Journal of Zhejiang University-SCIENCE B (Biomedicine & Biotechnology) 2021 22(12):1053-1059
www.jzus.zju.edu.cn; www.springer.com/journal/11585
E-mail: jzus_b@zju.edu.cn

Correspondence

https://doi.org/10.1631/jzus.B2100412

Development of a novel chemokine signaling-based multigene signature to predict prognosis and therapeutic response in colorectal cancer

Xin QI✉, Donghui YAN, Jiachen ZUO, Rui WANG, Jiajia CHEN

School of Chemistry and Life Sciences, Suzhou University of Science and Technology, Suzhou 215011, China

Colorectal cancer (CRC) is the most lethal gastrointestinal cancer in both males and females worldwide (Sung et al., 2021). Because of the high heterogeneity of tumors, robust prognostic biomarkers are urgently needed in CRC management (Koncina et al., 2020). Chemokine signaling is a well-known pivotal player in immunity, inflammation, and cancer metastasis (Lacalle et al., 2017; Poeta et al., 2019; Do et al., 2020), and multiple genes involved in chemokine signaling have been demonstrated as potential prognostic biomarkers for CRC (Cabrero-De Las Heras and Martínez-Balibrea, 2018; Ottaiano et al., 2020; Yu et al., 2020). Therefore, the aim of our study was to develop a chemokine signaling-based multigene signature (CSbMgSig) that could effectively predict overall survival (OS) and therapeutic response for patients with CRC.

Gene expression data from human CRC tissues and corresponding clinical data were retrieved from the public Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo) with the accession numbers GSE131418, GSE39582, and GSE17536. Given the critical role of distal metastasis in influencing patient survival, we first downloaded the GSE131418 dataset (with 332 primary and 184 metastatic human CRC samples) from the Moffitt Cancer Center (Kamal et al., 2019) in order to identify differentially expressed genes (DEGs) involved in CRC metastasis. The GSE39582 (n=582) and GSE17536 (n=145) cohorts were then respectively used as the training set and validation set for CSbMgSig construction. We also collected 192 genes in the chemokine signaling pathway from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (https://www.genome.jp/kegg) with the entry number hsa04062 to explore the prognostic roles of chemokine signaling-related genes (CSRGs) in CRC. The present research used the strategy (Fig. S1) modified from our previous study (Qi et al., 2021).

We found that 59 CSRGs were significantly differentially expressed between primary and metastatic human CRC samples in the GSE131418 dataset. Furthermore, univariate Cox regression analysis showed that the abnormal expression patterns of nine CSRGs, namely adenylate cyclase 2 (*ADCY2*), protein kinase B γ (*AKT3*), chemokine (C-C motif) ligand 7 (*CCL7*), *CCL8*, CXC motif chemokine ligand 2 (*CXCL2*), CXC chemokine receptor 3 (*CXCR3*), nuclear factor-κB (NF-κB) inhibitor β (*NFKBIB*), phosphatidylinositol 3-kinase catalytic subunit α (*PIK3CA*), and phospholipase C β4 (*PLCB4*), were significantly correlated with OS in CRC patients in the GSE39582 dataset (Fig. S2).

Least absolute shrinkage and selection operator (LASSO) Cox regression analysis was also performed to select the most robust prognostic genes from the above nine CSRGs. Based on the optimal λ value, we were able to establish the CSbMgSig (Figs. 1a–1c). Notably, most genes of the signature have been shown to be involved in the pathogenesis of CRC. For example, overexpression of *CCL7* is closely associated with liver metastasis and shorter OS of CRC patients (Kurzejamska et al., 2019). The CRC patients were divided into high- and low-risk groups based on the median value of risk scores, which were calculated based on the regression coefficient and expression level of each gene in the signature (Fig. 1d). The expression level of each signature-associated CSRG was obviously

✉ Xin QI, qixin@usts.edu.cn
Xin QI, https://orcid.org/0000-0001-9038-4287

Received May 9, 2021; Revision accepted Aug. 16, 2021;
Crosschecked Nov. 22, 2021

changed between the high- and low-risk patients (Fig. 1f), and patients with high risk scores had significantly shorter OS (Figs. 1e and 1g). Furthermore, we found that based on the time-dependent receiver operating characteristic (ROC) curve analysis results generated by the CSbMgSig, the area under the curve (AUC) value reached 0.700 at one year, 0.624 at two years, and 0.619 at three years (Fig. 1h). Therefore, the results indicated that the established CSbMgSig possessed a powerful predictive ability for the OS of CRC patients.

To investigate the relationship between the CSbMgSig and clinicopathological characteristics, we determined whether there were significant differences in risk scores among distinct subgroups stratified by TNM stage, T stage, N stage, M stage, tumor location, age, and sex. The results revealed that CRC patients

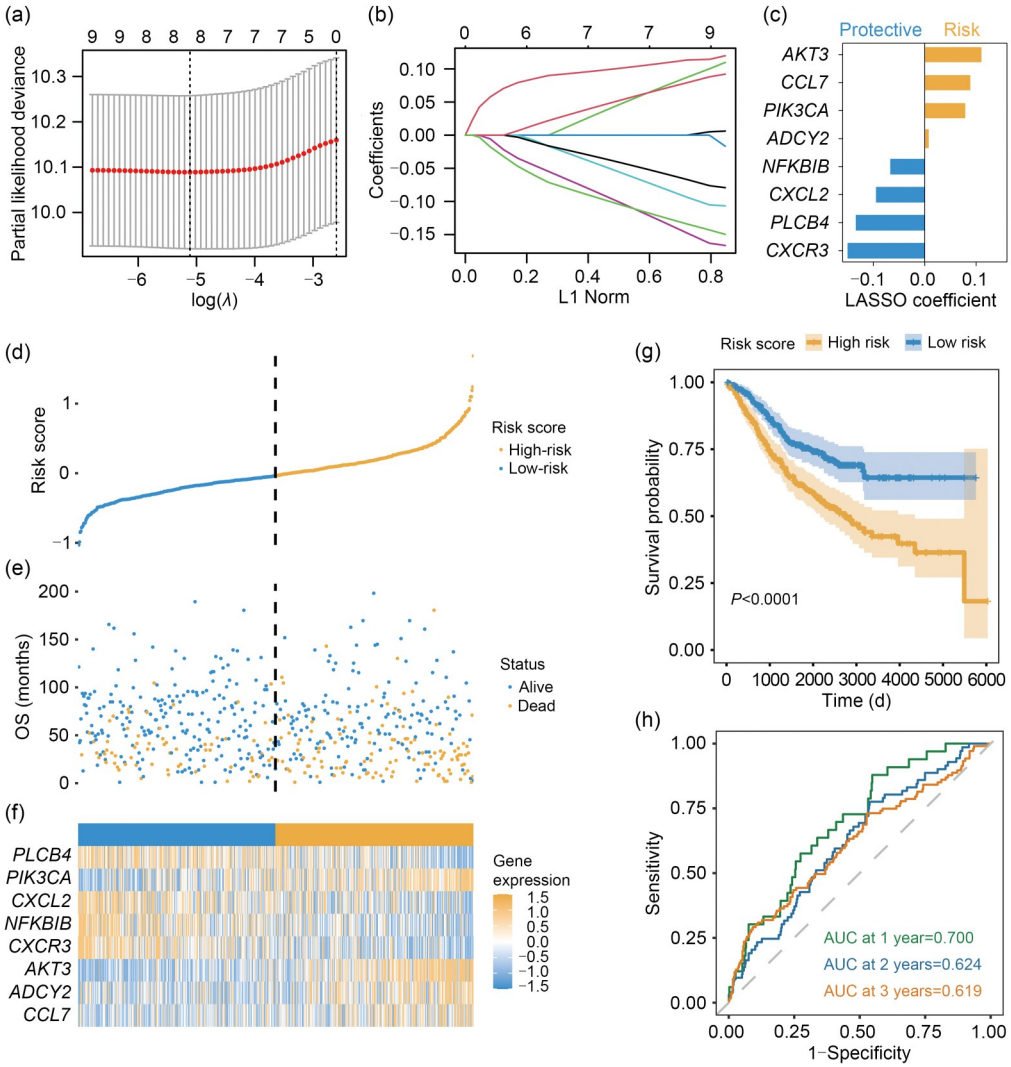

Fig. 1 Development of the CSbMgSig with the LASSO method using the training GSE39582 dataset. (a, b) LASSO Cox regression analysis used to identify the robust prognostic CSRGs; (c) LASSO coefficient profiles of the eight robust CSRGs with prognostic value; (d) Distribution of the CSbMgSig-based risk scores; (e) Distribution of survival status in CRC patients with high and low risk scores; (f) Expression patterns of the eight CSRGs that make up the CSbMgSig between the high- and low-risk groups; (g) Kaplan-Meier survival analysis of the OS difference between CRC patients in the high- and low-risk groups; (h) Time-dependent ROC curve analysis of the ability of CSbMgSig to predict the OS of CRC patients. CSbMgSig: chemokine signaling-based multigene signature; CSRGs: chemokine signaling-related genes; LASSO: least absolute shrinkage and selection operator; CRC: colorectal cancer; OS: overall survival; ROC: receiver operating characteristic; *AKT3*: protein kinase B γ; *ADCY2*: adenylate cyclase 2; *CCL7*: chemokine (C-C motif) ligand 7; *PIK3CA*: phosphatidylinositol 3-kinase catalytic subunit α; *NFKBIB*: nuclear factor-κB inhibitor β; *CXCL2*: CXC motif chemokine ligand 2; *PLCB4*: phospholipase C β4; *CXCR3*: CXC chemokine receptor 3.

with advanced TNM stage, T stage, N stage, M stage, or proximal tumor location exhibited higher risk scores (Fig. 2a). To further assess the prognostic power of the CSbMgSig, we performed a survival stratification analysis in different subgroups of CRC patients. As shown in Fig. 2b, high-risk patients had worse outcomes than

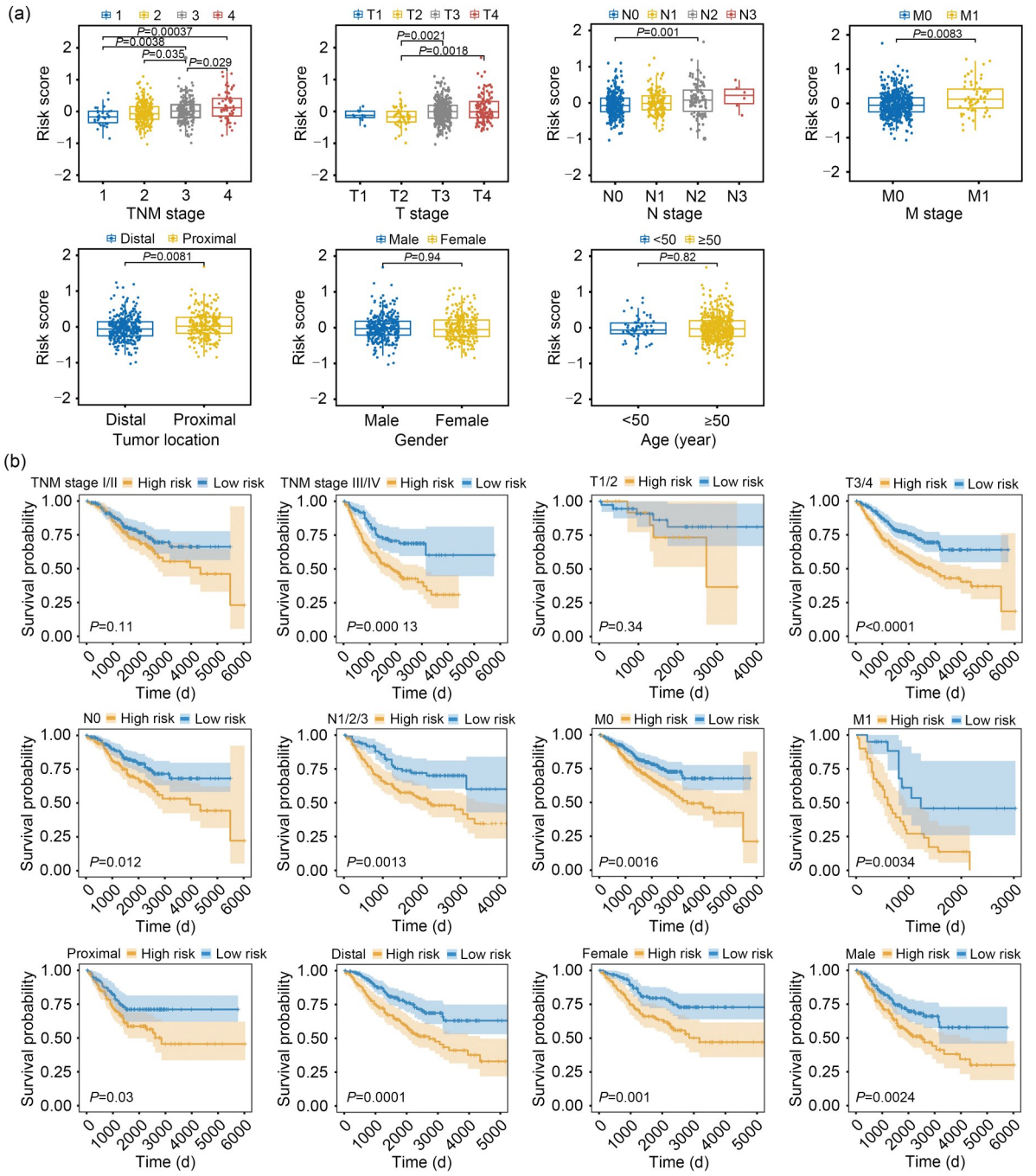

Fig. 2 Correlation between the CSbMgSig and common clinicopathological characteristics. (a) Boxplots show the correlation between CSbMgSig-derived risk scores and several common clinicopathological features of CRC patients, including TNM stage, T stage, N stage, M stage, tumor location, gender, and age. *P* values were calculated by the Wilcoxon test. (b) Survival stratification analysis shows the prognostic role of CSbMgSig-derived risk scores within different clinical subgroups divided by TNM stage (stage I/II and stage III/IV), T stage (T1/2 and T3/4), N stage (N0 and N1/2/3), M stage (M0 and M1), tumor location (distal and proximal), and gender (female and male). CSbMgSig: chemokine signaling-based multigene signature; CRC: colorectal cancer.

low-risk patients in the following subgroups: TNM stage III/IV, T3/4 stage, N stage (N0 and N1/2/3), M stage (M0 and M1), tumor location (primary and distal), and sex (female and male). These results indicated that the established CSbMgSig had a strong predictive ability for the prognosis of CRC patients in these clinical subgroups.

To confirm the prognostic performance of the constructed CSbMgSig, we divided the CRC patients from the GSE17536 cohort into high- and low-risk groups based on the median risk score calculated by the established signature formula (Fig. 3a). As expected, the survival status distribution plot and Kaplan-Meier survival curve indicated that CRC patients with higher risk scores had lower survival probability and shorter OS time (Figs. 3b and 3d). The time-dependent ROC curve analysis also indicated that the CSbMgSig had a strong prognostic ability for CRC patients in the GSE17536 dataset (Fig. 3e). Moreover, the expression

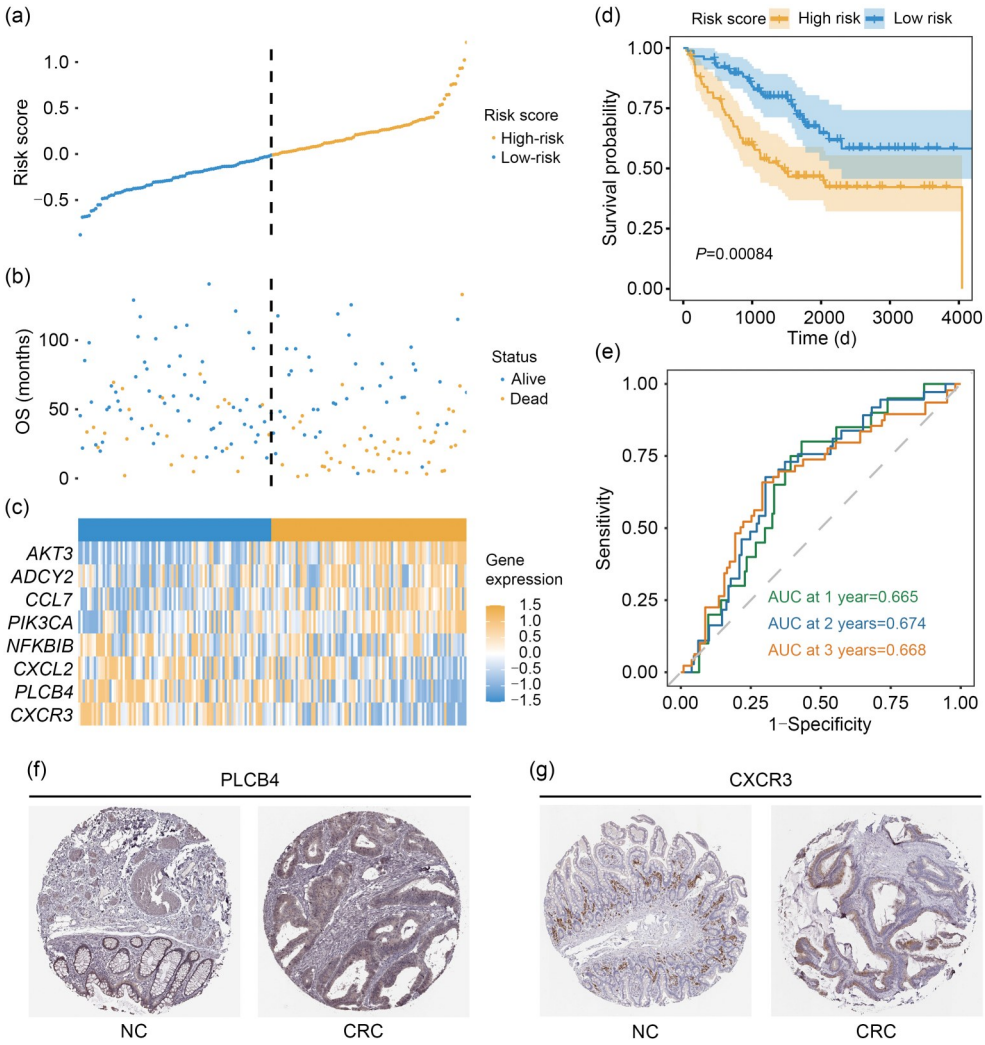

Fig. 3 Validation of the prognostic performance of the CSbMgSig in the GSE17536 dataset. (a) Distribution of the CSbMgSig-based risk scores; (b) Distribution of survival status in CRC patients with high and low risk scores; (c) Expression patterns of the eight CSRGs that make up the CSbMgSig in the high- and low-risk groups; (d) Kaplan-Meier survival analysis of the OS difference between CRC patients in the high- and low-risk groups; (e) Time-dependent ROC curve analysis of the ability of CSbMgSig to predict OS of CRC patients; (f, g) Protein expression levels of the representative CSRGs PLCB4 and CXCR3 in NC and CRC tissues from the Human Protein Atlas database. CSbMgSig: chemokine signaling-based multigene signature; CRC: colorectal cancer; CSRGs: chemokine signaling-related genes; OS: overall survival; ROC: receiver operating characteristic; *AKT3*: protein kinase B γ; *ADCY2*: adenylate cyclase 2; *CCL7*: chemokine (C-C motif) ligand 7; *PIK3CA*: phosphatidylinositol 3-kinase catalytic subunit α; *NFKBIB*: nuclear factor-κB inhibitor β; *CXCL2*: CXC motif chemokine ligand 2; *PLCB4*: phospholipase C β4; *CXCR3*: CXC chemokine receptor 3; NC: normal control.

levels of the eight genes involved in the signature were obviously different between high- and low-risk patients (Fig. 3c), and the expression patterns of the representative signature-related genes (PLCB4 and CXCR3) were validated in CRC and normal control tissues utilizing the Human Protein Atlas (HPA) database (https://www.proteinatlas.org) (Figs. 3f and 3g). These results implied that the CSbMgSig had a robust ability to predict clinical outcomes of CRC patients. In addition, to further examine whether the prognostic power of the CSbMgSig was independent of other clinicopathological factors, we conducted multivariate Cox regression analysis for the established signature and clinical factors including age, gender, and TNM stage. As shown in Fig. S3, the CSbMgSig-based risk score was strongly associated with OS of CRC patients in both the GSE39582 and GSE17536 datasets, indicating that the CSbMgSig was an independent prognostic factor for CRC patients.

To investigate the potential biological processes and pathways related to the CSbMgSig, we firstly identified 311 DEGs between high-and low-risk groups in the training cohort and performed KEGG pathway and gene ontology (GO) functional enrichment analyses on the identified DEGs. As shown in Fig. S4a, the results of GO analysis showed that these DEGs were significantly enriched in GO terms related to cell migration and chemotaxis. In addition, KEGG pathway analysis indicated that these DEGs were primarily enriched in multiple hallmark biological pathways involved in CRC pathogenesis, such as extracellular matrix (ECM)-receptor interaction and the NF-κB and chemokine signaling pathways (Fig. S4b). Moreover, gene set enrichment analysis (GSEA) indicated that multiple tumor hallmarks, e.g., angiogenesis, epithelial-mesenchymal transition, inflammatory response, interferon-γ (IFN-γ) response, and kirsten ratsarcoma viral oncogene homolog (KRAS) signaling, were closely associated with high-risk CRC patients (Fig. S4c), while cancer-related pathways, e.g., DNA repair, E2F targets, G2/M checkpoint, myelocytomatosis viral oncogene homolog (MYC) target, oxidative phosphorylation, and peroxisome, were significantly enriched in the low-risk group (Fig. S4d).

Considering the critical role of immunity in cancer development and progression, we first determined the correlation between the CSbMgSig-derived risk score and immune status in the GSE39582 cohort through single-sample GSEA (ssGSEA) enrichment analysis. As shown in Fig. 4a, the scores of multiple immune cell types, such as antigen-loaded dendritic cells (aDCs), $CD8^+$ T cells, B cells, helper T cell 2 (Th2) cells, macrophages, tumor infiltrating lymphocytes (TILs), and regulatory T cells (Tregs), were dramatically increased in the high-risk group (all adjusted $P<0.05$). Moreover, the scores of several pivotal immune-related functions, e.g., antigen presenting cell (APC) co-stimulation, check-point, inflammation-promoting, T cell co-inhibition, and IFN response, were also significantly elevated in the high-risk group (all adjusted $P<0.05$; Fig. 4b). The link between risk score and molecular subtype was determined in the GSE39582 dataset to investigate its association with immune status. Molecular subtypes C1 and C3 were associated with down-regulated immune pathways, while C2 was correlated with up-regulated immune pathways (Marisa et al., 2013). As shown in Fig. 4c, the risk score of the C2 subtype was significantly higher than those of the C1 and C3 subtypes, indicating the close relationship between risk score and immune status. Also, the results of correlation analyses showed that the risk score was significantly correlated with the immune score and stromal score (Figs. 4d and 4e), which were tightly linked to immune infiltration in the tumor microenvironment. In addition, as immune checkpoints have been demonstrated as an important target for CRC immunotherapy, we further examined the expression levels of well-known immune checkpoints between high- and low-risk groups in the GSE39582 dataset. As shown in Fig. 4f, cytotoxic T lymphocyte-associate protein-4 (*CTLA-4*) and programmed cell death-1 (*PD-1*) were dramatically increased in patients with low risk scores (both $P<0.05$), and programmed cell death-ligand 1 (*PD-L1*) and *PD-L2* were significantly decreased in patients in the low-risk group (both $P<0.05$). The results were further confirmed through correlation analyses (Fig. 4g), supporting the important role of the CSbMgSig in mediating immune response.

Considering the important role of chemotherapy in CRC treatment (Ichikawa et al., 2020; Glimelius et al., 2021), we assessed the chemotherapeutic responses of high- and low-risk patients based on the GDSC database, which contains sensitivities and genomic profiles of large-scale anti-tumor compounds (Yang et al., 2013). Notably, 53 chemotherapeutic drugs had significant differences in the half-maximal inhibitory concentration (IC_{50}) between high- and low-risk

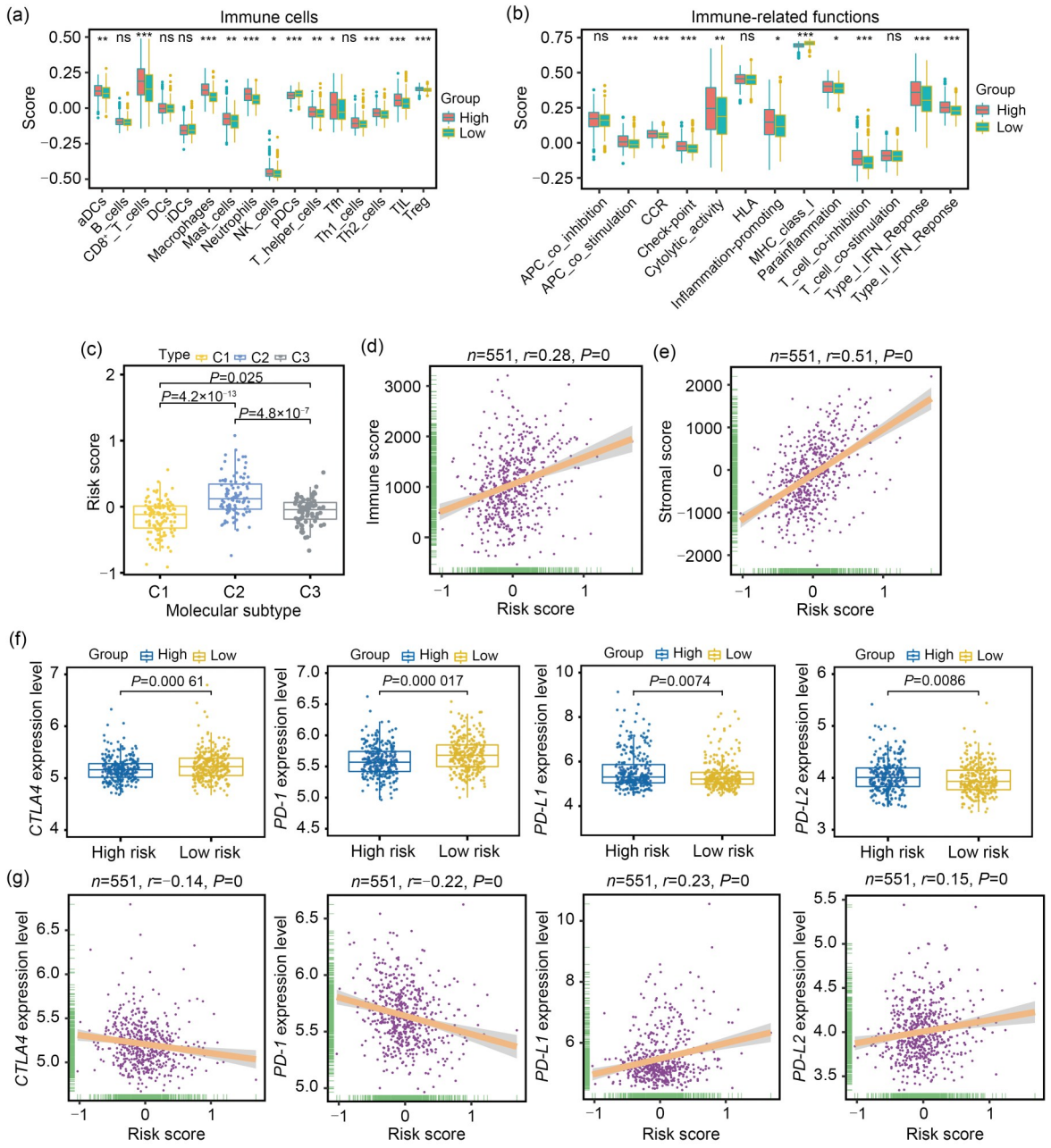

Fig. 4 Correlation between CSbMgSig and immune status of CRC patients in the GSE39582 dataset. (a, b) Comparison of the ssGSEA scores of immune cells and immune-related functions within each CSbMgSig risk group. *P* values were calculated with the Wilcoxon test. * *P*<0.05; ** *P*<0.01; * *P*<0.001; ns: not significant, *P*>0.05. (c) Comparison of the risk scores in different immune-related molecular subtypes. *P* values were calculated with the Wilcoxon test. (d, e) Pearson correlation analysis shows the relationship between risk score and immune score (d) or between risk score and stromal score (e). (f) Comparison of the expression levels of *CTLA-4*, *PD-1*, *PD-L1*, and *PD-L2* between different risk groups. *P* values were calculated by the Wilcoxon test. (g) Pearson correlation analysis shows the relationship between risk score and expression levels of *CTLA-4*, *PD-1*, *PD-L1*, and *PD-L2*. CSbMgSig: chemokine signaling-based multigene signature; CRC: colorectal cancer; ssGSEA: single-sample gene set enrichment analysis; CTLA-4: cytotoxic T lymphocyte associate protein-4; PD-1: programmed cell death-1; PD-L1: programmed cell death-ligand 1.**

patients in the GSE39582 and GSE17536 cohorts. Of these, 16 anti-tumor compounds have been proven to be chemotherapeutic drugs for CRC patients (Fig. S5). Therefore, these data indicated that the established CSRG signature provides potential candidates for personalized treatment of CRC patients.

In summary, we established a novel CSbMgSig as a promising tool for prognostic risk assessment in CRC patients. The prognostic signature was confirmed to be independently associated with OS of CRC patients in both training and validation cohorts. In particular, functional analysis demonstrated that the signature plays a key role in immune infiltration and drug response of CRC patients. Therefore, evaluating the CSbMgSig-based risk score of individual CRC patients will not only contribute to risk stratification and OS prediction, but also provide valuable insights into therapeutic efficacy.

Materials and methods

Detailed methods are provided in the electronic supplementary materials of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 31900490 and 31770903) and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 19KJB180027).

Author contributions

Xin QI designed the research and drafted the manuscript. Xin QI, Donghui YAN, Jiachen ZUO, and Rui WANG collected the data and performed the computational analyses. Xin QI, Donghui YAN, Jiachen ZUO, Rui WANG, and Jiajia CHEN revised the manuscript. All authors have read and approved the final manuscript, and therefore, have full access to all the data in the study and take responsibility for the integrity and security of the data.

Compliance with ethics guidelines

Xin QI, Donghui YAN, Jiachen ZUO, Rui WANG, and Jiajia CHEN declare that they have no conflict of interest.

This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Cabrero-de Las Heras S, Martínez-Balibrea E, 2018. CXC family of chemokines as prognostic or predictive biomarkers and possible drug targets in colorectal cancer. *World J Gastroenterol*, 24(42):4738-4749. https://doi.org/10.3748/wjg.v24.i42.4738

Do HTT, Lee CH, Cho J, 2020. Chemokines and their receptors: multifaceted roles in cancer progression and potential value as cancer prognostic markers. *Cancers*, 12(2):287. https://doi.org/10.3390/cancers12020287

Glimelius B, Stintzing S, Marshall J, et al., 2021. Metastatic colorectal cancer: advances in the folate-fluoropyrimidine chemotherapy backbone. *Cancer Treat Rev*, 98:102218. https://doi.org/10.1016/j.ctrv.2021.102218

Ichikawa N, Kamiyama T, Yokoo H, et al., 2020. Preoperative chemotherapy in colorectal cancer patients with synchronous liver metastasis. *Mol Clin Oncol*, 12(4):374-383. https://doi.org/10.3892/mco.2020.1992

Kamal Y, Schmit SL, Hoehn HJ, et al., 2019. Transcriptomic differences between primary colorectal adenocarcinomas and distant metastases reveal metastatic colorectal cancer subtypes. *Cancer Res*, 79(16):4227-4241. https://doi.org/10.1158/0008-5472.CAN-18-3945

Koncina E, Haan S, Rauh S, et al., 2020. Prognostic and predictive molecular biomarkers for colorectal cancer: updates and challenges. *Cancers*, 12(2):319. https://doi.org/10.3390/cancers12020319

Kurzejamska E, Sacharczuk M, Landázuri N, et al., 2019. Effect of chemokine (C-C motif) ligand 7 (CCL7) and its receptor (CCR2) expression on colorectal cancer behaviors. *Int J Mol Sci*, 20(3):686. https://doi.org/10.3390/ijms20030686

Lacalle RA, Blanco R, Carmona-Rodríguez L, et al., 2017. Chapter five─chemokine receptor signaling and the hallmarks of cancer. *Int Rev Cell Mol Biol*, 331:181-244. https://doi.org/10.1016/bs.ircmb.2016.09.011

Marisa L, de Reyniès A, Duval A, et al., 2013. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med*, 10(5):e1001453. https://doi.org/10.1371/journal.pmed.1001453

Ottaiano A, Scala S, Normanno N, et al., 2020. Prognostic and predictive role of CXC chemokine receptor 4 in metastatic colorectal cancer patients. *Appl Immunohistochem Mol Morphol*, 28(10):755-760. https://doi.org/10.1097/PAI.0000000000000828

Poeta VM, Massara M, Capucetti A, et al., 2019. Chemokines and chemokine receptors: new targets for cancer immunotherapy. *Front Immunol*, 10:379. https://doi.org/10.3389/fimmu.2019.00379

Qi X, Wang R, Lin YX, et al., 2021. A ferroptosis-related gene signature identified as a novel prognostic biomarker for colon cancer. *Front Genet*, 12:692426. https://doi.org/10.3389/fgene.2021.692426

Sung H, Ferlay J, Siegel RL, et al., 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 71(3):209-249. https://doi.org/10.3322/caac.21660

Yang W, Soares J, Greninger P, et al., 2013. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res*, 41(D1):D955-D961. https://doi.org/10.1093/nar/gks1111

Yu L, Yang XM, Xu CC, et al., 2020. Comprehensive analysis of the expression and prognostic value of CXC chemokines in colorectal cancer. *Int Immunopharmacol*, 89:107077. https://doi.org/10.1016/j.intimp.2020.107077

Supplementary information

Materials and methods; Figs. S1─S5