Computational and Structural Biotechnology Journal 20 (2022) 2153–2168

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Evaluation of stromal cell infiltration in the tumor microenvironment enable prediction of treatment sensitivity and prognosis in colon cancer

Rui Zhou [a,1], Zhaowei Wen [a,1], Yifu Liao [b,1], Jingjing Wu [a], Shaoyan Xi [c], Dongqiang Zeng [a], Huiying Sun [a], Jianhua Wu [a], Min Shi [a], Jianping Bin [d], Yulin Liao [d], Wangjun Liao [a,*]

[a] Department of Oncology, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong, PR China
[b] Department of Neurology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510080, PR China
[c] Department of Pathology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, Guangdong, PR China
[d] Department of Cardiology, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong, PR China

ARTICLE INFO

Article history:
Received 23 January 2022
Received in revised form 26 April 2022
Accepted 27 April 2022
Available online 30 April 2022

Keywords:
Stromal cell infiltration
Machine learnin
Gene signature
Treatment sensitivity
CRISPR library screen
Colon cancer

ABSTRACT

Current clinical factors for screening candidates that might benefit from adjuvant chemotherapy in colon cancer are inadequate. Tumor microenvironment, especially the stromal components, has the potential to determine treatment response. However, clinical translation of the tumor-associated stromal characterization into a practical biomarker for helping treatment decision has not been established. Using machine learning, we established a novel 31-gene signature, called stromal cell infiltration intensity score (SIIS), to distinguish patients characterized by the enrichment of abundant stromal cells in five colon cancer datasets from GEO (N = 990). Patients with high-SIIS were at higher risk for recurrence and mortality, and could not benefit from adjuvant chemotherapy due to their intrinsic drug resistance; however, the opposite was reported for patients with low-SIIS. The role of SIIS in detection of patients with high stromal cell infiltration and reduced drug efficiency was consistently validated in the TCGA-COAD cohort (N = 382), Sun Yat-sen University Cancer Center cohort (N = 30), and could also be observed in TCGA pan-cancer settings (N = 4898) and four independent immunotherapy cohorts (N = 467). Based on multi-omics data analysis and the CRISPR library screen, we reported that lack of gene mutation, hypomethylation in ADCY4 promoter region, activation of WNT-PCP pathway and SIAH2-GPX3 axis were potential mechanisms responsible for the chemoresistance of patients within high-SIIS group. Our findings demonstrated that SIIS provide an important reference for those making treatment decisions for such special patients.

1. Introduction

Colon cancer is a disease characterized by considerable molecular heterogeneity that is related to the prognosis and treatment outcomes [1]. Although several molecular classification strategies have been proposed to characterize distinct biological properties [2], more effective and clinically accessible classifiers that could complement the TNM system to enable the personalization of fluorouracil (FU)-based postsurgical adjuvant chemotherapy (ADJC) in localized colon cancer remain to be explored.

The extensive literature has underscored the clinical importance of tumor microenvironment (TME) features of various cancer types [3]. For colon cancer, a consensus immunoscore, has been demonstrated to be a robust prognostic indicator independent of TNM staging, thereby making colon cancer a paradigmatic tumor for immune classification [4]. Besides its prognostic value, immunoscore was also revealed to be a reference for ADJC performance [5]. The consensus immunoscore only quantifies the densities of $CD3^+$ and $CD8^+$ T cells, therefore it mainly reflects degree of the local intratumoral adaptive immune reaction. However, in addition to T cells, some other cellular components in TME might also play an important role in tumorigenesis and progression. For example, our previous studies showed that the infiltration intensity of stromal cells is greater than that of T cells with respect to differentiating distinct TME subtypes and predicting the likelihood of the chemotherapeutic benefit of patients with colon cancer [6,7]. However, to date, a potential clinical translation of the stromal contex-

* Corresponding author at: Department of Oncology, Nanfang Hospital, Southern Medical University, 1838 North Guangzhou Avenue, Guangzhou 510515, PR China.
E-mail address: nfyyliaowj@163.com (W. Liao).
[1] These authors are contributed equally.

https://doi.org/10.1016/j.csbj.2022.04.037
2001-0370/

ture into a marker for prognosis and therapeutic benefit prediction has not been established.

The consensus immunoscore is detected by immunohistochemical staining, a method that relies heavily on a limited repertoire of phenotypic markers and biopsy specimens of sufficient size. In other words, the immunohistochemical staining can only reflect a limited number of cell types so that most of the information regarding TME panoramic landscape features is ignored. Besides, immunohistochemical staining is typically applied to analyze a specific two-dimensional tissue section, so it cannot be used to estimate the TME content across three-dimensional tissue samples. To address these issues, several computational approaches have recently been developed to characterize TME compositions in a high-throughput manner directly from bulk tissue gene expression profiles [8]. Cell content data generated by these methods to establish clinical models has revealed that they are correlated with the clinical outcomes of tumor patients [9,10]. However, as the operation of these computational tools requires the whole transcriptome data as the input, they are not considered clinically practical. Hence, in this study, we first employed the enrichment methods single-sample gene set enrichment analysis (ssGSEA) to calculate the enrichment scores for 23 immune and stromal cells based on clinically annotated colon cancer gene expression profiles, and then identified a special patient population with high infiltration of stromal cells by cluster analysis. As a result, we established a transcriptome-based and easy-to-use scoring assay based on 31 informative genes to effectively quantify stromal cell components of the TME in colon cancer patients, which is a robust prognostic biomarker for distinguishing patients who are at high risk of relapse and cannot benefit from ADJC simultaneously because of their intrinsic drug resistance.

2. Materials and methods

2.1. Patient cohorts used in this study

Public transcriptome data on colon cancer samples were retrospectively collected from the GEO (https://www.ncbi.nlm.nih.gov/geo/) and TCGA-COAD (https://cancergenome.nih.gov/) datasets. For samples from the Sun Yat-sen University Cancer Center cohort, written informed consent was obtained from all patients. This study was approved by the Nanfang Hospital Ethics Review Board. RNA extraction and sequencing were performed as previously described [11]. All colon cancer samples included in the analysis in this study were obtained before ADJC performance. Details are provided in the Supplementary Materials and Methods.

2.2. Generation of the 31-gene-based quantitative indicator

A four-phase screening strategy was designed to filter out the informative genes to build the "stromal infiltration intensity score (SIIS)". Briefly, a four-phase screening strategy was designed to filter out the informative genes to build the "stromal infiltration intensity score (SIIS)", including differential gene expression analysis; robust prognostic gene identification; elastic net analysis (iteration = 1000); and Cox proportional hazards model (iteration = 1000) with the least absolute shrinkage and selection operator (LASSO) penalty analysis Differential expression gene (DEG) analysis among TMECS1 and TMECS2 clusters was conducted in the GSE39582 dataset using the "LIMMA" package. The significance criterion for DEGs was set as an absolute "log2FC" value >0.5 and an adjusted p value < 0.05. Then, robust prognostic DEGs were identified using the bootstrapping method (repeated 1000 times) as described previously, and DEGs that were incorporated in 70% of resample runs (achieved $p < 0.05$ in robustness testing) were considered as robust prognostic genes and selected for further analysis. The elastic net analysis (iteration = 1000) and LASSO-Cox analysis (iteration = 1000) were performed by utilizing a R package called "glmnet". The elastic net regularization for generalized linear models is the linear combination of LASSO and ridge regularization methods. An elastic net model was developed to screen out the most informative genes for distinguishing patients within TMECS1 and TMECS2. The alpha parameter of the elastic net was set at 0.9.

The penalized Cox regression model with LASSO penalty was used to select the most useful prognostic markers among informative DEGs picked by the elastic net model, and the optimal values of the penalty parameter λ were determined by tenfold cross-validations. The expression levels of informative DEGs were analyzed as binary variables in LASSO-Cox regression; the optimal cut-off values were evaluated based on the association between relapse-free survival and gene expression in the GEO cohort using the "survminer" package. Among the 31 genes finally selected by repeated LASSO-Cox regression, 11 genes had positive Cox coefficients (named $DEGs^{\beta+}$) and the other 20 genes had negative Cox coefficients (named $DEGs^{\beta-}$). The final score was defined as, SIIS = the average expression of $DEGs^{\beta+}$ – the average expression of $DEGs^{\beta-}$.

2.3. The SubMap and nearest template prediction (NTP) analysis

The SubMap analysis and NTP analysis were used to explore the internal relationship between SIIS and TMECS algorithm. The implementation of these two algorithms is achieved through the online module of GenePattern website (https://cloud.genepattern.org/). Details are provided in the Supplementary Materials and Methods.

2.4. Prediction of the chemotherapeutic response

Chemoresponse was predicted using the R package "pRRophetic" which implemented a built-in ridge regression model based on the CTRP database and GDSC database respectively. Details are provided in the Supplementary Materials and Methods.

2.5. Pathway, immune, and metabolic score estimation

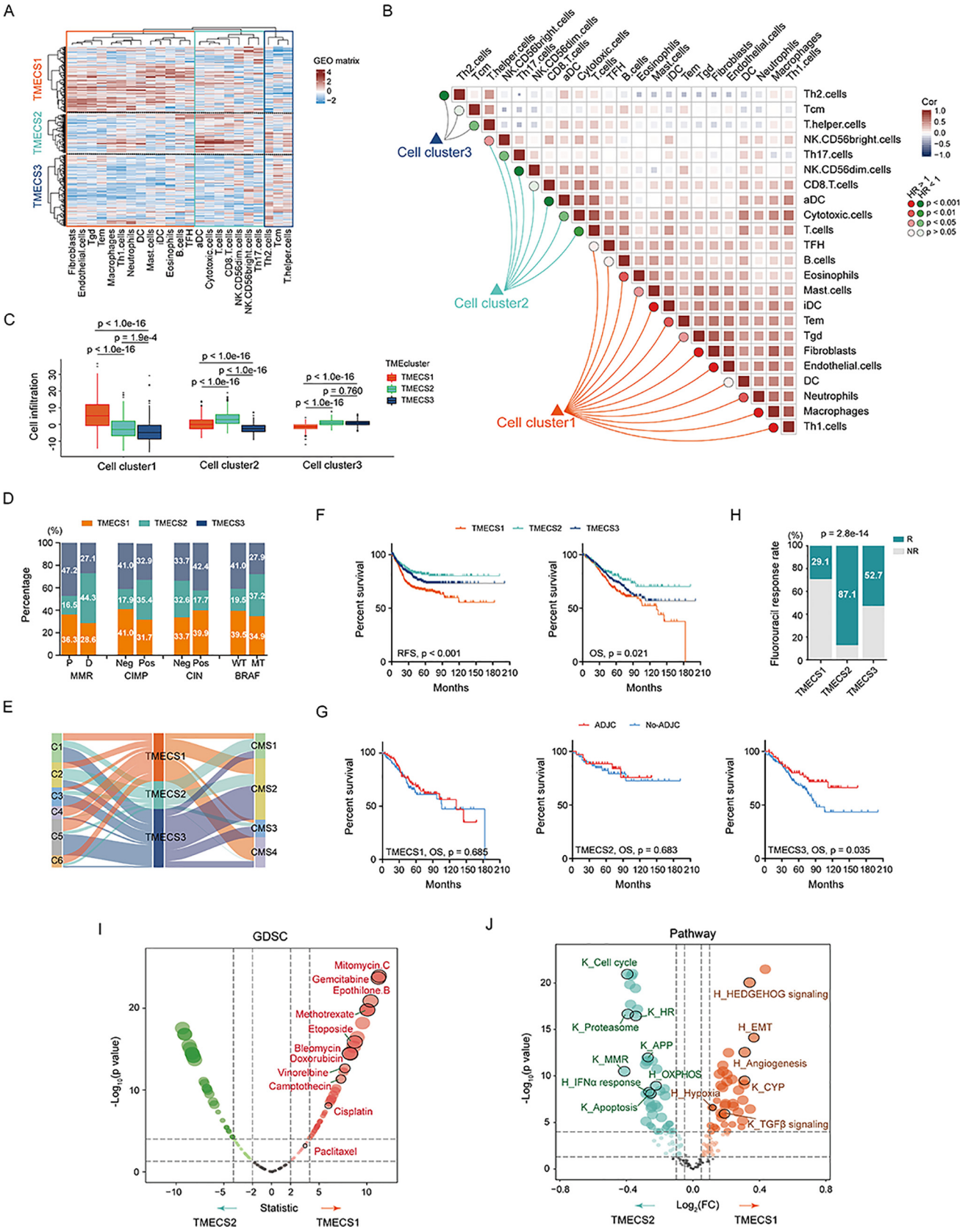
The pathway, immune, and metabolic score estimation were achieved by different R packages. Details are provided in the Supplementary Materials and Methods.

2.6. Multi-omics analysis

The genomic mutation, DNA methylation, and proteomic data were download from TCGA-COAD dataset, details are provided in the Supplementary Materials and Methods.

2.7. CRISPR synergistic activation mediator library screen for identification of chemoresistance genes

The CRISPR synergistic activation mediator (SAM) library which contains 70,290 unique sgRNA sequences targeting 23,430 human genes (3 sgRNAs per gene) was used to generate a mutant cell pool for high-throughput screening. The count value of sgRNAs was transformed using the "voom" algorithm, and differentially enriched sgRNAs between chemotherapy drug- and vehicle-treated groups were analyzed using the "LIMMA" package. The adjusted p-value for multiple testing was calculated using the Benjamini–Hochberg correction. The criteria for screening candidate sgRNAs were: 1) the average count values of candidate sgRNAs in


A
TMECS1
TMECS2
TMECS3
GEO matrix
Fibroblasts
Endothelial.cells
Tgd
Tem
Macrophages
Th1.cells
Neutrophils
DC
Mast.cells
iDC
Eosinophils
B.cells
TFH
aDC
Cytotoxic.cells
T.cells
CD8.T.cells
NK.CD56dim.cells
NK.CD56bright.cells
Th17.cells
Th2.cells
Tcm
T.helper.cells
B
Cell cluster3
Cell cluster2
Cell cluster1
Cor
HR > 1
HR < 1
C
Cell infiltration
TMEcluster
Cell cluster1
Cell cluster2
Cell cluster3
D
Percentage
MMR
CIMP
CIN
BRAF
E
CMS1
CMS2
CMS3
CMS4
F
Percent survival
RFS, p < 0.001
OS, p = 0.021
Months
G
ADJC
No-ADJC
TMECS1, OS, p = 0.685
TMECS2, OS, p = 0.683
TMECS3, OS, p = 0.035
H
Fluorouracil response rate
p = 2.8e-14
R
NR
I
GDSC
Mitomycin C
Gemcitabine
Epothilone B
Methotrexate
Etoposide
Bleomycin
Doxorubicin
Vinorelbine
Camptothecin
Cisplatin
Paclitaxel
Statistic
J
Pathway
K_Cell cycle
K_HR
K_Proteasome
K_APP
K_MMR
H_IFNα response
H_OXPHOS
H_Hypoxia
K_Apoptosis
H_HEDGEHOG signaling
H_EMT
H_Angiogenesis
K_CYP
K_TGFβ signaling
Log2(FC)

both the chemotherapy drug-treated group and the vehicle group should be greater than 1,000; 2) the absolute "log2FC" value calculated by difference analysis of sgRNA level of chemotherapy drug-treated group versus vehicle group was more than 0.5. Details are provided in the Supplementary Materials and Methods.

2.8. Connectivity map analysis

Connectivity map (Cmap) analysis was performed using the 150 genes with the most significant fold changes (up-regulated and down-regulated) according to the instructions provided by the Cmap website (https://clue.io/).

2.9. In vitro and in vivo experiments

To address the role of GPX3 in mediating chemoresistance in colon cancer cells, several in vitro and in vivo experiments, including cell culture, cell transfection, Methylthiazolyl-tetrazolium (MTT) assay, clonogenic assay, western blotting, real-time quantitative reverse-transcription polymerase chain reaction, immunohistochemistry staining, and subcutaneous tumor formation assays, were performed as previously described [12].

2.10. Statistical analysis

All statistical tests were performed using R software (version 3.6.0) or SPSS software (version 25.0). Details are provided in the Supplementary Materials and Methods.

3. Results

3.1. Identification of distinct TME clusters associated with chemotherapeutic benefit

Using an unsupervised hierarchical cluster analysis (Fig. 1A), we identified three distinct TME clusters based on 990 samples (Tables S1-Tables S2) in the meta-GEO cohort, denoted as tumor microenvironment cell subtype1 (TMECS1; n = 369), tumor microenvironment cell subtype2 (TMECS2; n = 218), and tumor microenvironment cell subtype3 (TMECS3; n = 403). As shown in Fig. 1B and 1C, TMECS1 was characterized by enrichment of the highest level of stromal cells and innate immune cells with immunosuppressive properties (these cell types are classified into cell cluster 1), moderate level of cytotoxic lymphocytes and activated dendritic cell (these cell types were classified into cell cluster 2), and the lowest level of Th2 cells, central memory T cells, and total T helper cells (these cell types were classified into cell cluster 3), which was recognized as immune-excluded phenotype [13]. Compared with TMECS1, TMECS2 had significantly higher abundance of cell cluster 2 and 3 infiltration, while cell cluster 1 infiltration was significantly decreased, indicating that the TME feature of TMECS2 was recognized as immune-inflamed phenotype [13]. Unlike TMECS1 and TMECS2, TMECS3 exhibited lowest abundance estimates of both cell cluster 1 and 2, which were closely linked to previously reported immune-desert phenotypes [13]. While the infiltration of cell cluster 3 was similar with that in TMECS2. In terms of the correlation between TMECS and molecular features of colon cancer, we found that TMECS2 was significant enriched in dMMR, CIMP (+), BRAF mutation, CMS1, and C2 (CIT) tumors (Fig. 1D, E; Table S3), while CIN (+), CMS4, C4 (CIT), and C6 (CIT) tumors were more common in the TMECS1 cluster than in the other two clusters (Fig. 1D, E; Table S3). However, there was no obvious difference in the KRAS mutation or P53 mutation among TME clusters. The survival analysis revealed that patients with TMECS2 exhibited a prominent survival advantage for both relapse-free survival (RFS) and overall survival (OS) among the three TME clusters, whereas TMECS1 was associated with the worst prognosis (Fig. 1F). Next, we evaluated the association between TME clusters and the chemotherapeutic response. The administration of ADJC significantly reduced the mortality risk of patients only in the TMECS3 cluster; it did not confer survival benefits to patients in the TMECS1 or TMECS2 clusters (Fig. 1G). Considering the high risk for recurrence of patients with TMECS1, we speculated that the failure of ADJC to provide benefits to these patients might be attributable to the existence of primary drug resistance. To validate this hypothesis, we compared the fluorouracil response rates and the predicted IC50 values of several chemotherapeutic drugs between TMECS1 and TMECS2 clusters. The results showed that patients in the TMECS1 cluster exhibited the lowest fluorouracil response rate among all the clusters (Fig. 1H) and significantly higher estimated IC50 values for the majority of chemotherapy drugs as compared to patients in the TMECS2 cluster (Fig. 1I). Finally, the biological analysis revealed that the TMECS2 cluster had enriched pathways of DNA damage repair and immune activation, whereas the TMECS1 cluster was characterized by multiple pathways related to stromal activation, chemoresistance, and cancer progression, such as hedgehog signaling, epithelial-mesenchymal transition (EMT), cytochrome P450 metabolism, and transforming growth factor (TGF)-β signaling (Fig. 1J).

3.2. Generation of the gene signature to distinguish patients within TMECS1 in the meta-GEO cohort

Through two-step regressions (Fig. 2A, Fig. S1A-B), we generated a quantitative indicator consisting of 31 genes, termed the SIIS (Fig. 2B). Boxplots show that the median SIIS was highest in the TMECS1 cluster and lowest in the TMECS2 cluster (Fig. 2C), and the subsequent TME correlation analysis revealed a strong positive correlation between the SIIS and infiltration of cell cluster 1

Fig. 1. Unsupervised clustering of TME cells in colon cancer and clinical characteristics of TME subtypes. (A) Heatmap showing the three TME subtypes (TMECS1, orange; TMECS2, turquoise; TMECS3, blue) developed by unsupervised analysis and hierarchical clustering of 23 TME cells. Rows represent the samples and columns represent TME cells. (B) Heatmap showing the correlation between infiltration levels of 23 tumor microenvironment cells in the meta-GEO cohort. Shading color represents the value of corresponding correlation coefficients. The survival impact of each cell was represented by the circle on the same row. Favor for relapse-free survival was colored as green, risk for relapse-free survival as red. (C) Box plot of the comparison of the infiltration level of each cell cluster between the different TME subtypes. (D) Bar charts summarizing the proportions of TME subtypes in and across different molecular characteristic subgroups in the GSE39582 cohort. (E) Sankey chart displaying the distribution of the TME subtypes in C1-C6 subtypes and CMS subtypes. (F) The survival curve of three TME subtypes for relapse-free survival (left) and overall survival (right) in the meta-GEO cohorts. (G) The survival curve of adjuvant chemotherapy performance for overall survival in three TME subtypes. (H) Bar charts summarize the proportions of patients with fluorouracil response signatures and those with non-response signatures within and across different TME subtypes. (I-J) Volcano plots representing the comparison of the estimated IC50 level of drugs in GDSC database (I) and activation levels of biological process (J) between patients in the TMECS1 and TMECS2 subtypes. *TMECS, tumor microenvironment cell subtype; TME, tumor microenvironment; Cor, correlation; Tcm, T central memory; Tem, T effector memory; Th, T helper; DC, dendritic cell; aDC, activated dendritic cell; iDC, immature dendritic cell; Tgd, T gamma delta; NK, natural killer; TFH, T follicular helper; CIMP, CpG island methylator phenotype; CIN, chromosome instability; MT, mutant type; WT, wild type; RFS, relapse-free survival; OS, overall survival; ADJC, adjuvant chemotherapy; R response; NR, nonresponse; CMS, consensus molecular subtypes; HR, homologous recombination; OXPHOS, oxidative phosphorylation; CYP, cytochrome P; MMR, mismatch repair; EMT, epithelial-mesenchymal transition; FC, fold change.* (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

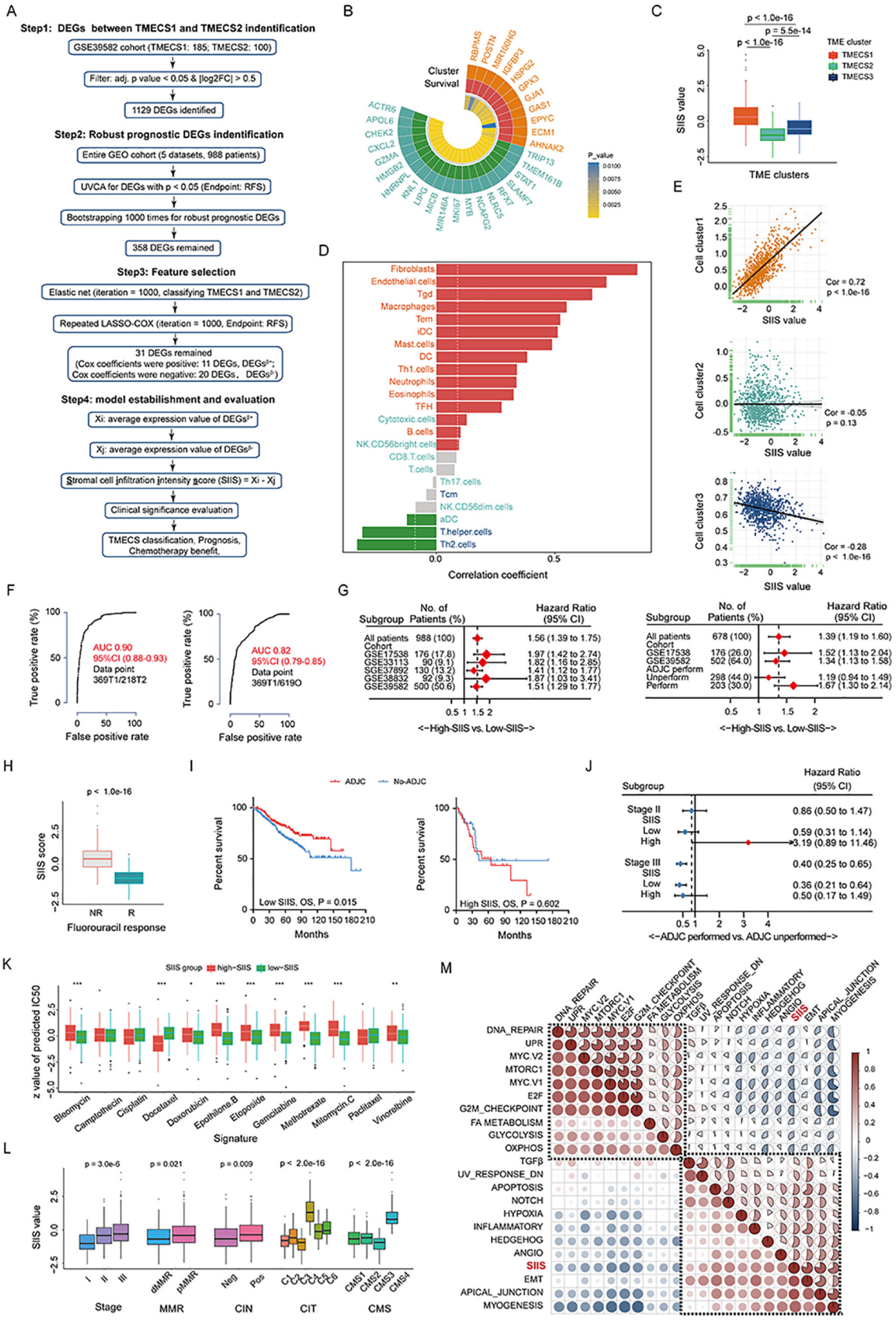

A
Step1: DEGs between TMECS1 and TMECS2 indentification
GSE39582 cohort (TMECS1: 185; TMECS2: 100)
Filter: adj. p value < 0.05 & |log2FC| > 0.5
1129 DEGs identified
Step2: Robust prognostic DEGs indentification
Entire GEO cohort (5 datasets, 988 patients)
UVCA for DEGs with p < 0.05 (Endpoint: RFS)
Bootstrapping 1000 times for robust prognostic DEGs
358 DEGs remained
Step3: Feature selection
Elastic net (iteration = 1000, classifying TMECS1 and TMECS2)
Repeated LASSO-COX (iteration = 1000, Endpoint: RFS)
31 DEGs remained
(Cox coefficients were positive: 11 DEGs, DEGs3+;
Cox coefficients were negative: 20 DEGs, DEGs3-)
Step4: model estabilishment and evaluation
Xi: average expression value of DEGs3+
Xj: average expression value of DEGs3-
Stromal cell infiltration intensity score (SIIS) = Xi - Xj
Clinical significance evaluation
TMECS classification, Prognosis, Chemotherapy benefit.
B
Cluster
Survival
P_value
C
p < 1.0e-16
p = 5.5e-14
p < 1.0e-16
TME cluster
TMECS1
TMECS2
TMECS3
SIIS value
TME clusters
D
Correlation coefficient
E
Cell cluster1
Cell cluster2
Cell cluster3
Cor = 0.72
p < 1.0e-16
Cor = -0.05
p = 0.13
Cor = -0.28
p < 1.0e-16
SIIS value
F
True positive rate (%)
False positive rate
AUC 0.90
95%CI (0.88-0.93)
Data point
369T1/218T2
AUC 0.82
95%CI (0.79-0.85)
Data point
369T1/619O
G
Subgroup
No. of Patients (%)
Hazard Ratio (95% CI)
<-High-SIIS vs. Low-SIIS->
H
p < 1.0e-16
SIIS score
NR
R
Fluorouracil response
I
ADJC
No-ADJC
Percent survival
Low SIIS, OS, P = 0.015
High SIIS, OS, P = 0.602
Months
J
Subgroup
Hazard Ratio (95% CI)
Stage II
SIIS
Low
High
0.86 (0.50 to 1.47)
0.59 (0.31 to 1.14)
3.19 (0.89 to 11.46)
Stage III
SIIS
Low
High
0.40 (0.25 to 0.65)
0.36 (0.21 to 0.64)
0.50 (0.17 to 1.49)
<-ADJC performed vs. ADJC unperformed->
K
SIIS group high-SIIS low-SIIS
z value of predicted IC50
Bleomycin
Camptothecin
Cisplatin
Docetaxel
Doxorubicin
Epothilone B
Etoposide
Gemcitabine
Melphalenate
Mitomycin C
Paclitaxel
Vinorelbine
Signature
L
p = 3.0e-5
p = 0.021
p = 0.009
p < 2.0e-16
p < 2.0e-16
SIIS value
Stage
MMR
CIN
CIT
CMS
M
DNA_REPAIR
UPR
MYC.V2
MTORC1
MYC.V1
E2F
G2M_CHECKPOINT
FA METABOLISM
GLYCOLYSIS
OXPHOS
TGFβ
UV_RESPONSE_DN
APOPTOSIS
NOTCH
HYPOXIA
INFLAMMATORY
HEDGEHOG
ANGIO
SIIS
EMT
APICAL_JUNCTION
MYOGENESIS

Table 1
Univariate and multivariate survival analyses of SIIS model and clinical variables.

	UVA (RFS)		UVA (OS)		MVA (RFS)		MVA (OS)	
	HR (95%CI)	p	HR (95%CI)	p	HR (95%CI)	p	HR (95%CI)	p
Age[a]	1.00 (0.99–1.01)	0.996	1.04 (1.02–1.05)	<0.001	1.01 (0.99–1.02)	0.239	1.04 (1.03–1.05)	<0.001
Gender (vs. Male)	0.73 (0.59–0.90)	0.003	0.82 (0.62–1.09)	0.176	0.81 (0.59–1.12)	0.199	0.63 (0.46–0.87)	0.005
SIIS[a]	1.52 (1.36–1.70)	<0.001	1.36 (1.18–1.56)	<0.001	1.58 (1.26–1.96)	<0.001	1.42 (1.14–1.78)	0.002
Stage (vs. stage I)								
Stage II	8.24 (2.03–33.39)	0.003	1.74 (0.91–3.34)	0.097	5.96 (1.45–24.47)	0.020	1.28 (0.63–2.57)	0.493
Stage III	17.97 (4.45–72.55)	<0.001	2.30 (1.20–4.41)	0.012	10.23 (2.50–41.90)	0.005	1.76 (0.87–3.56)	0.117
CMS (vs. CMS4)								
CMS1	0.55 (0.35–0.86)	0.009	0.96 (0.63–1.45)	0.846	1.14 (0.63–2.06)	0.660	1.53 (0.88–2.62)	0.129
CMS2	0.59 (0.41–0.84)	0.003	0.63 (0.44–0.92)	0.015	1.33 (0.78–2.25)	0.291	1.01 (0.61–1.68)	0.969
CMS3	0.42 (0.25–0.73)	0.002	0.38 (0.20–0.72)	0.003	1.22 (0.59–2.53)	0.592	0.72 (0.33–1.54)	0.369

[a]Continuous variable.

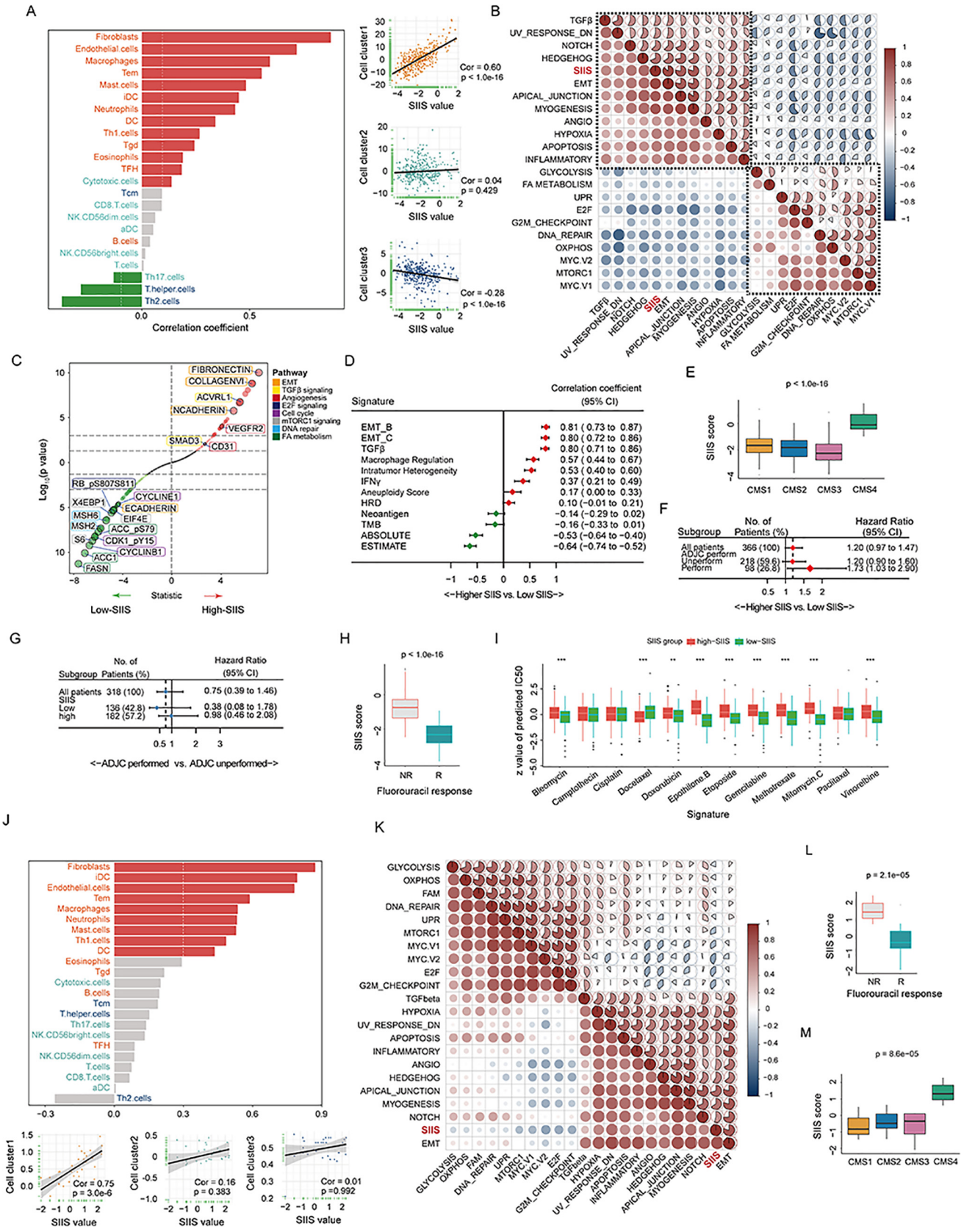
(Fig. 2D, E), especially for fibroblasts and endothelial cells, suggesting that the SIIS signature could be used as a quantitative index for measuring stromal cell infiltration intensity in colon cancer samples. Consistently, analyzing of a public colon cancer single-cell RNA-sequencing dataset also revealed higher SIIS value distributions in fibroblasts and endothelial cells when compared to other cell types (Fig. S2A-B). A receiver-operating characteristic curve analysis revealed that the area under curve of SIIS in distinguishing TMECS1 from TMECS2 and all other non-TMECS1 (including TMECS2 and TMECS3) patients was 0.90 and 0.82 respectively (Fig. 2F). To further clarify the internal relationship between SIIS and TMECS algorithm, we performed NTP analysis and SubMap analysis. As shown in Fig. S3A and S3B, the SIIS value was significant higher in samples of predicted TMECS1 subtypes than that in predicted non-TMECS1 subtypes (Fig. S3A) and samples of high-SIIS group shared significant similarity in marker gene expression with TMECS1 samples (Fig. S3B, adj p = 0.004).

Then, we sought to determine the clinical relevance of the SIIS model to the meta-GEO cohort. As expected, patients with higher SIIS had significantly increased relapse and mortality risks according to both univariate (Fig. 2G) and multivariate analyses (Table 1). Interestingly, the negative association between the SIIS and OS was obviously increased for patients who underwent ADJC (Fig. 2G), suggesting that this marker might have the ability to indicate the benefits of chemotherapy. To validate this speculation, we analyzed the relationships among the SIIS, ADJC benefits, and chemoresponses in the GSE39582 dataset. The median SIIS was significantly higher in the fluorouracil nonresponse group, and ADJC benefits obviously decreased with increased SIIS, especially in patients with stage II disease (Fig. 2H-J). Similarly, the high-SIIS group exhibited significantly higher estimated IC50 than low-SIIS group did regarding eight chemotherapy drugs in GDSC (Fig. 2K), and correlated with advanced stages, pMMR, CIN (+), C4 (CIT), CMS4, and activation of stromal and chemoresistant pathways (Fig. 2L-M). In terms of metabolic characteristics, we found that the majority of the items were highly upregulated in low-SIIS group, while the glycosaminoglycan metabolism, arachidonic acid metabolism, and retinoic metabolism were shown to be significantly activated in low SIIS tumors (Fig. S4A).

3.3. Validation of the role of the SIIS signature in the TCGA-COAD cohort

The performance of the SIIS signature was further evaluated using the TCGA-COAD dataset (n = 382) (Table S2). Consistent with the results of the meta-GEO (Fig. S3A-B), SIIS value was also significant elevated in predicted TMECS1 subtypes and the transcriptome characteristics of samples in high-SIIS group of TCGA-COAD dataset were closely linked to that of samples in TMECS1 subtypes of meta-GEO (adj p = 0.004). Moreover, the SIIS value also correlated positively with stromal component infiltration level in the TCGA-COAD cohort (Fig. 3A), and the high-SIIS group exhibited similar molecular traits and metabolic characteristics in the TCGA-COAD cohort as in the meta-GEO cohort (Fig. 3B, E and Fig. S4B). The pathway enrichment results at the transcriptome level were further confirmed at the protein level by analyzing the proteomic data of TCGA-COAD samples (Fig. 3C). We further tested the correlation of SIIS in the TCGA cohort using several known signatures obtained from the study by Thorsson et al [14]. Consistent with our previous findings, patients in the highSIIS group exhibited significantly higher EMT, TGF-β, and intratumoral heterogeneity scores, whereas scores associated with tumor purity were all down-regulated (Fig. 3D). However, there were no significant differences in terms of tumor mutation burden or the number of neoantigens among the different SIIS groups (Fig. 3D). Clinically, higher SIIS indicated not only increased mortality risk but also diminished survival benefits of chemotherapy (Fig. 3F, G). Finally, the drug response analysis indicated that higher SIIS was also nota-

Fig. 2. Construction and validation of the SIIS model in the meta-GEO cohort. (A) Workflow chart of the SIIS model establishment. (B) Circos plot showing the expression level and the survival impact of 31 stroma-related genes used for calculating SIIS (expressed higher in TMECS1, orange; expressed higher in TMECS2, turquoise; risk factor for relapse-free survival, red; protect factor for relapse-free survival, green). (C) Box plot of the comparison of infiltration level of each cell cluster between the different TME subtypes. (D) Bar chart showing the correlations between SIIS value and infiltration levels of TME cells in the meta-GEO cohort. Nonsignificant correlations are marked by gery. (E) Scatter plot showing the correlations between SIIS value and infiltration levels of cell clusters in the meta-GEO cohort. (F) Receiver operating characteristics curve of the SIIS model. (G) Forest plot showing the associations between SIIS and relapse-free survival (left) and the associations between SIIS and overall survival (right) in various subgroups. (H) Box plot of the comparison of the SIIS score between the fluorouracil response and nonresponse groups. (I) The survival curves of ADJC performance for overall survival in low-SIIS (left) and high-SIIS (right) subgroups. (J) Forest plot showing the benefit of ADJC in different SIIS subgroups. (K) Box plot of estimated IC50 values of chemotherapy drugs in high-SIIS and low-SIIS subgroup in the GSE39582 cohort. (L) Box plot of SIIS values in different clinical subgroups. (M) Correlation matrix of SIIS values and the activation levels of biological process in the meta-GEO cohort. The values of correlation coefficients were represented by both the shading of the color and sector size. *TMECS, tumor microenvironment cell subtype; TME, tumor microenvironment; DEG, deferentially expressed genes; FC, fold change; UVCA, univariate Cox analysis; SIIS, stromal cell infiltration intensity; Cor, correlation; Tcm, T central memory; Tem, T effector memory; Th, T helper; DC, dendritic cell; aDC, activated dendritic cell; iDC, immature dendritic cell; Tgd, T gamma delta; NK, natural killer; TFH, T follicular helper; AUC, area under curve; CI, confidence interval; CIN, chromosome instability; RFS, relapse-free survival; OS, overall survival; ADJC, adjuvant chemotherapy; R response; NR, nonresponse; CMS, consensus molecular subtypes; MMR, mismatch repair; EMT, epithelial-mesenchymal transition; OXPHOS, oxidative phosphorylation; UPR, unfolded protein response; FA, fatty acid.* (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)


A
Fibroblasts
Endothelial.cells
Macrophages
Tem
Mast.cells
iDC
Neutrophils
DC
Th1.cells
Tgd
Eosinophils
TFH
Cytotoxic.cells
Tcm
CD8.T.cells
NK.CD56dim.cells
aDC
B.cells
NK.CD56bright.cells
T.cells
Th17.cells
T.helper.cells
Th2.cells
Correlation coefficient
Cell cluster1
Cor = 0.60
p < 1.0e-16
Cell cluster2
Cor = 0.04
p = 0.429
Cell cluster3
Cor = -0.28
p < 1.0e-16
SIIS value
B
TGFβ
UV_RESPONSE_DN
NOTCH
HEDGEHOG
SIIS
EMT
APICAL_JUNCTION
MYOGENESIS
ANGIO
HYPOXIA
APOPTOSIS
INFLAMMATORY
GLYCOLYSIS
FA METABOLISM
UPR
E2F
G2M_CHECKPOINT
DNA_REPAIR
OXPHOS
MYC.V2
MTORC1
MYC.V1
C
FIBRONECTIN
COLLAGENVI
ACVRL1
NCADHERIN
VEGFR2
SMAD3
CD31
RB_pS807S811
CYCLINE1
X4EBP1
ECADHERIN
MSH6
EIF4E
MSH2
ACC_pS79
S6
CDK1_pY15
CYCLINB1
ACC1
FASN
Pathway
EMT
TGFβ signaling
Angiogenesis
E2F signaling
Cell cycle
mTORC1 signaling
DNA repair
FA metabolism
Log10(p value)
Statistic
Low-SIIS
High-SIIS
D
Signature
Correlation coefficient (95% CI)
EMT_B 0.81 (0.73 to 0.87)
EMT_C 0.80 (0.72 to 0.86)
TGFβ 0.80 (0.71 to 0.86)
Macrophage Regulation 0.57 (0.44 to 0.67)
Intratumor Heterogeneity 0.53 (0.40 to 0.60)
IFNγ 0.37 (0.21 to 0.49)
Aneuploidy Score 0.17 (0.00 to 0.33)
HRD 0.10 (−0.01 to 0.21)
Neoantigen −0.14 (−0.29 to 0.02)
TMB −0.16 (−0.33 to 0.01)
ABSOLUTE −0.53 (−0.64 to −0.40)
ESTIMATE −0.64 (−0.74 to −0.52)
<−Higher SIIS vs. Low SIIS−>
E
p < 1.0e-16
SIIS score
CMS1
CMS2
CMS3
CMS4
F
Subgroup
No. of Patients (%)
Hazard Ratio (95% CI)
All patients 366 (100) 1.20 (0.97 to 1.47)
ADJC perform
Unperform 218 (59.6) 1.20 (0.90 to 1.60)
Perform 98 (26.8) 1.73 (1.03 to 2.90)
<−Higher SIIS vs. Low SIIS−>
G
Subgroup
No. of Patients (%)
Hazard Ratio (95% CI)
All patients 318 (100) 0.75 (0.39 to 1.46)
SIIS
Low 136 (42.8) 0.38 (0.08 to 1.78)
high 182 (57.2) 0.98 (0.46 to 2.08)
<−ADJC performed vs. ADJC unperformed−>
H
p < 1.0e-16
SIIS score
NR
R
Fluorouracil response
I
SIIS group high−SIIS low−SIIS
z value of predicted IC50
Bleomycin
Camptothecin
Cisplatin
Docataxel
Doxorubicin
Epothilone.B
Etoposide
Gemcitabine
Methotrexate
Mitomycin.C
Paclitaxel
Vinorelbine
Signature
J
Fibroblasts
iDC
Endothelial.cells
Tem
Macrophages
Neutrophils
Mast.cells
Th1.cells
DC
Eosinophils
Tgd
Cytotoxic.cells
B.cells
Tcm
T.helper.cells
Th17.cells
NK.CD56bright.cells
TFH
NK.CD56dim.cells
T.cells
CD8.T.cells
aDC
Th2.cells
Cell cluster1
Cor = 0.75
p = 3.0e-6
Cell cluster2
Cor = 0.16
p = 0.383
Cell cluster3
Cor = 0.01
p =0.992
SIIS value
K
GLYCOLYSIS
OXPHOS
FAM
DNA_REPAIR
UPR
MTORC1
MYC.V1
MYC.V2
E2F
G2M_CHECKPOINT
TGFbeta
HYPOXIA
UV_RESPONSE_DN
APOPTOSIS
INFLAMMATORY
ANGIO
HEDGEHOG
APICAL_JUNCTION
MYOGENESIS
NOTCH
SIIS
EMT
L
p = 2.1e-05
SIIS score
NR
R
Fluorouracil response
M
p = 8.6e-05
SIIS score
CMS1
CMS2
CMS3
CMS4

bly linked to the fluorouracil nonresponse signature (Fig. 3H) and elevated estimated IC50 of multiple chemotherapy drugs (Fig. 3I) in analysis of the TCGA-COAD dataset. These results highlight the robustness of our SIIS signature to predict unfavorable prognoses for and chemoresistance in patients with colon cancer.

3.4. Validation of the role of the SIIS signature in the Sun Yat-sen University cancer Center (SYSUCC) cohort

In addition to the meta-GEO and TCGA-COAD cohorts, we also validated the role of SIIS in the SYSUCC cohort (Table S2). Consistent with the results of the meta-GEO and TCGA-COAD databases, the higher distribution of SIIS value in predicted TMECS1 (Fig. S3A), the transcriptome similarity between high-SIIS group and TMECS1 (Fig. S3B), the infiltration of cell cluster 1, activation of stromal-related and chemoresistant pathways, enrichment of fluorouracil nonponse status and CMS4 subtypes in high-SIIS group could also be observed in this independent cohort (Fig. 3J-M). These data underscore the performance of the SIIS to reflect the infiltration intensity of stromal cells in colon cancer, thus indicating strong reproducibility.

3.5. Exploration of the role of the SIIS signature in pan cancer setting

Next, we evaluated the role of the SIIS across cancers. Transcriptome data of 11 TCGA cohorts (BLCA, BRCA, CESC, ESCA, HNSC, LIHC, LUAD, LUSC, READ, SKCM, STAD) (Table S4) were selected for pan-cancer analyses. In terms of the TME characteristics and transcriptome traits of SIIS value, we found that the SIIS value was positively associated with infiltration abundance of stromal cell (fibroblasts, endothelial cells) and mast cell, as well as the activation of TGF-β response scores and EMT signatures, but negatively associated with total T helper cells in all cancer cohorts we analyzed, (Fig. 4A; Tables S5-S6). These results suggested that SIIS score can also be used as an effective quantitative tool of stromal component in pan cancer. However, different from the results of colon cancer cohort analysis, the SIIS value in BRCA, CESC, HNSC, LUAD, and SKCM was shown to be significant negatively correlated with cytotoxic cell infiltration and IFNγ signaling (Fig. 4A; Tables S5-S6). As for the relationship between SIIS and chemodrug sensitivity, the results showed in the BLCA, BRCA, CESC, HNSC, LUAD, LUSC, SKCM, and STAD cohorts, the predicted IC50 values of more than 6 drugs increased significantly in the high-SIIS group. Notably, among these tumor types, the indication effect of SIIS model on chemoresistance may be the strongest in HNSC, because t-statistics of all chemotherapeutic drugs are greater than 5 (t-statistics of 7 drugs are greater than 10). Interestingly, the down-regulation of predicted IC50 of docetaxel in high-SIIS groups could only be observed in BRCA, ESCA and LIHC cohorts, indicating that the role of docetaxel as a potential drug choice for patients of high-SIIS group might be highly dependent on cancer type. We then explored the survival impact of the SIIS model in these cancers. As shown in Fig. 4B, ten cohorts (except ESCA) experienced an increase in the mortality risk when patients had higher SIIS; however, only three of them (BLCA, SKCM, and STAD) reached statistical significance (Fig. 4B). Finally, we compared the distribution of SIIS across different tumor types (Fig. 4C). Intriguingly, we found that colon and breast cancers were the tumor types with the lowest and highest SIIS distributions, respectively. We further obtained the PAM50 subtyping information documented in the TCGA-BRCA database. Patients with the luminal A subtype had the highest median SIIS, whereas the median SIIS for patients with basal-like disease (also known as triple-negative breast cancer) was the lowest (Fig. 4D). The basal-like subtype has been reported to be sensitive to chemotherapy; therefore, these results further suggest that the SIIS might be capable of predicting chemotherapy efficacy for patients with breast cancer.

3.6. SIIS signature in predicting immunotherapeutic benefits

Recent evidence has supported that the immune-excluded phenotype triggered by stromal cell enrichment, activation of TGF-β signaling, and the EMT pathway are associated with a lack of response to immunotherapy [15]. Considering the strong connection between the SIIS and stromal activation in pan-cancer cohorts, we next investigated whether the SIIS signature could predict immunotherapeutic benefits for solid tumors. Four independent immunotherapy cohorts with more than 50 samples in each cohort were analyzed (Table S8). In all cohorts, higher SIIS were significantly associated with lower treatment response rates (Fig. 4E-H). Moreover, the survival data showed that patients treated with immunotherapy in the low-SIIS group had significantly prolonged progression-free survival (GSE176307) and overall survival (Imvigor210) than those in the high-SIIS group (Fig. S5). The comparison of the SIIS distribution of metastatic urothelial cancer patients with different molecular characteristics (Fig. 4I, J) showed that SIIS were significantly higher in the immune-excluded phenotype, low TMB, and low NEO groups than in the other groups. In terms of the TCGA subtype comparison (Fig. 4J), the SIIS was the highest for the "infiltrated" subtype patients and lowest for genomically unstable patients. However, although the SIIS tended to be decreased in the FGFR mutant subgroup (Fig. 4I) and in the subgroup with high PD-L1 expression on immune cells (Fig. 4J), statistical differences were not reached. Above all, the results of these four immunotherapy cohorts suggest that the SIIS is associated with the response to immunotherapies and can further predict the prognosis of patients.

Fig. 3. Validation of the SIIS model in the TCGA-COAD and SYSUCC cohorts. (A) Bar chart (left) showing the correlations between SIIS value and infiltration levels of TME cells and scatter plot (right) showing the correlation between SIIS value and cell clusters infiltration levels in the TCGA-COAD cohort. Nonsignificant correlations are marked by gery. (B) Correlation matrix of SIIS values and the activation levels of biological process in the TCGA-COAD cohort. (C) Scatter plot represents the comparison of the protein expression level of the pathway marker genes between patients in the high-SIIS and low-SIIS groupd. (D) Forest plot showing the associations between SIIS and other published signatures in the TCGA-COAD cohort. (E) Box plot of the comparison of SIIS values between the different CMS subtypes. (F) Forest plot showing the associations between SIIS and overall survival in various subgroups. (G) Forest plot showing the benefit of ADJC in different subtypes. (H) Box plot of the comparison of the SIIS score between the fluorouracil response and nonresponse groups. (I) Box plot of estimated IC50 values of chemotherapy drugs in high-SIIS and low-SIIS subgroup in the TCGA-COAD cohort. (J) Bar chart (upper) showing the correlations between SIIS value and infiltration levels of TME cells and scatter plot (down) showing the correlations between SIIS value and infiltration levels of cell clusters in the SYSUCC cohort. Nonsignificant correlations are marked by gery. (K) Correlation matrix of SIIS values and the activation levels of biological process in the SYSUCC cohort. The value of correlation coefficients was represented by both the shading of the color and sector size. (L-M) Box plot of SIIS values in the fluorouracil response and nonresponse groups (L) and in the CMS subgroups (M). *SIIS, stromal cell infiltration intensity; Cor, correlation; Tcm, T central memory; Tem, T effector memory; Th, T helper; DC, dendritic cell; aDC, activated dendritic cell; iDC, immature dendritic cell; Tgd, T gamma delta; NK, natural killer; TFH, T follicular helper; AUC, area under curve; CI, confidence interval; ADJC, adjuvant chemotherapy; R response; NR, nonresponse; CMS, consensus molecular subtypes; EMT, epithelial-mesenchymal transition; OXPHOS, oxidative phosphorylation; UPR, unfolded protein response; FA, fatty acid.*

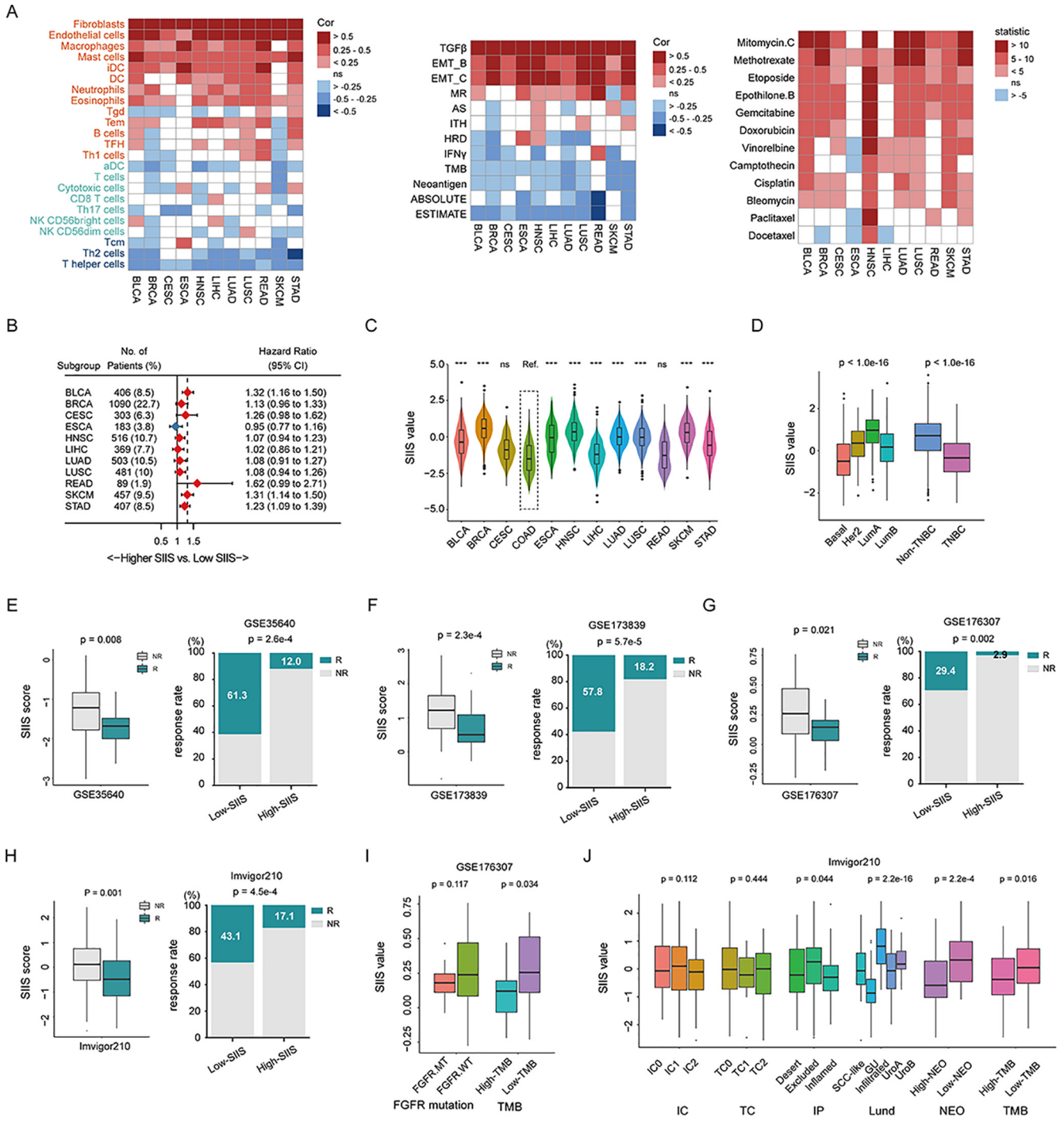
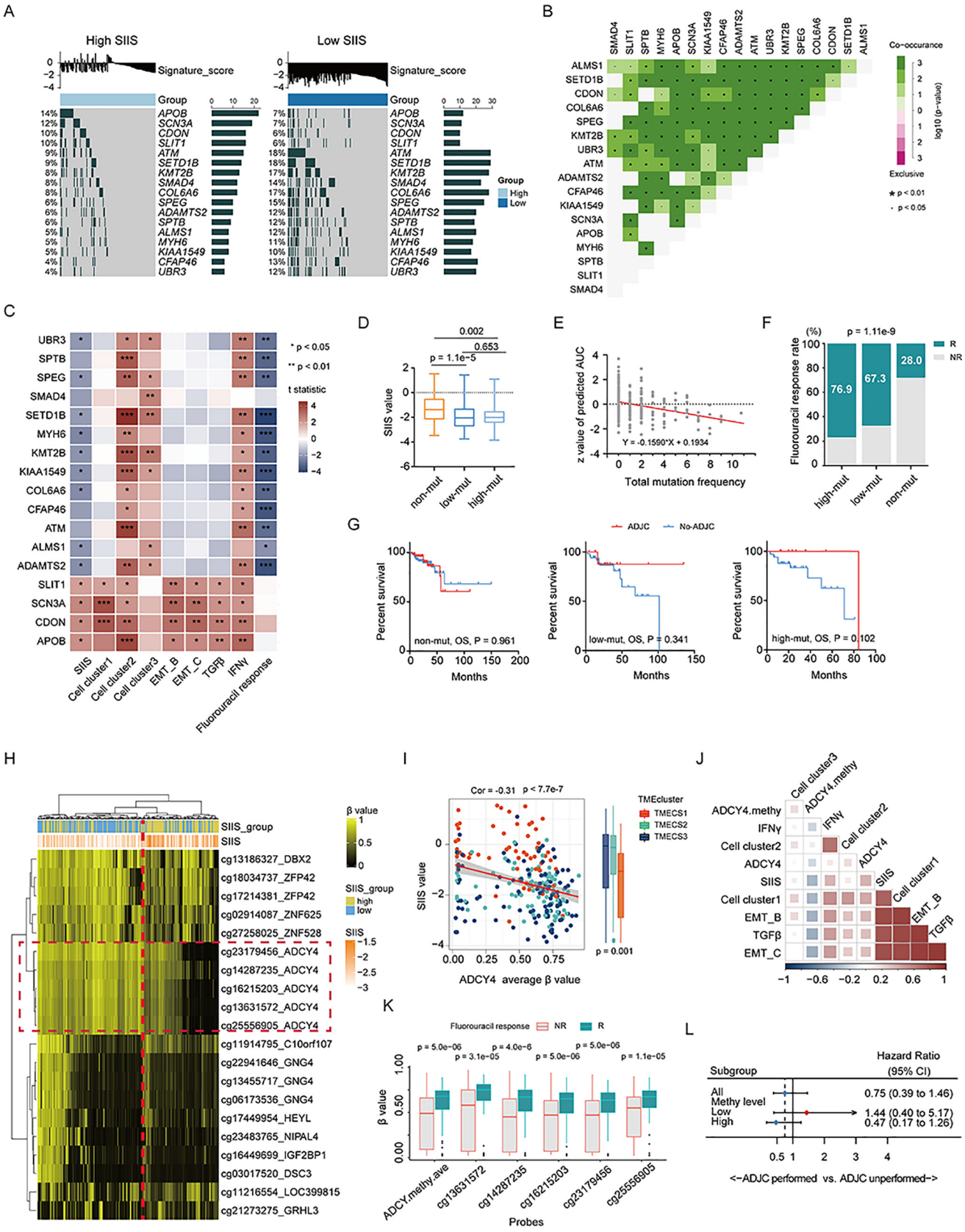


Fig. 4. Validation of the SIIS model in the TCGA pan cancer cohorts and the immunotherapy-treated cohorts. (A) Heatmaps showing the associations between SIIS and TME cells (left), the associations between SIIS and other published signatures (middle), and the associations between SIIS and esitimated IC50 values of chemotherapy drugs (right) in the TCGA pan cancer cohorts. (B) Forest plot showing the associations between SIIS and overall survival in TCGA pan cancer cohorts. (C) Violin plot of SIIS values in TCGA pan cancer cohorts. (D) Box plot of the comparison of SIIS values between the different PAM50 subtypes and between TNBC and non-TNBC group. (E-H) Box plot (left) of the comparison of SIIS values between the immunotherapy response and nonresponse group, and bar charts (right) summarizing the proportions of patients with immunotherapy response and those with non-response within and across low-SIIS and high-SIIS groups in the GSE35640 (E), GSE173839 (F), GSE176307 (G), and Imvigor210 (H) cohorts. (I) Box plot of the SIIS values in patients with different FGFR mutation and TMB status in the GSE176307 cohort. (J) Box plot of the SIIS values in patients with different immune phenotypes. *SIIS, stromal cell infiltration intensity; Cor, correlation; Tcm, T central memory; Tem, T effector memory; Th, T helper; DC, dendritic cell; aDC, activated dendritic cell; iDC, immature dendritic cell; Tgd, T gamma delta; NK, natural killer; TFH, T follicular helper; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; ESCA, esophageal carcinoma; HNSC, head and neck squamous cell carcinoma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; READ, rectum adenocarcinoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; ITH, intratumor heterogeneity; HRD, homologous recombination deficiency; AS, aneuploidy score; MR, macrophage regulation; CI, confidence interval; R response; NR, nonresponse; MT, mutant type; WT, wild type; TNBC, triple negative breast cancer; NEO, neoantigen, TMB, tumor mutation burden; IC, immune cell; TC, tumour cell; IP, immune phenotype; GU, genomically unstable; Uro, urobasal.*


A
High SIIS
Low SIIS
Signature_score
Group
APOB
SCN3A
CDON
SLIT1
ATM
SETD1B
KMT2B
SMAD4
COL6A6
SPEG
ADAMTS2
SPTB
ALMS1
MYH6
KIAA1549
CFAP46
UBR3
High
Low
B
Co-occurance
Exclusive
log10 (p-value)
C
t statistic
SIIS
Cell cluster1
Cell cluster2
Cell cluster3
EMT_B
EMT_C
TGFβ
IFNγ
Fluorouracil response
D
SIIS value
non-mut
low-mut
high-mut
E
z value of predicted AUC
Total mutation frequency
F
Fluorouracil response rate
R
NR
G
ADJC
No-ADJC
Percent survival
Months
non-mut, OS, P = 0.961
low-mut, OS, P = 0.341
high-mut, OS, P = 0.102
H
SIIS_group
β value
cg13186327_DBX2
cg18034737_ZFP42
cg17214381_ZFP42
cg02914087_ZNF625
cg27258025_ZNF528
cg23179456_ADCY4
cg14287235_ADCY4
cg16215203_ADCY4
cg13631572_ADCY4
cg25556905_ADCY4
cg11914795_C10orf107
cg22941646_GNG4
cg13455717_GNG4
cg06173536_GNG4
cg17449954_HEYL
cg23483765_NIPAL4
cg16449699_IGF2BP1
cg03017520_DSC3
cg11216554_LOC399815
cg21273275_GRHL3
I
Cor = -0.31 p < 7.7e-7
TMEcluster
TMECS1
TMECS2
TMECS3
p = 0.001
ADCY4 average β value
J
ADCY4.methy
IFNγ
Cell cluster2
ADCY4
SIIS
Cell cluster1
EMT_B
TGFβ
EMT_C
K
Fluorouracil response
Probes
ADCY.methy.ave
L
Subgroup
Hazard Ratio (95% CI)
All
Methy level
Low
High
0.75 (0.39 to 1.46)
1.44 (0.40 to 5.17)
0.47 (0.17 to 1.26)
<-ADJC performed vs. ADJC unperformed->

3.7. Multi-omic analysis

To uncover the potential mechanisms underlying the intrinsic chemoresistance in patients within high-SIIS group, the genomic analysis was performed in the TCGA-COAD cohort. There were 17 genes, the mutation frequencies of which were significantly different between patients in the high- and low-SIIS group, including 4 genes with higher mutation frequencies in the high-SIIS group and the mutaions of the remaining 13 genes were mainly enriched in the low-SIIS group (Fig. 5A). The co-occurrence status of these genes was shown in Fig. 5B. Through *t*-test analysis, we found that among these 13 low-SIIS related differentially mutated genes, the mutation status of 12 genes (except SMAD4) was significantly associated with fluorouracil sensitivity. We further constructed a combined mutation score based on the mutation states of these 12 genes: if the patient has mutation in any of these genes, 1 point will be counted. As shown in Fig. 5D-F, with the increase of mutation score, the SIIS value (Fig. 5D-E) and fluorouracil nonresponse rates (Fig. 5F) were both decreased. The following survival analysis showed that patients with non-mutation of these genes (the mutation score was 0) could not benefit from ADJC (Fig. 5G). Taken together, these results suggest that lack of genetic mutation is one of the mechanisms responsible for chemotherapy resistance in patients with high SIIS score.

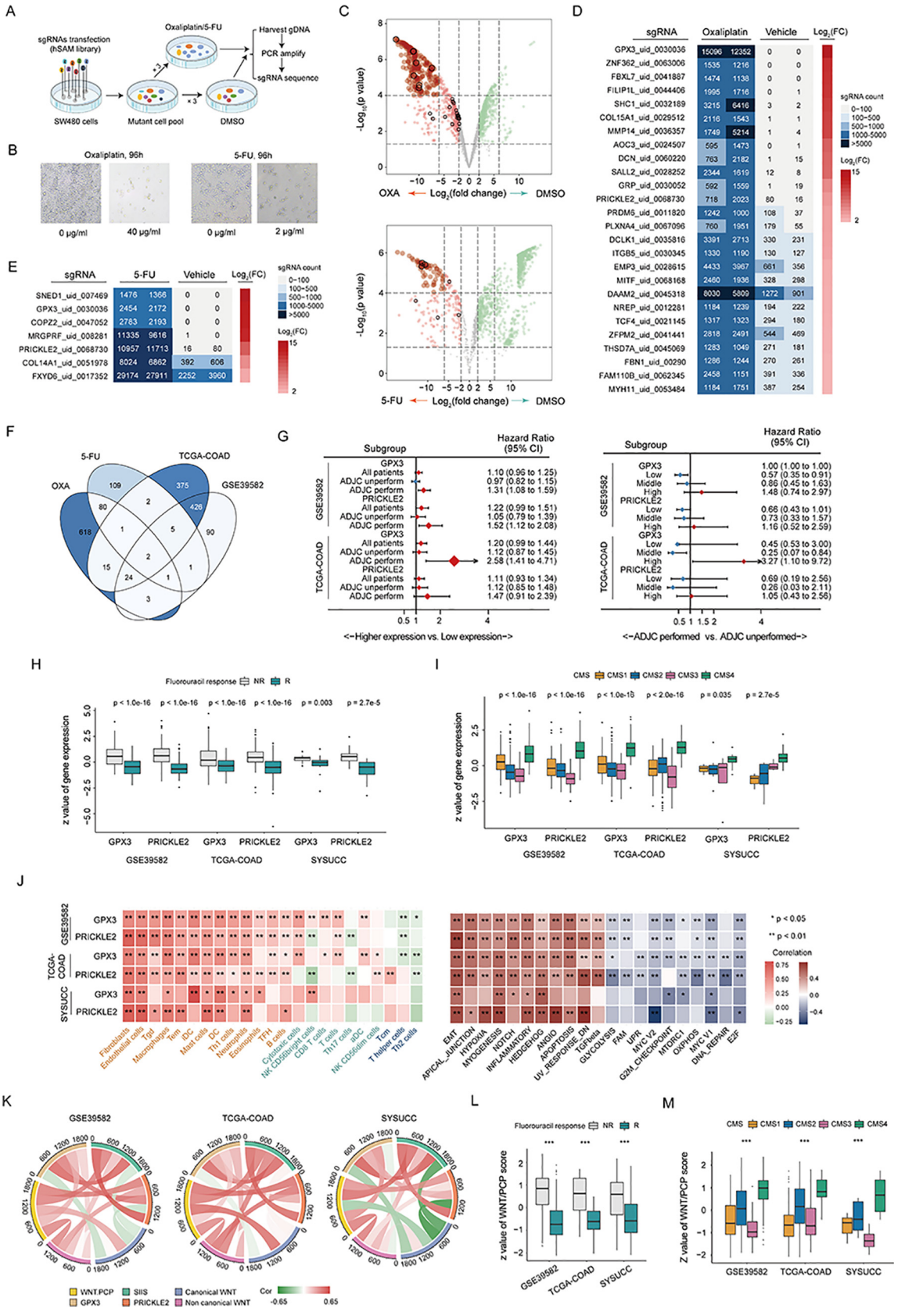
In addition to genomic mutations, we also explored the correlations between gene promoter methylation level and SIIS values. Through a comprehensive investigation into the DNA promoter methylation landscape, we found a total of 125 probes with high variability, the β values detected by which were significantly correlated with the SIIS value (adjusted p value < 0.05, Table S9). Among them, we noted that there are 5 probes targeting the promoter region of ADCY4 gene, and the methylation level detected by cg14287235 showed the strongest correlation with the SIIS value (Fig. 5H). Thus, we focused on ADCY4 in the subsequent analysis. As shown in Fig. 5I and 5J, the correlation analysis revealed that ADCY4 methylation was significantly negatively correlated with the level of stromal cell infiltration and activation of EMT and TGFβ pathways, while its transcriptional expression was positively correlated with these stromal-related characteristics. However, the association between transcriptional expression and methylation of ADCY4 was weak, suggesting that the effect of ADCY4 methylation on biological behavior may not be achieved by affecting its transcription. Finally, the results of chemotherapy efficacy analysis also demonstrated that the methylation level of ADCY4 decreased significantly in patients who did not respond to fluorouracil (Fig. 5K), and only patients with high ADCY4 methylation (β value > 0.5) tended to benefit from ADJC (Fig. 5L). Collectively, we proposed that DNA methylation, such as hypomethylation of ADCY4, may offer a lens into the complexity and diversity of the TME, especially for stromal caused chemoresistance.

3.8. GPX3 and PRICKLE2 were identified as driver genes for chemotherapy resistance of patients in the high-SIIS group

The CRISPR screen provides an efficient high-throughput technique and a powerful opportunity to access novel therapeutic biomarkers. To identify critical genes involved in chemoresistance of patients with high SIIS, we performed gain-of-function screening of SW480 cells utilizing the human genome-scale CRISPR/CAS9 SAM2 pooled library with 40 μM of oxaliplatin, and 2 μM of 5-fluorouracil (5-FU) used as an effective selection pressure (see Materials and methods, Fig. 6A, B). From this screen, we identified a subset of 488 sgRNAs that were significantly enriched in the oxaliplatin-treated cells (Fig. 6C, upper), whereas 279 sgRNAs were significantly enriched in the 5-FU-treated cells (Fig. 6C, bottom) when compared to the vehicle control, indicating that the genes targeted by these sgRNAs might be potential drivers of fluorouracil or oxaliplatin resistance. Through a correlation analysis of the SIIS and expressions of these potential driver genes adjusted for tumor purity, a total of 31 genes that were significantly positively correlated with SIIS were identified in the GSE39582 and TCGA-COAD cohorts (Fig. 6C-E). Among these genes, GPX3 and PRICKLE2 attracted special attention because they are intersection genes whose sgRNA were significantly increased in both oxaliplatin-treated and 5-FU-treated populations simultaneously, and whose expressions were increased in samples from the high-SIIS groups in the GSE39582 and TCGA-COAD cohorts (Fig. 6F). A clinical relevance analysis suggested that higher GPX3 and PRICKLE2 expressions reflected the adverse outcomes of RFS and OS, especially for patients who underwent ADJC (Fig. 6G, left). Moreover, ADJC increased the mortality risk for patients in the high GPX3 and high PRICKLE2 groups (Fig. 6G, right). Boxplots (Fig. 6H, I) show that mRNA expressions of GPX3 and PRICKLE2 were significantly elevated in the fluorouracil nonresponse and CMS4 subtype groups. Pathway and immune analyses confirmed that the infiltration abundance of members in cell cluster 1 and stroma pathway activation levels significantly increased as GPX3 and PRICKLE2 expressions increased in the GS39582, TCGA-COAD, and SYSUCC cohorts (Fig. 6J).

PRICKLE2 is one of the components of the WNT/PCP pathway, a branch of noncanonical WNT pathways controlling tissue polarity and cell movement [16]. Because mounting evidence has uncovered the role of WNT/PCP signaling deregulation in malignant phenotypes, including chemoresistance in cancer, we assessed the association between SIIS and WNT/PCP activation. Interestingly, we found that the SIIS was significantly positively correlated with WNT/PCP activity levels in all three databases that we analyzed (Fig. 6K), whereas the correlations between the SIIS and β-catenin-dependent canonical WNT signaling were weak (TCGA-COAD) or even manifested significantly negative correlations (GSE39582 and SYSUCC) (Fig. 6K). We also observed a consistently

Fig. 5. SIIS associated genomic alteration and methylation characteristics. (A) Oncoprints depicted the genomic alteration landscapes in the context of high- and low-SIIS value. (B) Matrix heatmap of interaction effect of genes mutating differentially in patients in the low- and the high-SIIS groups. (C) Matrix heatmap of differential value of multiple signatures between gene mutant and wild type group. Red marked square indicates higher value in the gene mutant group than in the wild type group, and blue marked square indicate higher value in the wild type group than in the gene mutant group. (D-E) Box plot (D) and scatter plot (E) of SIIS value distribution in groups of different mutation score. (F) Bar charts summarizing the proportions of patients with fluorouracil response and those with non-response within and across groups of different mutation score. (G) The survival curve of ADJC performance for overall survival in groups of different mutation score. (H) Heatmap exhibited the landscape of differentially methylated probes in high- and low-SIIS group. (I) Scatter plot showing the correlation between the average ADCY4 methylation value and SIIS value. Every single dot represents one sample, and corresponding TME subtypes are identified in different colors. (J) Correlation matrix showing the associations among ADCY4 methylation level, ADCY4 transcriptional expression, and multiple signatures in the TCGA-COAD cohort. (K) Box plot of ADCY4 methylation level in fluorouracil response and nonresponse groups. (M) Forest plot showing the benefit of ADJC in subgroups stratified by total ADCY4 methylation levels. *SIIS, stromal cell infiltration intensity; ADJC, adjuvant chemotherapy; R response; NR, nonresponse; EMT, epithelial-mesenchymal transition; CM, combined mutation; Cor, correlation; CI, confidence interva.* $^{*}p < 0.05$ *and,* $^{**}p < 0.01$, $^{***}p < 0.001$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)


A
sgRNAs transfection (hSAM library)
Oxaliplatin/5-FU
Harvest gDNA
PCR amplify
sgRNA sequence
SW480 cells
Mutant cell pool
DMSO
B
Oxaliplatin, 96h
5-FU, 96h
0 μg/ml
40 μg/ml
0 μg/ml
2 μg/ml
C
-Log10(p value)
OXA ← Log2(fold change) → DMSO
5-FU ← Log2(fold change) → DMSO
D
sgRNA
Oxaliplatin
Vehicle
Log2(FC)
sgRNA count
0–100
100–500
500–1000
1000–5000
>5000
GPX3_uid_0030036 15096 12352 0 0
ZNF362_uid_0063006 1535 1216 0 0
FBXL7_uid_0041887 1474 1138 0 0
FILIP1L_uid_0044406 1995 1716 0 1
SHC1_uid_0032189 3215 6416 3 2
COL15A1_uid_0029512 2116 1543 1 1
MMP14_uid_0036357 1749 5214 1 4
AOC3_uid_0024507 595 1473 0 1
DCN_uid_0060220 763 2182 1 15
SALL2_uid_0028252 2344 1619 12 8
GRP_uid_0030052 592 1559 1 19
PRICKLE2_uid_0068730 718 2023 80 16
PRDM6_uid_0011820 1242 1000 108 37
PLXNA4_uid_0067096 760 1951 179 55
DCLK1_uid_0035816 3391 2713 330 231
ITGB5_uid_0030345 1330 1190 130 127
EMP3_uid_0028615 4433 3967 661 356
MITF_uid_0068168 2460 1936 328 298
DAAM2_uid_0045318 8030 5809 1272 901
NREP_uid_0012281 1184 1239 194 222
TCF4_uid_0021145 1317 1323 294 180
ZFPM2_uid_0041441 2818 2491 544 469
THSD7A_uid_0045069 1283 1049 271 181
FBN1_uid_00290 1286 1244 270 261
FAM110B_uid_0062345 2458 1151 391 336
MYH11_uid_0053484 1184 1751 387 254
E
sgRNA
5-FU
Vehicle
Log2(FC)
SNED1_uid_007469 1476 1366 0 0
GPX3_uid_0030036 2454 2172 0 0
COPZ2_uid_0047052 2783 2193 0 0
MRGPRF_uid_008281 11335 9616 1 0
PRICKLE2_uid_0068730 10957 11713 16 80
COL14A1_uid_0051978 8024 6862 392 606
FXYD6_uid_0017352 29174 27911 2252 3960
F
5-FU
TCGA-COAD
OXA
GSE39582
109 375 426 80 2 618 1 5 90 15 24 1 2 1 3
G
Subgroup
Hazard Ratio (95% CI)
GSE39582
GPX3
All patients 1.10 (0.96 to 1.25)
ADJC unperform 0.97 (0.82 to 1.15)
ADJC perform 1.31 (1.08 to 1.59)
PRICKLE2
All patients 1.22 (0.99 to 1.51)
ADJC unperform 1.05 (0.79 to 1.39)
ADJC perform 1.52 (1.12 to 2.08)
TCGA-COAD
GPX3
All patients 1.20 (0.99 to 1.44)
ADJC unperform 1.12 (0.87 to 1.45)
ADJC perform 2.58 (1.41 to 4.71)
PRICKLE2
All patients 1.11 (0.93 to 1.34)
ADJC unperform 1.12 (0.85 to 1.48)
ADJC perform 1.47 (0.91 to 2.39)
<–Higher expression vs. Low expression–>
GSE39582
GPX3
Low 1.00 (1.00 to 1.00)
Middle 0.57 (0.35 to 0.91)
High 0.86 (0.45 to 1.63)
PRICKLE2
Low 1.48 (0.74 to 2.97)
Middle 0.66 (0.43 to 1.01)
High 0.73 (0.33 to 1.57)
TCGA-COAD
GPX3
Low 1.16 (0.52 to 2.59)
Middle 0.45 (0.53 to 3.00)
High 0.25 (0.07 to 0.84)
PRICKLE2
Low 3.27 (1.10 to 9.72)
Middle 0.69 (0.19 to 2.56)
High 0.26 (0.03 to 2.11)
1.05 (0.43 to 2.56)
<–ADJC performed vs. ADJC unperformed–>
H
Fluorouracil response NR R
p < 1.0e-16 p < 1.0e-16 p < 1.0e-16 p < 1.0e-16 p = 0.003 p = 2.7e-5
z value of gene expression
GPX3 PRICKLE2 GPX3 PRICKLE2 GPX3 PRICKLE2
GSE39582 TCGA-COAD SYSUCC
I
CMS CMS1 CMS2 CMS3 CMS4
p < 1.0e-16 p < 1.0e-16 p < 1.0e-16 p < 2.0e-16 p = 0.035 p = 2.7e-5
z value of gene expression
GPX3 PRICKLE2 GPX3 PRICKLE2 GPX3 PRICKLE2
GSE39582 TCGA-COAD SYSUCC
J
GSE39582
TCGA-COAD
SYSUCC
GPX3
PRICKLE2
Fibroblasts Endothelial cells Tgd Macrophages Tem iDC Mast cells DC Th1 cells Neutrophils Eosinophils TFH B cells Cytotoxic cells NK CD56bright cells CD8 T cells T cells Th17 cells aDC NK CD56dim cells Tcm T helper cells Th2 cells
EMT APICAL_JUNCTION HYPOXIA MYOGENESIS NOTCH INFLAMMATORY HEDGEHOG ANGIO APOPTOSIS UV_RESPONSE_DN TGFbeta GLYCOLYSIS FAM UPR MYC V2 G2M_CHECKPOINT MTORC1 OXPHOS MYC V1 DNA_REPAIR E2F
* p < 0.05
** p < 0.01
Correlation
K
GSE39582
TCGA-COAD
SYSUCC
WNT/PCP
GPX3
SIIS
PRICKLE2
Canonical WNT
Non canonical WNT
Cor
-0.65
0.65
L
Fluorouracil response NR R
z value of WNT/PCP score
GSE39582 TCGA-COAD SYSUCC
M
CMS CMS1 CMS2 CMS3 CMS4
Z value of WNT/PCP score
GSE39582 TCGA-COAD SYSUCC

increased GSVA score of the WNT/PCP pathway in the fluorouracil nonresponse and CMS4 groups of the GSE39582, TCGA-COAD, and SYSUCC colon cancer cohorts (Fig. 6L, M). Collectively, these results suggest that the WNT/PCP pathway, but not the canonical WNT/β-catenin pathway, mainly contributed to the biological features of patients with high SIIS.

Because GPX3 was one of the members of the SIIS signature, we moved forward with GPX3 in the following study. To validate the results of our CRISPR library screen, we performed an in vitro study by transfecting exogenous GPX3 plasmid or siRNA targeting GPX3 into SW480 (GPX3 low-expressing cells [17]) and CACO2 (GPX3 high-expressing cells [17]) colon cancer cell lines, respectively. Consistently, exogenous GPX3 overexpression recapitulated the chemoresistance-promoting effect of sgRNA-mediated transcriptional activation in SW480 cells toward 5-FU and oxaliplatin treatment, as determined by MTT (Fig. 7A) and cell colony formation assays (Fig. 7C), whereas downregulation of GPX3 in CACO2 cells resulted in a dramatic increase in response to 5-FU and oxaliplatin during in vitro (Fig. 7B, C) and in vivo experiments (Fig. 7D, E). We detected the expression and clinical significance of GPX3 in colon cancer using a paraffin-embedded tissue microarray (Fig. 7G, Table S2). Based on the protein level, we also confirmed that high GPX3 expression independently identified patients with an inferior disease-free survival (Fig. 7F, Table S10). In terms of chemotherapy benefits, we found that ADJC, especially high-intensity chemotherapy schemes (oxaliplatin combination, 6 months of chemotherapy, or 8 cycles of Xeloda), tended to increase the risk of relapse or death (Fig. 7H, I) for patients with high GPX3 protein expression. Similar trends were also observed with the stratification of the tumor stage (Fig. S6A-S6B).

3.9. Identification of candidate molecular targets and compounds for chemosensitization of patients in the high-SIIS group

To identify candidate molecular targets and compounds that may be options for achieving chemosensitization in patients with high SIIS, we used Cmap tools. After filtering the results, we identified several most enriched candidate molecular targets (Table S11) and compounds (Table S12) in the analysis performed in the GSE39582 and TCGA-COAD cohorts (Fig. 7J). The mechanisms of action of the compounds are shown in Fig. 7K. Among the intersection molecular targets, we noticed that SIAH2, an upstream regulator that is required for the maintenance of expression levels and activation of transcription activity of HIFα in cells under hypoxic conditions [18], had the highest enrichment score in both GSE39582 and TCGA-COAD cohorts. Coincidentally, hypoxia was identified as a strong transcriptional regulator of GPX3 expression through the presence of the HIF1α-binding site on the promoter region of GPX3 [19]. Accordingly, we tested GPX3 expression in CACO2 cells transfected with siRNA targeting SIAH2 (siSIAH2) and in SW480 cells transfected with SIAH2 overexpression plasmid (oxSIAH2). We found that GPX3 mRNA and protein levels were obviously decreased or increased by siSIAH2 and oxSIAH2 transfection (Fig. 7L). Moreover, siSIAH2 transfection significantly increased the apoptotic cell population induced by 5-FU or oxaliplatin in CACO2 cells (Fig. 7M), further suggesting that SIAH2 regulated GPX3-mediated chemoresistance. Additionally, we tested the chemosensitization effects of the compounds with the highest enrichment scores and reported that several inhibitors, such as the MLN8054, LY317615 (a PKCβ inhibitor), and SJ-72550, possessed the capability for chemosensitization to both 5-FU and oxaliplatin in CACO2 cells (Fig. 7M).

4. Discussion

Given the significant heterogeneity in survival benefits observed among patients with colon cancer who underwent ADJC, developing an effective classifier with clinical practicability may be helpful for accelerating the application of precision medicine. The present guidelines for ADJC in patients with colon cancer are based on relapse risks indicated by clinicopathological factors [20], MMR status [21], or immunoscore proposed by Galon et al [5]. However, recent evidence has suggested that intrinsic drug sensitivity related to molecular heterogeneity is another fundamental factor determining the clinical benefits of chemotherapy, independent of relapse risk. For example, although CMS4 [22], stem-like [23], or CCS3 subtype tumors [24] are associated with a high risk of relapse, no statistically significant benefits from oxaliplatin treatment could be observed in patients with these subtypes [25]. Similarly, during this study, we also identified a group of patients with colon cancer who were characterized by abundant stromal cell infiltration, had the shortest RFS and OS, lowest chemoresponse rate, and lacked the benefits of ADJC. These results suggest that formulating ADJC strategies according to the relapse risk alone is far from enough. Moreover, most molecular classifications are constructed based on global gene expression profiles derived from microarray or RNA sequencing platforms, which are impractical in actual clinical settings because of their exorbitant cost, long turnaround time, and reliance on bioinformatics expertise. In view of these facts, we designed and focused on the development of a quantitative indicator that closely associates with major cell types, especially stromal cells, within the TMECS1 group and can help provide a more individualized chemotherapeutic benefit assessment for patients with colon cancer. Our efforts allowed us to establish a novel 31-gene SIIS signature, capable of identifying colon cancer patients with both poor prognoses and intrinsic chemotherapeutic resistance, that was successful independently validated in multiple public datasets and in the cohort of patients

Fig. 6. CRISPR-CAS9 SAM screening of driver genes for chemoresistance in high-SIIS groups. (A) Experimental outline of screening and analysis. (B) Optical microscopic images of SW480 cells treated with oxaliplatin (0 μg/mL, 40 μg/mL) and 5FU (0 μg/mL, 2 μg/mL) for 96 h. (C) Volcano plots to compare differences in sgRNA abundance between oxaliplatin- and DMSO-treated cells (upper) and between 5-FU- and DMSO-treated cells (bottom); (D-E) Heatmap showing the counts of sgRNA-representing genes that mediating oxaliplatin (D) and 5-FU resistance (E). (F) Venn plot showing differentially enriched genes in four groups as following: oxaliplatin treated cells, 5-FU treated cells; patients with higher-SIIS value in TCGA-COAD and GSE39582 dataset; (G) Forest plots showing the correlation of gene expression of GPX3 and PRICKLE2 with overall survival in different subgroups (left) and the benefit of ADJC performance in patients with different expression levels of GPX3 and PRICKLE2 (right). (H-I) Box plot of the comparison of the GPX3 and PRICKLE2 expression between the fluorouracil response and nonresponse groups (H) and among various CMS subtypes in the GSE39582, TCGA-COAD, and SYSUCC cohorts (I). (J) Correlation matrix showing the correlations of GPX3 and PRICKLE2 expression with the infiltration level of TME cells (left) and with the activation level of biological process (right) in the GSE39582, TCGA-COAD, and SYSUCC cohorts. (K) The correlation chord chart showing the mutual correlation between SIIS value, the gene expression of GPX3 and PRICKLE2, and the GSVA score of non canonical WNT pathway, canonical WNT pathway, and WNT/PCP pathway in the GSE39582 cohort (left), in the TCGA-COAD cohort (middle), and in the SYSUCC cohort (right). (L-M) Box plot of the comparison of the GSVA score of WNT/PCP pathway between the fluorouracil response and nonresponse groups (L) and among various CMS subtypes (M) in the GSE39582, TCGA-COAD, and SYSUCC cohorts. *OXA, oxaliplatin; FC, fold change; Tcm, T central memory; Tem, T effector memory; Th, T helper; DC, dendritic cell; aDC, activated dendritic cell; iDC, immature dendritic cell; Tgd, T gamma delta; NK, natural killer; CI, confidence interval; ADJC, adjuvant chemotherapy; R response; NR, nonresponse; CMS, consensus molecular subtypes; EMT, epithelial-mesenchymal transition; OXPHOS, oxidative phosphorylation; UPR, unfolded protein response; FA, fatty acid; GSVA, gene set variation analysis.*

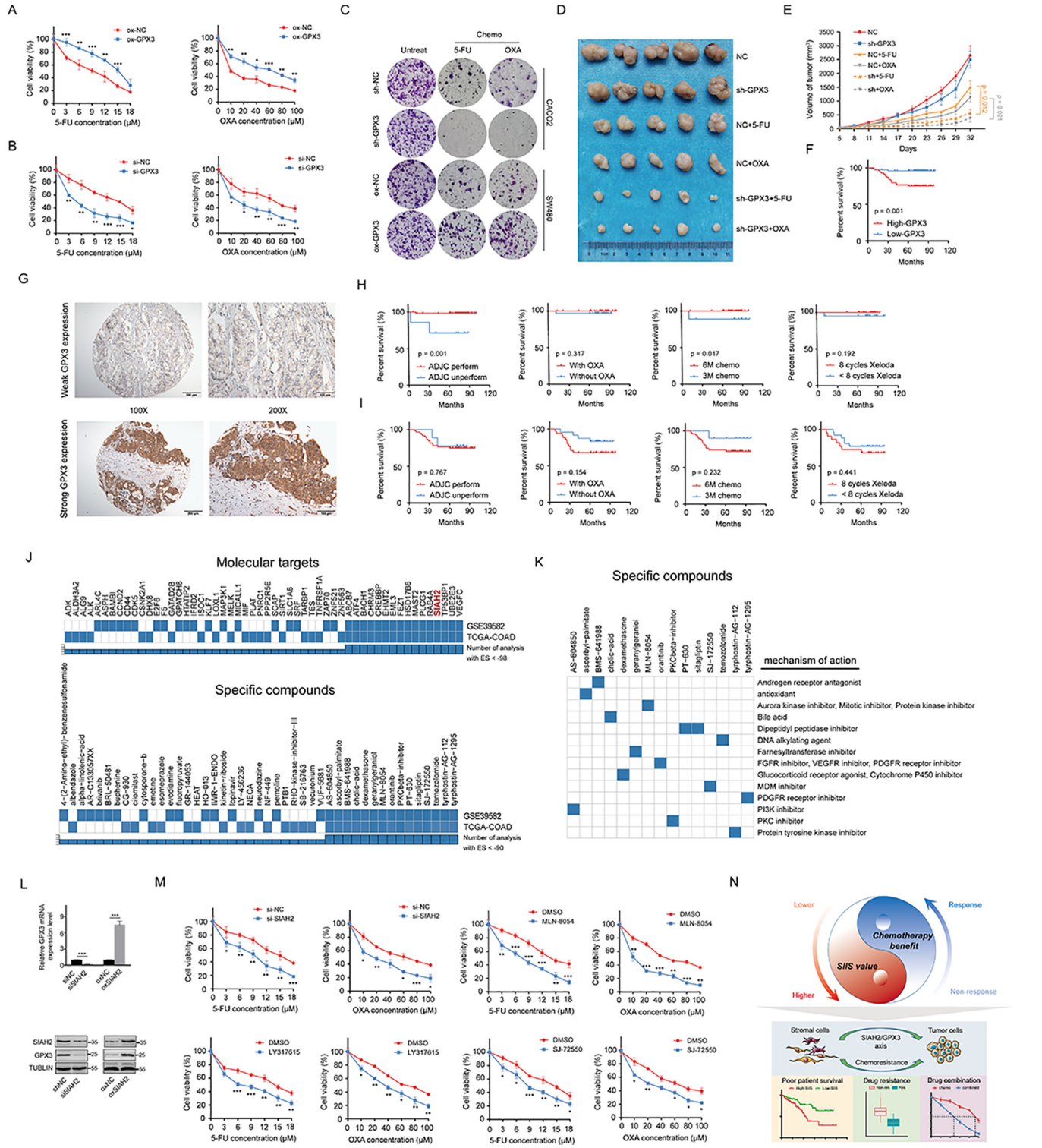


Fig. 7. Validation of the role of GPX3 in driving chemotherapy resistance in colon cancer. (A) Dose–response curves of SW480 cells transfected with empty vectors or GPX3 plasmid after treatment with 5-FU (left) or oxaliplatin (right) in different concentration for 24 h. (B) Dose–response curves of CACO2 cells transfected with empty vectors or GPX3 siRNA after treatment with 5-FU (left) or oxaliplatin (right) in different concentration for 24 h. (C) Colony formation ability of shGPX3, oxGPX3 and corresponding control cells with or without chemotherapy drug treatment. (D) Xenograft tumors of killed mice in different groups. (E) Growth curves of subcutaneous xenograft tumors in different groups. (F) The survival curve of for disease-free survival in the tissue microarray analysis. (G) Representative micrographs of GPX3 protein expression in colon cancer patients, as detected by immunohistochemistry. (H-I) The survival curve of ADJC performance, oxaliplatin combination, chemotherapy duration, and Xeloda cycles for disease-free survival in the low-GPX3 (H) and high-GPX3 (I) patients. (J) Heatmaps showing the significantly enriched molecular targets (upper, enrichment score <−96) and compounds (down, enrichment score <−90) based on the connectivity map analysis. (K) Heatmap showing the mechanisms of the action (rows) of each compound revealed by the connectivity map analysis. (L) The mRNA (upper) and protein levels (down) of GPX3 expression in cells transfected with siSIAH2 or oxSIAH2 cells as compared to the vehicle control. (M) Dose–response curves of CACO2 cells transfected with SIAH2 siRNA or in treatment groups based on different regimens. (N) Summary of the current study. *OXA, oxaliplatin; ADJC, adjuvant chemotherapy; SIIS, stromal cell infiltration intensity; ES, enrichment score.*

from SYSUCC. Intriguingly, through the GDSC analysis, we found that although the estimated IC50 values of the majority of chemotherapeutic drugs were significantly increased in the high-SIIS group, higher SIIS were consistently associated with improved efficiency of docetaxel. Docetaxel is mainly used to treat patients with platinum-resistant tumors in the clinic [26]. Our results suggest that docetaxel may be a potential treatment option for patients within high-SIIS group (also harboring platinum resistance). Collectively, the SIIS model has great potential to become an effective supplement to the current TNM staging system and could help to improve treatment decision for patients with colon cancer. More importantly, the SIIS structure consisted of only 31 genes, thereby offering a smaller panel of genes that can be easily translated into a simple, inexpensive, and quantitative PCR-based assay for routine clinical testing. Besides, the SIIS value is very easy to be calculated since it is just defined as difference between the average expression values of the two groups of genes.

During this research, we also evaluated the application value of the SIIS model at the pan-cancer level. The consistent positive correlations of SIIS value with level of stromal cell infiltration, stromal pathway activation, and predicted IC50 values of multiple chemotherapy drugs across a variety of tumors indicated that SIIS score is an universal tool for effective identification of patients whose tumors were characterized by high stromal components and chemoresistance. It is also encouraging to find that the SIIS model enable prediction of immunotherapy outcomes in solid tumors, and this may be partly due to the correlation between SIIS value and TGFβ pathway activation [15]. Therefore, trials to determine whether patients in the high-SIIS group could benefit from the combination of TGF-β inhibition and immune checkpoint blockade may have clinical significance.

Mounting evidence has shown that genomic alterations and DNA methylation status can reform TME and affect therapeutic efficacy. For example, the germline or somatic mutations in DNA damage repair genes usually confer improved long-term survival for patients treated with platinum-containing chemotherapy [27–29]. Similarly, the DNA methyltransferase inhibitors have also been shown to enhance chemosensitivity in multiple solid tumors [30]. In this study, we reported for the first time that the lack of multiple genetic mutation and the ADCY4 promoter hypomethylation were two potential mechanisms mediating chemoresistance in the background of stromal cell infiltration, and they could serve as novel biomarkers for predicting ADJC benefit in patients with colon cancer.

Another interesting aspect of our work was the identification of GPX3, one of the 31 genes that comprise the SIIS panel, as the driver gene contributing to chemotherapy drug resistance in the background of stromal cell infiltration using the CRISPR screen. GPX3 functions as a plasma antioxidant enzyme that protects cells against oxidative damage by inactivating reactive oxygen species [31]. Because inducing reactive oxygen species accumulation and oxidative stress is one of the main mechanisms by which chemotherapeutic agents exert cytotoxic effects [32], the antioxidant role has led to the exploration of GPX3 as a determinant of chemotherapeutic activities. For example, Saga et al. reported that high GPX3 expression was linked with cisplatin resistance in ovarian cancer cells [33]. Similarly, Pelosof et al. discovered that GPX3 promoter hypermethylation, which leads to reduced GPX3 mRNA and protein expression, increased oxaliplatin and cisplatin sensitivity in colon cancer cell lines [17]. Consistent with the findings of Pelosof et al., we also found that GPX3 is a core gene that mediates 5-FU and oxaliplatin resistance in colon cancer cells simultaneously. Moreover, using a tissue microarray analysis, we found that the indication for GPX3 in chemotherapy resistance was also maintained at the protein level, and that patients with high GPX3 protein expression experienced more adverse prognostic events when receiving ADJC, oxaliplatin combination, and longer durations of chemotherapy. These results are encouraging and highlight the potential clinical significance of using GPX3 as a simple biomarker for the identification of patients in the high-SIIS group and those who would suffer from more aggressive adjuvant treatment. However, expanding the sample size and prospective clinical studies are required to confirm these findings. Additionally, although our study revealed in vitro that some compounds, such as aurora kinase inhibitors, could increase the chemosensitivity of GPX3 highly-expressed colon cancer cells, whether these drugs could be used as effective combined treatment strategies is also warranted further investigations.

There were some limitations to the present study. First, although the comprehensive analysis in this study has yielded several important conclusions, it is still insufficient for clinical transformation of SIIS model because this study is based on retrospective datasets. Therefore, the results of our study should be further validated in prospective clinical trials. Moreover, an appropriate cutoff value of the SIIS model also needs to be determined for consequent clinical practice.

5. Conclusion

In conclusion, we developed and validated a novel individual scoring system based on quantifying the tumor-associated stromal components in the TME, that may serve as a practical and robust biomarker for the prediction of survival and providing more precise therapeutic options in colon cancer.

Availability of data and materials

The public data used in this study are available at:

GSE17536 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17536); GSE33113 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE33113); GSE37892 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37892); GSE38832 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE38832); GSE39582 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39582);

GSE188711 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE188711);

GSE35640 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE35640);

GSE173839 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE173839);

GSE176307 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE176307);

TCGA datasets (https://xenabrowser.net/datapages/);

IMvigor210 R package “IMvigor210”.

The SUSYCC colon cancer dataset generated and analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

Funding

This work was supported by the National Natural Science Foundation of China (No. 82102731 to RZ, No. 82073303 to WL), Natural Science Foundation of Guangdong Province of China (2020A1515110686 to RZ, 2022A1515012418 to RZ).

CRediT authorship contribution statement

Rui Zhou: Data curation, Formal analysis, Funding acquisition, Methodology, Visualization, Roles/Writing - original draft. **Zhaowei Wen:** Data curation, Formal analysis,Investigation, Visualiza-

tion, Roles/Writing - original draft. **Yifu Liao:** Formal analysis, Investigation, Visualization, Roles/Writing - original draft. **Jingjing Wu:** Validation, Visualization, Writing - review & editing. **Shaoyan Xi:** Validation, Resources. **Dongqiang Zeng:** Validation. **Huiying Sun:** Validation. **Jianhua Wu:** Validation. **Min Shi:** Writing - review & editing. **Jianping Bin:** Writing - review & editing. **Yulin Liao:** Writing - review & editing. **Wangjun Liao:** Conceptualization, Funding acquisition, Project administration, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.04.037.

References

[1] Roerink SF, Sasaki N, Lee-Six H, Young MD, Alexandrov LB, et al. Intra-tumour diversification in colorectal cancer at the single-cell level. Nature 2018;556:457–62.
[2] Wang W, Kandimalla R, Huang H, Zhu L, Li Y, et al. Molecular subtyping of colorectal cancer: recent progress, new challenges and emerging opportunities. Semin Cancer Biol 2019;55:37–52.
[3] Hui L, Chen Y. Tumor microenvironment: sanctuary of the devil. Cancer Lett 2015;368:7–13.
[4] Pages F, Mlecnik B, Marliot F, Bindea G, Ou FS, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. Lancet 2018;391:2128–39.
[5] Ascierto PA, Marincola FM, Fox BA, Galon J. No time to die: the consensus immunoscore for predicting survival and response to chemotherapy of locally advanced colon cancer patients in a multicenter international study. Oncoimmunology 2020;9:1826132.
[6] Zhou R, Zeng D, Zhang J, Sun H, Wu J, et al. A robust panel based on tumour microenvironment genes for prognostic prediction and tailoring therapies in stage I-III colon cancer. EBioMedicine 2019;42:420–30.
[7] Zhou R, Sun H, Zheng S, Zhang J, Zeng D, et al. A stroma-related lncRNA panel for predicting recurrence and adjuvant chemotherapy benefit in patients with early-stage colon cancer. J Cell Mol Med 2020;24:3229–41.
[8] Widodo SS, Hutchinson RA, Fang Y, Mangiola S, Neeson PJ, et al. Toward precision immunotherapy using multiplex immunohistochemistry and in silico methods to define the tumor immune microenvironment. Cancer Immunol Immunother 2021;70:1811–20.
[9] Zeng D, Zhou R, Yu Y, Luo Y, Zhang J, et al. Gene expression profiles for a prognostic immunoscore in gastric cancer. Br J Surg 2018;105:1338–48.
[10] Zhou R, Zhang J, Zeng D, Sun H, Rong X, et al. Immune cell infiltration as a biomarker for the diagnosis and prognosis of stage I-III colon cancer. Cancer Immunol Immunother 2019;68:433–42.
[11] Fang Y, Wang Y, Zeng D, Zhi S, Shu T, et al. Comprehensive analyses reveal TKI-induced remodeling of the tumor immune microenvironment in EGFR/ALK-positive non-small-cell lung cancer. Oncoimmunology 2021;10:1951019.
[12] Sun H, Zhou R, Zheng Y, Wen Z, Zhang D, et al. CRIP1 cooperates with BRCA2 to drive the nuclear enrichment of RAD51 and to facilitate homologous repair upon DNA damage induced by chemotherapy. Oncogene 2021;40:5342–55.
[13] Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. Nature 2017;541:321–30.
[14] Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS et al. The immune landscape of cancer, Immunity 2018;48:812-830 e814.
[15] Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, et al. TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature 2018;554:544–8.
[16] Daulat AM, Borg JP. Wnt/planar cell polarity signaling: new opportunities for cancer treatment. Trends Cancer 2017;3:113–25.
[17] Pelosof L, Yerram S, Armstrong T, Chu N, Danilova L, et al. GPX3 promoter methylation predicts platinum sensitivity in colorectal cancer. Epigenetics 2017;12:540–50.
[18] Qi J, Nakayama K, Gaitonde S, Goydos JS, Krajewski S, et al. The ubiquitin ligase Siah2 regulates tumorigenesis and metastasis by HIF-dependent and -independent pathways. Proc Natl Acad Sci U S A 2008;105:16713–8.
[19] Bierl C, Voetsch B, Jin RC, Handy DE, Loscalzo J. Determinants of human plasma glutathione peroxidase (GPx-3) expression. J Biol Chem 2004;279:26839–45.
[20] Argiles G, Tabernero J, Labianca R, Hochhauser D, Salazar R, et al. Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2020;31:1291–305.
[21] Ribic CM, Sargent DJ, Moore MJ, Thibodeau SN, French AJ, et al. Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. N Engl J Med 2003;349:247–57.
[22] Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, et al. The consensus molecular subtypes of colorectal cancer. Nat Med 2015;21:1350–6.
[23] Sadanandam A, Lyssiotis CA, Homicsko K, Collisson EA, Gibb WJ, et al. A colorectal cancer classification system that associates cellular phenotype and responses to therapy. Nat Med 2013;19:619–25.
[24] De Sousa EMF, Wang X, Jansen M, Fessler E, Trinh A, et al. Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. Nat Med 2013;19:614–8.
[25] Song N, Pogue-Geile KL, Gavin PG, Yothers G, Kim SR, et al. Clinical outcome from oxaliplatin treatment in stage II/III colon cancer according to intrinsic subtypes: secondary analysis of NSABP C-07/NRG oncology randomized clinical trial. JAMA Oncol 2016;2:1162–9.
[26] Fu S, Hennessy BT, Ng CS, Ju Z, Coombes KR, et al. Perifosine plus docetaxel in patients with platinum and taxane resistant or refractory high-grade epithelial ovarian cancer. Gynecol Oncol 2012;126:47–53.
[27] Chae H, Kim D, Yoo C, Kim KP, Jeong JH, et al. Therapeutic relevance of targeted sequencing in management of patients with advanced biliary tract cancer: DNA damage repair gene mutations as a predictive biomarker. Eur J Cancer 2019;120:31–9.
[28] Miron B, Hoffman-Censits JH, Anari F, O'Neill J, Geynisman DM, et al. Defects in DNA repair genes confer improved long-term survival after cisplatin-based neoadjuvant chemotherapy for muscle-invasive bladder cancer. Eur Urol Oncol 2020;3:544–7.
[29] Palacio S, McMurry HS, Ali R, Donenberg T, Silva-Smith R, et al. DNA damage repair deficiency as a predictive biomarker for FOLFIRINOX efficacy in metastatic pancreatic cancer. J Gastrointest Oncol 2019;10:1133–9.
[30] Hu C, Liu X, Zeng Y, Liu J, Wu F. DNA methyltransferase inhibitors combination therapy for the treatment of solid tumor: mechanism and clinical application. Clin Epigenetics 2021;13:166.
[31] Nirgude S, Choudhary B. Insights into the role of GPX3, a highly efficient plasma antioxidant, in cancer. Biochem Pharmacol 2021;184:114365.
[32] Dharmaraja AT. Role of reactive oxygen species (ROS) in therapeutics and drug resistance in cancer and bacteria. J Med Chem 2017;60:3221–40.
[33] Saga Y, Ohwada M, Suzuki M, Konno R, Kigawa J, et al. Glutathione peroxidase 3 is a candidate mechanism of anticancer drug resistance of ovarian clear cell adenocarcinoma. Oncol Rep 2008;20:1299–303.