to verify the credibility. Hence, more assays should be arranged to validate the mechanistic relations of these genes and CRC in the future,

## CONCLUSION

Five prognosis-associated T cell-inflamed related genes were linked to the OS of CRC patients. A prognostic model was built with effective effects. Of note, this study may provide novel targets for the treatment of CRC patients.

## ACKNOWLEDGMENT

*None.*

**Funding:** None.

**Conflicts of interests:** No potential conflict of interest was reported by the authors.

**Author contribution:** Xiaoran Shen, Yifeng Zhang, Yu Dong, Junbo Qian, \*Junbo Qian conceived and designed the experiments. Xiaoran Shen contributed significantly to the experiments and arranging data. Yifeng Zhang and Yu Dong performed data analyses. Junbo Qian wrote the draft manuscript. Junbo Qian

and Xiaoran Shen revised the manuscript. All authors read and approved the final manuscript.

## REFERENCES

- Li J, X Ma, D Chakravarti, S Shalapour and RA DePinho, Genetic and biological hallmarks of colorectal cancer. *Genes Dev*, 2021. **35(11-12)**: p. 787-820.
- Dekker E, PJ Tanis, JLA Vleugels, PM Kasi and MB Wallace, Colorectal cancer. *Lancet*, 2019. **394(10207)**: p. 1467-1480.
- Siegel RL, KD Miller, A Goding Sauer, SA Fedewa, LF Butterly, JC Anderson, *et al.* Colorectal cancer statistics, 2020. *CA Cancer J Clin*, 2020. **70(3)**: p. 145-164.
- Zygulska AL and P Pierzchalski, Novel Diagnostic Biomarkers in Colorectal Cancer. *Int J Mol Sci*, 2022. **23(2)**.
- Olenius T, L Koskenvuo, S Koskenсалo, A Lepistö and C Böckelman, Long-term survival among colorectal cancer patients in Finland, 1991-2015: a nationwide population-based registry study. *BMC Cancer*, 2022. **22(1)**: p. 356.
- Boukouris AE, M Theοchari, D Stefanou, A Papalambros, E Feleκouras, H Gogas, *et al.* Latest evidence on immune checkpoint inhibitors in metastatic colorectal cancer: A 2022 update. *Crit Rev Oncol Hematol*, 2022. **173**: p. 103663.
- Bagchi S, R Yuan and EG Engelman, Immune Checkpoint Inhibitors for the Treatment of Cancer: Clinical Impact and Mechanisms of Response and Resistance. *Annu Rev Pathol*, 2021. **16**: p. 223-249.
- Carlino MS, J Larkin and GV Long, Immune checkpoint inhibitors in melanoma. *Lancet*, 2021. **398(10304)**: p. 1002-1014.
- Liu Z, JP Li, M Chen, M Wu, Y Shi, W Li, *et al.* Detecting Tumor Antigen-Specific T Cells via Interaction-Dependent Fucosyl-Biotinylation. *Cell*, 2020. **183(4)**: p. 1117-1133.e19.
- Garris CS and JJ Luke, Dendritic Cells, the T-cell-inflamed Tumor Microenvironment, and Immunotherapy Treatment Response. *Clin Cancer Res*, 2020. **26(15)**: p. 3901-3907.
- Hu J, A Yu, B Othmane, D Qiu, H Li, C Li, *et al.* Siglec15 shapes a non-inflamed tumor microenvironment and predicts the molecular subtype in bladder cancer. *Theranostics*, 2021. **11(7)**: p. 3089-3108.
- Li X, Y Xiang, F Li, C Yin, B Li and X Ke, WNT/β-Catenin Signaling Pathway Regulating T Cell-Inflammation in the Tumor Microenvironment. *Front Immunol*, 2019. **10**: p. 2293.
- Baharom F, RA Ramirez-Valdez, A Khalilnezhad, S Khalilnezhad, M Dillon, D Hermans, *et al.* Systemic vaccination induces CD8(+) T cells and remodels the tumor microenvironment. *Cell*, 2022. **185(23)**: p. 4317-4332.e15.
- Ott PA, YJ Bang, SA Piha-Paul, ARA Razak, J Bennouna, JC Soria, *et al.* T-Cell-Inflamed Gene-Expression Profile, Programmed Death Ligand 1 Expression, and Tumor Mutational Burden Predict Efficacy in Patients Treated With Pembrolizumab Across 20 Cancers: KEYNOTE-028. *J Clin Oncol*, 2019. **37(4)**: p. 318-327.
- Shah MA, T Kojima, D Hochhauser, P Enzinger, J Raimbourg, A Hollebecque, *et al.* T cell-inflamed gene expression profile and PD-L1 expression and pembrolizumab efficacy in advanced esophageal cancer. *Future Oncol*, 2022.
- Yin H, TA Harrison, SS Thomas, CL Sather, AL Koehne, RC Malen, *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*, 2022.
- Niu D, Y Chen, H Mi, Z Mo and G Pang, The epiphany derived from T-cell-inflamed profiles: Pan-cancer characterization of CD8A as a biomarker spanning clinical relevance, cancer prognosis, immunosuppressive environment, and treatment responses. *Front Genet*, 2022. **13**: p. 974416.
- Wang R, Y Ma, S Zhan, G Zhang, L Cao, X Zhang, *et al.* B7-H3 promotes colorectal cancer angiogenesis through activating the NF-κB pathway to induce VEGFA expression. *Cell Death Dis*, 2020. **11(1)**: p. 55.
- Ingebrigtsen VA, K Boye, JM Nesland, A Nesbakken, K Flatmark and Ø Fodstad, B7-H3 expression in colorectal cancer: associations with clinicopathological parameters and patient outcome. *BMC Cancer*, 2014. **14**: p. 602.
- Wu Z, X Huang, X Han, Z Li, Q Zhu, J Yan, *et al.* The chemokine CXCL9 expression is associated with better prognosis for colorectal carcinoma patients. *Biomed Pharmacother*, 2016. **78**: p. 8-13.
- Verbeke H, S Struyf, G Laureys and J Van Damme, The expression and role of CXC chemokines in colorectal cancer. *Cytokine Growth Factor Rev*, 2011. **22(5-6)**: p. 345-58.
- Hu Y, J Ding, C Wu, H Gao, M Ge, Q Shao, *et al.* Differential Expression and Prognostic Correlation of Immune Related Factors Between Right and Left Side Colorectal Cancer. *Front Oncol*, 2022. **12**: p. 845765.
- Lythgoe MP, DSK Liu, NE Annels, J Krell and AE Frampton, Gene of the month: lymphocyte-activation gene 3 (LAG-3). *J Clin Pathol*, 2021. **74(9)**: p. 543-547.