**Discover** Oncology

Research

# Identification of DNA damage and repair gene-related markers in pancreatic ductal adenocarcinoma by single-cell and bulk RNA sequencing

**Chaoyi Zhang[1,2,3,4] · Rong Tang[1,2,3,4] · Jianhui Yang[1,2,3,4] · Yueyue Chen[1,2,3,4] · Yangyi Li[1,2,3,4] · Cong Zhou[1,2,3,4] · Wei Wang[1,2,3,4] · Xian-Jun Yu[1,2,3,4] · Jin Xu[1,2,3,4]**

Received: 2 December 2024 / Accepted: 2 April 2025
Published online: 08 April 2025
© The Author(s) 2025 OPEN

## Abstract

**Background** The DNA damage response (DDR) has a major impact on the development and progression of pancreatic ductal adenocarcinoma (PDAC). Investigating biomarkers linked to the DDR may facilitate prognostic assessment and prediction of immunological characteristics for patients with PDAC.

**Methods** The single-cell RNA sequencing (scRNA-seq) dataset GSE212966 was obtained from the GEO database, whereas the bulk RNA-seq data were sourced from The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) databases. Least absolute shrinkage and selection operator (LASSO) and univariate Cox regression analyses were used to select genes to construct a prognostic risk model. Finally, the correlations of the model score with drug sensitivity, immunological checkpoints, and immune infiltration were assessed.

**Results** We used 16 DDR marker genes to construct a predictive model. Furthermore, we established that the model had strong performance in both the training and validation cohorts. For PDAC, the model risk score served as an independent predictor of prognosis. There were notable differences in the proportions of the immune cells in the tumor microenvironment and drug sensitivity between the high and low risk score groups. The study confirmed that the risk score model is useful for predicting the immunotherapy response. Our experiments verified that knockdown of *LY6D* inhibits cell proliferation, promotes apoptosis and DNA damage.

**Conclusion** Our creative integration of bulk RNA sequencing and scRNA-seq data allowed us to construct a DDR-related prognostic model. Our model can be used to predict the immunological features, treatment response and prognosis of PDAC with a relatively high degree of accuracy.

**Keywords** Pancreatic ductal adenocarcinoma · DNA damage- and repair-related genes · Survival · Chemotherapeutic response

Chaoyi Zhang and Rong Tang have contributed equally to this work.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s12672-025-02293-w.

✉ Xian-Jun Yu, yuxianjun@fudanpci.org; ✉ Jin Xu, xujin@fudanpci.org | [1]Department of Pancreatic Surgery, Fudan University Shanghai Cancer Center, No. 270 Dong'An Road, Shanghai 200032, People's Republic of China. [2]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, People's Republic of China. [3]Shanghai Pancreatic Cancer Institute, No. 399 Lingling Road, Shanghai 200032, People's Republic of China. [4]Pancreatic Cancer Institute, Fudan University, Shanghai 200032, People's Republic of China.

## 1 Introduction

With a five-year survival rate of 10% across all stages, pancreatic ductal adenocarcinoma (PDAC) is one of the deadliest cancers worldwide [1, 2]. The high mortality rate and poor prognosis of PDAC stem from the absence of clinical signs in the disease's early stages, the high recurrence rate, and substantial resistance to treatment modalities, including chemotherapy and radiation. Many clinical studies on PDAC treatments have been unsuccessful, and thus alternative treatment options are limited. The molecular and functional heterogeneity of PDAC may partly explain the lack of clinical therapeutic advancements.

Genomic instability is defined as a high frequency of harmful alterations in the genomic structure that result from a compromised DNA repair response, which is a hallmark of cancers. The emergence of several cancers has been linked to the silencing of or to harmful mutation in tumor suppressor genes, including checkpoint and DNA repair genes such as *P53*, *P16*, and ataxia telangiectasia mutated (*ATM*). The DNA damage response (DDR), which encompasses signaling and repair, is necessary for genome stability. Multiple DDR mechanisms, including direct reversal, base excision repair, nucleotide excision repair, mismatch repair, single-strand break repair, and double-strand break repair, can occur [3–5]. DDR pathways play a role in tumor development and progression and the tumor response to treatment [6]. Certain studies have shown a correlation between DDR deficiency and the stimulation of anticancer immunity via the activation of the classical cGAS-STING pathway. In addition to causing alterations in the tumor immune milieu, DDR anomalies also impact the therapeutic efficacy of immunological checkpoint inhibitors (ICIs) [7]. Moreover, the main factor limiting the response to single drugs in cancer therapy is tumor acquisition of resistance. Drug resistance may result from the interplay of DDR mechanisms.

Personalized medicine for PDAC is a relatively new research area, and there are currently no predictive biomarkers available. Several studies have been performed to investigate numerous possible biomarkers for predicting the response to different treatments (chemotherapy, immunotherapy, or targeted therapy) and the prognosis of PDAC. The most recognized biomarkers are *BRCA*1 and *BRCA2* mutations, which are linked to advantages from platinum-based therapies and PARP inhibitors (PARPis). The only approved targeted maintenance treatment for pancreatic cancer is olaparib, a DNA-damaging drug designed for hormone receptor (HR)-deficient carriers of germline *BRCA (gBRCA)* mutations [8]. With the exception of the < 1% of patients with microsatellite instability-high (MSI-H) tumors, PDAC patients are mostly resistant to FDA-approved immunotherapies that have improved the prognosis of other solid tumors. The widespread resistance of PDAC to immune checkpoint blockade (ICB) is almost unparalleled among human cancers. One important feature of PDAC, which is often hereditary in nature, is impaired DNA damage repair. Sixty-three genetic changes were found in each PDAC case in a large-scale genomic study, and these genes are enriched in DNA damage control pathways [9]. Familial breast cancer is often linked to germline DDR gene mutations (*ATM, BRCA1, BRCA2, MLH1, MSH2, PALB2, PMS2, and STK11*) as well as mutations in traditional cancer risk-related genes such as *CDKN2 A* or *TP53* [10]. *ATM* is the most often mutated DDR gene in somatically altered sporadic PDAC. DDR deficiency may increase the susceptibility of PDAC patients to novel treatment interventions that increase the DNA damage burden above a threshold that can be tolerated.

Consequently, investigating the characteristic genes that impact the DDR and creating a predictive signature based on characteristic genes associated with the DDR would aid in the development of tailored therapy and monitoring regimens for patients with PDAC. Through the integration of bulk and single-cell RNA sequencing data, we performed multiomics studies to identify the pathways involved in malignant development and to predict cancer prognosis effectively. To identify DDR response-related DEGs in PDAC, we assessed single-cell and bulk RNA sequencing data from patients with PDAC, with a focus on genes associated with DNA damage repair. A risk score (RS) model using these genes was built to predict the prognosis of PDAC. Our findings suggest that the RS model shows stability and potential in predicting PDAC prognosis, providing a foundation for further exploration of its clinical utility.

## 2 Materials and methods

### 2.1 Data download

From the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo), we obtained the scRNA-seq dataset GSE212966 [11]. We also retrieved the expression profiles and patient survival data from the GSE62452 and GSE57495 datasets from GEO. The data were processed in the following ways: (1) samples lacking clinical follow-up information were eliminated; (2) samples with a unified unit of survival time of days, < 0 days, and unknown survival time were eliminated; (3)

probe identifiers were transformed into gene symbols; (4) samples corresponding to multiple genes were eliminated; and (5) expression values corresponding to multiple gene symbols were summed to determine the median expression value. A total of 10,526 DDR-related genes were identified via the term"DNA damage repair"in the GeneCards database (http://www.genecards.org). Among them, 9,723 genes associated with the DDR that had a correlation coefficient greater than 1.0 were selected.

### 2.2 Single-cell sequencing data analysis

Initially, we used the R package"Seurat"(5.0.2) [12] to generate a Seurat object for the scRNA-seq data. We added quality control criteria on the basis of the prior threshold. Cells meeting the following criteria were included: range of counts, 1800–20,000; number of features, 300–8000; proportion of mitochondrial genes, < 20%; percentage of ribosomal genes < 20%; and percentage of red blood cell reads < 5%. Next, we eliminated the cell cycle effects, normalized the Seurat objects, reduced their dimensionality (1:15), clustered them at a resolution of 0.5, and annotated them.

Seurat groups cells according to their principal component analysis (PCA) scores, with each PDAC cell basically serving as a"meta feature"that incorporates data from a correlated feature set to mitigate the significant technical noise in any single feature for scRNA-seq data. Data on the top 2000 hypervariable genes were used for the PCA. The"harmony"R package was used for batch correction to mitigate the batch effect on future analyses since single-cell sequencing data were obtained from many datasets. The FindClusters approach was utilized to find tumor cell subgroups with a resolution of 0.5 after refining and consulting the clustering findings in the initial contribution. UMAP was used to further decrease the dimension of the data that were acquired following PCA. The DimPlot function was used to depict the various cell types once they were separated into low-dimensional areas. The expression, landscape distribution, and distinctive genes of tumor cells were visualized via the FeaturePlot, DotPlot, DoHeatmap, and VlnPlot tools. In each cell subgroup, marker genes associated with particular DDR reactions were identified via the DotPlot tool. The"FindAllMarkers"tool was utilized to locate cell cluster marker genes. Finally, we carried out manual annotation. The reference cell markers were obtained from a published study [13] and the CellMarker database (https://xteam.xbio.top/CellMarker/).

### 2.3 Identification of active subgroups

Marker gene intersections were chosen for each cell subgroup on the basis of DDR-related parameters and substantial population specificity (adj.p < 0.05& |avg_log2 FoldChange|> 1.5 & pct.1 > 0.5 & pct.2 < 0.5). The"AUCell"R program was utilized to determine each cell's activity score based on the expression levels of the intersecting genes. The threshold for this analysis was established via the AUCell_explore thresholds function. The area under the ROC curve (AUC) score of every cell was used to color-code the UMAP plot of cell clusters, allowing researchers to determine which fraction of cells in a given subgroup was active with respect to particular genes connected to the DDR. With the default settings, the FindAllMarkers tool was used to identify the DEGs between the active and inactive subgroups. Marker genes with significant specificity (avg_log2 FC > 1) were used to identify the functions of the related cell population.

### 2.4 Enrichment analysis

To rank genes according to the degree of differential expression between two samples, gene set enrichment analysis (GSEA) was performed with predetermined gene sets. The most highly upregulated and downregulated genes in predetermined gene sets were subsequently used for analysis. These specific gene sets are often obtained via functional annotation or derived from earlier studies. GSEA was used to identify the distinctive genes enriched by active cell subgroups via the"ClusterProfiler"[14] and"hallmark"pathway gene set R packages. Significant enrichment was defined as gene sets with $p < 0.01$. Additionally, we used the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) analyses to conduct pathway enrichment studies.

### 2.5 Identification of differentially expressed genes (DEGs)

To improve the comparability of genomes between cancer and paraneoplastic disorders, we merged normal samples from the Genotype-Tissue Expression Project (GTEx) database with samples from the TCGA-PDAC cohort. To identify DEGs between the tumor and normal groups (log2 FoldChange > 1.5 & P < 0.05), the"limma"R program was utilized.

**Fig. 1** Identification of cell subgroups and analysis of marker gene expression via single-cell RNA sequencing (scRNA-seq) data. **A** The distribution of pancreatic ductal adenocarcinoma (PDAC) cell subgroups is displayed on the UMAP plot. **B** PDAC cell subgroup annotation findings are displayed on the UMAP plot. **C** The cell types distributed in PDAC patients are displayed in a cumulative histogram. **D** Gene expression associated with distinct cell subgroups is depicted in a violin diagram. **E** The expression of genes linked to DNA damage and repair responses in each cell subgroup is depicted in the bubble diagram

### 2.6 Development of a prognostic model and survival analysis

We used the DNA damage response-related marker genes (DDRMGs) (log2 FoldChange > 0.25) to intersect the DEGs between the tumor and normal samples. The predictive model was built using clinical data for the TCGA-PDAC cohort, including age, TNM stage, and gene expression (in transcripts per million mapped reads (TPM) format). Univariate Cox regression analysis was then used to identify genes linked to the prognosis of PDAC. The gene set was further reduced via least absolute shrinkage and selection operator (LASSO) regression analysis. In the end, 16 DDRMGs were chosen to construct the prognostic model. This following formula was used to obtain the risk score (RS). Gene expression × coefficient = RS. Patients were divided into high- and low-risk groups on the basis of the median RS. We next used the GSE62452 and GSE57495 cohorts to validate the model. The overall survival (OS) time of patients with PDAC was predicted via Kaplan—Meier (K—M) survival analysis. To assess the specificity and sensitivity of the model, a receiver operating characteristic (ROC) curve was generated.

### 2.7 Calculating the proportions of invading immune cells and the immune score for DEG identification

The"CIBERSORT"R program was used to determine the degree of immune infiltration in PDAC. Initially, we computed the proportions of 21 classical immune cell types and compared them between the high- and low-risk groups. Additionally, correlation analysis was performed to clarify the association between the proportions of infiltrating immune cells and the model RS. Immune checkpoints can be used to ascertain a patient's responsiveness to immunotherapy. We thus assessed the differences in immune checkpoint gene expression between the two risk groups.

### 2.8 Drug sensitivity and immunotherapy efficacy

Gene expression profiles and half-maximum inhibitory concentration (IC50) values of many anticancer medications were obtained via the R program"OncoPredict"[15]. To predict the medication sensitivity of the high- and low-risk categories, we employed the"calcPhenotype"tool. We also investigated the relationship between the IC50 values and the activity of our model genes. Immunotherapy efficacy in the two risk groups was predicted via the"IMvigor210 CoreBiologies"R package.

### 2.9 Pathway enrichment analysis

Several pathways that were enriched in two risk categories were evaluated via gene set variation analysis (GSVA). The"HALLMARK"and"KEGG"reference gene sets were obtained from the Molecular Signature Database (MSigDB; https://www.gseamsigdb.org/gsea/msigdb/).

### 2.10 Immunohistochemical staining

To determine the expression of LY6D, 25 PDAC tissue samples from Fudan University Shanghai Cancer Center were subjected to immunohistochemistry. After two hours of drying at 65 degrees Celsius, the tissue slices were deparaffinized in xylene and then rehydrated in ethanol. The sections were blocked with 10% goat serum for 10 min and washed with PBS following antigen extraction in 0.1 mol/L citrate buffer solution. After a 4 °C overnight incubation period, the sections were incubated with a primary antibody (82,094–1-RR, Proteintech, China;1:1000), and then, the sections were incubated with biotin-labeled goat anti-rabbit IgG for half an hour. The sections were treated in a consecutive manner at room temperature, first with diaminobenzidine for 1 min and then with hematoxylin for 2 min. The proportion of stained cells and the intensity of the cell staining on each slide were independently graded for the assessment of LY6D expression. Two separate pathologists evaluated the immunohistochemistry data in a blinded fashion. A score of 0–4 was assigned

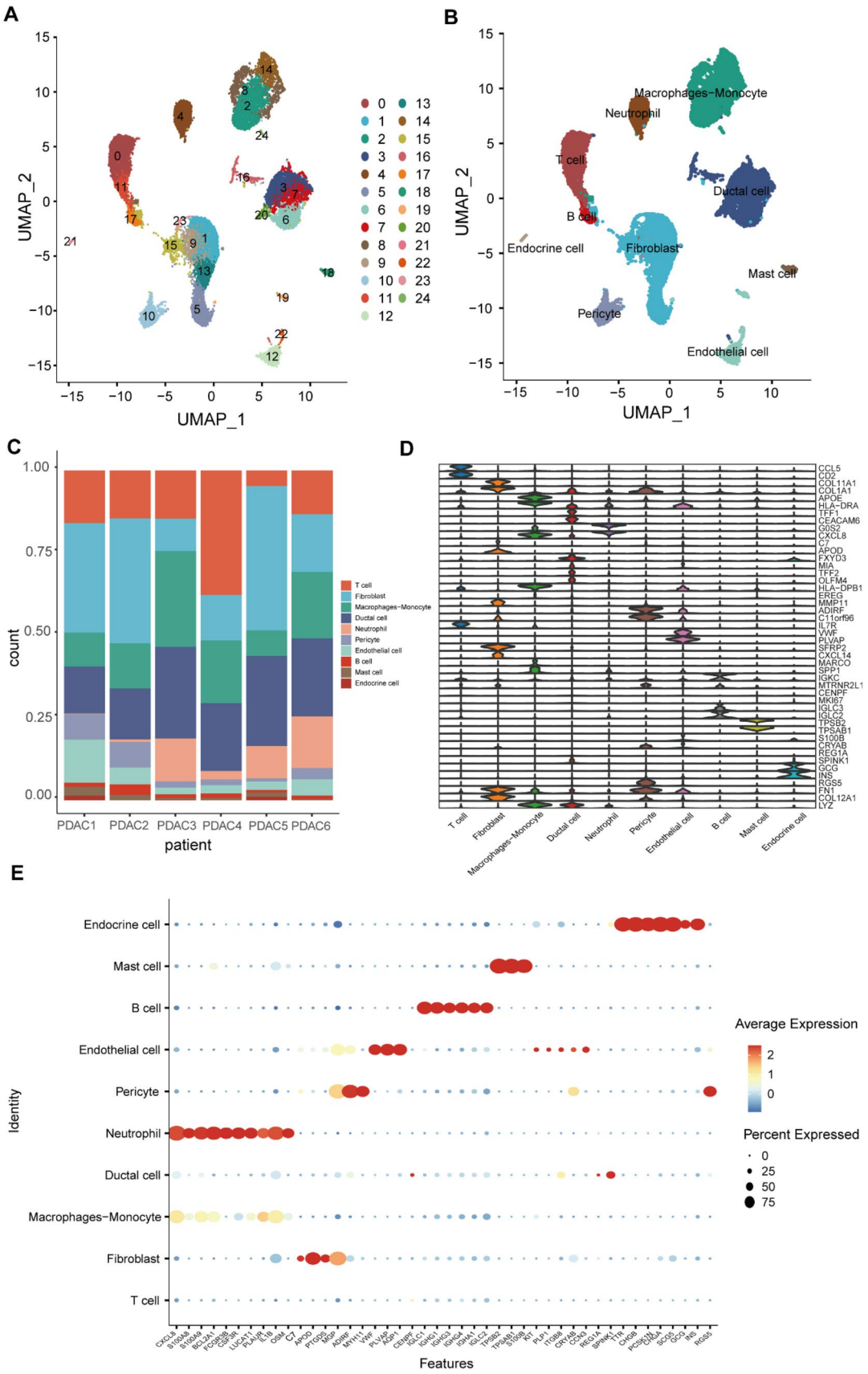

A
UMAP_2
UMAP_1
0 13
1 14
2 15
3 16
4 17
5 18
6 19
7 20
8 21
9 22
10 23
11 24
12
B
Macrophages−Monocyte
Neutrophil
T cell
Ductal cell
B cell
Endocrine cell
Fibroblast
Mast cell
Pericyte
Endothelial cell
C
count
patient
PDAC1 PDAC2 PDAC3 PDAC4 PDAC5 PDAC6
T cell
Fibroblast
Macrophages−Monocyte
Ductal cell
Neutrophil
Pericyte
Endothelial cell
B cell
Mast cell
Endocrine cell
D
E
Identity
Features
Average Expression
Percent Expressed

**Fig. 2** Determination of subgroups of activated cells. **A** The UMAP plots show the expression of the top ten most active DDRMGs in different cell subpopulations. **B** The cutoff value for the AUC score of the DNA damage and repair marker genes was 0.14. **C** The cell activity score is displayed on the color-coded UMAP plot. Brighter colors are associated with higher activity levels. **D** The histogram illustrates the cumulative distribution of active and inactive cell populations. **E** GSEA findings for DDR-activated genes. There were four repressed and four activated signaling pathways among the eight differentially enriched pathways

for the percentage of positive cells: 0, < 5%; 1, 5–25%; 2, 26–50%; 3, 51–75%; and 4, > 75%. The following system was used to score stain intensity: 0, no staining; 1, weakly positive; 2, moderately positive; and 3, strongly positive. The percentage of positive cells and the staining intensity were multiplied to determine the final staining score.

### 2.11 Cell culture

The SW1990 and PATU- 8988 T cell lines were purchased from the National Infrastructure of Cell Line Resources. All cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 100 units/mL penicillin, and 100 µg/mL streptomycin.

### 2.12 Cell transfection

Small interfering RNA (siRNA) targeting LY6D were obtained from Genomeditech (Shanghai, China). The target sequences of the siRNAs were as follows: siLY6D#1:5′-GCAAUGAGAAGCUGCACAA- 3′; siLY6D#2:5′-GUUUAUUUUUGUACUCAAA – 3′. Transfection of SW1990 and PATU- 8988 T cells was carried out using Lipofectamine 3000 (Invitrogen, Waltham, MA, USA).

### 2.13 RT-qPCR

Total RNA was extracted using the TRIzol method. PrimeScript RT Master Mix (RR036 A, Takara) was used to generate cDNA from 500 ng RNA. The ChamQ Universal SYBR qPCR Master Mix (RR820 A, Takara) was used to generate the reactive solution for RT-qPCR, and the Quantastudio 7 device captured the data. Gene expression was computed using the $2 - \Delta\Delta CT$ method, and the results were normalised to the mRNA expression of ACTB in each sample. The specific primers used were: human LY6D forward 5′-GGAATCTGGTGAAGAAGGACTGTG- 3′ and reverse 5′-GGTCCTCCTGGCAGCACTG- 3′; human GAPDH forward 5′-CTGGGCTACACTGAGCACC- 3′ and reverse 5′′-AAGTGGTCGTTGAGGGCAATG- 3′.

### 2.14 Immunofluorescence staining of γ-H2 AX and 53BP1

The assessment of γ-H2 AX immunofluorescence labeling was performed via a DNA damage test kit from Beyotime(C2038S) in Shanghai, China. After three cold PBS washes, the transfected SW1990 cells and PATU- 8988 T cells were blocked with QuickBlock™ Blocking Buffer for Immunol Staining for one hour at 4 °C. A mouse monoclonal antibody (1:200) against γ-H2 AX and a rabbit monoclonal antibody (1:200) against TP53BP1-Specific antibody (83,809–1-RR, Proteintech) were then added to the cells, which were then incubated for 1 h at room temperature. After being rinsed with Immunol Staining Wash Buffer containing 0.5% Triton X- 100, SW1990 cells and PATU- 8988 T cells were treated for 80 min at 4 °C with Coralite®488-Conjugated Goat Anti-Rabbit IgG(H + L) (SA00013 - 2, Proteintech;1:200) and Cy3-conjugated anti-mouse secondary antibody (C2038S,Beyotime; 1:200). Images were obtained using a confocal microscope after the nuclei were labeled with 4,6-diamidino- 2-phenylindole (DAPI).

### 2.15 5-Ethynyl- 2′-deoxyuridine (EdU) cell proliferation assay

An EdU cell proliferation kit (C0078S, Beyotime) was used to measure the capacity for cell proliferation. The process was completed according to the kit's instructions. EdU was used at a working concentration of 10 µM.

### 2.16 Cell cycle assays

A cell cycle and apoptosis analysis kit (C1052, Beyotime) was used to measure cell cycle changes. The process was completed according to the kit's instructions.

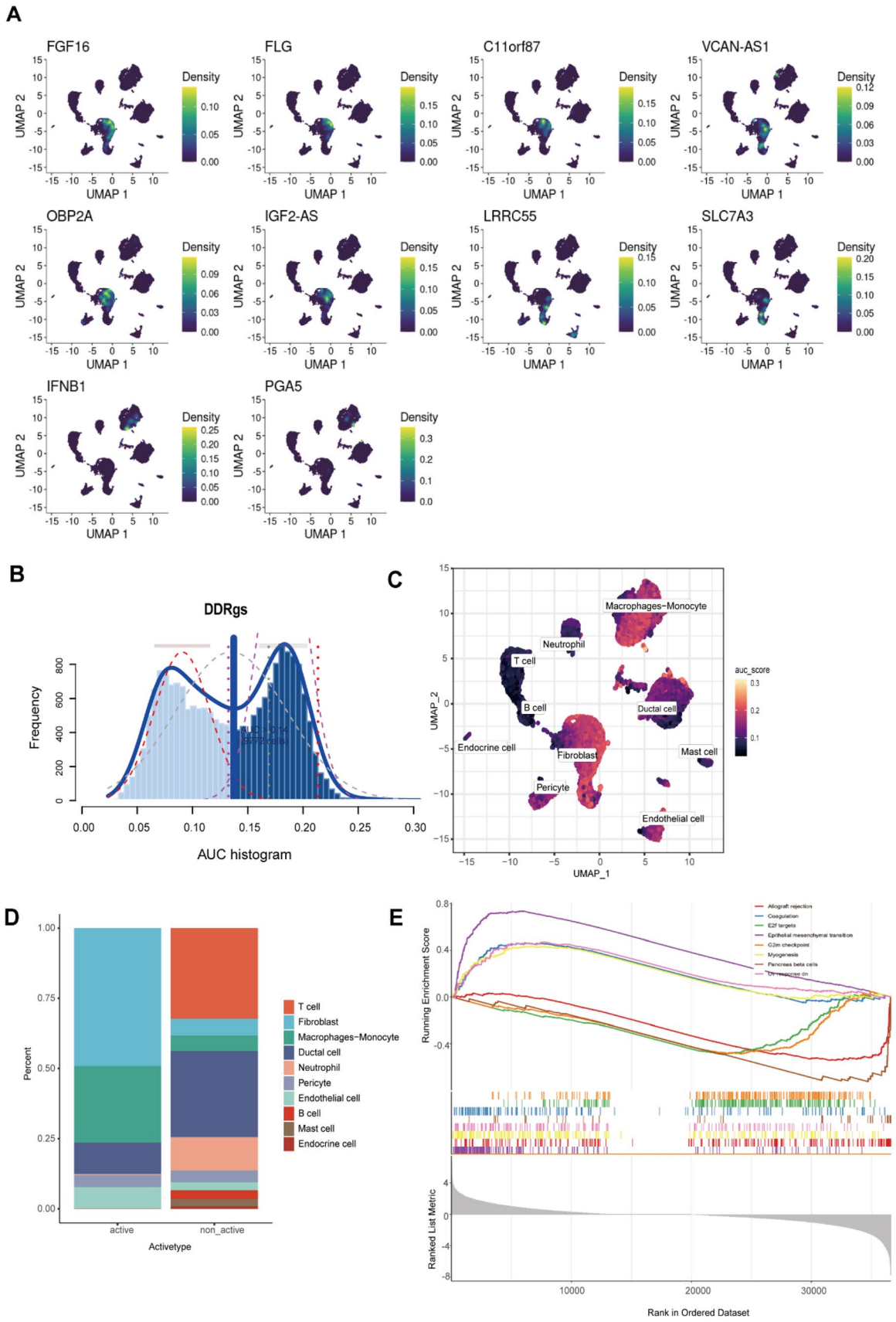

A
FGF16
FLG
C11orf87
VCAN-AS1
OBP2A
IGF2-AS
LRRC55
SLC7A3
IFNB1
PGA5
UMAP 1
UMAP 2
Density
B
DDRgs
Frequency
AUC histogram
C
Macrophages−Monocyte
Neutrophil
T cell
B cell
Ductal cell
Endocrine cell
Fibroblast
Mast cell
Pericyte
Endothelial cell
auc_score
UMAP_1
UMAP_2
D
Percent
active
non_active
Activetype
T cell
Fibroblast
Macrophages−Monocyte
Ductal cell
Neutrophil
Pericyte
Endothelial cell
B cell
Mast cell
Endocrine cell
E
Running Enrichment Score
Ranked List Metric
Rank in Ordered Dataset

**Fig. 3** Cox and LASSO regression analyses were conducted on the TCGA-PDAC dataset. **A** The logarithm of the independent variable lambda ▶ is represented by the abscissa of the LASSO regression independent variable, while its coefficient is represented by the ordinate. **B** Confidence interval for LASSO regression for each lambda. **C** Significant prognostic gene LASSO regression coefficient. **D**–**S** Kaplan–Meier (K–M) survival curves for patients grouped according to the expression of the prognostic genes identified via Cox regression analysis. Red signifies the high-expression group, whereas blue represents the low-expression group, and significant differences are indicated

### 2.17 Apoptosis detection

An caspase- 3 activity and apoptosis detection kit for live cell (C1077S, Beyotime) was used to measure the caspase- 3 activity and apoptosis in cultured cells. The process was completed according to the kit's instructions. The concentration of GreenNuc™ Caspase- 3 Substrate is 1 mM.

### 2.18 Statistical analyses

The differences between the two groups of samples were analyzed via the Wilcoxon signed-rank test, whereas the differences among several groups of samples were assessed via the Kruskal—Wallis test. NS indicates $p > 0.05$, $^{*}$ indicates $p \leq 0.05$, $^{**}$ indicates $p \leq 0.01$, $^{***}$ indicates $p \leq 0.001$, and $^{****}$ indicates $p \leq 0.0001$.

## 3 Results

### 3.1 Identification of cellular subgroups exhibiting an active DDR

#### 3.1.1 Tumor single-cell landscape

The single-cell sequencing dataset GSE212966 obtained from the Gene Expression Omnibus (GEO) database comprises tumor samples from six individuals diagnosed with PDAC. On the basis of the findings of"VlnPlot"analysis, thresholds were set to filter the cells (Supplementary Fig. S1 A). After filtering and normalization, a total of 30,483 cells remained for use in the subsequent studies. The results of PCA revealed that cells can be successfully segregated from one another, and the top 14 principal components were identified as important principal components for further research via the"ElbowPlot"tool (Supplementary Fig. S1B).

With UMAP dimensionality reduction, the expression patterns of individual cells were visualized on the basis of clustering outcomes based on the single-cell sequencing dataset and the annotation data acquired with SingleR. The cells were split into 25 subgroups, labeled 0 to 24, according to the data (Fig. 1A). We manually annotated the data and identified 10 primary cell types. Figure 1B displays the distribution of various PDAC cell subgroups. The cell percentage bar graph revealed that the tumor microenvironment is an intricately interwoven system composed of immune cells, ductal cells, endothelial cells, and tumor-associated fibroblasts, among other cell types (Fig. 1C). The two most prominent marker genes for every cell cluster are displayed in the violin diagram (Fig. 1D). Additionally, there was considerable heterogeneity in the cell components of the various PDAC samples. By intersecting DDR-related genes with cell cluster marker genes, we were able to obtain 478 genes to serve as the input gene set for the following stage. The bubble diagram shows how they manifested in the designated cell subgroups (Fig. 1E).

#### 3.1.2 Identification of cell subgroups with an active DDR

Using the"AUCell"approach, we then calculate the activation level of each cell and ranked the cells according to AUC score. We demonstrated the expression of the top 10 most active DDRMGs in various cell subpopulations using UMAP plots (Fig. 2A). A total of 9807 single cells were defined as active, with a threshold of 0.14 (Fig. 2B). The detailed layout of these active cells is displayed in a UMAP plot (Fig. 2C). The results indicate that tumor-associated fibroblasts and macrophages may play significant roles in the DDR process, as evidenced by the composition and proportions of active and inactive cells in Fig. 2D. DDR-active cells contribute to PDAC progression and therapeutic resistance by maintaining genomic stability through enhanced DNA repair mechanisms (e.g., homologous recombination), enabling tumor cells to

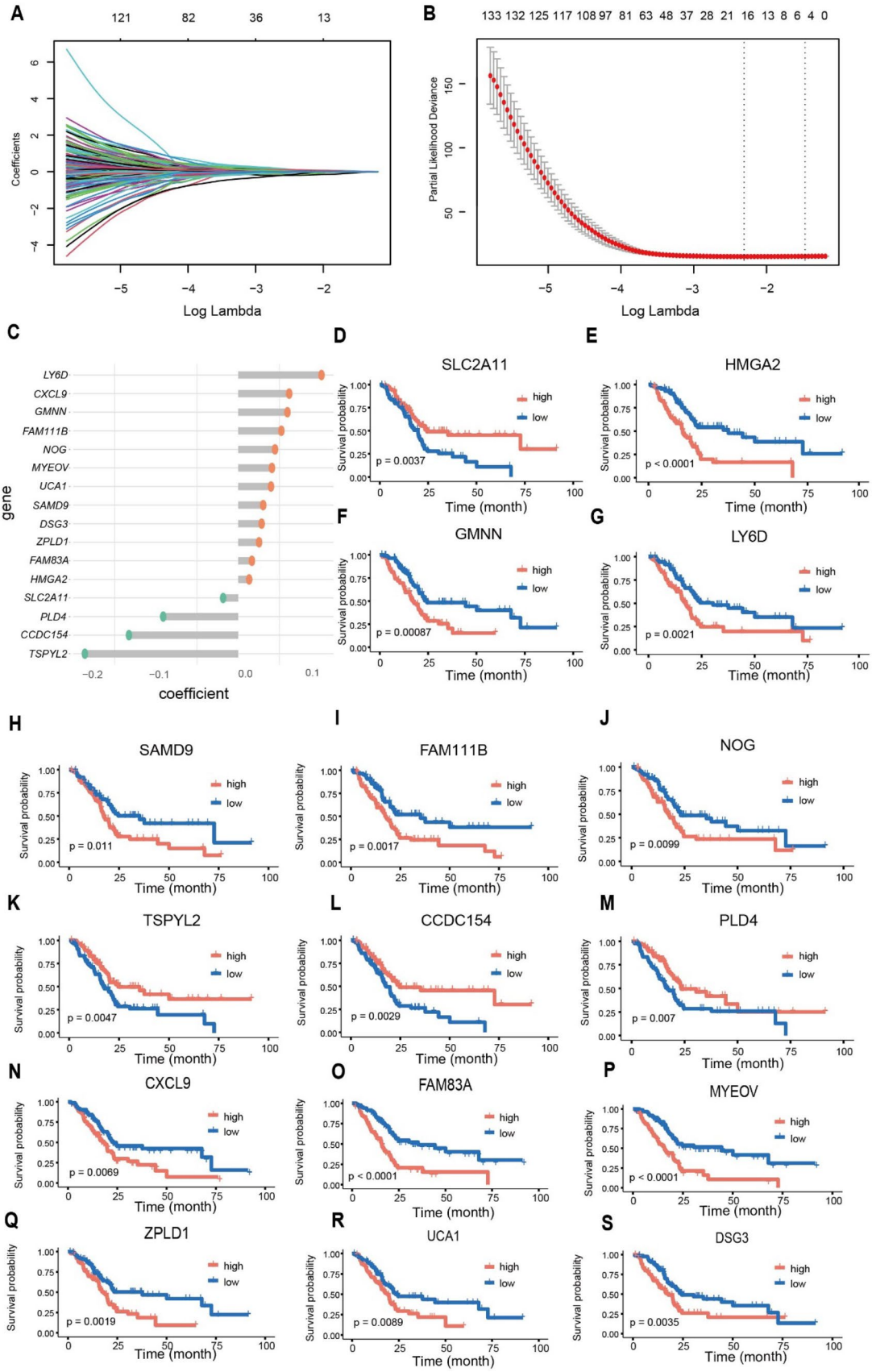

A
121 82 36 13
Coefficients
Log Lambda
B
133 132 125 117 108 97 81 63 48 37 28 21 16 13 8 6 4 0
Partial Likelihood Deviance
Log Lambda
C
LY6D
CXCL9
GMNN
FAM111B
NOG
MYEOV
UCA1
SAMD9
DSG3
ZPLD1
FAM83A
HMGA2
SLC2A11
PLD4
CCDC154
TSPYL2
gene
coefficient
D
SLC2A11
p = 0.0037
E
HMGA2
p < 0.0001
F
GMNN
p = 0.00087
G
LY6D
p = 0.0021
H
SAMD9
p = 0.011
I
FAM111B
p = 0.0017
J
NOG
p = 0.0099
K
TSPYL2
p = 0.0047
L
CCDC154
p = 0.0029
M
PLD4
p = 0.007
N
CXCL9
p = 0.0069
O
FAM83A
p < 0.0001
P
MYEOV
p < 0.0001
Q
ZPLD1
p = 0.0019
R
UCA1
p = 0.0089
S
DSG3
p = 0.0035
high
low
Survival probability
Time (month)

**Fig. 4** In terms of clinical features, the RS is an independent predictor of outcome. **A** Risk score graphs for the TCGA. The dashed line at the top of the graphic represents the median risk score for each cohort. Patients were classified into several risk groups on the basis of these criteria. The appropriate patient data are displayed in the middle and bottom of the diagram. **B** The findings of the Cox regression analysis conducted on the clinical features of the TCGA cohort are displayed in a forest map. **C** The nomogram based on the prediction model. The clinical factor's contribution to outcome events is depicted by the line segment, whereas the likelihood of 1-, 3-, and 5-year survival is illustrated by the lower three lines. The total points represent the cumulative score derived from the individual scores of all the variable values

evade apoptosis triggered by chemotherapy or radiation. These cells also remodel the tumor microenvironment (TME) by promoting stromal interactions (e.g., tumor-associated fibroblasts) and immune evasion, as shown by their high DDR activity in our single-cell analysis (Fig. 2C). Such cells likely drive resistance to genotoxic therapies and support tumor survival, aligning with studies linking DDR activation to chemoresistance in PDAC [16].

The identification of 478 DDR-related genes across cell clusters reveals key molecular mechanisms underlying PDAC heterogeneity. These genes are enriched in pathways critical for DNA repair (e.g., BRCA1/2, ATM), cell cycle regulation (e.g., CDK1), and immune modulation (e.g., CXCL12). This highlights potential therapeutic targets, such as PARP inhibitors for BRCA-deficient tumors or checkpoint inhibitors targeting DDR-immunomodulatory crosstalk. By mapping these genes to specific cell types (e.g., fibroblasts and macrophages), our study provides insights into stromal-driven resistance mechanisms and personalized treatment strategies.

#### 3.1.3 Identification of functions of DDR-reactive cell subgroups

By employing GSEA, the expression of marker genes with high specificity was used to infer the function of the population of active DDR cells. After the marker gene sets were subjected to GO and KEGG pathway enrichment analyses, the ten most significantly enriched terms and pathways were used to create a bubble diagram. The analysis revealed that marker gene sets were considerably enriched in extracellular tissue-related activities and pathways linked to neuroactive ligand–receptor interactions and cancer (Supplemental Fig. 1C–1D). On the basis of the log2 FC value, GSEA was performed on the DEGs with hallmark gene sets. Eight pathways were enriched. The G2M checkpoint pathway was notably enriched (Fig. 2E).

Enrichment of pathways such as"Allograft rejection,""Coagulation,""E2 F targets,""Epithelial-mesenchymal transition (EMT),""G2/M checkpoint,""Myogenesis,""Pancreas beta cells,"and"UV response dn"significantly influences the tumor microenvironment (TME) and patient prognosis. Allograft rejection pathways indicate immune activation, which can enhance anti-tumor immunity [16]. Coagulation pathways are linked to tumor progression and metastasis by promoting angiogenesis and fibrin deposition [17]. E2 F targets regulate cell cycle progression, and their dysregulation is associated with uncontrolled proliferation [18]. EMT facilitates tumor invasion and metastasis by enhancing cell motility and resistance to apoptosis [19]. Myogenesis and pancreas beta cell pathways may reflect stromal interactions and metabolic reprogramming in the TME [20]. UV response dn pathways are often downregulated in tumors, potentially reducing DNA repair capacity [21].

The G2/M checkpoint is crucial for maintaining genomic stability by ensuring proper cell division. Dysregulation of this checkpoint, often observed in cancers, leads to mitotic errors and chromosomal instability, promoting tumor aggressiveness and poor prognosis [22]. C646 has been shown to impede G2/M cell cycle-related proteins and augment the antitumor efficacy against pancreatic cancer [23]. G2/M checkpoint is also triggered by the ATR-CHK1-WEE1 pathway in response to DNA damage [24]. Understanding these pathways provides insights into TME dynamics and potential therapeutic strategies.

#### 3.1.4 Gene expression pattern and bulk RNA-sequencing analyses of PDAC samples

We acquired TCGA-PDAC raw count data to perform an analysis of differential expression. For differential expression analysis, 179 PDAC samples and 171 normal samples were acquired by merging normal samples from the GTEx project with TCGA-PDAC data. With the use of the"limma"R package, 3671 DEGs—2108 upregulated genes and 1563 downregulated genes—were found. The PCA diagram revealed the effective separation of tumor samples from normal samples (Supplementary Fig. S2 A). The DEGs between tumor and normal tissues were visualized in a volcano plot and a gene expression heatmap (Supplementary Fig. S2B-C). After the DEGs and the marker genes of the active cell subgroup intersected, 1422 differentially active marker genes were identified.

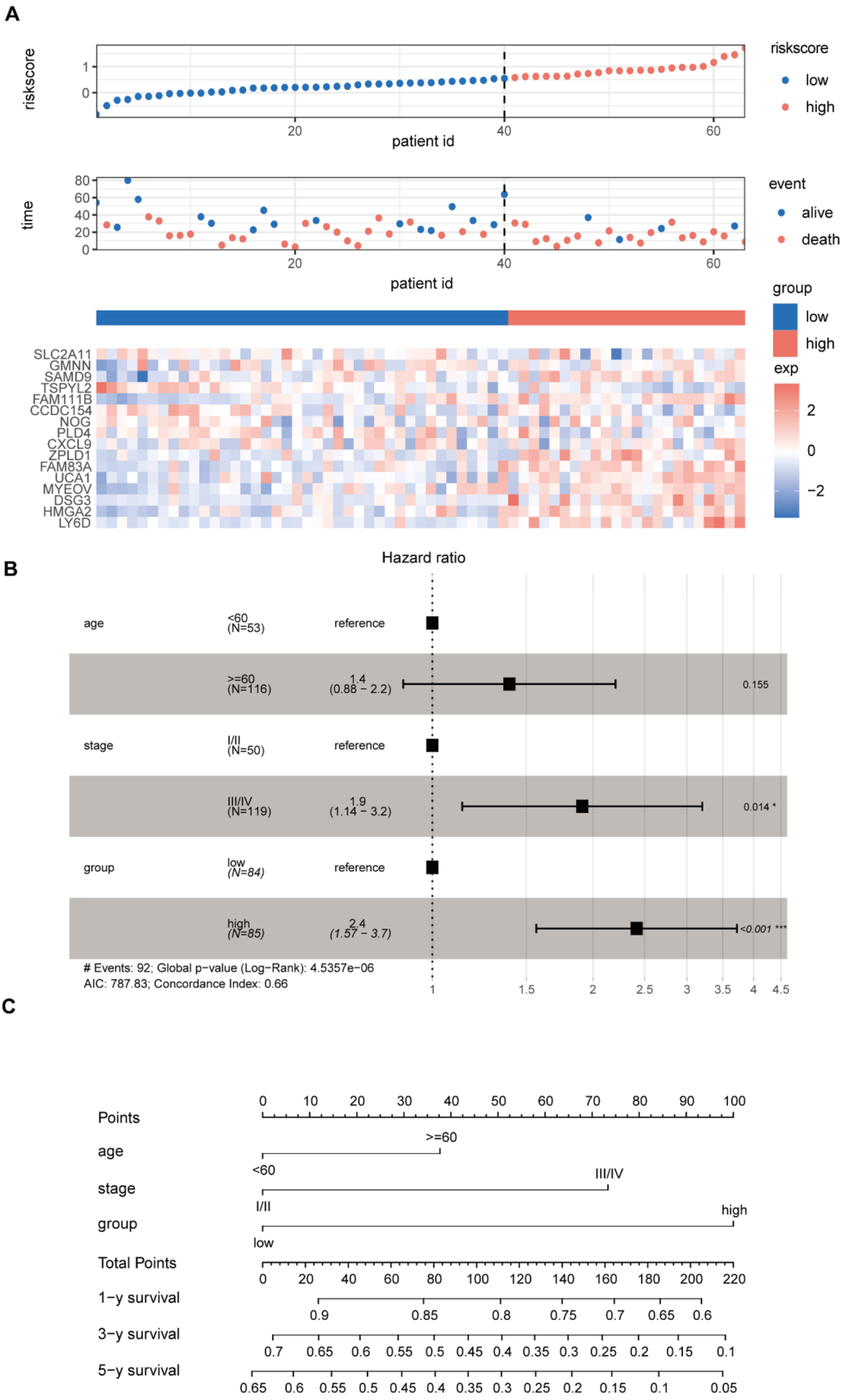

A
riskscore
patient id
riskscore
low
high
time
event
alive
death
group
low
high
exp
SLC2A11
GMNN
SAMD9
TSPYL2
FAM111B
CCDC154
NOG
PLD4
CXCL9
ZPLD1
FAM83A
UCA1
MYEOV
DSG3
HMGA2
LY6D
B
Hazard ratio
age
<60 (N=53)
reference
>=60 (N=116)
1.4 (0.88 − 2.2)
0.155
stage
I/II (N=50)
reference
III/IV (N=119)
1.9 (1.14 − 3.2)
0.014 *
group
low (N=84)
reference
high (N=85)
2.4 (1.57 − 3.7)
<0.001 ***
# Events: 92; Global p−value (Log−Rank): 4.5357e−06
AIC: 787.83; Concordance Index: 0.66
C
Points
age
stage
group
Total Points
1−y survival
3−y survival
5−y survival
>=60
<60
III/IV
I/II
high
low

**Fig. 5** Assessment of the associations of the model risk score with immune cell infiltration and immune markers. **A** The variation in the percentage of 21 infiltrating immune cells between the patients in the high- and low-risk groups is displayed in the box diagram. The high-risk group is shown in yellow, whereas the low-risk group is shown in blue. **B** The proportions of immune cells in the high-risk and low-risk groups. **C** Relationships between immunological checkpoint genes and model gene expression

## 3.2 Development and validation of the prognostic risk model

### 3.2.1 Identifying the gene signatures for prognosis in PDAC

To determine whether marker genes in the active cell population were associated with prognosis ($p < 0.05$), univariate Cox analysis was employed. After removing duplicate genes from the training set via LASSO regression analysis, a seed value of 10,210 was established. Sixteen genes were ultimately associated with the prognosis of PDAC, and the results are shown in Figs. 3A–3 C. The 16 prognostic genes were systematically selected from 1422 active markers and 3671 DEGs through multi-step filtering (intersection analysis, survival validation, and LASSO regression), ensuring statistical rigor and clinical relevance to PDAC prognosis. To categorize the patients into high- and low-risk groups, the median expression value of each of the sixteen genes—which were identified via Cox regression analysis—was utilized as the cutoff value. K–M survival curves were generated for patients from the TCGA-PDAC cohort to identify genes that reliably predict patient survival. The KM curves of these 16 genes showed notable variations from one another. The prognosis was poorer for patients in the high-risk group than for those in the low-risk group (Fig. 3D- 3S).

### 3.2.2 Verification of the robustness of the model with both internal and external datasets

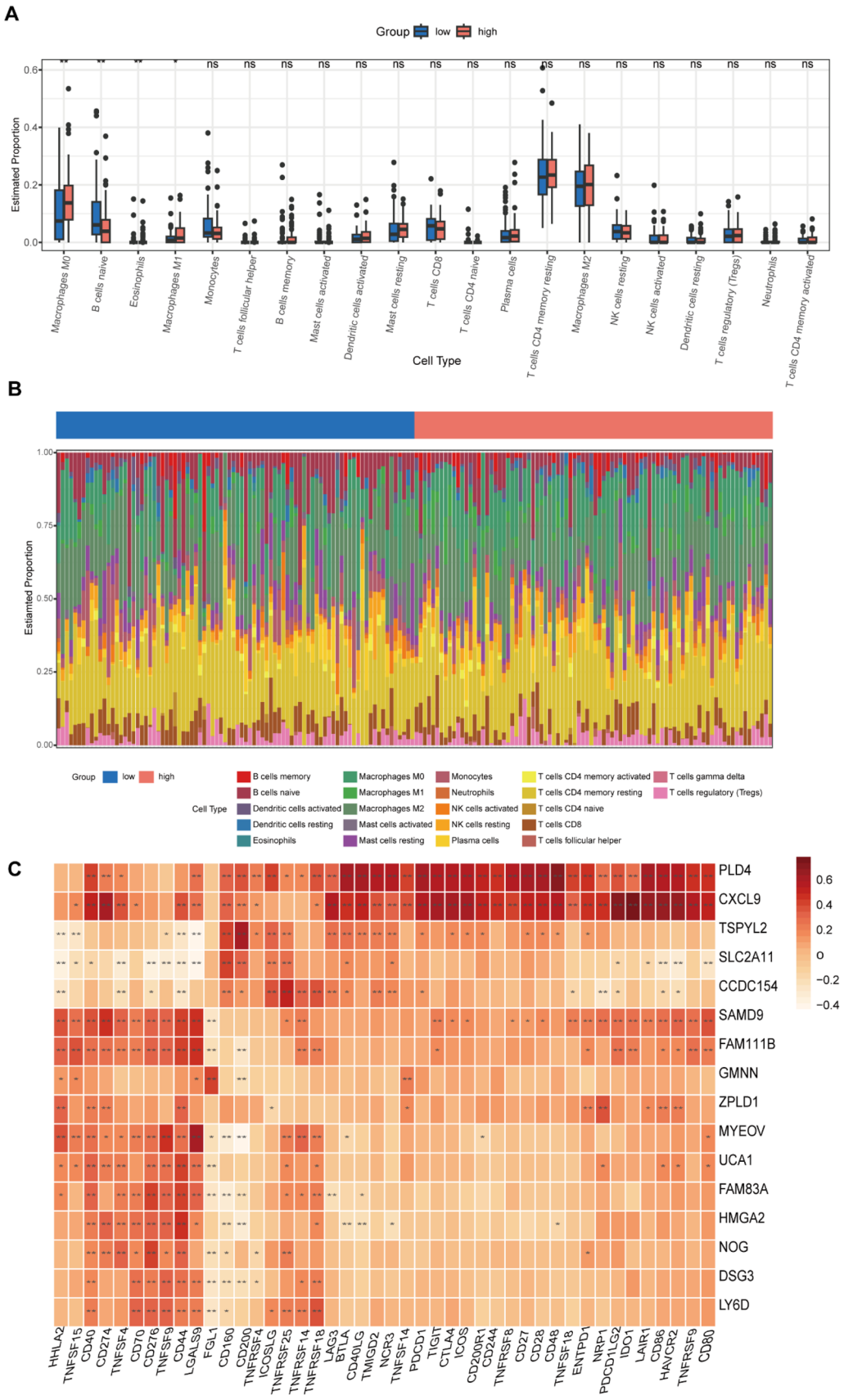
By dividing the patients into high-risk and low-risk groups using the median RS value as the cutoff, we assessed the robustness of the models created using the sixteen gene signatures. We selected the TCGA-PDAC cohort as the model training cohort. We also selected GSE62452 and GSE57495 as validation cohorts to confirm the validity of our model. The K−M analysis revealed that, across all datasets, the OS time of the high-risk group was significantly shorter than that of the low-risk group (Supplementary Fig. S3 A–3 C). The model's ability to predict prognosis was assessed using ROC curves (Supplementary Fig. S3D–3 F). In the TCGA-PDAC training cohort, the AUCs for 3- and 5-year survival were 0.87 and 0.85, respectively. In the GSE62452 cohort, the AUCs of the GSE62452 cohort was 0.82 at 3 years and 0.62 at 5 years. The AUCs of the GSE57495 cohort for 3- and 5-year survival were 0.74 and 0.70, respectively. Collectively, these findings show that our predictive model performed well across several cohorts.

## 3.3 Relationships between clinical characteristics and the model risk score in PDAC

To determine whether the RS can be used as an independent prognostic indicator, multivariate Cox regression analysis was performed on patient clinical characteristics, including age and stage (Fig. 4A–4 C). The heatmap in Fig. 4A displays prognostic genes (e.g., *LY6D*, *TSPYL2*) identified through single-cell and bulk RNA-seq integration. While *LY6D and TSPYL2* [25] show preliminary links to DDR processes (see Results), further validation is needed to confirm their functional roles. LY6D knockdown experiments demonstrated increased DNA damage (Fig. 8D, E), supporting its potential DDR relevance.

## 3.4 Examination of the model's immune cell infiltration features

The incidence and progression of malignancies are impacted by immune cell infiltration in the TME. Consequently, we used the CIBERSORT method to assess differences in the TME features between patients in the high-risk and low-risk groups. The immune cell bar chart provides a detailed representation of the 21 immune cell groups and their proportions (Supplementary Fig. S2 A). The data revealed notable differences in the proportions of immune cells infiltrating tumors between the high- and low-risk patient groups (Fig. 5A). Specifically, significant differences were observed for immune cells including M0 and M1 macrophages, naive B cells and eosinophils. Figure 5B shows the distribution of 21 immune cell types among patients in the high-risk and low-risk groups. Immunological cells secrete chemicals called immunological checkpoints that have the ability to control the level of immune activation. They have a significant effect on the development of autoimmune illnesses. Consequently, we examined the relationships among the expression levels of


A
Group low high
Estimated Proportion
Cell Type
Macrophages M0
B cells naive
Eosinophils
Macrophages M1
Monocytes
T cells follicular helper
B cells memory
Mast cells activated
Dendritic cells activated
Mast cells resting
T cells CD8
T cells CD4 naive
Plasma cells
T cells CD4 memory resting
Macrophages M2
NK cells resting
NK cells activated
Dendritic cells resting
T cells regulatory (Tregs)
Neutrophils
T cells CD4 memory activated
B
Estiamted Proportion
Group low high
Cell Type
B cells memory
B cells naive
Dendritic cells activated
Dendritic cells resting
Eosinophils
Macrophages M0
Macrophages M1
Macrophages M2
Mast cells activated
Mast cells resting
Monocytes
Neutrophils
NK cells activated
NK cells resting
Plasma cells
T cells CD4 memory activated
T cells CD4 memory resting
T cells CD4 naive
T cells CD8
T cells follicular helper
T cells gamma delta
T cells regulatory (Tregs)
C
PLD4
CXCL9
TSPYL2
SLC2A11
CCDC154
SAMD9
FAM111B
GMNN
ZPLD1
MYEOV
UCA1
FAM83A
HMGA2
NOG
DSG3
LY6D
HHLA2
TNFSF15
CD40
CD274
TNFSF4
CD70
CD276
TNFSF9
CD44
LGALS9
FGL1
CD160
CD200
TNFRSF4
ICOSLG
TNFRSF25
TNFRSF14
TNFRSF18
LAG3
BTLA
CD40LG
TMIGD2
NCR3
TNFSF14
PDCD1
TIGIT
CTLA4
ICOS
CD200R1
CD244
TNFRSF8
CD27
CD28
CD48
TNFSF18
ENTPD1
NRP1
PDCD1LG2
IDO1
LAIR1
CD86
HAVCR2
TNFRSF9
CD80

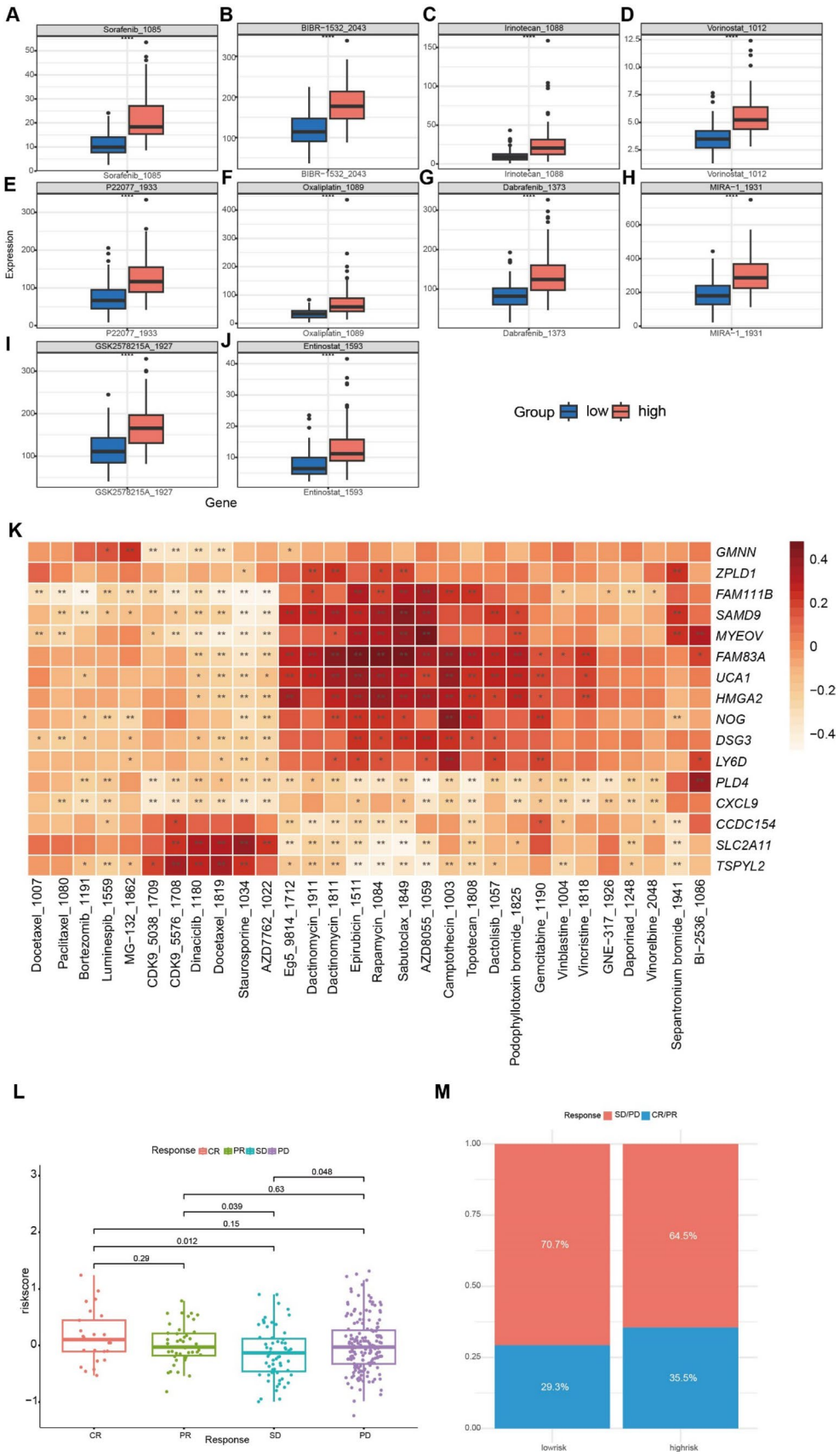



**Fig. 6** The treatment results for PDAC patients predicted by the RS model. **A**–**J** The variation in the IC50 values of ten chemotherapeutic medicines between the low-risk and high-risk cohorts. Red indicates the high risk group, whereas blue indicates the low risk group. **K** The heatmap indicates the relationship between the IC50 values of chemotherapeutic medications and the expression of model genes. The color of the dots indicate the value of the correlation coefficient (positive or negative), whereas the symbol * denotes significance. **L** Violin plot illustrating the distribution of RSs across several immunotherapy response categories. The numerical number indicates the substantial disparity in the replies of the two groups. Immunotherapy response was classified into four groups: complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD). **M** The cumulative distribution of immunotherapy response for patients in the high-risk and low-risk groups is displayed in a histogram

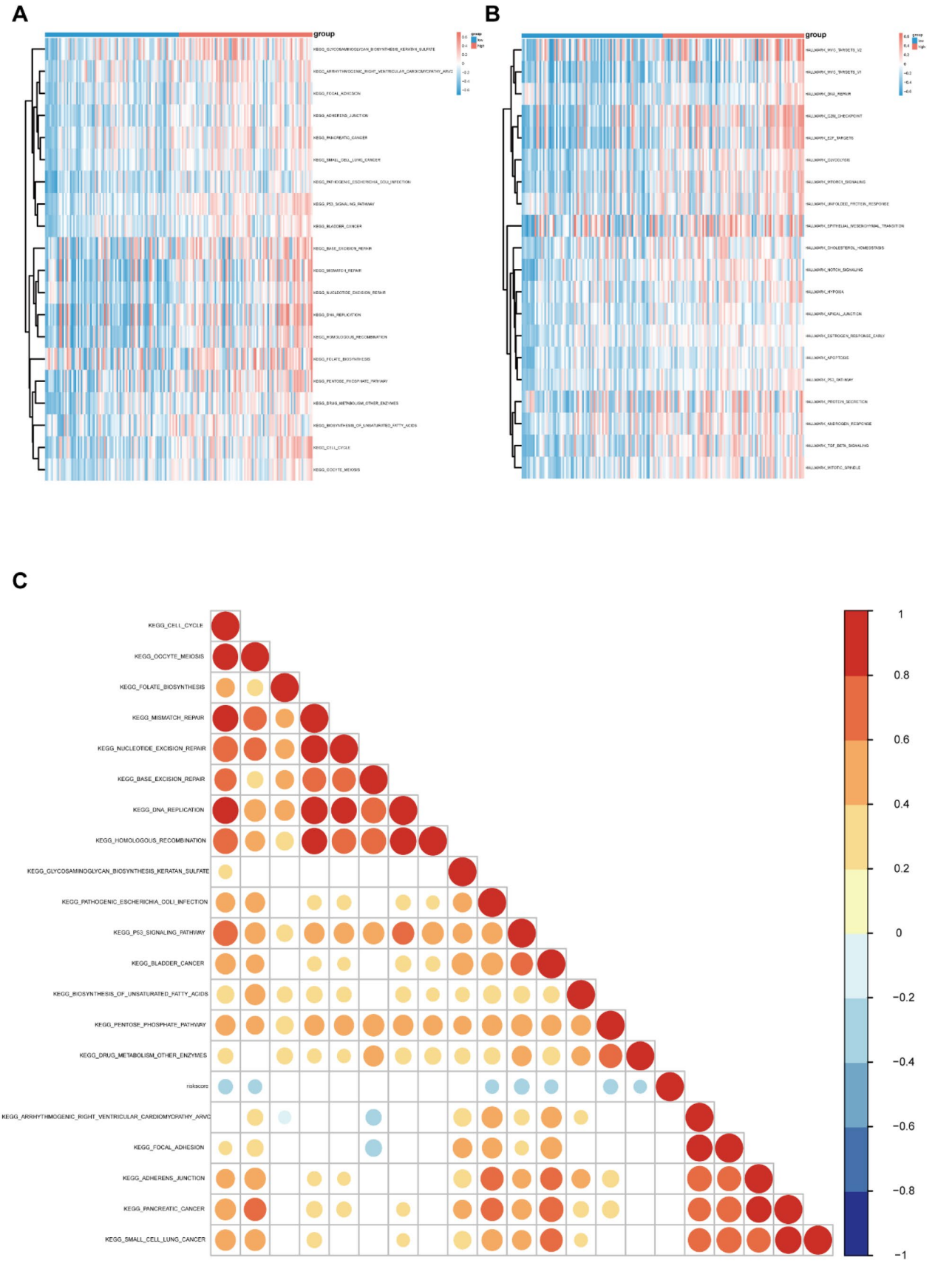



**Fig. 7** Variations in pathway enrichment between the high and low risk groups. **A** KEGG pathways. **B** HALLMARK pathways. **C** Research into the correlation between RS and the KEGG pathway enrichment scores

**Fig. 8** *LY6D* affects DNA damage repair and proliferation in PDAC cells. **A** Immunohistochemical staining was used to detect the expression of LY6D in PDAC tissues and adjacent tissues. The dotted diagram shows a significant increase in LY6D expression in PDAC tissues compared to normal tissues ($p < 0.0001$). **B** IHC scores for each tissue type. **C** qPCR results showing the knockdown efficiency of LY6D in SW1990 cells and PATU- 8988 T cells. **D**, **E** Co-staining of 53BP1 and γ-H2 AX confirmed that LY6D knockdown induces DSBs, with colocalized foci observed in both SW1990 and PATU- 8988 T cells. **F**–**H** EdU assays demonstrated the proliferation ability of different SW1990 and PATU- 8988 T treatment groups

16 model genes, cytokines, and five kinds of immunomodulators. The results revealed a robust relationship between immunological checkpoints and CXCL9 (Fig. 5C).

### 3.5 Estimating drug sensitivity and the immunotherapy response

In the TCGA-PDAC cohort, we computed the IC50 values and anticipated drug sensitivity for the high- and low-risk groups. The nine medications with the lowest P values are displayed individually in the box plot (Fig. 6A–J). Compared with the high-risk group, the low-risk group exhibited increased drug sensitivity, as indicated by lower IC50 values. On the basis of these data, we next carried out a correlation study between the IC50 values of various medications and the model gene expression levels. Figure 6K shows a good correlation between the IC50 values of these medicines and the expression level of the gene FAM83 A. In addition, the IMvigor210 dataset was divided into groups on the basis of the median value of the RS and was used to ascertain the model's ability to predict the response to immunotherapy. We discovered that, as shown in Fig. 6L, the RS determined by our model roughly demonstrated a favorable trend toward an immunotherapy response. Additionally, there were discernible variations in their overall response rates (ORRs). Compared with the low-risk group, the high-risk group presented greater complete response (CR) and partial response (PR) rates and lower progressive disease (PD) and stable disease (SD) rates (Fig. 6M). In conclusion, immunotherapy efficacy data revealed that high-risk group patients often benefited more from ICI treatment. In conclusion, these results may lead to a new understanding of the benefit of treating high-risk PDAC patients with ICIs and low-risk PDAC patients with chemotherapeutic medication.

### 3.6 Distinctions in enriched pathways among patients in various risk groups

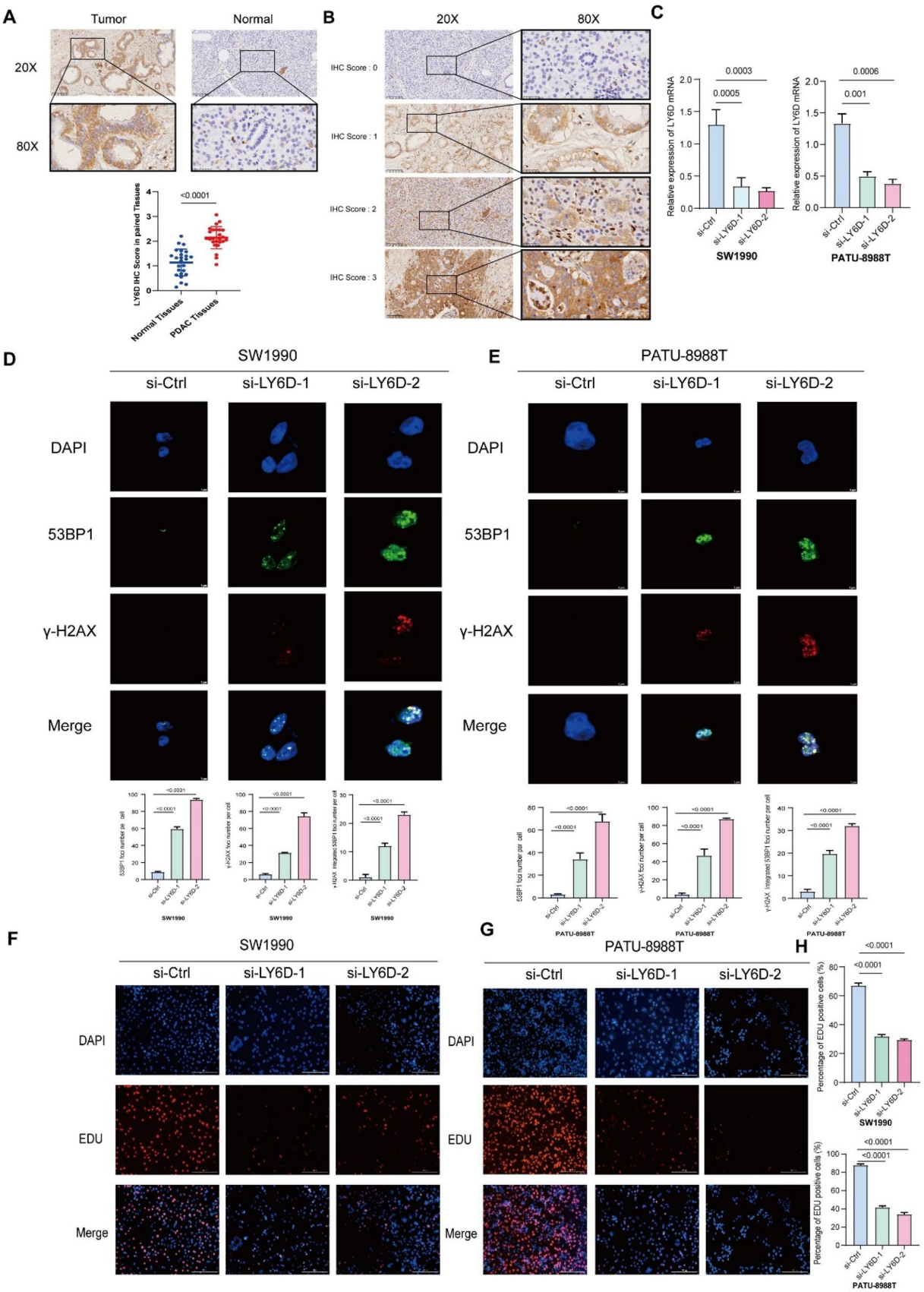
Finally, we delineated distinct signaling pathways that were enriched in both the high- and low-risk TCGA-PDAC groups. The"GSVA"algorithm was applied for HALLMARK and KEGG pathway enrichment analyses. There were notable differences in the pathways enriched in the various risk groups. Heatmaps were used to display the top 20 pathways with the highest log2 FoldChange values for both the KEGG analysis and the HALLMARK analysis (Fig. 7A, B). Additionally, Fig. 7C shows the results of an investigation of the relationship between RS and enrichment scores determined by our approach. These results will be useful in understanding the fundamental mechanism of the DDR in tumorigenesis.

### 3.7 LY6D knockdown induces DSBs

Figure 3C shows that *LY6D* made the greatest contribution to the RS computation. Consequently, we decided to carry out further experimental validation using *LY6D*. The IHC results revealed higher LY6D expression in pancreatic ductal adenocarcinoma tissues than in normal tissues (Fig. 8A). We used the H score for quantitative immunohistochemical analysis and Fig. 8B indicates the basis of our score. We knocked down LY6D in SW1990 cells and PATU- 8988 T cells using siRNAs (Fig. 8C). Then, we performed co-staining of 53BP1 (a marker of DNA double-strand breaks, DSBs) and γ-H2 AX (a marker of DNA damage signaling) in LY6D-knockdown PDAC cells. The merged images (DAPI/53BP1/γ-H2 AX) clearly demonstrate colocalization of 53BP1 foci and γ-H2 AX signals, confirming that *LY6D* depletion induces DSBs(Fig. 8D, E).

### 3.8 The expression of LY6D influences the cycle, apoptosis, and proliferation of PDAC cells

To further validate the effect of LY6D on pancreatic cancer cells, we performed EDU assays to detect the effect of LY6D on cell proliferation (Fig. 8F-H), as well as apoptosis(Fig. 9A-D) and cycle assays(Fig. 9E). These results clearly show that LY6D promotes PDAC cell proliferation by enhancing the S-phase and G2/M transition while decreasing apoptosis.


A
Tumor
Normal
20X
80X
<0.0001
LY6D IHC Score in paired Tissues
Normal Tissues
PDAC Tissues
B
20X
80X
IHC Score : 0
IHC Score : 1
IHC Score : 2
IHC Score : 3
C
0.0003
0.0005
0.0006
0.001
Relative expression of LY6D mRNA
si-Ctrl
si-LY6D-1
si-LY6D-2
SW1990
PATU-8988T
D
SW1990
si-Ctrl
si-LY6D-1
si-LY6D-2
DAPI
53BP1
γ-H2AX
Merge
E
PATU-8988T
F
SW1990
EDU
G
PATU-8988T
H
Percentage of EDU positive cells (%)

## 4 Discussion

PDAC ranks as one of the most detrimental types of cancer in humans, as it has a very poor prognosis. PDAC is the most common type of pancreatic cancer, accounting for approximately 96% of cases [26]. PDAC whole-exome and whole-genome sequencing has revealed changes in the key driver genes of most pancreatic tumors [27, 28]. Genetic mutations in germline genes related to the DDR have been found in patients with sporadic PDAC [29]. The *BRCA2* gene has the most frequent germline mutations (1.4–7.4%). A recent prospective phase III study (the POLO trial, Pancreas Cancer Olaparib Ongoing, NCT02184195) was conducted to assess the effectiveness of olaparib in patients with BRCA-mutant metastatic PDAC [8]. PARPi have been approved for use as maintenance therapy in patients with metastatic pancreatic cancer who harbor pathogenic germline mutations in *BRCA1/2*. Recently, our team's latest findings reveal the dual functional significance of TPX2 in controlling DNA DSB repair pathway selection and mitotic progression, providing a potential PARPi therapeutic strategy for pancreatic cancer patients [30]. Studies have shown that DDR deficiency is present in as many as 24% of patients with PDAC [31]. Other studies have revealed the critical roles that DDR malfunction plays in PDAC carcinogenesis and treatment outcomes [32]. Therefore, discovering more about the underlying processes of the DDR will aid in the determination of the best treatment option for patients and the understanding of PDAC characteristics.

In this work, we combined bulk and single-cell RNA sequencing data to develop a predictive model for PDAC. On the basis of single-cell sequencing data from PDAC patients from the GEO database, we annotated and distinguished several cell subgroups. The findings revealed notable variations in the TME of various PDAC patients. The marker genes with significant specificity for each cell subgroup overlapped with a total of 478 DDR genes. On the basis of the expression patterns of these 478 DDR genes, we were able to identify active cell subgroups and their respective roles. Afterward, we used the"AUCell"technique to identify DDRMGs. GO and KEGG pathway enrichment analyses revealed notable enrichment of pathways linked to the ECM-receptor interaction and focal adhesion. The ECM, a critical component of the TME, undergoes significant remodeling in PDAC, contributing to tumor progression and treatment resistance. Desmoplasia, characterized by excessive fibroblast activity and ECM deposition, is a hallmark of PDAC histology and plays a key role in its fibrotic microenvironment [33]. Our AUCell score analysis further highlighted the importance of tumor-associated fibroblasts, which exhibited the highest DDRMG scores. Recent studies have also demonstrated a close relationship between ECM remodeling and DDR mechanisms [34–36]. Our findings provide a fresh view of the DDR mechanism, revealing how it modifies the ECM and induces treatment resistance in PDAC.

Next, we determined which marker genes were differentially expressed in PDAC patients via bulk RNA-sequencing data. The expression patterns of sixteen genes—*LY6D, CXCL9, GMNN, FAM111B, NANOG, MYEOV, UCA1, SAMD9, DSG3, ZPLD1, FAM83 A, HMGA2, SLC2 A11, PLD4, CCDC154 and TSPYL2*—that are linked to prognosis were used to construct prognostic risk models. Modifications in gene expression patterns mostly impact the immunological landscape and tumor features of various individuals. According to the latest research reports, LY6 proteins, such as LY6D, are becoming well known as superior tumor-associated antigens that, when combined with effective mismatch repair, may significantly suppress antitumor immunity in patients with malignancies [37]. The lymphocyte antigen 6 complex, locus D (LY6D, commonly known as E48), is a GPI-anchored protein that is expected to be present on the plasma membrane and cell surface. It is located on human chromosome 8 in the 8q24 region [38]. Recent studies have revealed that *LY6D* is associated with cell adhesion and has elevated expression in some types of malignancies [39, 40]. *LY6D* is a CRC antigen that is expressed preferentially. By utilizing a TcE, one may effectively reroute and activate cytotoxic T cells to lyse CRC cells that express *LY6D* [41]. Our study demonstrated that LY6D is highly expressed in pancreatic ductal adenocarcinoma and that knockdown of LY6D affects DNA damage repair capacity and cell proliferation in PDAC. According to a recent significant study, the CXCL9:SPP1 ratio can be used to describe the antitumor immune cell abundance in cancer, the gene expression patterns of each type of tumor-infiltrating cell, the control of communication networks that determine tumor control or progression, and the response to immunotherapy [42].

When the median RS value was used to categorize patients into high- and low-risk groups, the ROC curve demonstrated that the model performed well in prognostic prediction. According to the survival study, individuals in the high-risk subgroup had a poorer prognosis than those in the low-risk subgroup did. Furthermore, our GSEA findings revealed a strong correlation between DDRMGs and the"MYC_TARGETS_V2"and"MYC_TARGETS_V1"pathways. MYC functions as a transcription factor for numerous DDR genes, such as WRN helicase, which is involved in the repair of replication forks that have stopped and abnormal replication intermediates. According to one study, MYC functions downstream in the DDR, regulating the *TP53/MDM2 (MDM2)*/ubiquitin-protein ligase tumor suppressor *P19 ARF/E3* [43]. This further validates the accuracy and guiding value of our models.

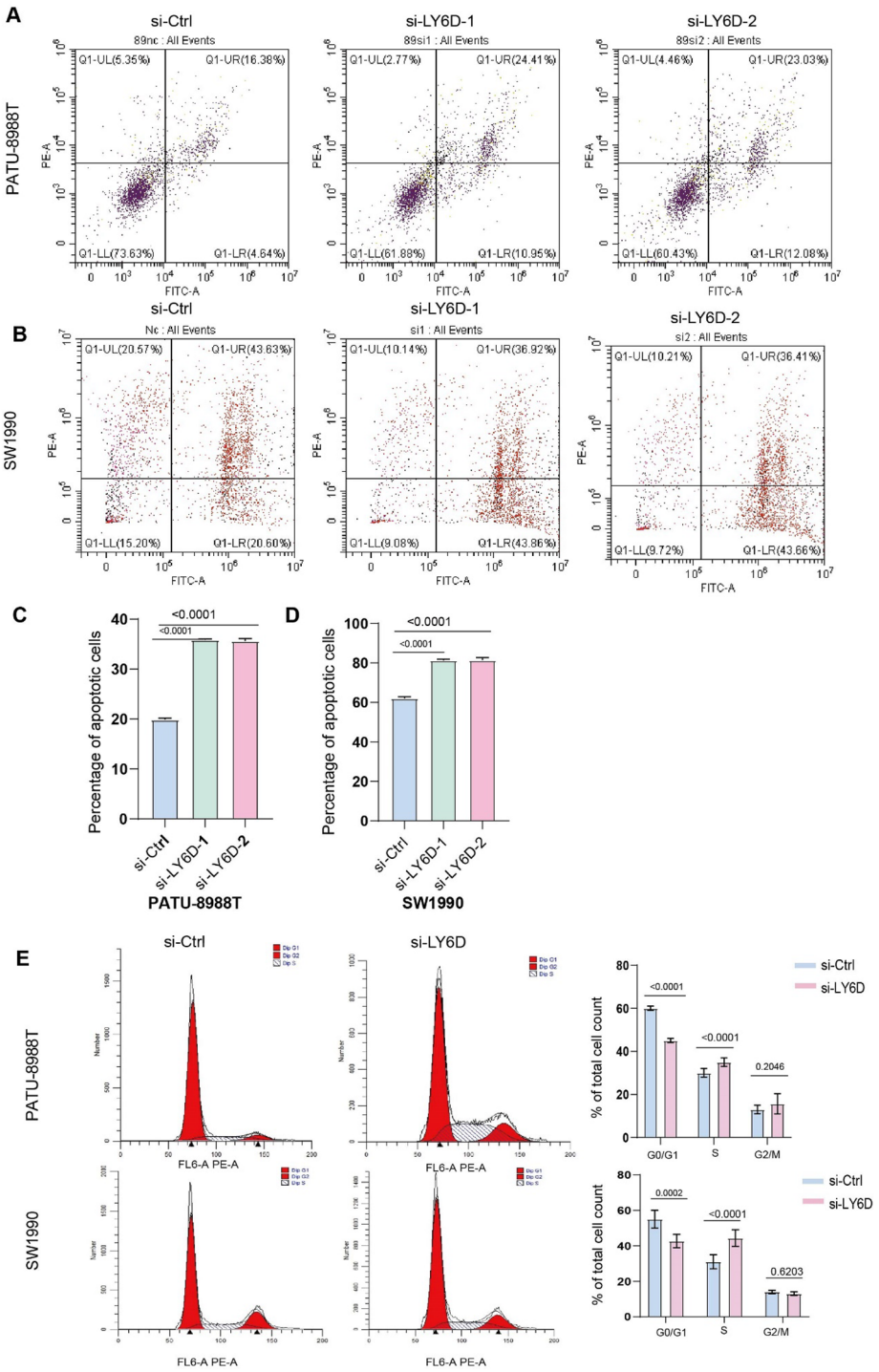



**Fig. 9** Effect of LY6D on the phenotype of PDAC cells. **A**, **B** Flow cytometry analysis of the effect of knockdown of LY6D on apoptosis in SW1990 and PATU- 8988 T cells. **C**, **D** The percentage of apoptosis in the experimental and control groups was compared by calculating the sum of Q1-UR and Q1-LR, which was validated by three repetitions of the experiment showing $p < 0.0001$. **E** Impact of LY6D knockdown on pancreatic cancer cell cycle as determined by flow cytometry

Immune checkpoints are crucial for helping immune cells avoid immune surveillance, according to recent research [44]. We also investigated the relationships between immunological checkpoints and sixteen signature genes that were incorporated into the prognostic model. Immune checkpoints and CXCL9 have been shown to be strongly correlated [45]. A study demonstrated that, by suppressing CD8 + cytotoxic T cell infiltration, CXCL9 mediates antitumor immune responses in PDAC [46]. Studies have demonstrated that tumor cell and TME immune features play major roles in the determining the efficacy of

immunotherapy [47]. DDR pathway abnormalities can alter immunogenicity and immune checkpoint levels, which directly impacts the efficacy of ICI therapy [48].

Additionally, we analyzed the sensitivity of the two risk groups to routinely used anticancer medications. The IC50 values of these medications were positively correlated with the RS, according to our data, and the low-risk group responded better to the treatments. The effectiveness of immunotherapy showed that patients in high-risk groups often benefited more from ICI treatment. Our results indicate that the most beneficial treatment differed between risk groups; thus, our model can provide guidance for improving PDAC treatment options.

In summary, our DDR-associated prognostic model offers valuable insights into the molecular mechanisms that may influence PDAC prognosis and treatment efficacy. This study presents a promising approach for integrating bulk RNA sequencing with single-cell sequencing data, which may contribute to advancing research in pancreatic cancer. The immunological environment, cell-specific markers, and cell heterogeneity can all be revealed by single-cell RNA sequencing methods, whereas bulk RNA sequencing can be used to detect significant changes in gene expression and signaling pathways. However, there are some limitations to our study. Since this work is retrospective in nature, additional validation of the model's accuracy with larger sample sizes and prospective cohorts is needed. Additionally, further investigation is needed to validate the regulatory roles of sixteen genes in the initiation and progression of PDAC and the DDR.

## 5 Conclusion

A novel prognostic model created on the basis of the sixteen DDRGs identified in PDAC may have practical applications, particularly in predicting immunotherapy efficacy.

**Acknowledgments** We would like to thank TCGA database for providing the PAAD data set and the GEO database for providing the GSE212966、GSE62452 and GSE57495 data set.

**Author contributions** CZ, RT, JY, YC, and YL conceived and designed the study and drafted the manuscript. CZ, WW, and XY provided analytical technical support. CZ, JX, and RT participated in the production of charts and pictures. All authors have read and approved the final manuscript. All authors contributed to the article and approved the submitted version.

**Funding** This study was jointly supported by the National Natural Science Foundation of China (U21 A20374 and 82072698), Shanghai Municipal Science and Technology Major Project (21 JC1401500), Scientific Innovation Project of Shanghai Education Committee (2019–01 - 07–00 - 07-E00057), and Natural Science Foundation of Shanghai (23ZR1479300).

**Data availability** The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

### Declarations

**Ethics approval and consent to participate** All studies involving human tissues were conducted in strict accordance with the Declaration of Helsinki and were approved by the Ethics Committee of Fudan University Shanghai Cancer Center. All participants were informed and consented to sample collection, intended research, and publication usage. Informed Consent was obtained from all the participants involved in the study.

**Competing interests** 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.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

## References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA. 2024;74(1):12–49.
2. Tavakkoli A, Singal AG, Waljee AK, Elmunzer BJ, Pruitt SL, McKey T, et al. Racial disparities and trends in pancreatic cancer incidence and mortality in the United States. Clin Gastroenterol Hepatol. 2020;18(1):171-8.e10.

3. Lai Y, Beaver JM, Laverde E, Liu Y. Trinucleotide repeat instability via DNA base excision repair. DNA Repair. 2020;93:102912.
4. Sassa A, Odagiri M. Understanding the sequence and structural context effects in oxidative DNA damage repair. DNA Repair. 2020;93:102906.
5. Szewczuk M, Boguszewska K, Kaźmierczak-Barańska J, Karwowski BT. The role of AMPK in metabolism and its influence on DNA damage repair. Mol Biol Rep. 2020;47(11):9075–86.
6. Jeggo PA, Pearl LH, Carr AM. DNA repair, genome stability and cancer: a historical perspective. Nat Rev Cancer. 2016;16(1):35–42.
7. Sun W, Zhang Q, Wang R, Li Y, Sun Y, Yang L. Targeting DNA damage repair for immune checkpoint inhibition: mechanisms and potential clinical applications. Front Oncol. 2021;11:648687.
8. Golan T, Hammel P, Reni M, Van Cutsem E, Macarulla T, Hall MJ, et al. Maintenance Olaparib for germline BRCA-mutated metastatic pancreatic cancer. N Engl J Med. 2019;381(4):317–27.
9. Notta F, Chan-Seng-Yue M, Lemire M, Li Y, Wilson GW, Connor AA, et al. A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. Nature. 2016;538(7625):378–82.
10. Perkhofer L, Gout J, Roger E, Kude de Almeida F, Baptista Simões C, Wiesmüller L, et al. DNA damage repair as a target in pancreatic cancer: state-of-the-art and future perspectives. Gut. 2021;70(3):606–17.
11. Chen K, Wang Q, Liu X, Tian X, Dong A, Yang Y. Immune profiling and prognostic model of pancreatic cancer using quantitative pathology and single-cell RNA sequencing. J Transl Med. 2023;21(1):210.
12. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573-87.e29.
13. Oh K, Yoo YJ, Torre-Healy LA, Rao M, Fassler D, Wang P, et al. Coordinated single-cell tumor microenvironment dynamics reinforce pancreatic cancer subtype. Nat Commun. 2023;14(1):5226.
14. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7.
15. Maeser D, Gruener RF, Huang RS. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. Brief Bioinform. 2021;22(6):bbab260.
16. Diorio C, Teachey DT, Grupp SA. Allogeneic chimeric antigen receptor cell therapies for cancer: progress made and remaining roadblocks. Nat Rev Clin Oncol. 2025;22(1):10–27.
17. Falanga A, Marchetti M, Vignoli A. Coagulation and cancer: biological and clinical aspects. J Thromb Haemost. 2013;11(2):223–33.
18. Kent LN, Leone G. The broken cycle: E2F dysfunction in cancer. Nat Rev Cancer. 2019;19(6):326–38.
19. Gundamaraju R, Lu W, Paul MK, Jha NK, Gupta PK, Ojha S, et al. Autophagy and EMT in cancer and metastasis: who controls whom? Biochim Biophys Acta. 2022;1868(9):166431.
20. Sahai E, Astsaturov I, Cukierman E, DeNardo DG, Egeblad M, Evans RM, et al. A framework for advancing our understanding of cancer-associated fibroblasts. Nat Rev Cancer. 2020;20(3):174–86.
21. Lee JW, Ratnakumar K, Hung KF, Rokunohe D, Kawasumi M. Deciphering UV-induced DNA damage responses to prevent and treat skin cancer. Photochem Photobiol. 2020;96(3):478–99.
22. Löbrich M, Jeggo PA. The impact of a negligent G2/M checkpoint on genomic instability and cancer induction. Nat Rev Cancer. 2007;7(11):861–9.
23. Ono H, Kato T, Murase Y, Nakamura Y, Ishikawa Y, Watanabe S, et al. C646 inhibits G2/M cell cycle-related proteins and potentiates anti-tumor effects in pancreatic cancer. Sci Rep. 2021;11(1):10078.
24. Smith HL, Southgate H, Tweddle DA, Curtin NJ. DNA damage checkpoint kinases in cancer. Expert Rev Mol Med. 2020;22:e2.
25. Liu Z, Li C, Yu C, Chen Z, Zhao C, Ye L. TSPYL2 reduced gefitinib resistance and DNA damage repair via suppressing SIRT1-mediated FOXO3 deacetylation. Future Med Chem. 2022;14(6):407–19.
26. Wood LD, Canto MI, Jaffee EM, Simeone DM. Pancreatic cancer: pathogenesis, screening, diagnosis, and treatment. Gastroenterology. 2022;163(2):386-402.e1.
27. Wood LD, Yurgelun MB, Goggins MG. Genetics of familial and sporadic pancreatic cancer. Gastroenterology. 2019;156(7):2041–55.
28. Park W, Chen J, Chou JF, Varghese AM, Yu KH, Wong W, et al. Genomic methods identify homologous recombination deficiency in pancreas adenocarcinoma and optimize treatment selection. Clin Cancer Res. 2020;26(13):3239–47.
29. Shindo K, Yu J, Suenaga M, Fesharakizadeh S, Cho C, Macgregor-Das A, et al. Deleterious germline mutations in patients with apparently sporadic pancreatic adenocarcinoma. J Clin Oncol. 2017;35(30):3382–90.
30. Xiao M, Tang R, Pan H, Yang J, Tong X, Xu H, et al. TPX2 serves as a novel target for expanding the utility of PARPi in pancreatic cancer through conferring synthetic lethality. Gut. 2024;74(3):410–23.
31. Waddell N, Pajic M, Patch AM, Chang DK, Kassahn KS, Bailey P, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. Nature. 2015;518(7540):495–501.
32. Dreyer SB, Upstill-Goddard R, Paulus-Hock V, Paris C, Lampraki EM, Dray E, et al. Targeting DNA damage response and replication stress in pancreatic cancer. Gastroenterology. 2021;160(1):362-77.e13.
33. Bachem MG, Schünemann M, Ramadani M, Siech M, Beger H, Buck A, et al. Pancreatic carcinoma cells induce fibrosis by stimulating proliferation and matrix synthesis of stellate cells. Gastroenterology. 2005;128(4):907–21.
34. Walter L, Canup B, Pujada A, Bui TA, Arbasi B, Laroui H, et al. Matrix metalloproteinase 9 (MMP9) limits reactive oxygen species (ROS) accumulation and DNA damage in colitis-associated cancer. Cell Death Dis. 2020;11(9):767.
35. Deng M, Lin J, Nowsheen S, Liu T, Zhao Y, Villalta PW, et al. Extracellular matrix stiffness determines DNA repair efficiency and cellular sensitivity to genotoxic agents. Sci Adv. 2020;6(37):eabb2630.
36. Hodkinson PS, Elliott T, Wong WS, Rintoul RC, Mackinnon AC, Haslett C, et al. ECM overrides DNA damage-induced cell cycle arrest and apoptosis in small-cell lung cancer cells through beta1 integrin-dependent activation of PI3-kinase. Cell Death Differ. 2006;13(10):1776–88.
37. Giordano G, Pancione M. MHC class III lymphocyte antigens 6 as endogenous immunotoxins: unlocking immunotherapy in proficient mismatch repair colorectal cancer. WIREs Mech Dis. 2024;16(1):e1631.

38. Brakenhoff RH, Gerretsen M, Knippels EM, van Dijk M, van Essen H, Weghuis DO, et al. The human E48 antigen, highly homologous to the murine Ly-6 antigen ThB, is a GPI-anchored molecule apparently involved in keratinocyte cell-cell adhesion. J Cell Biol. 1995;129(6):1677–89.
39. Luo L, McGarvey P, Madhavan S, Kumar R, Gusev Y, Upadhyay G. Distinct lymphocyte antigens 6 (Ly6) family members Ly6D, Ly6E, Ly6K and Ly6H drive tumorigenesis and clinical outcome. Oncotarget. 2016;7(10):11165–93.
40. Romeu C, Farré X, Cardesa A, Nadal A. Expression of Ep-CAM, but not of E48, associates with nodal involvement in advanced squamous cell carcinomas of the larynx. Histopathology. 2013;62(6):954–61.
41. Corrales L, Hipp S, Martin K, Sabarth N, Tirapu I, Fuchs K, et al. LY6G6D is a selectively expressed colorectal cancer antigen that can be used for targeting a therapeutic T-cell response by a T-cell engager. Front Immunol. 2022;13:1008764.
42. Bill R, Wirapati P, Messemaker M, Roh W, Zitti B, Duval F, et al. CXCL9:SPP1 macrophage polarity identifies a network of cellular programs that control human cancers. Science (New York, NY). 2023;381(6657):515–24.
43. Kciuk M, Gielecińska A, Kołat D, Kałuzińska Ż, Kontek R. Cancer-associated transcription factors in DNA damage response. Biochim Biophys Acta. 2022;1877(4):188757.
44. Zheng H, Li Y, Zhao Y, Jiang A. Single-cell and bulk RNA sequencing identifies T cell marker genes score to predict the prognosis of pancreatic ductal adenocarcinoma. Sci Rep. 2023;13(1):3684.
45. Nuñez NG, Berner F, Friebel E, Unger S, Wyss N, Gomez JM, et al. Immune signatures predict development of autoimmune toxicity in patients with cancer treated with immune checkpoint inhibitors. Med (New York, NY). 2023;4(2):113-29.e7.
46. Gao HF, Cheng CS, Tang J, Li Y, Chen H, Meng ZQ, et al. CXCL9 chemokine promotes the progression of human pancreatic adenocarcinoma through STAT3-dependent cytotoxic T lymphocyte suppression. Aging. 2020;12(1):502–17.
47. Llosa NJ, Cruise M, Tam A, Wicks EC, Hechenbleikner EM, Taube JM, et al. The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints. Cancer Discov. 2015;5(1):43–51.
48. Shi C, Qin K, Lin A, Jiang A, Cheng Q, Liu Z, et al. The role of DNA damage repair (DDR) system in response to immune checkpoint inhibitor (ICI) therapy. J Exp Clin Cancer Res. 2022;41(1):268.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.