IVYSPRING INTERNATIONAL PUBLISHER

Journal of Cancer

2025; 16(7): 2375-2387. doi: 10.7150/jca.108188

Research Paper

Construction of a Prognostic Model based on CSC-related Genes in Patients with Colorectal Cancer

Zi-Yue Li[1,2*], Ming-Feng Li[1*], Ying-Ying He[3*], Guan-Sheng Zheng[4], Jie-Rong Chen[4], Yun-Miao Guo[1], Qizhou Lian[2]✉, Cai-Feng Yue[5]✉

1. Zhanjiang Institute of Clinical Medicine, Central People's Hospital of Zhanjiang, Guangdong Medical University Zhanjiang Central Hospital, Zhanjiang 524045, China.
2. Cord Blood Bank, Guangzhou Institute of Eugenics and Perinatology, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou 510623, China.
3. Department of Anesthesiology, Central People's Hospital of Zhanjiang, Guangdong Medical University Zhanjiang Central Hospital, Zhanjiang 524045, P. R. China.
4. Department of Laboratory Medicine, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou 510080, China.
5. Department of Laboratory Medicine, Central People's Hospital of Zhanjiang, Guangdong Medical University Zhanjiang Central Hospital, Zhanjiang 524045, China.

*These authors contributed equally to this work.

✉ Corresponding authors: Cai-Feng Yue: caifengyue@163.com and Qizhou Lian: qz.lian@siat.ac.cn.

Received: 2024.12.04; Accepted: 2025.03.06; Published: 2025.04.13

Abstract

Colorectal cancer (CRC) is one of the most common and deadly malignancies. Lack of efficient biomarkers for prognosis has limited the improvement of survival outcome in patients with CRC. Numerous studies have demonstrated the important roles of cancer stem cells (CSCs) in both treatment resistance and disease recurrence of CRC. Thus, the current study aims to construct a prognostic model based on expression level of CSC-related genes for precise molecular subtyping of CRC patients with different prognoses, TME infiltration patterns and therapeutic responses. The RNA sequencing data and clinical information were obtained from UCSC Xena database, followed by identification of differential expressed genes, univariate Cox regression, and LASSO regression to identify prognostic CSC-related genes and construct a novel prognostic risk scoring model consisting of 21 CSC-related genes. The patients in high-risk group suffered poor survival outcome (P<0.0001). Moreover, the performance of CSC-related prognostic model was validated in individual GEO datasets including GSE41258 and GSE39582 (P<0.05). Furthermore, patients with high-risk score exhibited lower response rate to immune checkpoint inhibitors as compared to those in low-risk group (17.4% vs. 28.2%), indicating the potential of CSC-related prognostic model to predict the immunotherapy response. Collectively, our findings provide an effective model to predict the immunotherapy response and survival outcome in patients with CRC.

Keywords: Colorectal cancer, Cancer stem cells, Prognosis, Tumor microenvironment, Immune checkpoint inhibitors

Introduction

Colorectal cancer (CRC) is currently the third most common malignancy and the second leading cause of cancer-related mortality worldwide, with 1,931,590 newly diagnosed cases and 935,173 deaths from cancer in 2020[1]. The long-term outcomes of patients with CRC have substantially improved due to considerable evolvement of surgical treatment, chemotherapy, and immunotherapy[2]. However, approximately one-fourth of patients present with distant metastases at the time of diagnosis and additional 25-50% of patients diagnosed at early stages subsequently develop metastatic diseases, which are major causes of treatment failure and thus poor prognosis[3]. Therefore, there is an urgent need

to identify effective biomarkers or indicators for treatment guidance and prognosis prediction in patients with CRC.

Cancer stem cells (CSCs) are a small subset of cancer cells with the ability to self-renew and dedifferentiate, which are critical for initiating and sustaining the growth of tumor[4]. The aberrant expression of CSC-related genes is supposed to play important roles in regulating the proliferation, metastasis, and therapeutic resistance of tumor cells, and thus shed a new light on the CSC-targeted therapies of tumor[5]. Recently, the gene expression-based stemness index (mRNAsi) was utilized for identification of therapeutic targets and precise prognosis in multiple cancers including gastric cancer[6,7], head and neck squamous cell carcinomas[8,9], prostate adenocarcinoma[10], esophageal cancer[11], bladder cancer[12], lung cancer[13], and glioma[14]. Collectively, the findings have revealed the potential of CSC-related genes as biomarkers to satisfy the unmet need for risk stratification and treatment optimization in patients with cancer. However, the relationship between the CSC-related genes and outcome in patients with CRC has been rarely explored.

In the current study, a prognostic model based on expression level of CSC-related genes was established for precise treatment planning and accurate prognosis of patients with CRC.

Materials and Methods

Data acquisition

The gene-level copy number data (SNP6.0 array), DNA methylation (Methylation 450K array), mRNA and miRNA expression data (z-score normalized), list of somatic mutations (including SNPs and INDELs) and copy number variations (CNV, including AMP and DEL), reverse phase protein array (RPPA) data, stemness scores (DNA methylation based and RNA expression-based), immune signature scores, and corresponding phenotype data of TCGA Pan-Cancer (PANCAN) cohort were collected by using UCSC Xena[15]. The different sets of transcript expression data were re-calculated and normalized by using UCSC TOIL recompute pipeline. After exclusion of subjects diagnosed under age 18, a total of 12,591 subjects and 33 cancer types were enrolled for further analyses. In the TCGA-COAD cohort, patients with age under 18, relapsed/secondary tumors, ambiguous and/or missing clinical and follow-up data were excluded. Eventually, a total of 450 patients were included for subsequent analyses. Moreover, eight expression profile datasets including GSE13507, GSE4412, GSE21653, GSE41258, GSE84437, GSE42127, GSE23554, GSE57495, and GSE39582 were downloaded from Gene Expression Omnibus (GEO) database[16] as validation sets. Furthermore, the clinical characteristics and RNA expression data of patients with urothelium carcinoma in IMvigor 210 cohort were downloaded by using R package IMvigor210CoreBiologies to assess the response to immunotherapy[17]. Responders are referred to as patients with complete remission (CR), or partial remission (PR), whereas non-responders are defined as those with stable disease (SD) or progressive disease (PD). The infiltration of immune cells and response to therapies were evaluated by using CIBERSORT[18] algorithm, ESTIMATE[19] algorithm, TIDE[20] algorithm, and GDSC[21] database, respectively. The list of CSC-related genes was obtained by searching in the molecular signatures database (MSigDB)[22], cancer stem cells database (CSCdb)[23], and published literatures.

Classification of molecular subtypes

CSC-related genes associated with survival of cancer patients were identified by univariate Cox regression analysis, followed by consensus clustering using R package ConsensusClusterPlus[24]. The relative change in area under the CDF curve was evaluated to determine the optimal k value and thus the number of clusters. The difference in length of survival time between distinct molecular subtypes was assessed by weighted log-rank test, and Kaplan-Meier (K-M) curves were plotted by using R package survival. The hazard ratio (HR) and P were calculated by using Cox regression analyses between molecular subtypes with the most favorable or poorest prognosis in multiple cancers, followed by validation in additional GEO datasets.

Identification of Differentially Expressed Genes (DEGs) among different subtypes

DEGs between molecular subtypes with either best or poorest prognosis were identified by using R package limma[25] according to the threshold of $|\log_2$ fold change (FC)$| \geq 1$ and false discovery rate (FDR) < 0.05. Subsequently, R packages EnhancedVolcano and pheatmap were employed to visualize the results of differential expression analyses.

Construction and validation of prognostic model based on CSC-related genes

Candidate genes represented in both lists of DEGs among different subtypes and CSC-related genes were further analyzed by using univariate Cox regression, and genes with $P \leq 0.01$ were identified as prognosis-associated genes. Subsequently, Lasso

regression was applied to perform dimensionality reduction and establish the prognostic model. The risk score for each patient with CRC was calculated according to the following formula, in which C_j represents the regression coefficient for gene j and exp_{ij} represents the expression of gene j in sample i.

$$Risk\ score_i = \sum_{j=1}^{n} C_j * exp_{ij}$$

The same formula was used in both the training set and external validation cohorts. Patients were assigned to low-risk or high-risk subset using the median of risk scores as threshold. Kaplan-Meier (K-M) curves and log-rank tests were applied to assess the difference in outcome of patients. R package timeROC was employed to generate ROC curve and calculate area under time dependent ROC curve (AUC). The risk scores of 298 patients from IMvigor210CoreBiologies dataset were calculated to evaluate their predictive ability of immunotherapeutic responsiveness by using Kruskal-Wallis test.

Statistical analysis

Statistical analyses were carried out by using R (version 4.1.2). Statistical significance between two groups was tested using Student's t-test. For variables more than three groups, a one-way analysis of variance or the Kruskal–Wallis test was used, depending on the type of data. Correlation coefficients were calculated using Spearman's correlation analysis. $P < 0.05$ was considered to indicate a significant difference, unless otherwise stated.

Results

Identification of Molecular Subtypes based on CSC-related Genes

Thirty-four out of 206 selected CSC-related genes were differentially expressed in CRC tissues as compared with normal tissues, as well as in other cancers cataloged in TCGA database (Figure S1, Supporting Information). On the basic of the expression level of 34 differentially expressed genes (DEGs), patients with CRC were divided into four

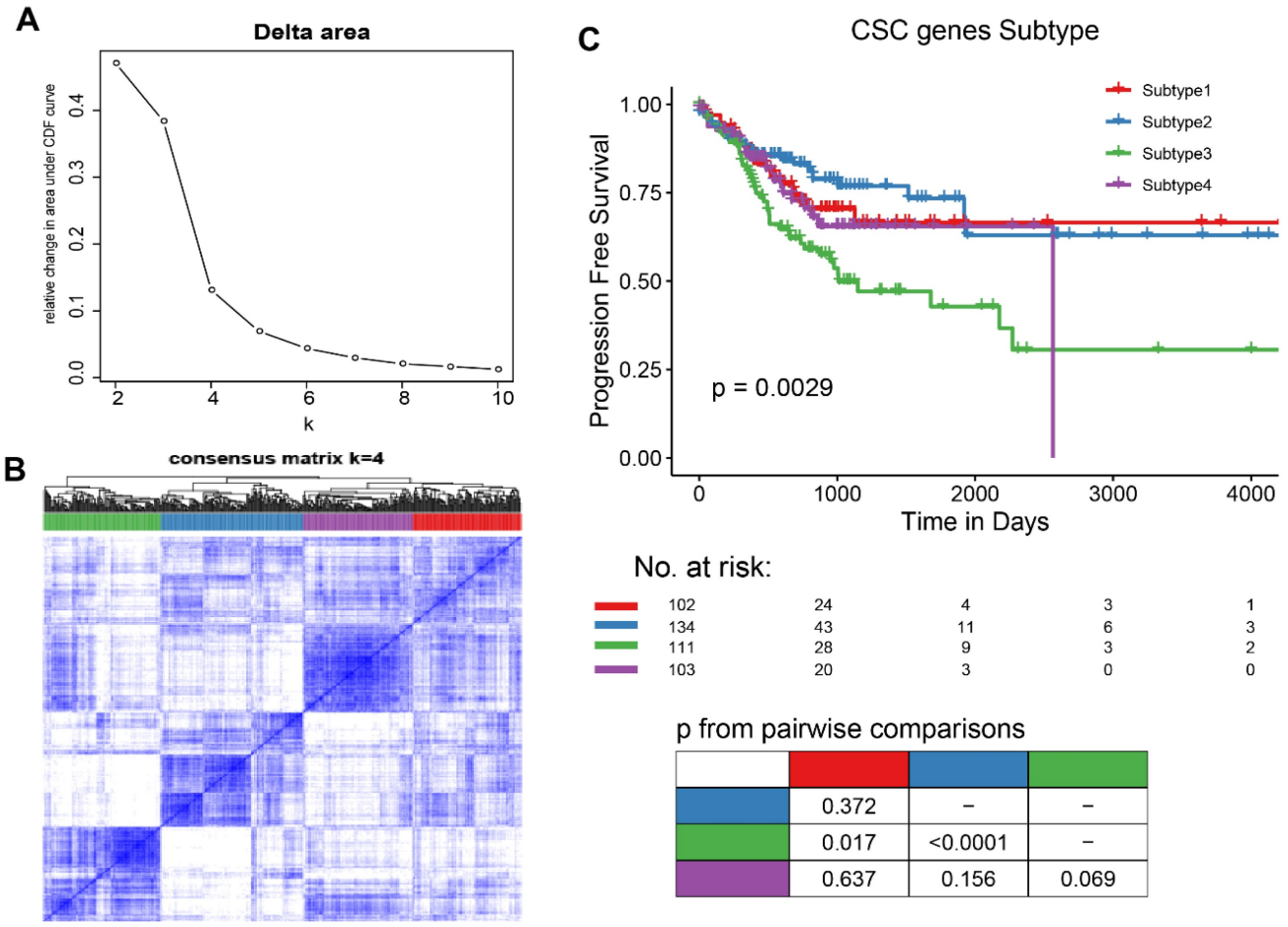

Figure 1. Identification of CSC-related subtypes by K-means analysis. (A-B). K = 4 was identified as the optimal value for consensus clustering, the patients were divided into 4 distinct gene clusters. C. Kaplan-Meier survival curve showing survival probability for the 4 subtypes.

molecular subtypes with different lengths of progression free survival time (Figure 1A-B), patients in subtype 2 had significantly better clinical outcome than those in subtype 3(Figure 1C). Furthermore, the differences in survival outcome were observed between subtypes classified based on the expression of DEGs in patients with diverse types of cancer in TCGA database (Figure S2A, Supporting Information) and GEO datasets (Figure S2B, Supporting Information).

Characteristics of CSC-related clusters for COAD

To explore the characteristics of CSC-related clusters, immune infiltration levels of 22 immune cells among the 4 subtypes in COAD were obtained from known studies and shown in Figure 2A. Moreover, clinical characteristics including the number of lymph nodes, whether MMR is deficient (dMMR), MSI statues treatment statues, pathological stage, TNM stage, age, and gender were interrogated among different subtypes (Figure 2B). A total of 1065 differentially expressed genes (DEGs) were identified ($|\log_2 FC| \geq 1$, $P < 0.05$), and the volcano map accurately reflected the gene expression differences between subtype 2 and subtype 3 (Figure 2C). The top 50 differentially expressed genes among different subtypes was shown in heatmap (Figure 2D).

Construction of CSC-related genes signature for COAD

The differential expressed genes were combined with CSC-related genes. After Univariate Cox regression analysis and least absolute shrinkage and selection operator (LASSO) regression analysis (Figure 3A-B), 15 potential pro-oncogenes (HR > 1, *INTS3, LINGO1, GRB7, PLXNB3, PTPRN, GRP, FABP4, C6orf15, DKK1, CALB2, RARG, PCOLCE2, GADD458, L1CAM, INHBA*) and 6 potential suppressor genes (HR < 1, *HSPB7, TNS1, DPYSL4, ISM1, FABP5, SPEG*) were identified (Figure 3C). The above CSC-related genes were used to develop the risk score prognostic signature, and the risk score for each COAD sample was calculated according to the following formula: coefficient × Expr (*INTS3, LINGO1, GRB7, PLXNB3, PTPRN, GRP, FABP4, C6orf15, DKK1, CALB2, RARG, PCOLCE2, GADD458, L1CAM, INHBA, HSPB7, TNS1, DPYSL4, ISM1, FABP5, SPEG*). Patients were divided into high-risk and low-risk groups according to the median risk

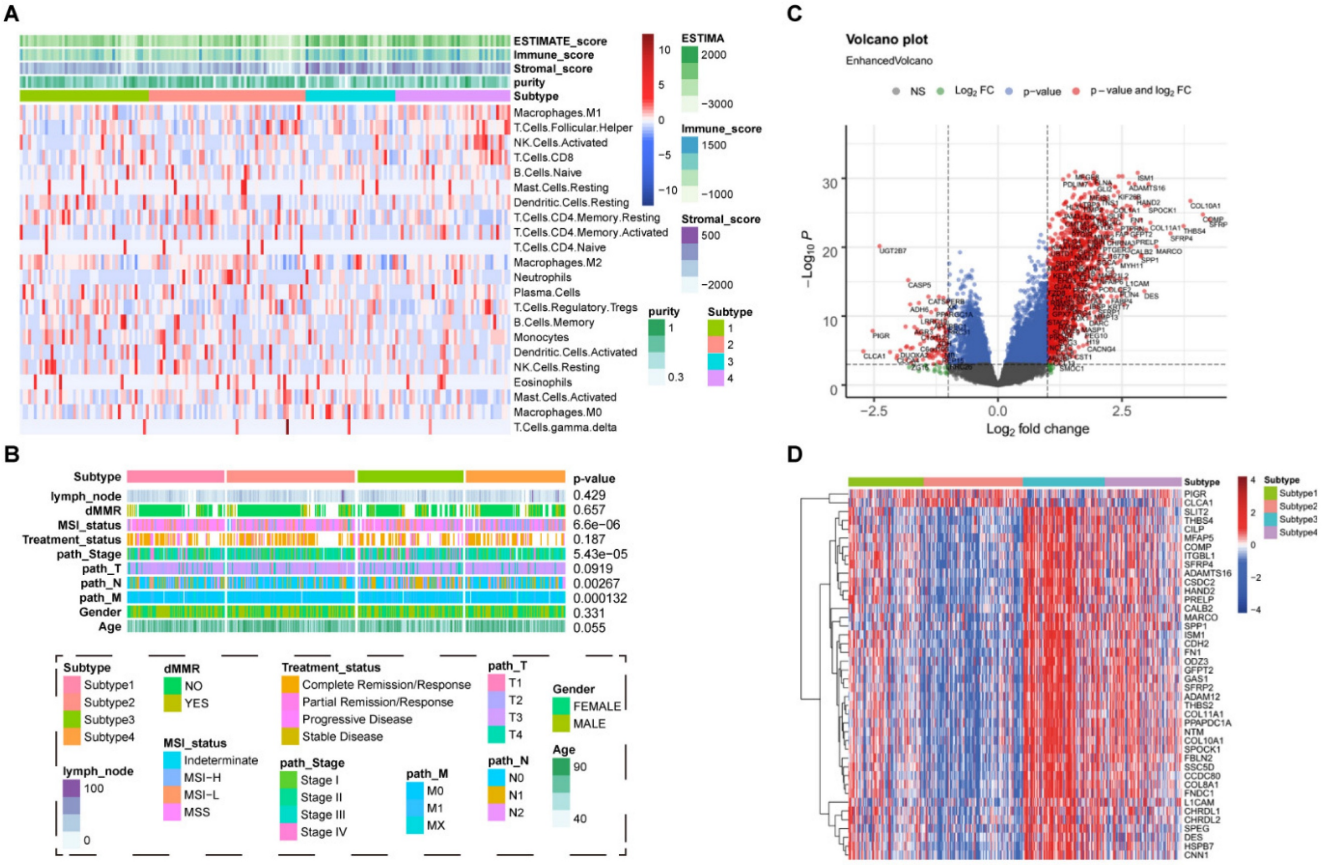

Figure 2. Characteristics of CSC-related clusters for COAD. (A). The immune infiltration levels of 22 immune cells among the 4 subtypes in COAD. (B). Heatmap showing the 4 subtypes in different clinical characteristics and clusters. (C). The volcano map reflects the differential expressed genes identified ($|\log_2 FC| > 1$ and $P < 0.05$). (D). Heatmap showing the top 50 differential expressed genes among the 4 subtypes.

score. The high-risk group had significant worse clinical outcomes (PFS: $P<0.0001$, Figure 3D; OS: $P=0.0038$, Figure S3, Supporting Information). Risk score curve plot and curve plot were shown in Figure 3E-F. The survival ROC curves predicted by the signature showed that the AUCs were all greater than 0.8, indicating the effectiveness of the CSC-related signature in predicting prognosis for COAD at the 1-year (AUC=0.64), 3-year (AUC=0.7), and 5-year (AUC=0.67) time points (Figure 3G). Heatmap displayed the distribution of 21 genes in the prognostic signature between the two groups (Figure 3H).

Validation of the prognostic signature in GEO Cohort

To validate the performance of the CSC-related signature in predicting OS, risk scores were calculated with the same formula for patients in GSE41258 and GSE39582. Similarly, the survival curve in GEO cohort also demonstrated that the high-risk group showed a poor overall survival compared to the low-risk group (Figure 4). Moreover, the survival ROC curves showed good effectiveness in predicting prognosis (Figure 4).

Immunotherapy response prediction

The results based on the use of the Imvigor210CoreBiologies dataset showed that patients in the high-risk group exhibited no adverse OS compared to those in the low-risk group ($\boldsymbol{P} = 0.31$, log rank test; Figure 5A). However, the response rate to ICIs was significantly higher in the low-risk group than that in the high-risk group (28.2% vs. 17.4%, respectively; Figure 5B). Concurrently, non-responders to ICIs (SD + PD) presented with higher risk scores than responders (CR + PR, Figure 5C). This finding indicates that the risk score can be used as a prognostic marker of the immune response.

The landscape of CSC-related score in pan-cancers

The CSC-related score was calculated among all types of cancers and shown in Figure 6A. Samples with CNV had significantly higher CSC-related score than those without (Figure 6B). The CSC-related score showed a correlation with CNV in pan-cancers (Figure 6C). For example, the KIRP patients with AMP had a significant higher CSC-related score. Meanwhile, the HNSC patients with DEL had a significant higher CSC-related score. Genome-wide variation with CNV and somatic mutation was shown as the CSC-related score increased in GI cancers, including COAD (Figure 6D) and STAD (Figure 6E).

Survival analyses of CSC-related signature in pan-cancers

Univariate Cox regression analysis and multivariate Cox regression analysis (adjusted for age, gender, and tumor grade) were applied to calculate the risk of CSC-related score on patient survival time (including PFS and OS, Figure 7A-B). The samples were divided into high-risk and low-risk groups according to the median CSC-related score. Kaplan-Meier curves for progression-free survival (PFS, Figure 7C-J) and overall survival (OS, Figure S4, Supporting Information) in pan-cancers were significant, the low-risk group had a higher survival rate.

Correlation of CSC-related score with immune characteristics and stemness score in pan-cancer

The Immune characteristics between high-risk and low-risk groups were demonstrated, including differences in Immune score, TIDE score and TMB (Figure 8A-C). Meanwhile, the correlation between CSC-related score and tumor stemness index (including mRNAsi, EREG-mRNAsi and mDNAsi) were presented in scatter plots (Figure 8D).

Marker counts of mRNA, miRNA, protein, mutation, SCNV and drug sensitivity analysis between groups based on CSC-related score

We utilized a performance score algorithm in pan-cancer using logistic regression analysis, corrected for clinical factors (including confounding factors such as age), and screened genes for which confounding factors were balanced between the two groups (Methods and Materials). For mRNA, 19,793 marker genes were screened; For miRNA, 743 marker genes were screened. For protein level, 214 marker genes were screened; For the mutation level, 135 marker genes were screened; For SCNV, 1671 marker genes were screened. Then, we calculated and obtained differentially expressed marker genes at mRNA, miRNA, protein, mutation and SCNV levels according to the high-risk and low-risk groups (Methods and Materials) (Figure 9A). The Ratio of characteristic marker genes with significant differences between high and low groups was calculated.

For mRNA marker genes in the high CSC-related score group (genes present in at least 10 cancer types), we performed drug sensitivity analysis based on cell line drug response data. A total of 104 marker genes was associated with 141 significant drugs ($|R| \geq 0.3$ and FDR ≤ 0.05, Figure 9B). In addition, we found that most of the mRNA marker genes showed a positive

correlation with drug small molecules (R ≥ 0.3 and FDR ≤ 0.05), such as *FN1* and *FLNA* (Figure 9B). The corresponding signaling pathways of drug targets were explored, and a total of 23 were involved, such as DNA replication and WNT signaling (Figure 9B). In addition, we found that drugs related to chromatin histone acetylation pathway showed correlation with the most marker genes (Figure 9B).

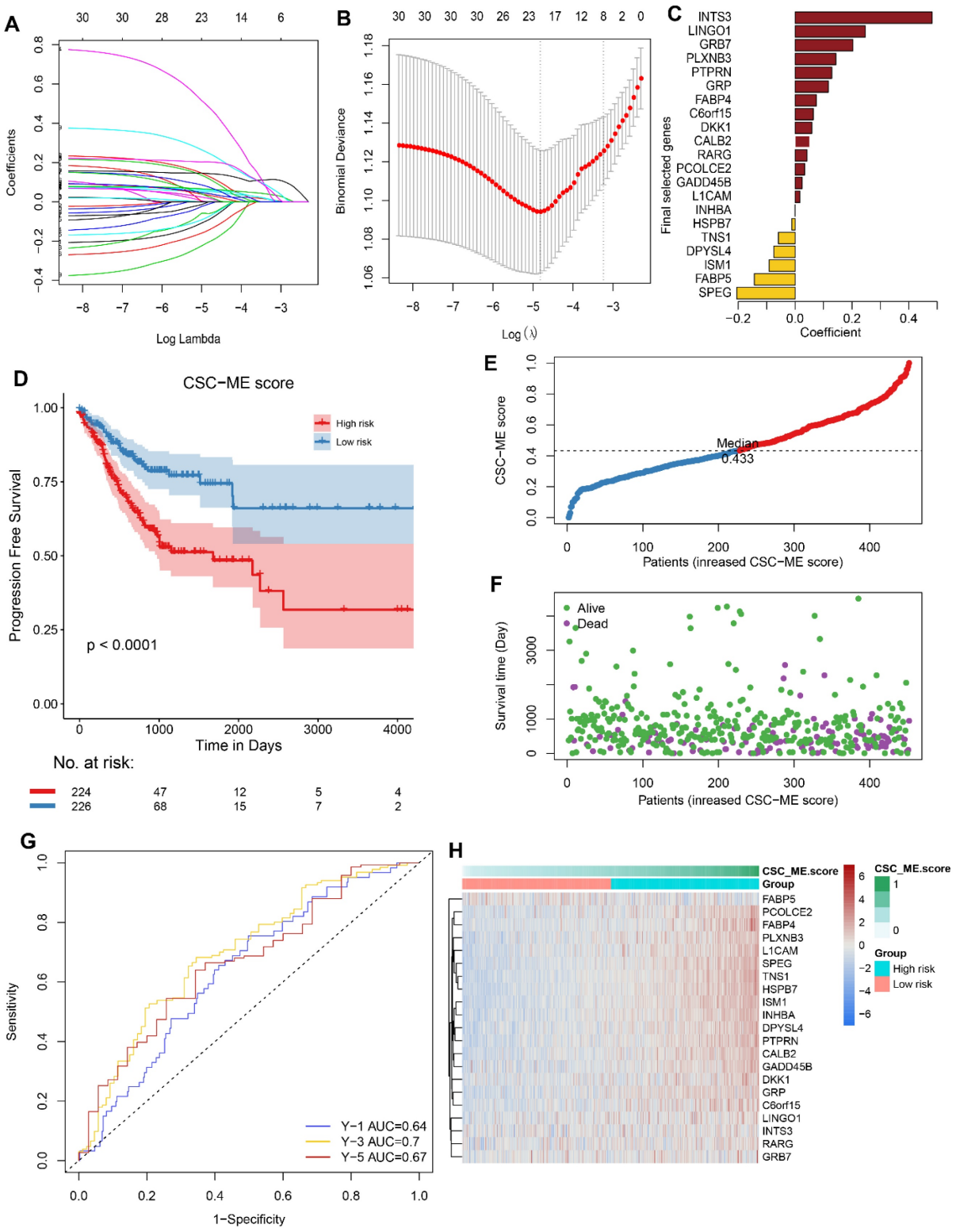

Figure 3. Construction of CSC-related genes signature for COAD. (A, B) The LASSO regression analysis of CSC-related genes associated with prognosis. (C). The coefficient score of the final selected genes. (D). Kaplan-Meier survival curve showing survival probability of high-risk or low-risk subgroups. (E). Risk score curve plot. The dotted line indicates the individual distribution of risk score, and the patients are categorized into low-risk (blue) and high-risk (red) groups. (F). Risk score scatter plot. Purple dots indicate the dead patients, and green dots indicate the alive. With the increase in risk score, more patients died. (G). The 1-year, 3-year, and 5-year survival ROC curves are predicted by the signature. (H). Heatmap showing the distribution of 21 genes in the prognostic signature between the two groups.

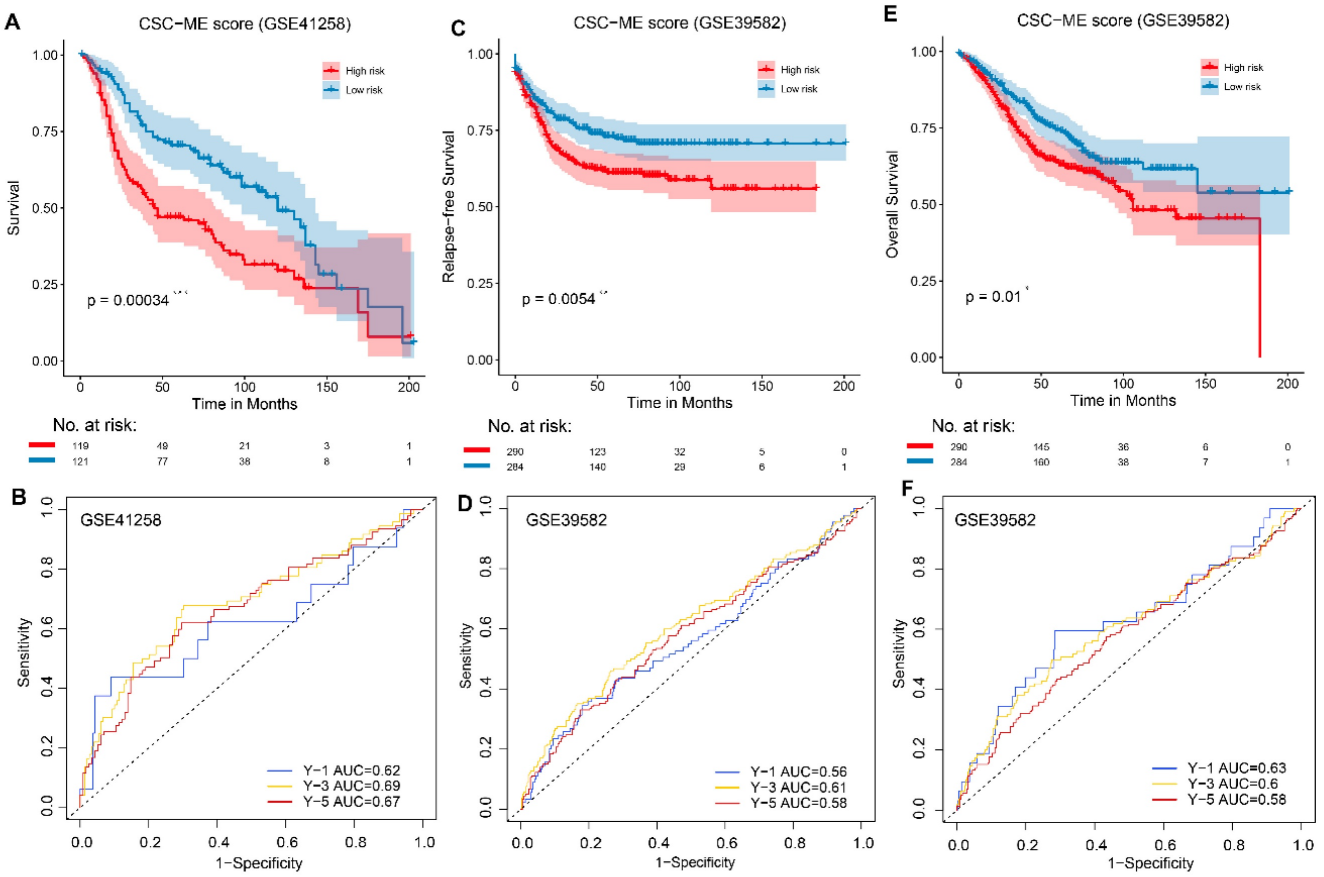

Figure 4. Validation of the constructed prognostic signature in GEO cohorts. (ACE, *, ** and *** stands for ***P*** < 0.05, 0.01 and 0.001 respectively). Kaplan-Meier survival curve showing survival probability of high-risk or low-risk subgroups. (BDF). The 1-year, 3-year, and 5-year survival ROC curves are predicted by the signature.

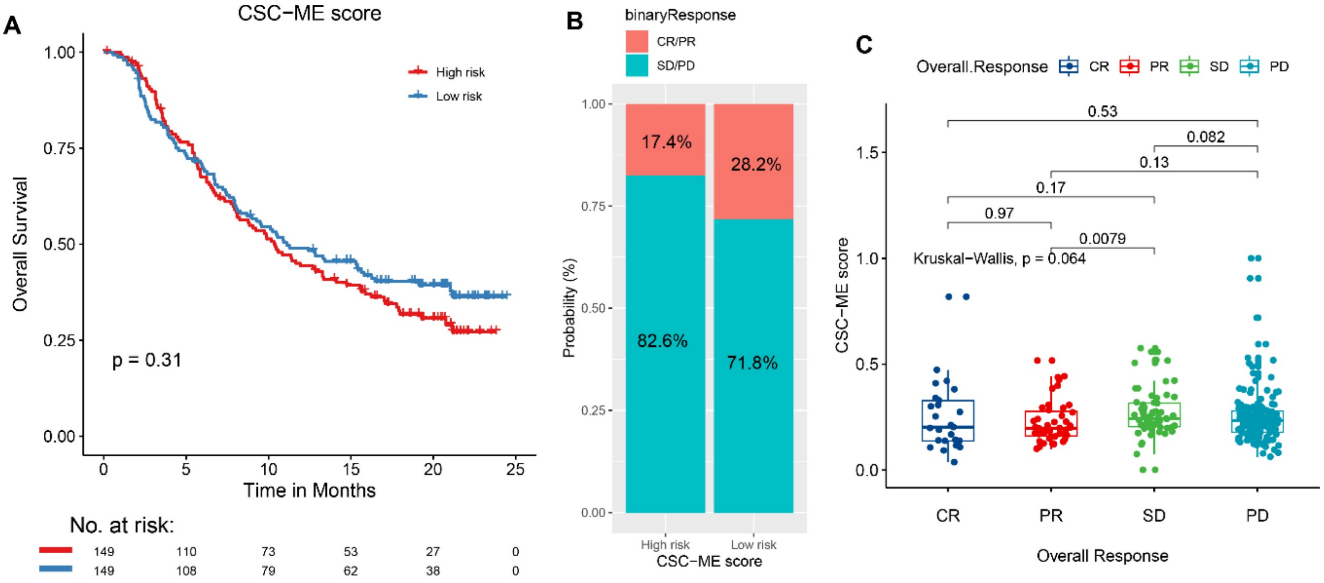

Figure 5. A high-risk score predicts poor response to immune checkpoint inhibitors (ICIs). (A). Overall survival (OS) analysis of high-risk and low-risk groups. (B). Comparison of Immunotherapy response ratio between high-risk and low-risk groups. (C). Comparison of risk scores between different immune response states.

Discussion

CRC is one of the most prevalent malignant tumors worldwide, resulting in high morbidity and mortality[26]. Although CRC might be cured by radical surgery combined with chemo- and radiotherapy, drug resistance, recurrence and metastasis are still the main causes of CRC-associated mortality. Accumulating evidence showing CRC originates from cancer stem cells (CSCs)[27,28]. CSCs

are capable of forming metastatic tumors owing to their proliferative capability[4], and it is acknowledged that CSCs are the main reasons resulting in treatment resistance and disease recurrence in CRC[29], which make them as promising therapeutic targets. In this study, we focused on the treatment planning and prognosis prediction value of CSC-related genes in pan-cancers especially CRC. We suggest that precise molecular subtyping of CSC-related genes would prospectively stratify CRC patients with different prognoses, TME infiltration patterns and therapeutic responses.

Among the 4 diverse molecular subtypes identified by consensus clustering based on the CSC-related genes, subtype 2 had significant better clinical outcome than subtype 3. To further elucidate the expression characteristics of the two subtypes, we performed differential expressed gene analysis, and a total of 1065 differential expressed genes were identified. The DEGs were intersected with CSC-related genes that connected with the lengh of survival by Univariate Cox regression analysis. Then LASSO regression analysis was performed, and a prognosis signature comprising 21 CSC-related genes in CRC was construted. The risk score of each patient was calculated and divided into high-risk and low-risk groups. The high-risk group had significant lower survival time than the low-risk group. With respect to immunotherapy, low-risk patients received better clinical benefits from ICIs when applying our signature to IMvigor210. Then we explored the application of the signature in pan-cancer. Patients with CNV exhibited significantly higher CSC-related scores compared to those without. Patients in high-risk group had better survival rate, immune score and TIDE score.

Previous research has partially elucidated the roles of CSC-related genes in cancer occurrence and development, as well as their potential as targets for cancer treatment. *GRB7*, growth factor receptor-bound protein 7, played an important role in MEKi resistance in CRC cells with KRAS mutations[30]. The overexpression of Protein Tyrosine Phosphatase Receptor Type N (*PTPRN*)

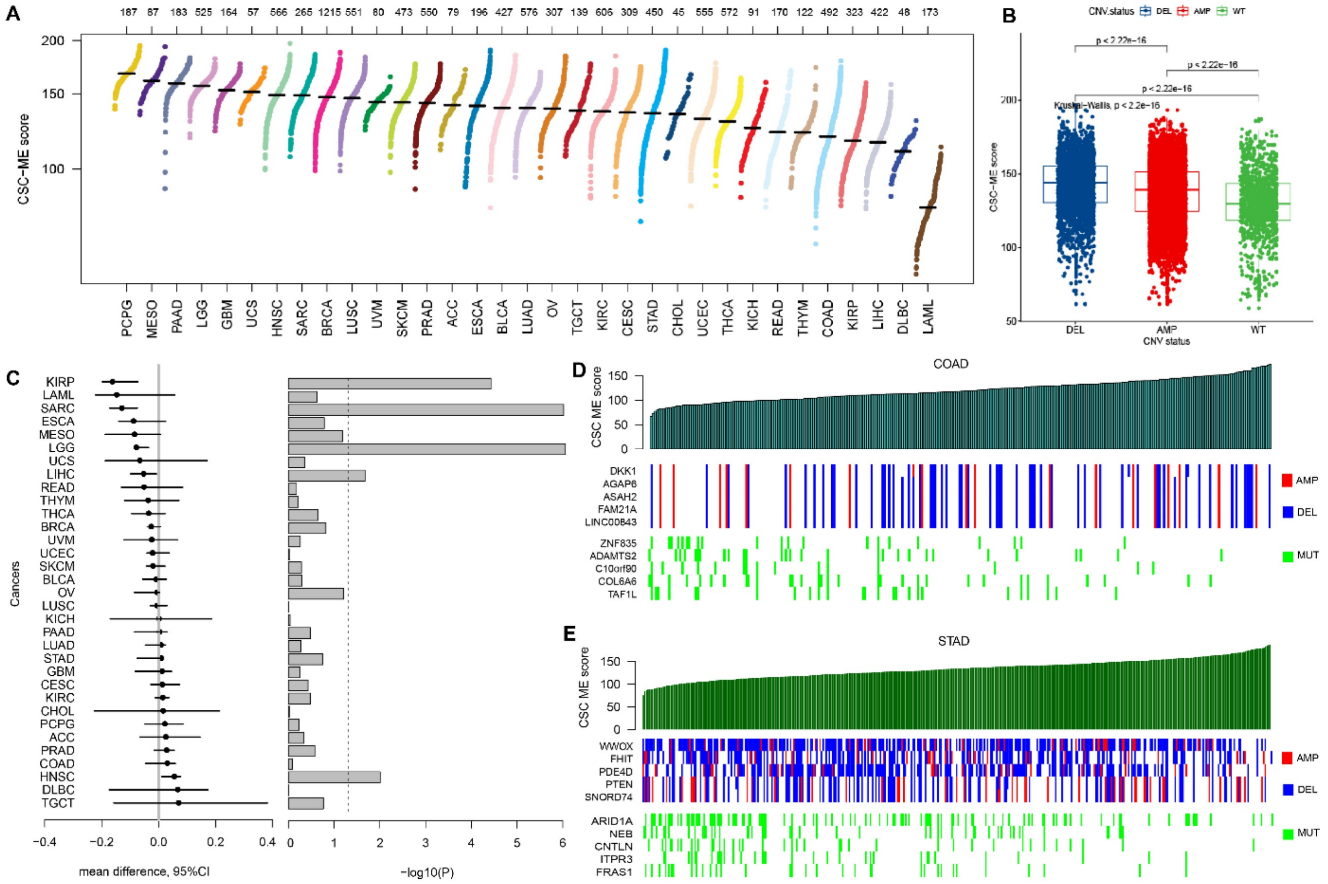

Figure 6. CSC-related score in pan-cancer and the relation with CNV. (A). Scatter plot of CSC-related score in pan-cancer. (B). Boxplot of samples with CNV (including AMP and DEL) versus wild type (WT). (C). Relative difference value and significance distribution of samples with CNV in pan-cancer. Genes with significant CNV and genes with somatic mutations in COAD (D) and STAD (E) as CSC-related score increased.

promoted LUAD cell migration and the expression of EMT markers by influencing MEK/ERK and PI3K/AKT signaling[31]. Gastrin Releasing Peptide (*GRP*) is a kind of secretory protein and regulates numerous functions of gastrointestinal and central nervous system. *GRP* exerted mitogenic effect to accelerate proliferation of CRC and head and neck squamous cancer cells[32]. Fatty acid-binding protein 4 (*FABP4*), as a carrier protein for fatty acids, is widely expressed in adipocytes, macrophages, dendritic cells, and microvascular endothelial cells. It participates in lipid transport, metabolism, and intracellular signal transduction. *FABP4* may promote CRC progression related to epithelial-mesenchymal transition (EMT)[33]. Wnt signalling inhibitor *DKK1* Promotes tumor immune evasion and impedes Anti-PD-1 treatment[34]. The L1 cell adhesion molecule (*L1CAM*) promotes tumor growth and metastasis[35]. As a secretory protein, Inhibin βA (*INHBA*) is a member of the TGF-β superfamily. *INHBA* was aberrant overexpression in CRC tissues and closely related to the poor prognosis of CRC patients[36]. *TNS1* encodes cytoskeletal protein that maintains structural integrity and mediates signal transduction. Elevated *TNS1* expression in CRC cells had been revealed to increase cell proliferation and invasiveness[37,38]. *DPYSL4* is a member of the collapsin response mediator protein family, which is involved in cancer invasion and progression. *DPYSL4* plays a key role in the tumor-suppressor function of p53 by regulating oxidative phosphorylation and the cellular energy supply via its association with mitochondrial supercomplexes, possibly linking to the pathophysiology of both cancer and obesity[39]. *ISM1* promoted EMT and colon cancer cell migration and proliferation[40]. Different from the above, Fatty Acid Binding Protein 5 (*FABP5*) suppresses colorectal cancer progression[41].

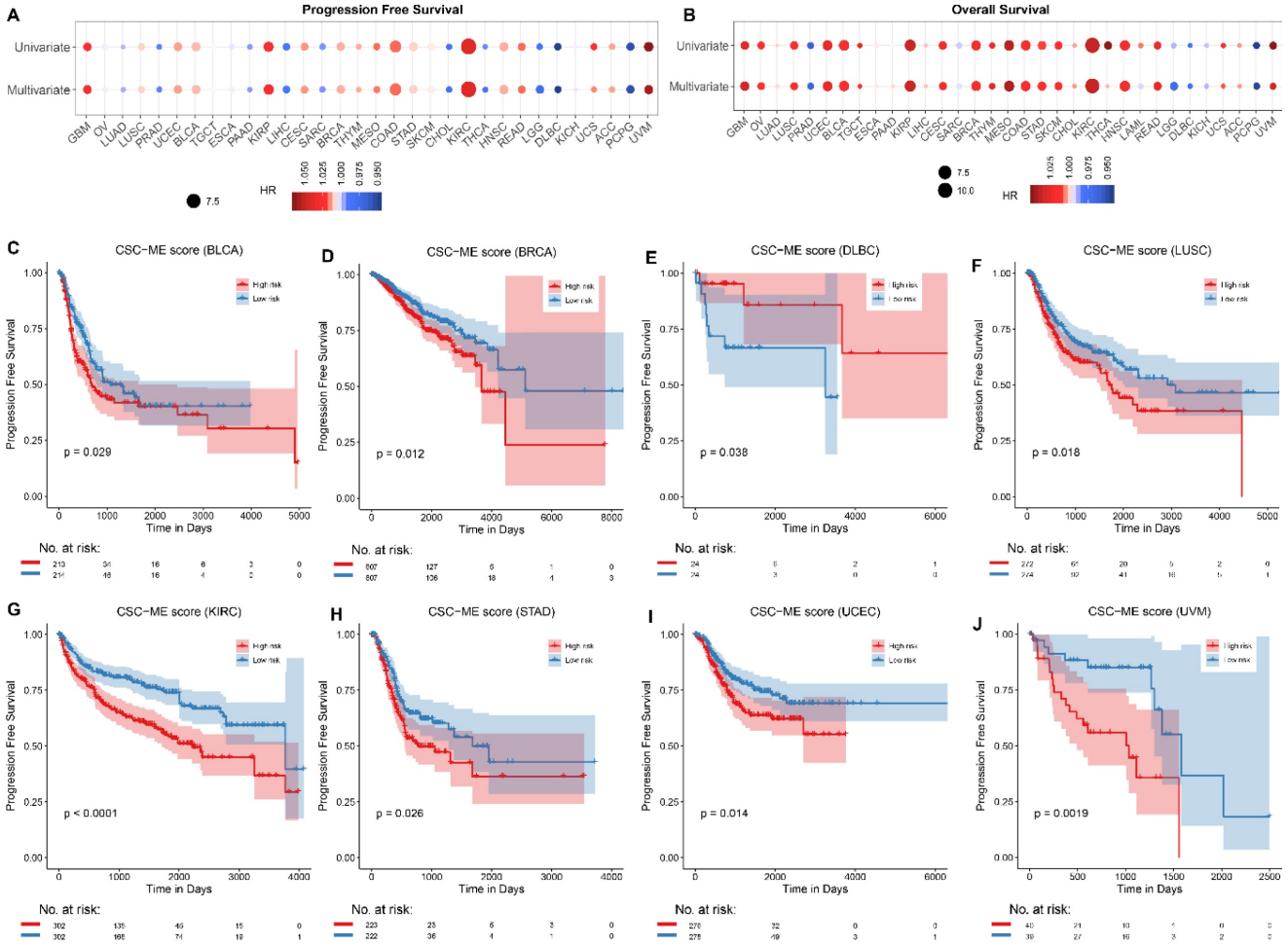

Figure 7. Survival analyses of CSC-related signature in pan-cancers. Univariate and multivariate Cox regression analysis of CSC-related score on the risk of PFS (A) and OS (B) in pan-cancer. (C-J). Survival curves (PFS) of high-risk and low-risk groups in pan-cancer.

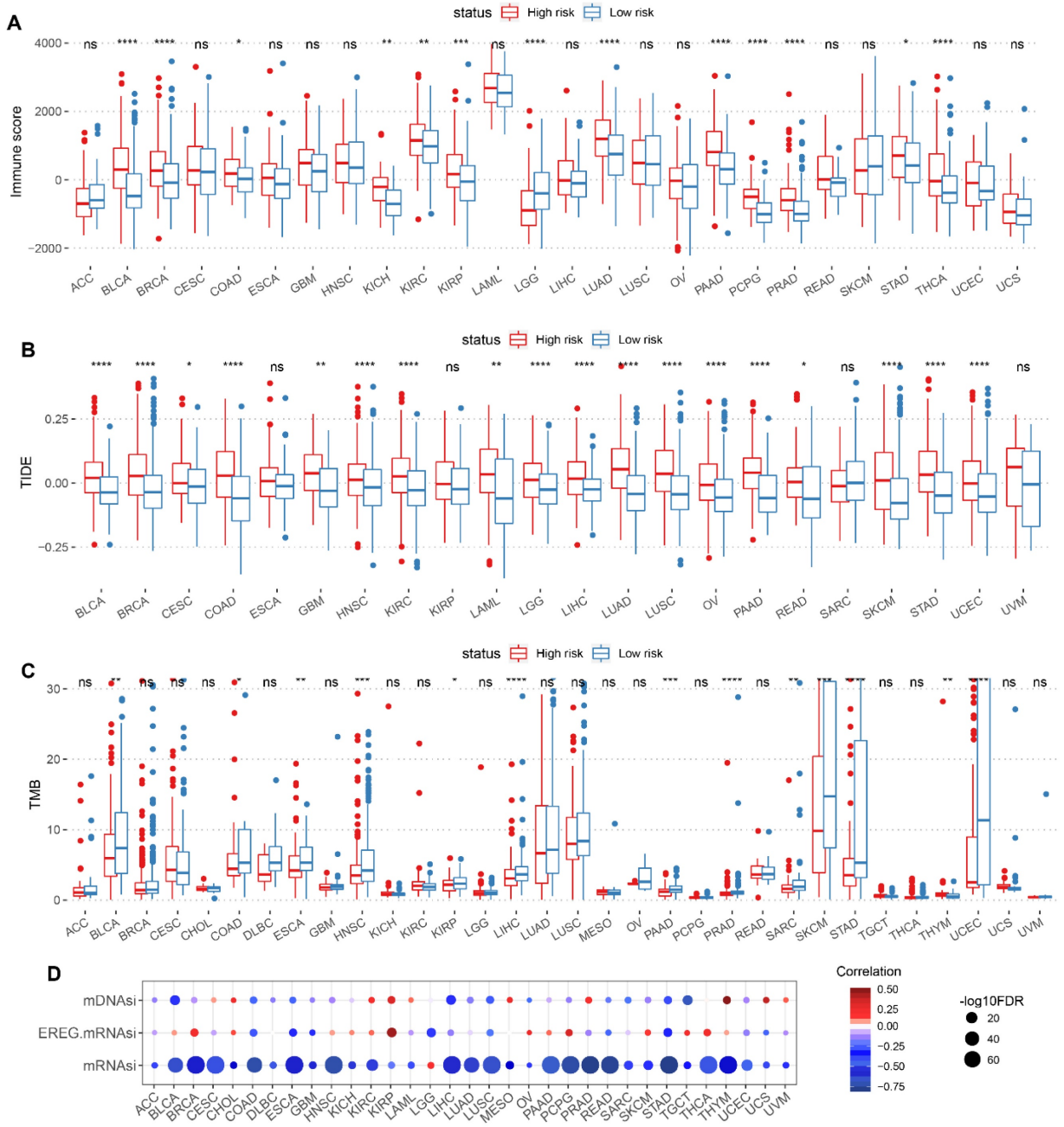

Figure 8. Correlation of CSC-related score with immune characteristics and stemness score in pan-cancer. Boxplot displaying immune characteristics including immune score (A), TIDE score (B), and TMB (C) between high and low CSC-related score groups. (D) Correlation of CSC-related score with mDNAsi, EREG mRNAsi and mRNAsi in pan-cancer. The redder color indicates a stronger positive correlation, and the bluer color indicates a stronger negative correlation.

Taken together, through consensus clustering on CSC-related genes in CRC, 4 subtypes with diverse prognosis, immune infiltration levels and clinical characteristics were identified. By applying Univariate Cox regression analysis and LASSO analysis, a 21-gene CSC-related signature was constructed and validated in GEO cohorts of CRC patients. The model has prospective clinical implications for prognosis evaluation and and preferential use of ICIs in CRC. Furthermore, the expression levels of CSC-related genes in tumor cells are also related to prognosis, tumor mircroenvironment, treatment outcome, stemness score and the efficacy of different chemotherapy-related drugs in pan-cancer. These results thus provide a reference for future research on CSC-related genes as potential pan-cancer targets. Our study also has some limitations including lack of internal or external laboratorial validation of the newly developed prognostic model, as well as comparison with other existing prognostic markers/models, which is warranted in the future study.

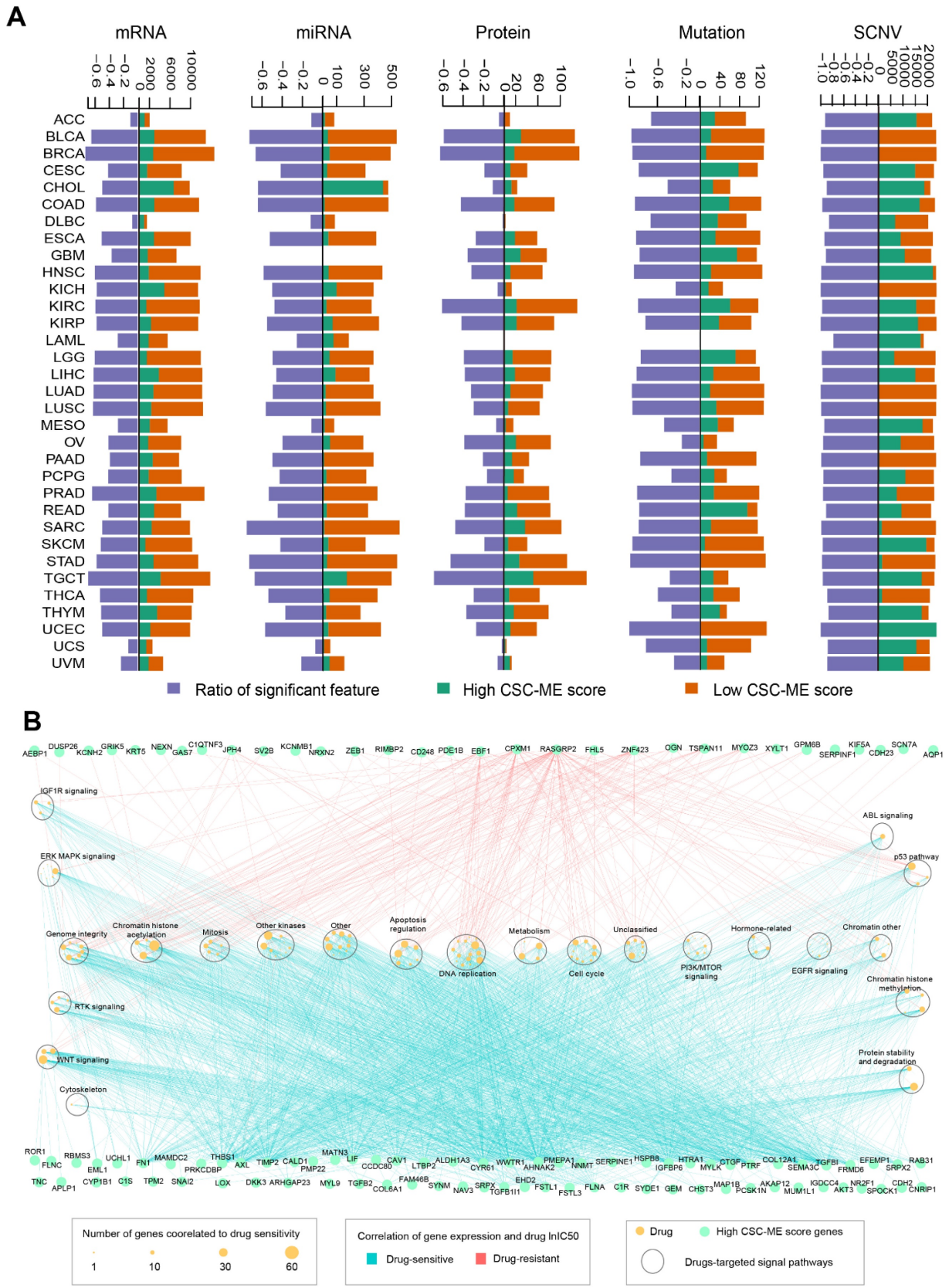

Figure 9. Marker counts of mRNA, miRNA, protein, mutation, SCNV and drug sensitivity analysis between groups based on CSC-related score. (A). Marker counts of mRNA, miRNA, protein, mutation and SCNV between groups based on CSC-related score. (B). mRNA markers, drug sensitivity and pathway enrichment analysis based on cell line data. The light blue line indicates a significant positive correlation between marker expression level and lnIC50 of the drug (R ≥ 0.3 and FDR ≤ 0.05), and the red line indicates a significant negative correlation between marker expression level and lnIC50 of the drug (R ≥ 0.3 and FDR ≤ 0.05). Green dots indicate marker genes with high CSC-related scores, yellow dots indicate drugs (size indicates the number of genes associated), and large oval boxes indicate pathways associated with drug targets.

Abbreviations

CNV: copy number variation; COAD: colon adenocarcinoma; CR: complete remission; CRC: colorectal cancer; CSC: cancer stem cell; DEGs: differentially expressed genes; FDR: false discovery rate; GEO: gene expression omnibus; HR: hazard ratio; ICIs: immune-checkpoint inhibitors; LASSO: least absolute shrinkage and selection operator; OS: overall survival; PD: progressive disease; PFS: progression-free survival; PR: partial remission; ROC: receiver operating characteristic; SD: stable disease; TCGA: the cancer genome atlas; TIDE: tumor immune dysfunction and exclusion; TMB: tumor mutational burden; TME: tumor microenvironment.

Supplementary Material

Supplementary figures.
https://www.jcancer.org/v16p2375s1.pdf

Acknowledgements

We thank the staffs of the Bioinformatics Platform in the Zhanjiang Institute of Clinical Medicine for data acquisition and processing.

Funding

This work was supported by the Guangdong Basic and Applied Basic Research Foundation (2023A1515030063) and Science and Technology Plan Project of Zhanjiang city (2021A05141, 2022A01101).

Author contributions

Z.-Y. L., M.-F. L., Q.-Z. L., and C.-F. Y. designed and oversaw the study. Z.-Y. L., M.-F. L., Y.-Y. H., and G.-S. Z. initiated or participated in data acquisition and analyses. Y.-Y. H., J.-R. C., and Y.-M. G. contributed to interpret and annotate the results. Z.-Y. L., M.-F. L., Q.-Z. L., and C.-F. Y. drafted the manuscript. All authors reviewed and approved the final version of manuscript.

Availability of data and materials

The datasets collected and analyzed in the current study are available from the online repositories as described in the section of MATERIALS AND METHODS.

Competing Interests

The authors have declared that no competing interest exists.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021; 71: 209–49.
2. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. Lancet. 2019; 394: 1467–80.
3. Ganesh K, Stadler ZK, Cercek A, et al. Immunotherapy in colorectal cancer: rationale, challenges and potential. Nat Rev Gastroenterol Hepatol. 2019; 16: 361–75.
4. Pardal R, Clarke MF, Morrison SJ. Applying the principles of stem-cell biology to cancer. Nat Rev Cancer. 2003; 3: 895–902.
5. Zhao W, Li Y, Zhang X. Stemness-Related Markers in Cancer. Cancer Transl Med. 2017; 3: 87–95.
6. Wang G, Wu Z, Huang Y, et al. Identification of the key genes controlling stomach adenocarcinoma stem cell characteristics via an analysis of stemness indices. J Gastrointest Oncol. 2022; 13: 593–604.
7. Chen X, Zhang D, Jiang F, et al. Prognostic Prediction Using a Stemness Index-Related Signature in a Cohort of Gastric Cancer. Front Mol Biosci. 2020; 7: 570702.
8. Luo Y, Xu W-B, Ma B, Wang Y. Novel Stemness-Related Gene Signature Predicting Prognosis and Indicating a Different Immune Microenvironment in HNSCC. Front Genet. 2022; 13: 822115.
9. Feng J, Li Y, Wen N. Characterization of Cancer Stem Cell Characteristics and Development of a Prognostic Stemness Index Cell-Related Signature in Oral Squamous Cell Carcinoma. Dis Markers. 2021; 2021: 1571421.
10. Zhang D, Zhang M, Zhang Q, Zhao Z, Nie Y. Identification of Prognostic Biomarkers Associated with Cancer Stem Cell Features in Prostate Adenocarcinoma. Med Sci Monit. 2020; 26: e924543.
11. Yi L, Huang P, Zou X, et al. Integrative stemness characteristics associated with prognosis and the immune microenvironment in esophageal cancer. Pharmacol Res. 2020; 161: 105144.
12. Kang Y, Zhu X, Wang X, et al. Identification and Validation of the Prognostic Stemness Biomarkers in Bladder Cancer Bone Metastasis. Front Oncol. 2021; 11: 641184.
13. Cancer Stemness-Based Prognostic Immune-Related Gene Signatures in Lung Adenocarcinoma and Lung Squamous Cell Carcinoma - PubMed [Internet]. [cited 16 May 2024]. Available at: https://pubmed.ncbi.nlm.nih.gov/34745015/
14. Tan J, Zhu H, Tang G, et al. Molecular Subtypes Based on the Stemness Index Predict Prognosis in Glioma Patients. Front Genet. 2021; 12: 616507.
15. Goldman MJ, Craft B, Hastie M, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020; 38: 675–8.
16. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 2013; 41: D991-995.
17. Mariathasan S, Turley SJ, Nickles D, et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature. 2018; 554: 544–8.
18. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015; 12: 453–7.
19. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013; 4: 2612.
20. Jiang P, Gu S, Pan D, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med. 2018; 24: 1550–8.
21. Yang W, Soares J, Greninger P, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 2013; 41: D955-961.
22. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015; 1: 417–25.
23. Shen Y, Yao H, Li A, Wang M. CSCdb: a cancer stem cells portal for markers, related genes and functional information. Database (Oxford). 2016; 2016: baw023.
24. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010; 26: 1572–3.
25. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43: e47.
26. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin. 2022; 72: 7–33.
27. Ricci-Vitiani L, Fabrizi E, Palio E, De Maria R. Colon cancer stem cells. J Mol Med (Berl). 2009; 87: 1097–104.
28. Humphries A, Wright NA. Colonic crypt organization and tumorigenesis. Nat Rev Cancer. 2008; 8: 415–24.
29. Das PK, Islam F, Lam AK. The Roles of Cancer Stem Cells and Therapy Resistance in Colorectal Carcinoma. Cells. 2020; 9: 1392.
30. Yu C, Luo D, Yu J, et al. Genome-wide CRISPR-cas9 knockout screening identifies GRB7 as a driver for MEK inhibitor resistance in KRAS mutant colon cancer. Oncogene. 2022; 41: 191–203.
31. Song X, Jiao X, Yan H, et al. Overexpression of PTPRN Promotes Metastasis of Lung Adenocarcinoma and Suppresses NK Cell Cytotoxicity. Front Cell Dev Biol. 2021; 9: 622018.
32. Saurin JC, Némoz-Gaillard E, Sordat B, et al. Bombesin stimulates adhesion, spreading, lamellipodia formation, and proliferation in the human colon carcinoma Isreco1 cell line. Cancer Res. 1999; 59: 962–7.
33. Y Z, W Z, M X, et al. High expression of FABP4 in colorectal cancer and its clinical significance. Journal of Zhejiang University Science B [Internet]. 2021 [cited 16 May 2024]; 22. Available at: https://pubmed.ncbi.nlm.nih.gov/33615754/

34. Shi T, Zhang Y, Wang Y, et al. DKK1 Promotes Tumor Immune Evasion and Impedes Anti-PD-1 Treatment by Inducing Immunosuppressive Macrophages in Gastric Cancer. Cancer Immunol Res. 2022; 10: 1506–24.
35. L1CAM defines the regenerative origin of metastasis-initiating cells in colorectal cancer - PubMed [Internet]. [cited 16 May 2024]. Available at: https://pubmed.ncbi.nlm.nih.gov/32656539/
36. Xiao Q, Xiao J, Liu J, Liu J, Shu G, Yin G. Metformin suppresses the growth of colorectal cancer by targeting INHBA to inhibit TGF-β/PI3K/AKT signaling transduction. Cell Death Dis. 2022; 13: 202.
37. Wang Z, Ye J, Dong F, Cao L, Wang M, Sun G. TNS1: Emerging Insights into Its Domain Function, Biological Roles, and Tumors. Biology (Basel). 2022; 11: 1571.
38. Zhou H, Zhang Y, Wu L, et al. Elevated transgelin/TNS1 expression is a potential biomarker in human colorectal cancer. Oncotarget. 2018; 9: 1107–13.
39. Nagano H, Hashimoto N, Nakayama A, et al. p53-inducible DPYSL4 associates with mitochondrial supercomplexes and regulates energy metabolism in adipocytes and cancer cells. Proc Natl Acad Sci U S A. 2018; 115: 8370–5.
40. Effect of ISM1 on the Immune Microenvironment and Epithelial-Mesenchymal Transition in Colorectal Cancer - PubMed [Internet]. [cited 16 May 2024]. Available at: https://pubmed.ncbi.nlm.nih.gov/34350177/
41. Ye M, Hu C, Chen T, et al. FABP5 suppresses colorectal cancer progression via mTOR-mediated autophagy by decreasing FASN expression. Int J Biol Sci. 2023; 19: 3115–27.