BMC Cancer

RESEARCH Open Access

Multi-omics analysis unveils a four-gene prognostic signature in esophageal squamous carcinoma and the therapeutic potential of PKP1

Xiuzhi Zhang[1], Zhi Wang[2], Yutong Zhao[2], Hua Ye[1,3], Tiandong Li[1], Han Wang[1], Guiying Sun[1], Feifei Liang[2], Liping Dai[2*], Peng Wang[1,3*] and Xiaoli Liu[4*]

Abstract

Background Esophageal squamous cell carcinoma (ESCC) is one of the most common malignancies, characterized by high heterogeneity and poor outcomes. Effective classification for patient stratification and identifying reliable markers for prognosis prediction and treatment choice are crucial.

Methods Integration of single-cell RNA-sequencing (RNA-seq) and bulk RNA-seq analyses were used to characterize ESCC. Non-negative matrix factorization (NMF) clustering was performed to stratify the ESCC patients into different subtypes and the clinical and pathological features of the ESCC subtypes were compared. Cox regression analysis and LASSO regression analysis were used to select key genes and construct a risk model for ESCC. The associations of the key genes with anti-cancer drug sensitivities in ESCC cell lines were investigated. RT-qRCR experiments, proteomics analysis, and multiplex immunohistochemistry (mIHC) experiments were used to validate the results. Furthermore, one identified gene was selected to investigate its correlation with EGFR expression and the gene effect scores of various potential gene targets across pan-cancer.

Results The study identified the dysregulated distributions of epithelial cells and fibroblasts as characteristic of ESCC. ESCC patients could be classified into four distinct subtypes with unique cell type features and prognoses. With the gene makers of the cell type features, a four-gene prognostic signature for ESCC was constructed. The CCND1-PKP1-JUP-ANKRD12 model could effectively discriminate the survival status of ESCC patients, independent of various pathological and clinical features. The risk score for the samples was correlated with the expression levels of immunoregulatory genes. The prognostic effects of CCND1, PKP1, and JUP were confirmed at the protein level. The phosphorylation levels of PKP1, JUP, and ANKRD12 were found to be dysregulated in ESCC tumors. Their expression dysregulation and heterogeneity were demonstrated in ESCC cell lines. All four genes were significantly correlated

*Correspondence:
Liping Dai
lpdai@zzu.edu.cn
Peng Wang
wangpeng1658@hotmail.com
Xiaoli Liu
lxlzts@126.com

Full list of author information is available at the end of the article

with at least one of the anti-cancer drug sensitivities in ESCC cell lines. PKP1 expression was significantly correlated with EGFR expression and gene effect scores in multiple cancers.

Conclusions We conclude that the CCND1-PKP1-JUP-ANKRD12 signature may serve as a novel indicator for ESCC prognosis and diagnosis. PKP1 expression might provide new clues for gene therapy efficacy in multiple cancers.

Keywords Esophageal squamous cell carcinoma, scRNA-seq, Heterogeneity, PKP1, JUP, Gene effect

Introduction

Esophageal cancer (EC) is one of the most malignant tumors and one of the top leading causes of cancer death in the world [1, 2]. With significant geographical variations in its incidence and prevalence, EC is particularly prevalent in East Asia and Eastern Africa. In fact, about 53% of new EC cases worldwide occur in China. There are two primary histological types of EC, namely squamous cell carcinoma (ESCC) and adenocarcinoma (EAC). Globally, ESCC is the most common type and accounts for about 85% of the EC cases [3]. In China, more than 90% of EC tumors are ESCC. Despite advances in treatment strategies, the prognosis of ESCC is not satisfying and the 5-year survival rate remains below 20%. As poor prognosis of ESCC is mainly due to late diagnosis and tumor heterogeneity [4], patient stratification for personalized therapy and early detection are very important.

Recent advancements in single-cell RNA sequencing (scRNA-seq) technologies have revolutionized our understanding of tumor heterogeneity and provided valuable insights into cancer biology. The advantages of scRNA-seq are evident, particularly in enabling the identification of distinct subpopulations and molecular profiles that may drive tumor progression and drug resistance [5]. In hepatocellular carcinoma (HCC), scRNA-seq analysis in rat model revealed the promoting roles of $GPNMB^{+}$ and $Gal\text{-}3^{+}$ hepatic parenchymal cells in HCC immunosuppression and tumorigenesis [6]. In colorectal cancer, scRNA-seq analysis uncovered the activation of mast cells within the tumor microenvironment [7]. The heterogeneity of ESCC has been confirmed in many studies. In fact, a recent study using scRNA-seq analysis revealed significant differences in the densities of tumor-infiltrating immune cells (such as T cells and dendritic cells) and their associations with tertiary lymphoid structures in ESCC [8]. The cell-cell interactions between malignant and non-malignant tissues in ESCC were found to be distinct [9] and the epithelial cells were shown to activate fibroblasts through downregulation of ANXA1-FPR2 interactions which in turn promote ESCC development [10]. The complexity of cell subpopulations and their communications in ESCC highlights the necessity for in-depth exploration of the cell compositions and their roles during ESCC occurrence and progression.

At present, many existing studies on prognostic biomarkers for ESCC mainly focus on analyzing certain biological processes or metabolic pathways [11, 12]. However, the impact of the intra-tumoral cellular heterogeneity of ESCC is not yet fully understood. In this study, to capture the heterogeneity of individual cell populations while maintaining scalability and cost-effectiveness, we integrated scRNA-seq and bulk RNA-seq analyses to comprehensively characterize ESCC. Using the cell types and their marker genes deduced from scRNA-seq data of ESCC, we estimated the abundances of these cell types with CIBERSORT (Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts) method. This analysis identified four cell-type-based ESCC clusters with different prognoses in the bulk RNA-seq data. Based on the gene expression profiles of the four ESCC clusters, we constructed a four-gene prognostic signature model for ESCC. The prognostic effects and dysregulations of the four genes were also validated at protein level. Additionally, their associations with anti-cancer drug sensitivities in ESCC cell lines were investigated. Based on the significant correlations of PKP1(one of the four genes) expression and the half maximal inhibitory concentrations (IC50s) of several EGFR inhibitors in ESCC cell lines, we speculated that there might be associations between PKP1 and EGFR in ESCC. In proteomics analysis and multiplex immunohistochemistry (mIHC) analysis, we found the co-expression of PKP1 and EGFR. Pan-cancer analysis was also performed to investigate the correlation of PKP1 expression with EGFR expression and multiple gene effect scores. We hope that the results will provide new directions for ESCC prognostic prediction, diagnosis, and personalized therapy.

Materials and methods

Cell type annotation and gene expression analysis in ESCC scRNA-seq data

A total of 64 samples (tumor: $n = 60$, adjacent normal, $n = 4$) in GSE160269 [13] were included for scRNA-seq analysis. The clinical features of the patients were shown in Supplementary Table S1. The cells with fewer expressed genes ($n \leq 300$) and the genes expressed in fewer cells ($n \leq 5$) were excluded. With "Seurat" R package, low quality cells with mitochondrial gene counts proportion > = 20% were also removed. After quality control, a total of 208,659 cells were included for further analysis. To depict the cell type compositions of the samples, cluster analysis of single-cell transcriptome was performed

using "FindNeighbors" and "FindClusters" functions with dimensions of 10 and a resolution of 1.0. With the cell markers obtained from CellMarker (http://xteam.xbio.top/CellMarker/) and previous studies [14–16], the cells in different clusters were annotated into different cell types. The results were visualized by the uniform manifold approximation and projection (UMAP) method which projected the cells into two dimensions for visualization. Marker genes were visualized for their expressions in different cell types using "dotplot" function. The proportions of different epithelial cell subtypes in ESCC and normal esophageal tissues were calculated and presented with heatmap plot. The representative maker genes (100 genes for each cell type) of different cell types were deduced with "cosg" R package. For further CIBERSORT analysis [17] in ESCC bulk-seq dataset, a gene expression matrix of different cell types with their maker genes was constructed to be a background for the cell type evaluations.

Bulk RNA-seq analysis of the cell type compositions with single-cell makers in ESCC dataset

CIBERSORT algorithm was used to estimate the relative abundance of cell types in the ESCC bulk RNA-seq dataset GSE53625 [18], based on the cell type signature matrix deduced from GSE160269 dataset. The clinical features of the 179 patients were shown in Supplementary Table S1. To estimate the potential effects of age, sex, tumor grade, tumor stage, tobacco-use, and alcohol-use on ESCC heterogeneity, the cell type compositions among the samples with different clinical characters were compared. To investigate the heterogeneity of the cell type compositions in different tumor samples, non-negative matrix factorization (NMF) was performed to stratify the ESCC patients into different clusters. Kaplan-Meier survival analysis was used to present the prognostic differences among the ESCC clusters. Age, sex, tumor grade, tumor stage, tumor location, alcohol-use, and tobacco- use were also considered to analyze the independent prognostic effects of ESCC clustering on the patients' survival.

Table 1 The primers for the genes

Gene	Forward primers (5′-3′)	Reverse primers (5′-3′)
CCND1	GCTGCGAAGTGGAAACCATC	CCTCCTTCTGCACACATTTGAA
PKP1	CCACTCCAATCGAGGTTCCAT	AGGGTTCCATTGTAGATCGGATA
JUP	TCTCCAACCTGACATGCAACA	CATAGTTGAGACGCACAGAGTTC
ANKRD12	AGGACAAGCTATATTCGCATCAC	TCCATCAAGTGCTTTTCTTTGCT
GADPH	CGGAGTCAACGGATTTGGTCGTAT	AGCCTTCTCCATGGTGGTGAAGAC

Key maker gene identification and risk model construction for ESCC

The representative features (cell types) of the clusters were extracted. Univariable Cox regression analysis was used to investigate the age- and gender-adjusted prognostic roles of marker genes of the representative cell types. The marker genes with $p<0.05$ were then applied to LASSO regression analysis to identify independent gene makers and construct risk model for ESCC overall survival (OS). Receiver Operating Curve (ROC) analysis was used to evaluate the efficiency of the risk model in discriminating of the OS status of the patients.

According to the risk model, the risk score of each patient was calculated, and the ESCC patients were divided into high- and low-risk groups using the median risk score as the threshold. The difference in OS between the high- and low-risk patients was visualized using Kaplan-Meier survival analysis with log-rank test. The independence of the risk model, the prognostic effects of the risk score were also investigated in different ESCC subgroups according to their clinical features (tumor stage, tumor grade, age, sex, alcohol-use, tobacco-use, tumor location, and adjuvant therapy).

The immunoregulatory genes were obtained from the Cancer Immunome Atlas (TCIA) (https://www.tcia.at) and the correlations of the key genes with the immune regulators were investigated.

Key gene expressions in ESCC cell lines by reverse transcription-qPCR (RT-qPCR)

RT-qPCR has been used to investigate gene expression levels in many studies [19, 20]. In this study, the relative expressions of the key genes were investigated in three ESCC cell lines (KYSE150, Eca109, and TE1) and the normal esophageal cell line HEEC. The four cell lines were obtained from Procell Life Science & Technology Co., Ltd (China). They were maintained in a 5% CO_2 incubator (Thermo Fisher, USA) at 37 °C in T25 culture flasks. According to the manufacturer's instructions, the total RNA was extracted using TRIzol (LEAGENE, China). With the RevertAid First Strand cDNA Synthesis Kits (Novoprotein), the extracted RNA was reverse transcribed. The primers for the genes were shown in Table 1. The relative expressions of the genes were calculated by the $2^{-\Delta\Delta Ct}$ method and normalized by GADPH. The experiments were performed in triplicate. For the gene expression comparisons, T-test were used and $p<0.05$ was considered significant.

Validation of the key gene expression dysregulations and their prognostic effects at protein level

The expressions of the key maker genes were then investigated for their dysregulations in ESCC at protein level with the proteomic and phosphoproteomic data of 124

ESCC tumors and their paired adjacent non-tumor tissues in *Liu* study [21] (clinical features were shown in Supplementary Table S1). The expressions of the proteins and their phosphorylation levels (of different site) were compared. The prognostic effects of the proteins were validated.

Further exploration of associations of the key gene expressions with anti-cancer drug sensitivities of ESCC cell lines

The gene expression profiles in ESCC cell lines were download from Depmap portal (https://depmap.org/portal/) and the half maximal inhibitory concentrations (IC50s) of the anti-cancer drugs in Cancer Cell Line Encyclopedia (CCLE) cell lines were used (CCLE_NP24.2009_Drug_data_2015.02.24). A total of 24 anti-cancer drug data in 15 ESCC cell lines (with the key gene expressions) were included for analyses. The Spearman correlations of the anti-cancer drug IC50s with the key gene expressions were estimated.

Multiplex immunohistochemistry (mIHC) analysis of PKP1 and EGFR in EC

The EC tissues from 48 EC patients were obtained from Henan People's Hospital of Henan Province, with the consent of the patients and the approval of the hospital's ethics committee (approval ID: 2019-48). The inclusion criteria for patients are as follows: aged 18 years or older; clinically diagnosed with esophageal cancer; no history of other malignant tumors; planned to receive standard-of-care treatment; and willing to provide informed consent. Exclusion criteria include: having other primary malignancies within the past 5 years; distant metastasis not amenable to curative treatment; severe comorbidities; pregnancy or lactation; significant psychiatric or cognitive impairment; inability to provide consent; and concurrent participation in another interventional clinical trial. A total of 48 patients were ultimately included, comprising 40 ESCC cases, two esophageal adenocarcinoma and six other esophageal malignancies. The clinical features of the samples were shown in Supplemental Table S1. With these tumor tissues, a tumor tissue microarray was constructed by Shanghai Weiao Biotech Company (Shanghai, China). Two-antibody triple-color mIHC experiment was performed to investigate the expression and location of PKP1 and EGFR in the EC tissues. According to previous studies [22–25], we performed deparaffinization, rehydration, antigen retrieval, cell permeabilization, endogenous enzyme inactivation, blocking, antibody incubation, washing and fluorescence staining on the tissue sections. Fluorescence dye tyramide signal amplification (TSA) (1:100) was used to label the proteins and 4,6-diamidino2-phenylindole dihydrochloride (DAPI, cat: SKU FP1490; Akoya Biosciences) was used to stain nuclei. Finally, the experiment was terminated with stop solution. Notably, the steps from primary antibody binding to fluorescence staining were carried out in two rounds. The first round was the staining of PKP1, and the second round was the staining of EGFR. During the experiments, the PKP1-specific antibody (Cat. 22632-1-AP, Proteintech, USA) and EGFR-specific antibody (Cat. BX50092C3) were diluted into 1:100 for use. Novolink Polymer Detection System (anti-rabbit Poly-HRP-IgG, Leica Biosystems) was used to detect the primary antibodies. Novo-Light TSA620 (red) and TSA520 (green) was used for fluorescent staining of PKP1 and EGFR, respectively.

Visiopharm software (Hoersholm, Denmark), a leading AI-driven precision pathology software was used for mIHC analysis. According previous study [25], the PKP1 intensity and EGFR intensity were evaluated. Generally, the tissue array is initially segmented into single cores based on the "tissuearray" App. Following tissue array segmentation, individual cells were identified based on the presence of DNA (DAPI) and membrane proteins where available. The unsupervised phenotyping algorithm was trained across multiple images resulting in identification of cellular phenotypes. The trained APP were used for division of identified cells into the identified phenotypes. The phenotypes were counted and further visualized. To improve visual clarity, the original images were enhanced with color-coded overlays, which made the location and distribution of the recognized phenotypes more noticeable.

Correlation analysis between PKP1 and EGFR expressions in pan-cancer tissues

The Cancer Genome Atlas (TCGA) database provides an excellent data source for studying the roles of genes in various types of tumors. Currently, many pan-cancer studies have suggested that the abnormal expression of certain genes may be a common feature among multiple types of tumors or a unique characteristic of a specific type of tumor [26–28]. These findings offer valuable directions for research into tumor biomarkers. In this study, the gene expression profiles of pan-cancer tissues in TCGA database and the corresponding basic clinical data were obtained from XENA browser datahub (https://gdc.xenahubs.net). The fragments per kilobase million (FPKM) data was transformed to transcripts per million (TPM). The expressions of PKP1 and EGFR were extracted and their Spearman's correlations in different tumors were estimated.

The correlations of PKP1 expression with gene effect scores of multiple genes in pan-cancer cell lines

Single gene or gene set pan-cancer analysis can reveal the common and distinct molecular mechanisms across

different types of cancers [27, 29]. Here, pan-cancer analysis of PKP1 was also performed. Gene effects (perturbation effects) data of CCLE cell lines in the pan-cancer CRISPR-Cas9 dataset was downloaded from Depmap portal (https://depmap.org/portal/). The gene effect score was evaluated according to the speed change after the knockout of a specific gene [30] and the negative effect scores indicated positive associations of the genes with the growth of the cell lines. Usually, effect scores less than −0.5 represented depletion in most cell lines, while less than −1 indicated strong killing effects of the gene knockout. As plenty of genes have different effects or effects of different level in different cell lines, it is essential to find the indicators of their gene effects in different samples. Here, to uncover the indicative potential of PKP1 expression for the potential gene targets in pan-cancer, the gene effect score of the genes with pronounced genetic effect variations (range > 2, $n = 929$) were investigated for their correlations with PKP1 expression CCLE cell lines ($n = 1005$).

Statistical analysis

All the analyses were performed with R 4.3.2 software. Seurat package (V4.4.0) was used for scRNA-seq analysis. The comparisons between two groups were conducted using the Wilcoxon test, while comparisons among multiple groups were conducted using the Kruskal-Wallis test. Cox regression, LASSO regression and Kaplan-Meier survival analysis with log rank test were used for survival analysis. Spearman correlation analysis was used to estimate the correlations between two variables. For these analyses, $p < 0.05$ was considered significant.

Results

Cell type identification and their heterogeneity in ESCC through scRNA-seq analysis

After dimensionality reduction and clustering analyses, 37 original cell clusters were identified in ESCC and the para-normal tissues (Fig. 1A). With the cell marker above and the representative genes of the cells, these clusters were classified into 18 cell types/subtypes including four epithelial subtypes and two fibroblast subtypes (Fig. 1B). The gene expression profiles were shown in Fig. 1C. When comparing the cell type compositions between ESCC and their paired normal tissues, significant differences were shown (Supplementary Figure S1). Interestingly, it was found that the epithelial_1 cells accounted for more than a half of the epithelial cells in the normal tissues while the epithelial_4 cells were uniquely presented in the tumor tissues (Fig. 1D). For fibroblast cells, their major types in tumor and normal tissues were also different. Obviously, fibroblast_1 accounted for almost 80% of the fibroblast cells in the tumor tissues, while fibroblast_2 was the major type in the normal tissues (Fig. 1D). The proportions of all the cell types were also investigated and significant heterogeneity among the tumor tissues was also presented (Fig. 1E).

Through "COSG" analysis, the marker genes of the 18 cell types (100 genes for each cell type) were identified (Supplementary Table S2) and the heatmap of the top-3 gene expressions of the cell types/subtypes were shown in Supplementary Figure S2.

Relative abundances of the cell types in ESCC bulk RNA-seq dataset

Using the cell type signature matrix derived from the scRNA-seq data, we estimated the abundances of cell types in the GSE56325 samples through CIBERSORT analysis. The heterogeneity in cell proportions among the ESCC samples was evident (Fig. 2A). As shown in Fig. 2B, the abundance of proliferating NK cells (CC_NK cells) was higher, while the abundances of epithelial_1 cells and epithelial_2 cells were lower in high-grade (grade III) tumors compared to low-grade (grade I) tumors ($p < 0.05$), indicating their associations with ESCC differentiation. For their relations to ESCC stage, only epithelial_1 cells presented significant difference among ESCC samples of different stages and their lower abundance was shown in later stage (stage II/III) tumors than early stage (stage I) ones ($p < 0.05$, Fig. 2C).

The associations of age, sex, alcohol-use and tobacco-use with the cell abundances were also investigated. As shown in Supplementary Figure S3A, none of the cell types presented abundance difference between the young (age < = 60y) and the old (age > 60y) samples ($p > 0.05$). In Supplementary Figure S3B, a higher level of epithelial_2 cells was shown in male patients while no significance was shown for the other cell types. Interestingly, compared to ESCC samples without a tobacco-use history, B cells ($p < 0.05$) and regulatory T cells ($p < 0.05$) were less enriched, while CC_NK/T cells were more enriched in ESCC samples with a tobacco-use history (Supplementary Figure S3C). Similarly, in ESCC samples with an alcohol-use history, B cells and regulatory T cells were less enriched, while CC_NK/T cells were more enriched compared to samples without an alcohol-use history (Supplementary Figure S3D). Besides, alcohol-use also presented potential impacts on the abundances of epithelial_2 cells, fibroblast_1 cells, fibroblast_2 cells, pericytes, and plasma cells (Supplementary Figure S3D), indicating its intimate relations to ESCC microenvironment.

NMF clustering of ESCC samples and its associations with ESCC prognosis

Based on the decline of the cophenetic coefficients and the RSS curve in the NMF analysis, the optimal number of clusters was determined to be four (k = 4) (Fig. 3A). Consequently, the ESCC samples were categorized into

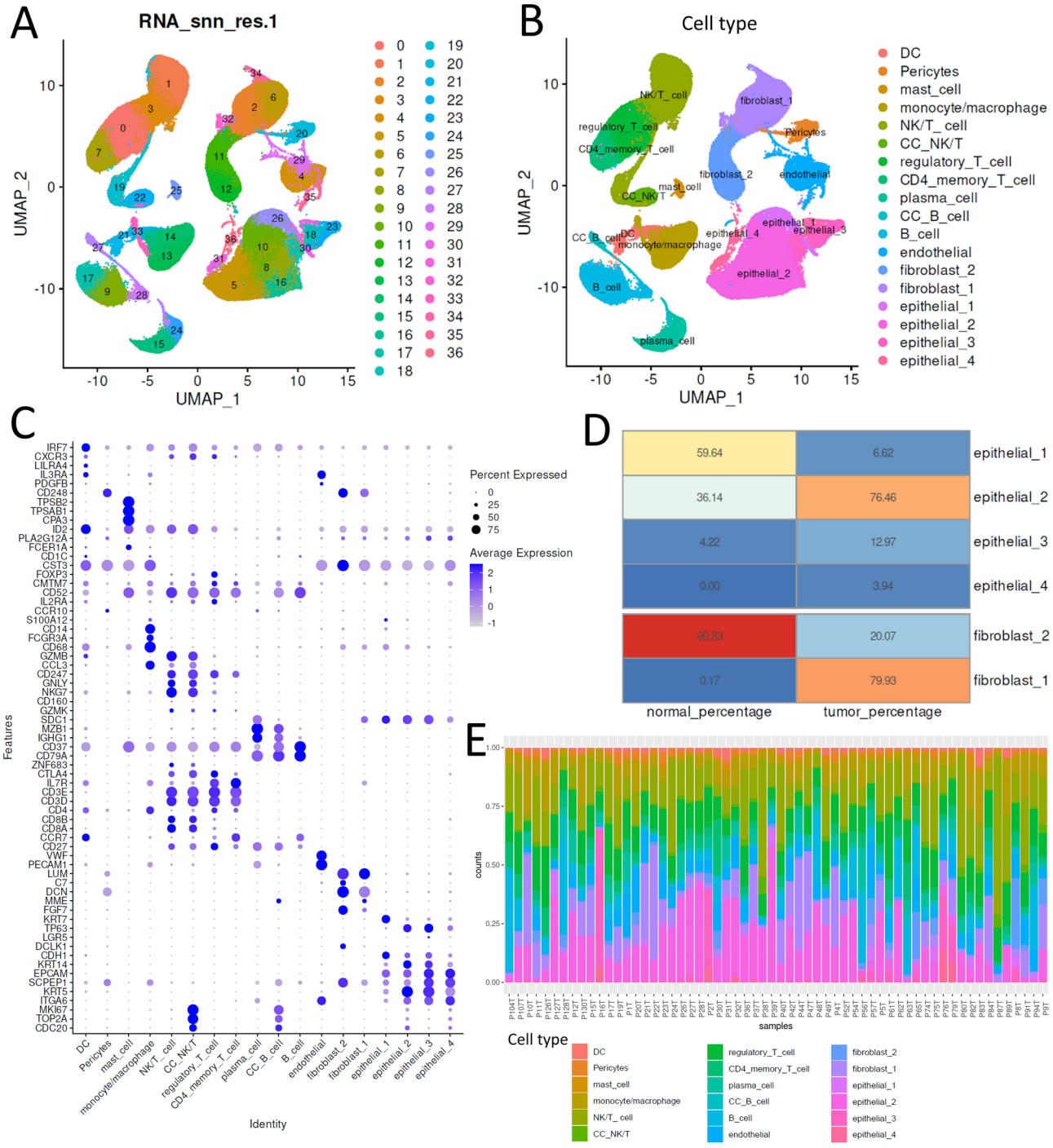

Fig. 1 Cell type annotation of ESCC single-cell samples. (**A**) UMAP plots showing the 37 cell clusters identified among a total of 64 study participants, including ESCC (n = 60) and normal tissues (n = 4). (**B**) The cell clusters were annotated as 18 cell type/subtypes with known gene signatures. (**C**) Expression profiles of marker genes for the different cell types/subtypes. (**D**) Proportions of different epithelial and fibroblast subtypes within the epithelial cells and fibroblasts, respectively. (**E**) The proportions of all the cell types in different ESCC samples. For the cell type annotations, CC_ represented cell cycle-related/ proliferating cells. "Seurat" package (version 4.4.0) was used for single-cell analysis

four groups (C1, C2, C3, and C4) for further analysis (Fig. 3B). Kaplan-Meier analysis revealed significant differences in overall survival (OS) among these clusters, with C1 patients having the shortest OS and C4 patients having the longest OS (Fig. 3C). Notably, the prognostic impact of ESCC clustering on OS was more pronounced in younger patients (age ≤ 60 years) (Fig. 3D) and male patients (Fig. 3E) compared to older patients (age > 60

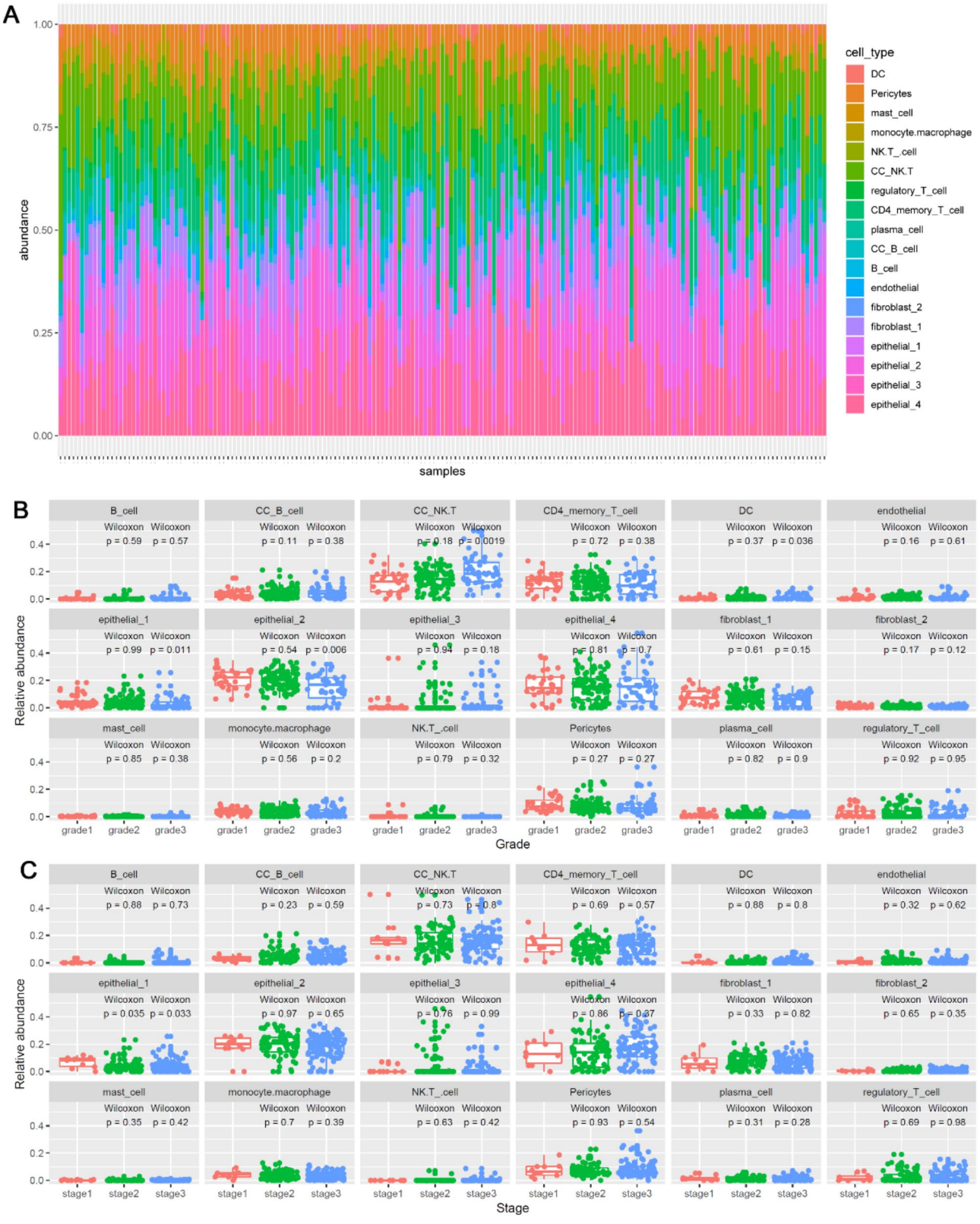

Fig. 2 Cell type compositions of ESCC samples and their difference across different ESCC clinical features. (**A**) Cell type compositions of ESCC samples in GSE53625 dataset. (**B**) Differences in cell type abundance between ESCC samples of different tumor grades. (**C**) Differences in cell type abundance between ESCC samples of different tumor stages

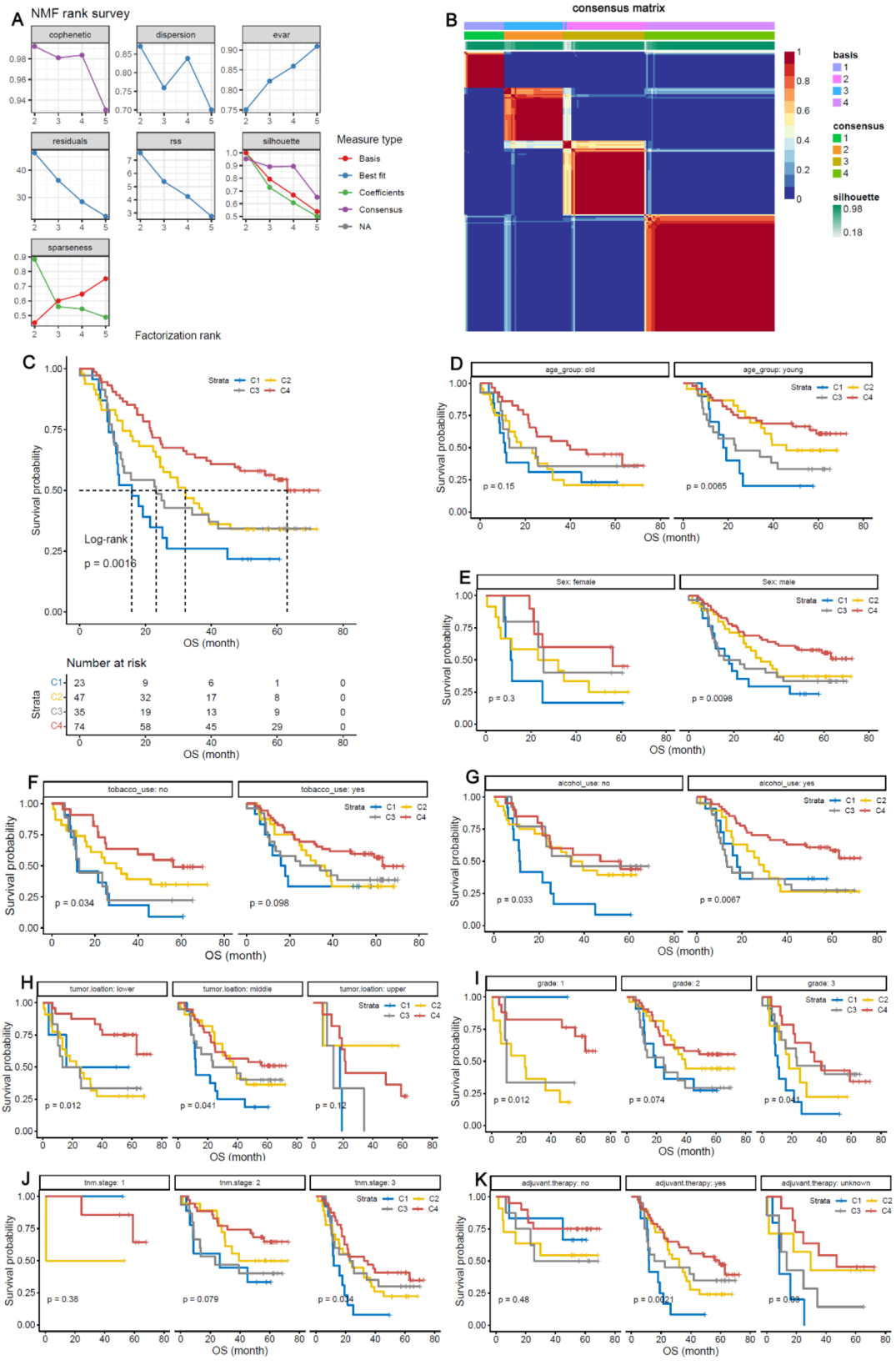

Fig. 3 Single-cell proportion-based NMF clustering of 159 ESCC samples from GSE53625. (**A**-**B**) NMF analysis identified four distinct clusters (C1-C4). (**C**) Significant OS differences were observed among patients in different clusters. (**D**-**E**) OS comparisons across age groups (**D**) and between genders (**E**). (**F**) OS comparisons in patients with and without tobacco use. (**G**) OS comparisons in patients with and without alcohol use. (**H**-**K**) OS comparisons across varying tumor locations, grades, stages, and adjuvant therapy statuses. OS, overall survival

years) and female patients. Significant OS differences were also observed in patients without a tobacco-use history (Fig. 3F), while a trend toward significance ($p < 0.1$) was seen in those with a tobacco-use history (Fig. 3F). Similarly, ESCC clustering had a significant prognostic effect in patients with and without alcohol use (Fig. 3G), as well as in those with tumors located in the lower and middle regions (Fig. 3H). In patients with grade I and III tumors, the prognostic effects were significant, while a trend toward significance was observed in grade II patients (Fig. 3I). Significant OS differences were also found among clusters in late-stage patients, although no significant difference was seen in stage-I patients. However, a trend ($p < 0.1$) was observed in stage-II patients and a significant difference ($p < 0.05$) in stage-III patients (Fig. 3J). Even in patients receiving adjuvant therapy, ESCC clustering retained its prognostic significance (Fig. 3K). Overall, these findings demonstrate the robust and independent prognostic value of ESCC clustering across various patient subgroups.

ESCC cluster-based gene risk model for ESCC OS

The five representative cell types (epithelial_2, epithelial_3, epithelial_4, CC_NK/T, and CD4_memory_T_cell) were identified via NMF clustering, and their marker genes were subjected to Cox regression analysis. After correcting for age and sex, 17 marker genes were found to have prognostic effects on ESCC OS (Supplementary Table S3). Through LASSO regression analyses, an eight-gene model and a four-gene model were constructed (Fig. 4A). As shown in Fig. 4B, the two models could predict the survival status of the patients effectively, with

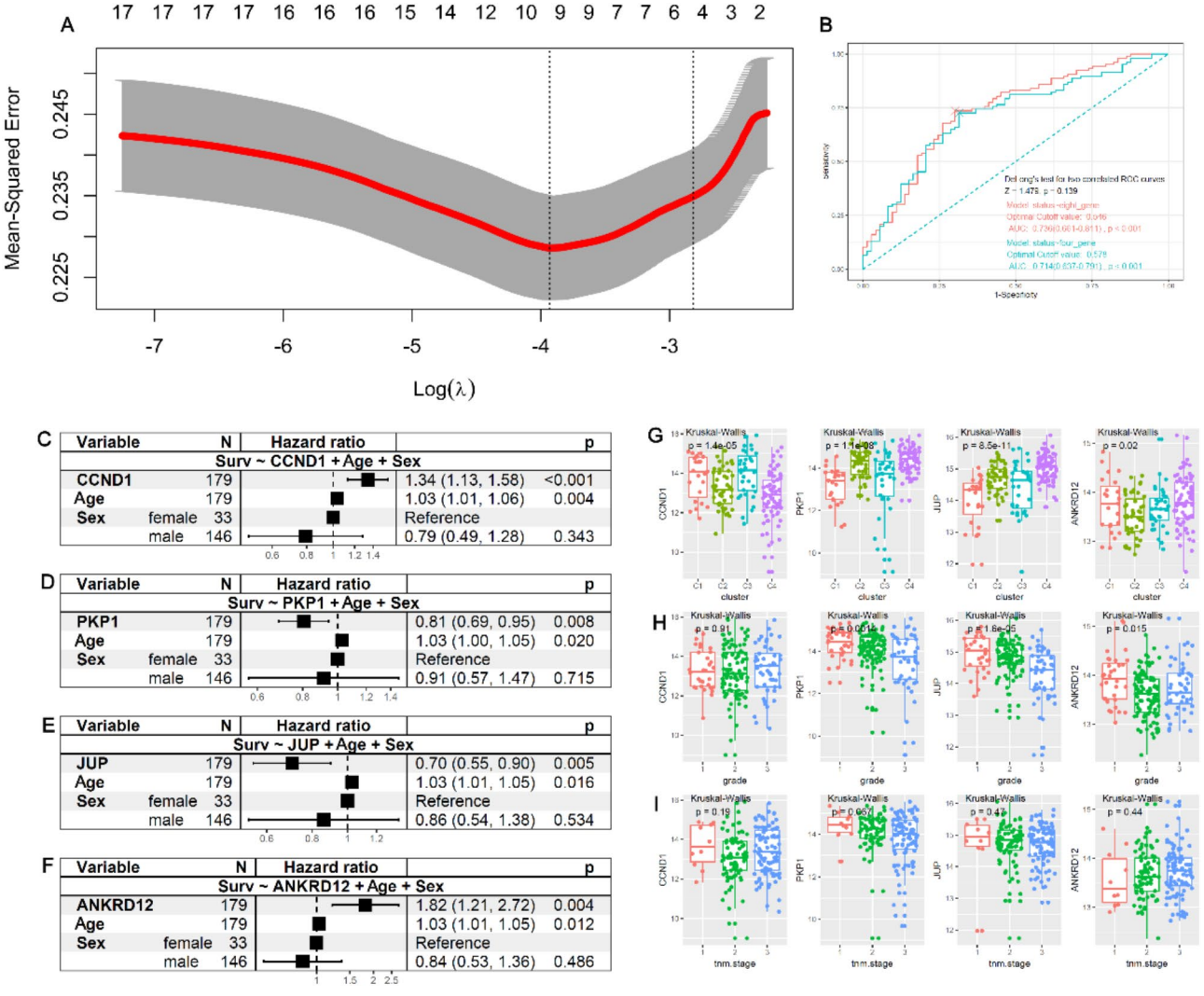

Fig. 4 Key genes and their prognostic effects and their associations with ESCC grade and stage. (**A**-**B**) The four-gene risk model and its efficiency in predicting ESCC status. (**C**-**F**) Independent prognostic effects of the four key genes. (**G**-**H**) Expression differences of CCND1, PKP1, JUP, and ANKRD12 among different ESCC clusters and ESCC samples of different grades. (**I**) Expression comparisons of CCND1, PKP1, JUP, and ANKRD12 among ESCC samples of different tumor stages

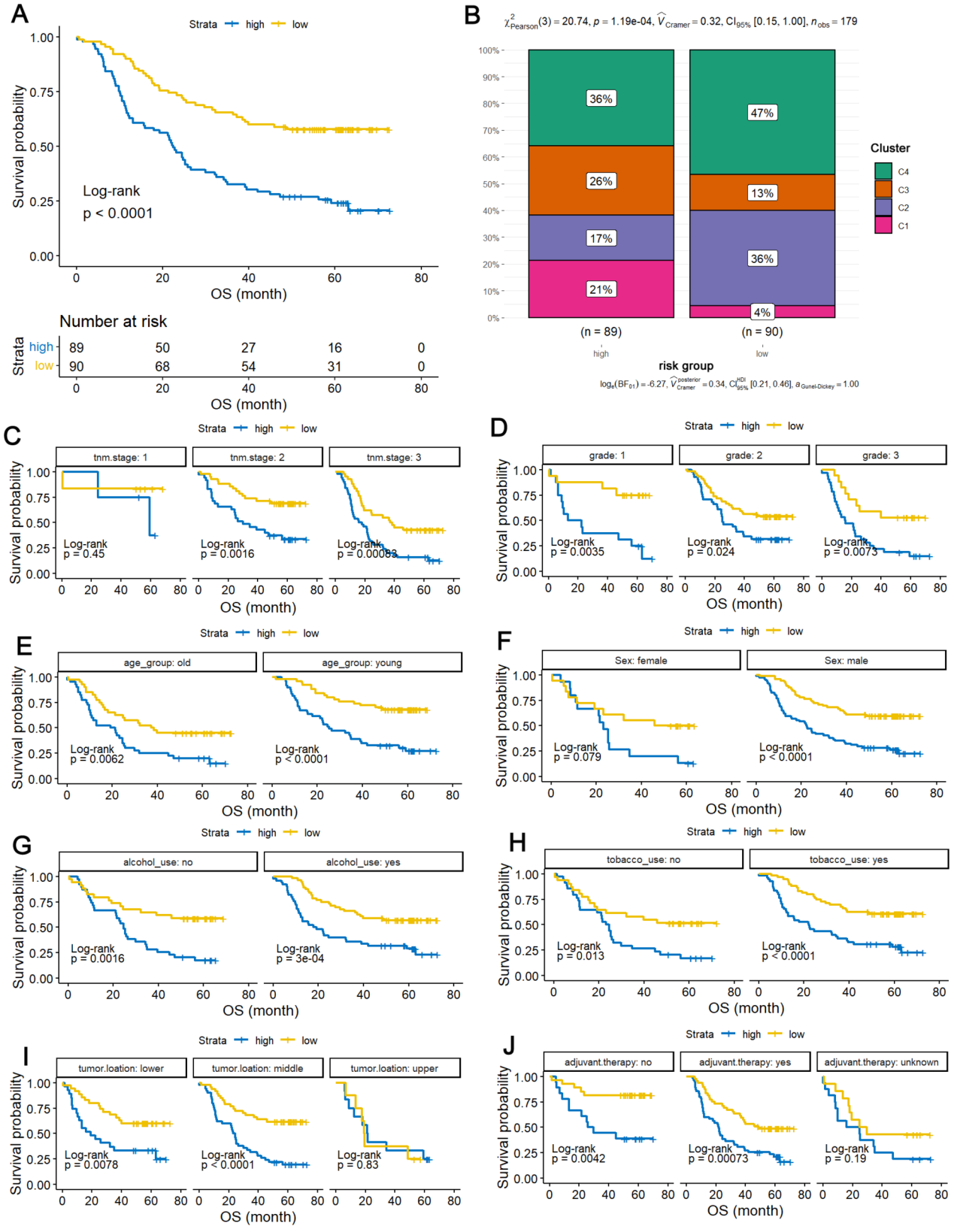

Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 Prognostic effects of risk score in ESCC patients. (**A**) Significant differences in OS between high- and low-risk groups. (**B**) Differences in ESCC cluster composition between high- and low-risk patients. (**C**-**D**) OS differences between high- and low-risk ESCC patients across tumor stages (**C**) and grades (**D**). (**E**-**F**) OS differences between high- and low-risk ESCC patients across age groups (**E**) and genders (**F**). (**G**-**H**) The OS difference between high- and low-risk ESCC patients with different alcohol (**G**) and tobacco (**H**) use histories. (**I**) OS differences among ESCC patients with different tumor locations (**J**) OS difference among ESCC patients with different adjuvant therapy histories. Kaplan-Meier survival analysis with log-rank test and Chi-square test were used and $p < 0.05$ was considered significant

AUCs of 0.735 and 0.714, respectively. While, the two risk models presented no significant difference between their effectiveness ($p > 0.05$). Given this equivalence in performance, we opted for the four-gene model due to its simplicity and parsimony. Generally, a more concise gene signature can be advantageous in practical applications, such as clinical settings, where it may be easier to implement and interpret. Additionally, a smaller number of genes can potentially reduce the complexity and cost associated with subsequent validation studies and clinical use. Therefore, the genes in the four-gene model were focused for further analysis The age and sex-corrected prognostic effects of the four genes were visualized in Fig. 4C-F, with significant differences between HCC clusters depicted in Fig. 4G. As shown in Fig. 4H, three of the four genes (PKP1, JUP, and ANKRD12) were differentially expressed across ESCC samples of varying grades, indicating their associations with ESCC differentiation. However, no significant differences were observed among ESCC samples of different stages (Fig. 4I). The prognostic relevance and differential expression patterns of the four genes suggest their potential utility in ESCC differentiation and clinical applications.

According to the coefficients of the genes in the model, the risk scores of the ESCC patients were estimated with the following formular:

$$\begin{aligned} Risk\ score\ =\ & 0.0350 * CCND1\ -\ 0.0188 * PKP1 \\ & -\ 0.0186 * JUP\ +\ 0.0998 * ANKRD12 \end{aligned}$$

With the median risk score as the threshold, the samples were separated into high- and low-risk groups. As shown in Fig. 5A-B, compared with patients in the low-risk group, those in the high-risk group exhibited significantly shorter survival. Additionally, there were more C1 patients (characterized by longer OS) and fewer C4 patients (characterized by shorter OS) among the high-risk group. The prognostic effects of the four-gene risk model were also evident across different ESCC subgroups, based on various clinical features of the patients (Fig. 5C-J). This suggests that the model has the potential to serve as an independent prognostic indicator for ESCC, regardless of tumor stage, tumor grade, age, sex, alcohol and smoking history, tumor location, and adjuvant therapy history.

There were also significant associations of ESCC risk score with immunoregulator expressions. As shown in Supplementary Table S4, eleven immunostimulators and five immunoinhibitors presented significant positive or negative correlations with ESCC risk score, respectively. Interestingly, most of the MHC-risk correlations were positive, indicating the intimate relations of MHC regulations to ESCC prognosis.

Expression dysregulation and heterogeneity of the key genes in ESCC cell lines

As shown in Fig. 6A, CCND1 was higher expressed in the three ESCC cell lines (KYSE150, TE1, and Eca109) than the normal esophageal cell line HEEC. Obvious heterogeneity of CCND1 expression in ESCC cell lines was shown, with highest expression in Eca109 cells while lowest expression in KYSE150 cells. PKP1 was also heterogeneously expressed in ESCC cell lines (Fig. 6B). In contrast to the higher expression of PKP1 in KYSE150 cells, a lower expression trend of PKP1 in TE1 cells than HEEC cells was presented. However, there was no significant difference in PKP1 expression between HEEC cells and TE1 and Eca109 cells. JUP was downregulated, while ANKRD12 was upregulated in ESCC cell lines relative to HEEC cells (Fig. 6C and D). These findings indicate that the dysregulation and heterogeneity of key genes in ESCC cell lines mirror the patterns observed in ESCC tumors through scRNA-seq and bulk RNA-seq analyses, highlighting their potential roles in ESCC development.

Prognostic effects and dysregulations and of the key genes at protein level

At the protein level, CCND1 exhibited unfavorable prognostic effects on ESCC OS (Fig. 7A), while PKP1 (Fig. 7B) and JUP (Fig. 7C) demonstrated favorable prognostic effects, corroborating the mRNA-level results. However, no significant prognostic effects of ANKRD12 protein were shown (Fig. 7D). As shown in Fig. 7E-H, three of the four proteins presented significant differences among ESCC tumors of different grades. Lower levels of PKP1 (Fig. 7F) and JUP (Fig. 7G) in lower-grade ESCC tumors indicate their positive association with ESCC differentiation, whereas higher levels of ANKRD12 (Fig. 7H) in higher-grade tumors suggest a negative correlation with differentiation. Although CCND1 expression did not significantly differ between ESCC tumors and normal tissues, its heterogeneity among ESCC samples was evident, with both the highest and lowest levels detected within ESCC samples (Fig. 7I). For PKP1 and JUP, they

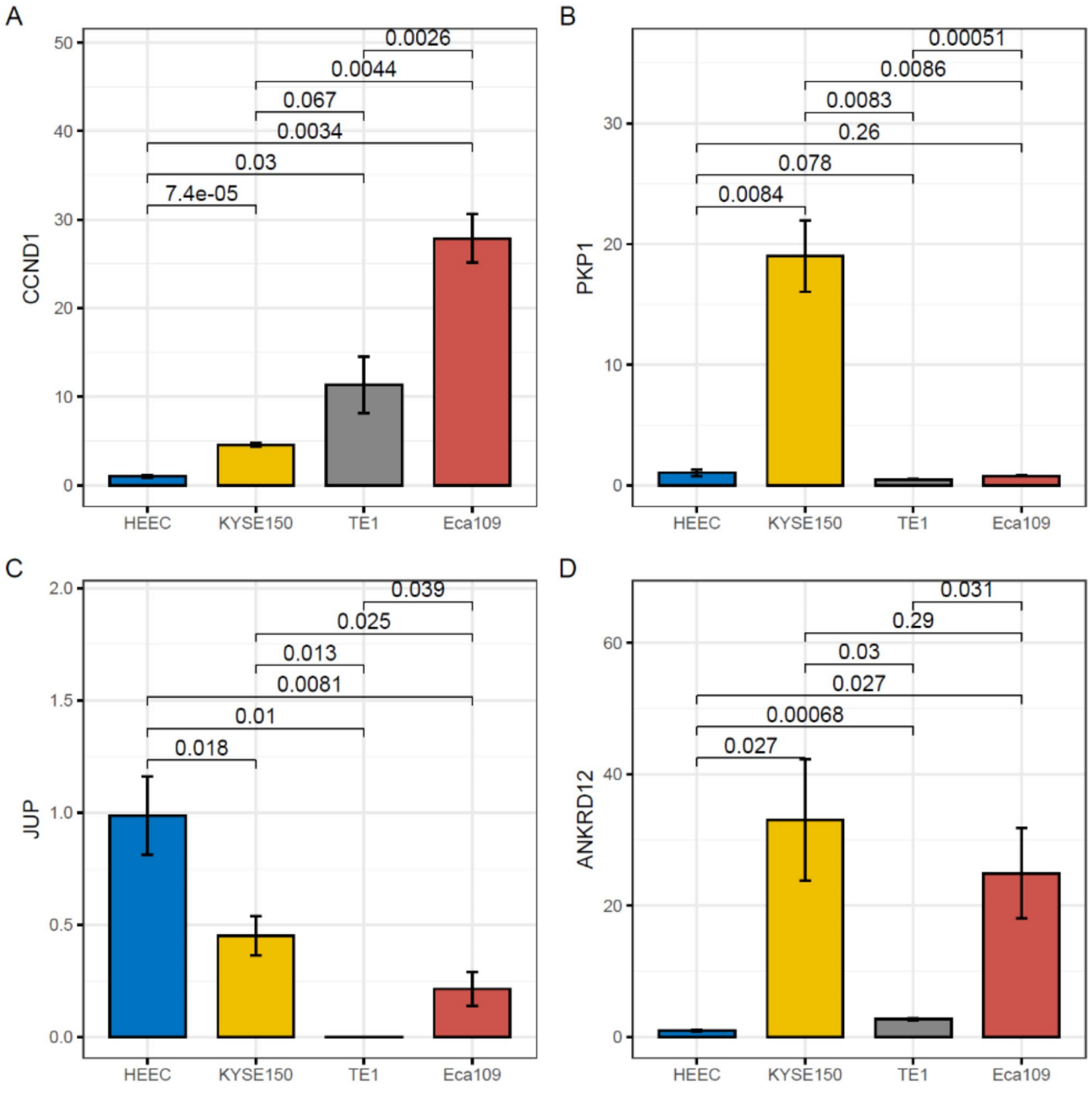

Fig. 6 RT-qPCR analysis of the key gene expressions in different ESCC cell lines. (**A**) Higher CCND1 expression and its heterogeneity in ESCC cell lines (KYSE150, TE1, and Eca109). (**B**) Comparisons of PKP1 expression between three ESCC cell lines (KYSE150, TE1, and Eca109) and the normal esophageal cell line HEEC. (**C**) Lower JUP expression and its heterogeneity in ESCC cell lines. (**D**) Higher ANKRD12 expression and its heterogeneity in ESCC cell lines. Wilcoxon test was used for the gene expressions between different groups and $p < 0.05$ was considered significant

were significantly lower expressed in the ESCC tumors than their paired normal controls (Fig. 7J-K). However, no significant expressional difference of ANKRD12 was shown between the tumor and normal groups (Fig. 7L). The phosphorylation levels of the four proteins were also investigated. The phosphorylation levels of CCND1 (two sites: CCND1:T286 and CCND1:T288) (Fig. 7M) presented no significant difference between the tumor and normal groups. For PKP1, 77 phosphorylation sites were detected and 80.5% of them (62/77) presented significant lower phosphorylation level in the ESCC tissues than their normal controls (Supplementary Figure S4). Similarly, comparing with the normal controls, among the 34 phosphorylation sites of JUP protein, 17 sites were lower phosphorylated while only one site was higher phosphorylated in ESCC tumors (Supplementary Figure S5). In contrast to the lower phosphorylation of PKP1 and JUP, six of the 24 sites of ANKRD12 were shown to be

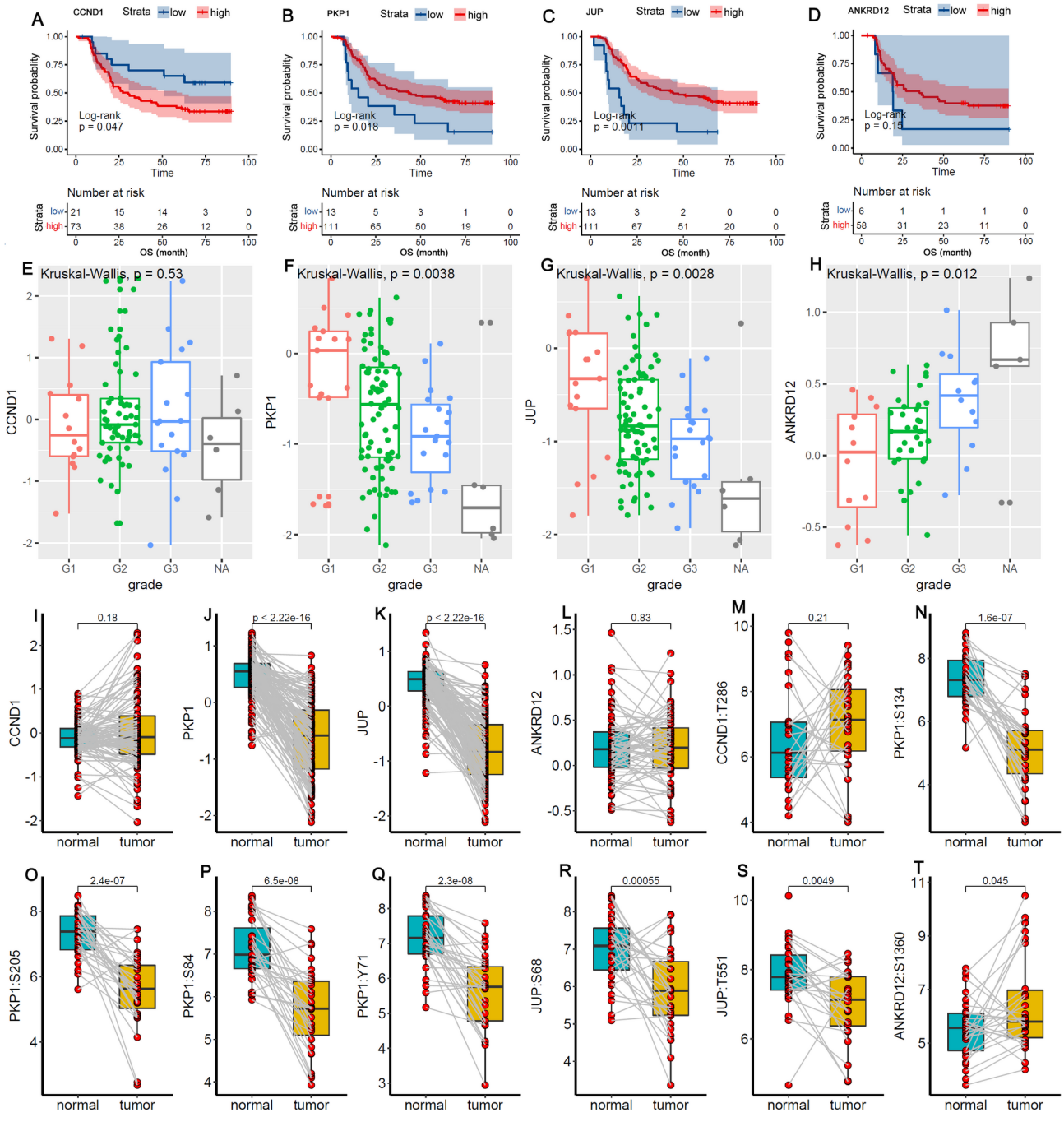

Fig. 7 Prognostic effects and dysregulations of the four key genes at protein level in ESCC. (**A**-**D**) Prognostic effects of CCND1, PKP1, JUP, and ANKRD12. (**E**-**H**) Expression levels of CCND1, PKP1, JUP, and ANKRD12 in ESCC tumors of different grades. (**I**-**L**) Expression levels of CCND1, PKP1, JUP, and ANKRD12 in ESCC tumors vs. paired normal controls. (**M**-**T**) Phosphorylation levels of CCND1, PKP1, JUP, and ANKRD12 at different sites in ESCC tumors vs. paired normal tissues. Kaplan-Meier analysis with log rank test, Kruskal-Wallis test, and Wilcoxon test were used and $p < 0.05$ was considered significant

higher phosphorylated in the ESCC tumors than the normal controls (Supplementary Figure S6). The representative phosphorylation sites with significant differences for PKP1, JUP, and ANKRD12 were shown in Fig. 7N-Q and R-S, and Fig. 7T, respectively. These results underscore the roles of key genes in tumor differentiation and progression, clarify the molecular mechanisms of ESCC, and offer insights for future therapies.

The associations of the four gene expressions with anti-cancer drug sensitivity in ESCC cell lines

The four gene expressions presented significant correlations with at least one of the 24 anti-cancer drug

sensitivities in ESCC cell lines (Supplementary Table S5). There was a significant negative correlation of CCND1 expression with the IC50 of Lapatinib (an EGFR inhibitor) in ESCC cell lines (Fig. 8A). PKP1 was negatively correlated with the IC50s of seven anti-cancer drugs (AZD0530, Erlotinib, Lapatinib, RAF265, TAE684, TKI258, ZD-6474) in ESCC cell lines (Fig. 8B-H). Similarly, there was a significant negative correlation of JUP expression with Topotecan IC50 (Fig. 8I). With regard to ANKRD12, a positive correlation of its expression with IC50 of PF2341066 (c-met inhibitor) was shown (Fig. 8J). The negative correlations suggested that CCND1, PKP1, and JUP expressions were potential positive indicators of the implication of the corresponding anti-cancer drugs in ESCC. While, higher ANKRD12 expression might be a negative indicator for PF2341066 use.

The correlation between PKP1 expression and EGFR expression

The negative correlations between the IC50s of EGFR inhibitors (ZD-6474, Erlotinib, and Lapatinib) and EGFR expression levels in Fig. 8 suggested a positive correlation between EGFR and PKP1. Proteomic analysis of ESCC tissues confirmed this positive correlation ($R = 0.45$, $p < 0.01$, Fig. 9A). mIHC analysis showed that PKP1 was located in the cell plasma and nucleus, while EGFR was predominantly membranous (Fig. 9C). Despite their different locations, the positive correlation between their expression levels was still evident ($R = 0.38$, $p < 0.01$, Fig. 9B-C).

The correlations of PKP1 expression in other tumors were also investigated. As shown in Fig. 9D, positive correlations between PKP1 and EGFR expression were observed in 24 out of 36 tumor types, demonstrating the generalizability of this relationship. In contrast, negative correlations were observed in three tumor types, highlighting the heterogeneity among different tumors.

Significant correlations of PKP1 expression with gene effects of multiple genes in pan-cancer cell lines

Among the 929 genes, 440 exhibited significant correlations between their gene effect scores and PKP1 expression in pan-cancer cell lines (Supplementary Table S6). Specifically, 232 genes showed negative correlations, indicating that higher PKP1 expression was associated with increased lethality of gene knockout. Conversely, 208 genes showed positive correlations, indicating that higher PKP1 expression was associated with decreased lethality of gene knockout. Figure 10A displays the correlations of these genes, as well as their effect scores across various cell lines. The top 20 genes with the strongest negative or positive correlations are labeled in Fig. 10A, and their gene effect scores are visualized in relation to PKP1 expression in Figs. 10B-C. Consistent with the negative correlation between PKP1 expression and the IC50 values of EGFR inhibitors in ESCC cell lines, a similar negative correlation was observed between PKP1 expression and EGFR effect score in pan-cancer cell lines (Fig. 10A-B). These results collectively highlight the potential role of PKP1 in modulating gene knockout lethality and its association with EGFR-related effects across multiple cancer types.

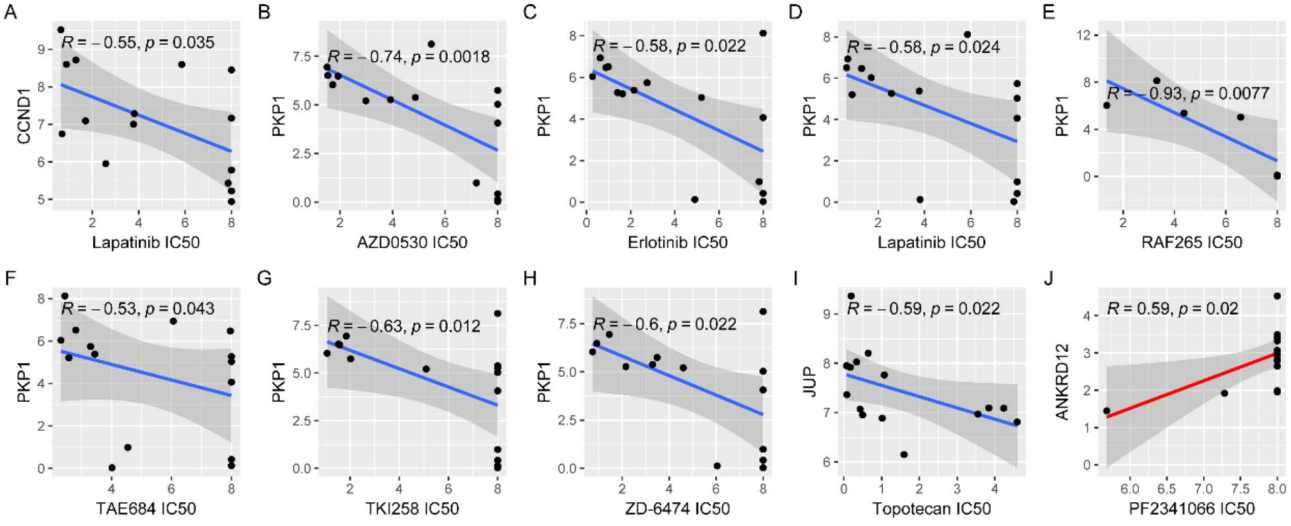

Fig. 8 Correlations of the key gene expressions with anti-cancer drug sensitivities in ESCC cell lines. (**A**) CCND1 expression negatively correlates with Lapatinib IC50. (**B**-**H**) PKP1 expression negatively correlates with IC50 values of AZD0530, Erlotinib, Lapatinib, RAF265, TAE684, TKI258, and ZD-6474. (**I**) JUP expression negatively correlates with Topotecan IC50. (**J**) JUP expression positively correlates with PF2341066 IC50. Spearman correlation analysis was used and $p < 0.05$ was considered significant

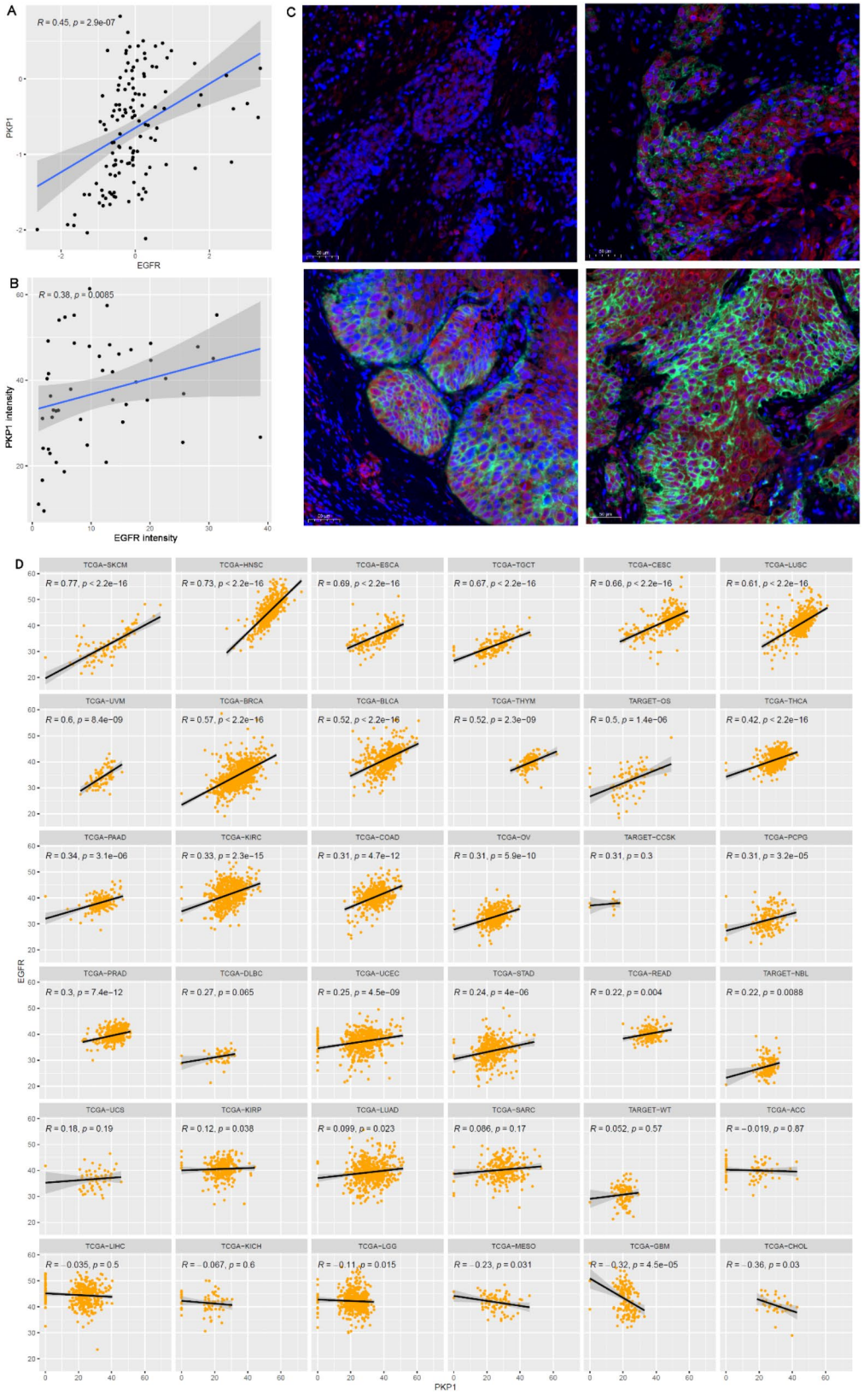

Fig. 9 (See legend on next page.)

(See figure on previous page.)

Fig. 9 The co-expression of PKP1 and EGFR in ESCC and pan-cancer tissues. (**A**) Significant positive correlation between PKP1 and EGFR protein expression in *Liu*-ESCC dataset. (**B**) Significant positive correlation between PKP1 and EGFR expression in ESCC tissues by mIHC analysis. (**C**) The representative images of PKP1 and EGFR expressions in ESCC tissues. (**D**) The correlation of PKP1 expression with EGFR expression in different cancers at mRNA level. In the mIHC image (**C**), red indicated the expression of PKP1 in the cytoplasm, green indicated the expression of EGFR in the cell membrane, and the purple color indicated the expression of PKP1 in the nucleus. Spearman correlation analysis was used and $p < 0.05$ was considered significant

Discussion

Tumor cell heterogeneity could promote tumor progression and therapy resistance [31, 32]. The contributions of inter-tumor heterogeneity to prognosis discrepancy were uncovered in many malignancies. To improve the patients' prognosis, it's crucial to identify the key cells and genes associated with intra- and inter-tumor heterogeneity. In this study, we focused on ESCC primary tumors to uncover the cell composition differences among the tumors and identify the key genes associated with ESCC prognosis. Through the integration of scRNA-seq and bulk RNA-seq analysis, we identified four ESCC clusters with different cell type features and a four-gene signature with independent prognostic effects. The dysregulations and heterogeneity of the four genes were confirmed at mRNA level in ESCC cell lines. At protein and/or phosphorylated protein level, their dysregulations were also shown. Furthermore, their associations with anti-cancer drug sensitivity in ESCC cell lines were also indicated. These results might provide new clues for ESCC study.

In a recent study, an integrative analysis of multi-omics data presented the complex heterogeneity between primary and metastatic ESCC in mRNA/protein levels as well as cell abundances [33]. Here, the cell abundance heterogeneity among the primary ESCC tumors were also shown in the scRNA-seq data and the single cell data derived from the bulk RNA-seq data. Tobacco smoking and alcohol consumption were reported to be important risk factors for ESCC in many studies [34–36]. Compared with existing studies [37, 38], we also took into account the potential impact of smoking, alcohol consumption and tumor location on patient survival. Interestingly, besides the associations of the cell abundances with ESCC grade and stage, we also found the impacts of tobacco-use and alcohol-use on the cell abundances of the cells in this study, indicating the involvements of tobacco-use and alcohol-use in regulating ESCC microenvironment.

NMF clustering was confirmed to effective in tumor stratification [39–41]. In this study, through NMF clustering, we found ESCC patients could be separated into four clusters with different prognosis. The independent prognostic effects of the ESCC clusters indicated the different roles of different cell types during ESCC progression, in consistent with the cell heterogeneity in ESCC reported in previous studies [8, 42]. To simplify the stratification model of the ESCC clusters while retaining its prognostic implications, we analyzed the marker gene expressions of the representative cell types and identified a four-gene prognostic signature. The CCND1-PKP1-JUP-ANKRD12 signature could separate ESCC patients into high- and low-risk groups effectively. Even in ESCC patients of different pathological or clinical features, the OS difference between the high- and low-risk patients also existed. The independence of the prognostic effects of the risk model indicated the important roles of the four genes during ESCC development.

Among the four genes, CCND1 was reported to be a cancer-related gene and dysregulated or mutated in many malignancies [43]. The tumor-promoting effects of CCND1 were shown in colorectal cancer [44], lung cancer [45], and hepatocellular carcinoma [46]. In melanoma, the upregulation of CCND1 was found to be associated the immune exclusion [47]. Here, we found the unfavorable prognostic effects of CCND1 in ESCC both at mRNA and protein level. However, in contrast to the upregulation of CCND1 in other tumors, here, no significant difference of total CCND1 protein level was shown between ESCC tumors and their paired normal tissues. Instead, the obvious heterogeneity of CCND1 and phosphorylated CCND1 among the ESCC samples was presented, highlighting the intimate relations of its inter-tumor heterogeneity on ESCC prognosis. In addition, the significant negative correlation of CCND1 to Lapatinib IC50 suggested that high expression of CCND1 might be an indicator for Lapatinib therapy in ESCC patients.

PKP1 gene encodes Plakophilin 1 (PKP1), a member of the arm-repeat family of catenins. As a structural component of desmosomes, PKP1 could stabilize cell-cell adhesion in epithelial cells [48]. The specific roles of PKP1 in tumor development seems to be complicated. In prostate cancer [49], its anti-proliferative properties were shown and its downregulation was found to be associated with shorter patient survival. In contrast, its overexpression and tumor-promoting roles were shown in breast cancer [50]. While in lung cancer, both oncogenic [51] and tumor-suppressive [52] functions of PKP1 were reported. In this study, the downregulation and favorable prognostic effects of PKP1 was shown in ESCC, in consistent with *Fujita* study [53]. It was found that phosphorylation of PKP1's N-terminal is essential for epidermal differentiation while loss of function PKP1 could impair skin differentiation and enhance epidermal carcinogens [54]. Here, we also found the downregulation of PKP1 and its phosphorylation and the lower expression in higher-grade

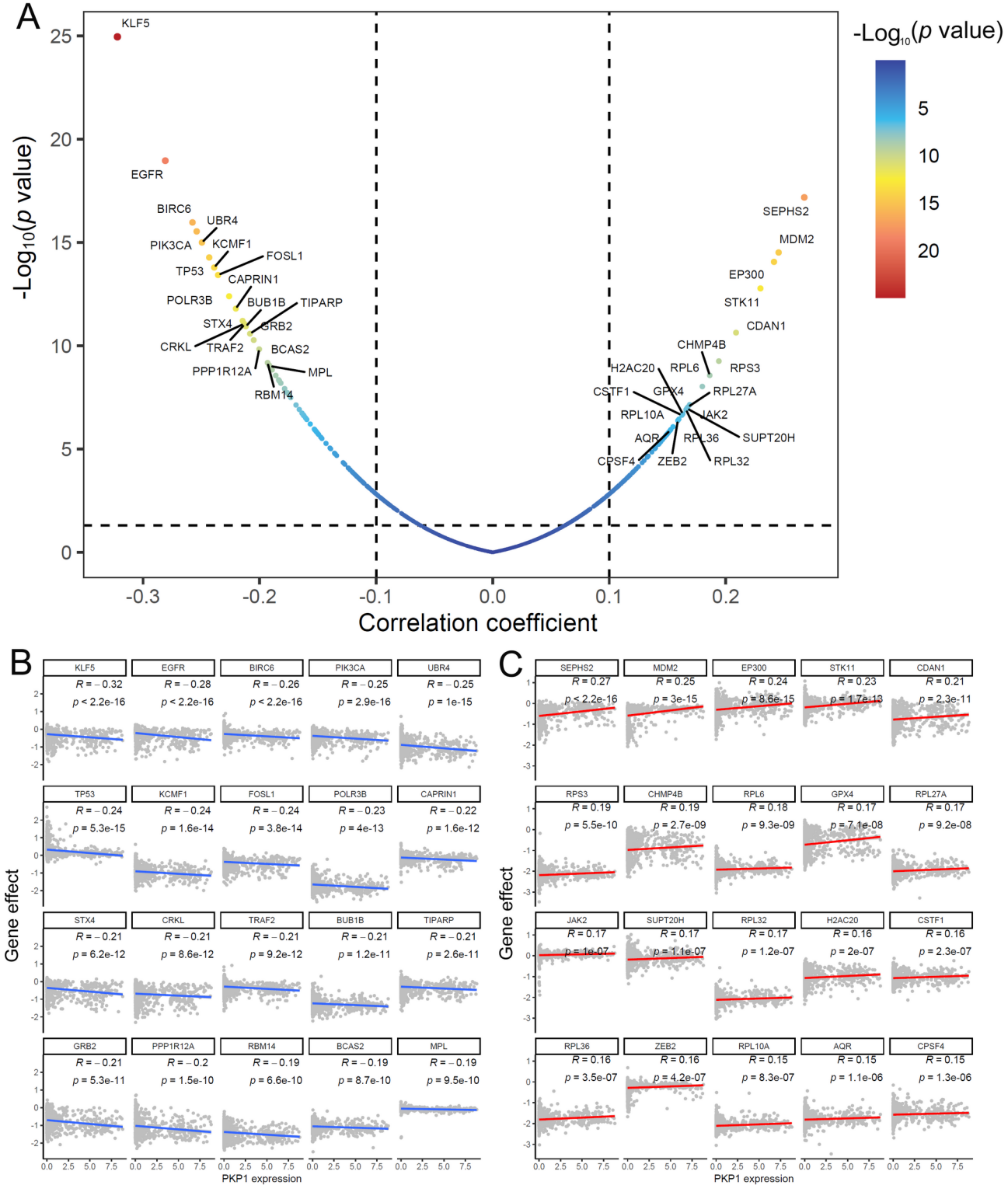

Fig. 10 The correlations of PKP1 expression with multiple gene effect scores in pan-cancer cell lines (**A**-**C**). Spearman correlation analysis was used and $p < 0.05$ was considered significant

ESCC tumors. It was indicated that PKP1 decrease might be an indicator for ESCC and its poor outcome.

In addition, through the associations of PKP1 expression with the sensitivity of ESCC cells to EGFR inhibitors, we uncovered the positive correlations between PKP1 and EGFR expressions in ESCC. TCGA provides an extensive and comprehensive dataset that allows researchers to identify potential cancer-specific therapeutic targets and enhance biomarker discovery. It also offers valuable insights into the molecular characterization of various cancers. In this study, pan-cancer analysis also indicated the positive correlations of the two gene expressions in multiple tumor tissues, indicating that they might be involved in similar processes or share similar regulatory mechanisms. We also investigated the associations of PKP1 expression with the effects of multiple genes and it presented significant correlations with the effect score of multiple genes in pan-cancer cell lines, highlighting its potential as an indicator for the efficacy of targeted gene therapy in the tumors.

While, some studies have highlighted potential biases in TCGA data, such as technical biases from sequencing methods and biological biases like tumor heterogeneity and sample purity [55, 56]. Therefore, multiple validations are important. In this study, the positive correlation between PKP1 and EGFR expression was confirmed in different ESCC cohorts. However, their correlation in other tumors requires further validation.

For JUP, it is also known as gamma-Catenin and can form distinct complexes with cadherins and desmosomal proteins. Both tumor-promoting and anti-tumor functions of JUP have been reported. In prostate cancer, high JUP expression has been reported to be associated with tumor progression and an unfavorable prognosis [57]. In contrast, its tumor suppressor properties were shown in several other tumors. In lung cancer, JUP expression was decreased, and its re-expression could inhibit transformed cell growth [58]. In bladder cancer, the restoration of its expression could suppress the tumorigenic potential and tumor cell migration [59]. In renal cell carcinoma, JUP knockdown could promote tumorigenicity, while its overexpression could suppress it [60]. The antitumor roles of JUP in ESCC and its favorable prognostic effects were also reported [61]. Consistently, in this study, we found the downregulation of JUP and its independent prognostic role in ESCC. Additionally, we noticed lower JUP levels in higher-grade ESCC tumors, downregulation of phosphorylated JUP, and a negative correlation with Topotecan IC50 in ESCC cell lines. These results indicate the close relationship between JUP dysregulation and ESCC progression.

ANKRD12 gene, also named ANCO-1, encodes ankyrin repeat domain 12 (ANKRD12), a member of the ankyrin repeats-containing cofactor family. Several studies, its associations with cognitive function were reported [62–65]. Unlike the three genes above, the clinical significance of ANKRD12 in tumor was less studied. In colorectal cancer, lower ANKRD12 expression (mRNA) was shown to be associated with shorter survival [66]. While, in ESCC, ANKRD12 mRNA was identified to be progression-associated mRNA [67]. Here, we also found the tumor-promoting potential of ANKRD12 in ESCC. Although no significant difference of its total protein level was shown between ESCC and normal controls, its positive correlation with ESCC grade was obvious and its upregulated phosphorylation level in the tumor tissues were presented.

Beyond their individual functions, these genes may also work in concert to regulate the progression of ESCC through several shared biological processes. CCND1 regulates the cell cycle by promoting progression from the G1 to S phase [68]. Silencing of ANKRD12 circRNA, which was derived from Exon 2 and Exon 8 of the ANKRD12 gene, can reduce cell proliferation, induce cell death, and downregulate CCD1 expression [69]. JUP dysregulation can also impact cell cycle progression through MYC signaling pathway activation [70]. In addition, JUP and PKP1 directly interact within the desmosome complex and participate in shared pathways related to desmosome assembly, Hippo signaling, and cell adhesion [71, 72]. These shared processes highlight the potential interplay between these genes in driving ESCC progression.

In this study, we combined the advantages of scRNA-seq, bulk RNA-seq to identify and validate the gene prognostic gene model for ESCC. The cell type compositions and gene expression heterogeneity were also considered. The prognostic effects of the risk model were independent of eight clinical features including tumor stage, tumor grade, age, sex, tumor location, tobacco-use, alcohol-use, and adjuvant therapy. The gene prognostic effects and dysregulations were validated at total protein level and phosphorylated protein level. We also investigate the associations of the key genes with anti-cancer drug sensitivity and their effects in ESCC cell lines. The positive correlation between PKP1 and EGFR expressions and the significant correlations of PKP1 expression with other gene effect scores were also uncovered, indicating the indicator potential of PKP1 expression as gene therapy indicator in cancerous diseases. The work also has limitations. First, in this study, all the samples in this study were from Chinese cohorts and the results need to be validated in the datasets from other regions and/or races. Second, the regulations and the interactions of the genes in ESCC were not clear enough. Further study is needed to clarify their functions and regulations in ESCC in the future. Additionally, the significance of liquid biopsy has been confirmed in several studies [73, 74]. Given the ease and feasibility of detecting relevant

biomarkers in liquid biopsies, it is necessary to examine the protein expression of the four key genes in the blood to assess their diagnostic and prognostic value. This will also be the direction of our future research.

Conclusions

In conclusion, we systematically analyzed the scRNA-seq and bulk RNA-seq data of ESCC and presented the cell proportional heterogeneity among the tumor tissues and their associations with ESCC prognosis. The four-gene signature was indicated as an independent indicator for ESCC prognosis and could effectively classify the ESCC tumors into high- and low-risk groups. The four genes were dysregulated at the protein and/or phosphorylated protein level in ESCC. They might serve as new markers for ESCC prognosis and diagnosis. Their correlations with ESCC grade and anti-cancer drug sensitivity in ESCC cell lines indicate their potential as therapeutic targets or biomarkers in ESCC. The positive correlations between PKP1 and EGFR expression levels suggest their close association with multiple cancers. Additionally, PKP1 expression is correlated with the gene effect scores of many genes across various cancers. These genes may represent potential new targets for gene therapy in multiple cancers, and PKP1 expression could serve as an indicator for targeted gene therapy in these cancers.

Abbreviations

EC	Esophageal cancer
ESCC	Esophageal squamous cell carcinoma
EAC	Esophageal adenocarcinoma
scRNA-seq	Single-cell RNA sequencing
UMAP	Uniform manifold approximation and projection
NMF	Non-negative matrix factorization
OS	Overall survival
CCLE	Cancer Cell Line Encyclopedia
IC50s	The half maximal inhibitory concentrations

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-025-14150-8.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

The authors are very grateful to Prof. Liyan Xu and her team group for sharing the proteomic data.

Author contributions

Conceptualization: Liping Dai, Peng wang and Xiaoli Liu; Methodology: Xiuzhi Zhang, Yutong Zhao, and Zhi Wang; Formal analysis and investigation: Xiuzhi Zhang, Tiandong Li, Han Wang, Guiying Sun, and Feifei Liang. Writing - original draft preparation: Xiuzhi Zhang and Hua Ye; Writing - review and editing: Lipng Dai and Peng Wang; Funding acquisition: Peng Wang. All authors made contributions to the article and have approved the final submitted version.

Funding

This work was supported by the Funded Project of International Training of High-level Talents in Henan Province and Zhengzhou Major Project for Collaborative Innovation (18XTZX12007).

Data availability

The scRNA-seq and bulk-seq datasets analyzed during the current study are available in GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE160269; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53625). The proteomic data can be obtained from a previous study (https://www.nature.com/articles/s41467-021-25202-5) and the raw files of proteome and phosphoproteome datasets can be obtained from PRIDE database (accession number PXD021701) or iProX database (accession number IPX0002501000). The TCGA pan-cancer data were available in XENA browser (https://gdc.xenahubs.net). The gene expression and the gene effect data in the cancer cell lines are available in Depmap data portal (https://depmap.org/portal/data_page/?tab=allData). The anti-cancer drug sensitivity data (IC50s) in the cell lines can be downloaded from Broad institute (https://data.broadinstitute.org/ccle_legacy_data/pharmacological_profiling/CCLE_NP24.2009_Drug_data_2015.02.24.csv).

Declarations

Ethics approval and consent to participate

The study was approved by the ethics committee of Henan People's Hospital of Henan Province (approval ID: 2019-48). Written informed consent was obtained from all subjects in this study. This study adhered to the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]College of Public Health, Zhengzhou University, Zhengzhou 4500001, China
[2]Henan Institute of Medical and Pharmaceutical Sciences, Zhengzhou University, Zhengzhou, Henan Province 450052, China
[3]Henan Key Laboratory of Tumor Epidemiology and State Key Laboratory of Esophageal Cancer Prevention & Treatment, Zhengzhou University, Zhengzhou, Henan Province 450052, China
[4]Laboratory Department, Henan Provincial People's Hospital, Zhengzhou 450003, China

Received: 10 August 2024 / Accepted: 14 April 2025
Published online: 25 April 2025

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Cancer J Clin. 2021;71(3):209–49.
2. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. Cancer J Clin. 2024;74(1):12–49.
3. Morgan E, Soerjomataram I, Rumgay H, Coleman HG, Thrift AP, Vignat J, Laversanne M, Ferlay J, Arnold M. The global landscape of esophageal squamous cell carcinoma and esophageal adenocarcinoma incidence and mortality in 2020 and projections to 2040: new estimates from GLOBOCAN 2020. Gastroenterology. 2022;163(3):649–e658642.
4. Hou G, Niu T, Jia A, Zhang Y, Chen X, Wei H, Jia Y, Xu Y, Li Y, Wang P, et al. NRG1 promotes tumorigenesis and metastasis and Afatinib treatment efficiency is enhanced by NRG1 Inhibition in esophageal squamous cell carcinoma. Biochem Pharmacol. 2023;218:115920.
5. Pellecchia S, Viscido G, Franchini M, Gambardella G. Predicting drug response from single-cell expression profiles of tumours. BMC Med. 2023;21(1):476.

6. Meng Y, Zhao Q, Sang Y, Liao J, Ye F, Qu S, Nie P, An L, Zhang W, Jiao S et al. GPNMB(+) Gal-3(+) hepatic parenchymal cells promote immunosuppression and hepatocellular carcinogenesis. The EMBO journal. 2023:e114060.
7. Xie Z, Niu L, Zheng G, Du K, Dai S, Li R, Dan H, Duan L, Wu H, Ren G, et al. Single-cell analysis unveils activation of mast cells in colorectal cancer microenvironment. Cell Bioscience. 2023;13(1):217.
8. Nakamura S, Ohuchida K, Hayashi M, Katayama N, Tsutsumi C, Yamada Y, Hisano K, Okuda S, Ohtsubo Y, Iwamoto C, et al. Tertiary lymphoid structures correlate with enhancement of antitumor immunity in esophageal squamous cell carcinoma. Br J Cancer. 2023;129(8):1314–26.
9. Zhou B, Guo W, Guo L, Li Y, Zheng Z, Huai Q, Tan F, Li Y, Xue Q, Ying J, et al. Single-cell RNA-sequencing data reveals the genetic source of extracellular vesicles in esophageal squamous cell carcinoma. Pharmacol Res. 2023;192:106800.
10. Chen Y, Zhu S, Liu T, Zhang S, Lu J, Fan W, Lin L, Xiang T, Yang J, Zhao X, et al. Epithelial cells activate fibroblasts to promote esophageal cancer development. Cancer Cell. 2023;41(5):903–e918908.
11. Cheng X, Zhao H, Li Z, Yan L, Min Q, Wu Q, Zhan Q. Integrative analysis of T cell-mediated tumor killing-related genes reveals KIF11 as a novel therapeutic target in esophageal squamous cell carcinoma. J Translational Med. 2025;23(1):197.
12. Yu C, Bian Y, Gao Y, Jiao Y, Xu Y, Wang W, Xin L, Lin H, Wang L. Machine learning-based lactate-related genes signature predicts clinical outcomes and unveils novel therapeutic targets in esophageal squamous cell carcinoma. Cancer Lett. 2025;613:217458.
13. Zhang X, Peng L, Luo Y, Zhang S, Pu Y, Chen Y, Guo W, Yao J, Shao M, Fan W, et al. Dissecting esophageal squamous-cell carcinoma ecosystem by single-cell transcriptomic analysis. Nat Commun. 2021;12(1):5291.
14. Wu H, Leng X, Liu Q, Mao T, Jiang T, Liu Y, Li F, Cao C, Fan J, Chen L, et al. Intratumoral microbiota composition regulates chemoimmunotherapy response in esophageal squamous cell carcinoma. Cancer Res. 2023;83(18):3131–44.
15. Liu ZQ, Dai H, Yao L, Chen WF, Wang Y, Ma LY, Li XQ, Lin SL, He MJ, Gao PT, et al. A single-cell transcriptional landscape of immune cells shows disease-specific changes of T cell and macrophage populations in human achalasia. Nat Commun. 2023;14(1):4685.
16. Jia Y, Zhang B, Zhang C, Kwong DL, Chang Z, Li S, Wang Z, Han H, Li J, Zhong Y, et al. Single-Cell transcriptomic analysis of primary and metastatic tumor ecosystems in esophageal squamous cell carcinoma. Adv Sci (Weinh). 2023;10(7):e2204565.
17. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. Methods in molecular biology (Clifton, NJ). 2018;1711:243–259.
18. Li J, Chen Z, Tian L, Zhou C, He MY, Gao Y, Wang S, Zhou F, Shi S, Feng X, et al. LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. Gut. 2014;63(11):1700–10.
19. Wu Z, Ou L, Wang C, Yang L, Wang P, Liu H, Xiong Y, Sun K, Zhang R, Zhu X. Icaritin induces MC3T3-E1 subclone14 cell differentiation through Estrogen receptor-mediated ERK1/2 and p38 signaling activation. Biomed pharmacotherapy = Biomedecine Pharmacotherapie. 2017;94:1–9.
20. Asa TA, Singh CD, Singh TS, Salahi S, Alom KM, Seo YJ. Nonenzymatically modified mRNA for regulating translation and apoptosis by modulating Cancer epigenetics. Bioorg Chem. 2025;157:108328.
21. Liu W, Xie L, He YH, Wu ZY, Liu LX, Bai XF, Deng DX, Xu XE, Liao LD, Lin W, et al. Large-scale and high-resolution mass spectrometry-based proteomics profiling defines molecular subtypes of esophageal cancer for therapeutic targeting. Nat Commun. 2021;12(1):4961.
22. Lin Y, Wang Y, Xue Q, Zheng Q, Chen L, Jin Y, Huang Z, Li Y. GPRC5A is a potential prognostic biomarker and correlates with immune cell infiltration in non-small cell lung cancer. Translational Lung cancer Res. 2024;13(5):1010–31.
23. Wong CN, Zhang Y, Ru B, Wang S, Zhou H, Lin J, Lyu Y, Qin Y, Jiang P, Lee VH et al. Identification and Characterization of Metastasis-Initiating Cells in ESCC in a Multi-Timepoint Pulmonary Metastasis Mouse Model. Advanced science (Weinheim, Baden-Wurttemberg, Germany) 2024:e2401590.
24. Xie W, Lu J, Chen Y, Wang X, Lu H, Li Q, Jin N, He J, Ou L, Ni J, et al. TCL1A-expressing B cells are critical for tertiary lymphoid structure formation and the prognosis of oral squamous cell carcinoma. J Translational Med. 2024;22(1):477.
25. Li Y, Jiang M, Aye L, Luo L, Zhang Y, Xu F, Wei Y, Peng D, He X, Gu J, et al. UPP1 promotes lung adenocarcinoma progression through the induction of an immunosuppressive microenvironment. Nat Commun. 2024;15(1):1200.
26. Liu H, Weng J, Huang CL, Jackson AP. Is the voltage-gated sodium channel B3 subunit (SCN3B) a biomarker for glioma? Funct Integr Genom. 2024;24(5):162.
27. Liu H, Dilger JP, Lin J. A pan-cancer-bioinformatic-based literature review of TRPM7 in cancers. Pharmacol Ther. 2022;240:108302.
28. Keller P, Dawood M, Chohan BS, Minhas F. HistoKernel: whole slide image level maximum mean discrepancy kernels for pan-cancer predictive modelling. Med Image Anal. 2025;101:103491.
29. Liu H, Weng J. A Pan-Cancer bioinformatic analysis of RAD51 regarding the values for diagnosis, prognosis, and therapeutic prediction. Front Oncol. 2022;12:858756.
30. Dempster JM, Boyle I, Vazquez F, Root DE, Boehm JS, Hahn WC, Tsherniak A, McFarland JM. Chronos: a cell population dynamics model of CRISPR experiments that improves inference of gene fitness effects. Genome Biol. 2021;22(1):343.
31. Marusyk A, Janiszewska M, Polyak K. Intratumor heterogeneity: the Rosetta stone of therapy resistance. Cancer Cell. 2020;37(4):471–84.
32. Maleki EH, Bahrami AR, Matin MM. Cancer cell cycle heterogeneity as a critical determinant of therapeutic resistance. Genes Dis. 2024;11(1):189–204.
33. Huang H, Li N, Liang Y, Li R, Tong X, Xiao J, Tang H, Jiang D, Xie K, Fang C, et al. Multi-omics analyses reveal Spatial heterogeneity in primary and metastatic oesophageal squamous cell carcinoma. Clin Translational Med. 2023;13(11):e1493.
34. Simba H, Kuivaniemi H, Abnet CC, Tromp G, Sewram V. Environmental and life-style risk factors for esophageal squamous cell carcinoma in Africa: a systematic review and meta-analysis. BMC Public Health. 2023;23(1):1782.
35. Abnet CC, Arnold M, Wei WQ. Epidemiology of esophageal squamous cell carcinoma. Gastroenterology. 2018;154(2):360–73.
36. Chetwood JD, Garg P, Finch P, Gordon M. Systematic review: the etiology of esophageal squamous cell carcinoma in low-income settings. Expert Rev Gastroenterol Hepatol. 2019;13(1):71–88.
37. Tian K, Yao Z, Pan D. Leveraging single-cell and multi-omics approaches to identify MTOR-centered deubiquitination signatures in esophageal cancer therapy. Front Immunol. 2024;15:1490623.
38. Huang Z, Cong Z, Luo J, Qiu B, Wang K, Gao C, Xu Y, Yang N, Zou Z, Hu L, et al. Association between cancer-associated fibroblasts and prognosis of neoadjuvant chemoradiotherapy in esophageal squamous cell carcinoma: a bioinformatics analysis based on single-cell RNA sequencing. Cancer Cell Int. 2025;25(1):74.
39. Yang Y, Shi Z, Bai R, Hu W. Heterogeneity of MSI-H gastric cancer identifies a subtype with worse survival. J Med Genet. 2021;58(1):12–9.
40. Lin BB, Lei HQ, Xiong HY, Fu X, Shi F, Yang XW, Yang YF, Liao GL, Feng YP, Jiang DG, et al. MicroRNA-regulated transcriptome analysis identifies four major subtypes with prognostic and therapeutic implications in prostate cancer. Comput Struct Biotechnol J. 2021;19:4941–53.
41. Wu Z, Li G, Wang W, Zhang K, Fan M, Jin Y, Lin R. Immune checkpoints signature-based risk stratification for prognosis of patients with gastric cancer. Cell Signal. 2024;113:110976.
42. Wang Z, Zhang M, Liu L, Yang Y, Qiu J, Yu Y, Li J. Prognostic and immunological role of cancer-associated fibroblasts-derived Exosomal protein in esophageal squamous cell carcinoma. Int Immunopharmacol. 2023;124Pt A:110837.
43. Ngamphaiboon N, Pattaranutaporn P, Lukerak S, Siripoon T, Jinawath A, Arsa L, Shantavasinkul PC, Taonam N, Trachu N, Jinawath N et al. A phase I study of the CDK4/6 inhibitor, Palbociclib in combination with cetuximab and radiotherapy (IMRT) for locally advanced head and neck squamous cell carcinoma. Clin cancer Research: Official J Am Association Cancer Res 2023.
44. Sun Y, Gong W, Zhang S. METTL3 promotes colorectal cancer progression through activating JAK1/STAT3 signaling pathway. Cell Death Dis. 2023;14(11):765.
45. Zhang N, Liu X, Huang L, Zeng J, Ma C, Han L, Li W, Yu J, Yang M. LINC00921 reduces lung cancer radiosensitivity by destabilizing NUDT21 and driving aberrant MED23 alternative polyadenylation. Cell Rep. 2023;42(12):113479.
46. Liao J, Chen Z, Chang R, Yuan T, Li G, Zhu C, Wen J, Wei Y, Huang Z, Ding Z, et al. CENPA functions as a transcriptional regulator to promote hepatocellular carcinoma progression via cooperating with YY1. Int J Biol Sci. 2023;19(16):5218–32.
47. Augustin RC, Newman S, Li A, Joy M, Lyons M, Pham MP, Lucas P, Smith K, Sander C, Isett B et al. Identification of tumor-intrinsic drivers of immune exclusion in acral melanoma. J Immunother Cancer 2023, 11(10).
48. Nekrasova O, Green KJ. Desmosome assembly and dynamics. Trends Cell Biol. 2013;23(11):537–46.
49. Kim M, Reidenbach S, Schlechter T, Rothmann AC, Will R, Hofmann I. Plakophilin 1 deficiency in prostatic tumours is correlated with immune cell

recruitment and controls the up-regulation of cytokine expression post-transcriptionally. FEBS J. 2023;290(7):1907–19.
50. Li K, Wu R, Zhou M, Tong H, Luo KQ. Desmosomal proteins of DSC2 and PKP1 promote cancer cells survival and metastasis by increasing cluster formation in circulatory system. Sci Adv. 2021;7(40):eabg7265.
51. Boyero L, Martin-Padron J, Fárez-Vidal ME, Rodriguez MI, Andrades Á, Peinado P, Arenas AM, Ritoré-Salazar F, Alvarez-Perez JC, Cuadros M, et al. PKP1 and MYC create a feedforward loop linking transcription and translation in squamous cell lung cancer. Cell Oncol (Dordrecht). 2022;45(2):323–32.
52. Haase D, Cui T, Yang L, Ma Y, Liu H, Theis B, Petersen I, Chen Y. Plakophilin 1 is methylated and has a tumor suppressive activity in human lung cancer. Exp Mol Pathol. 2019;108:73–9.
53. Fujita J, Nakajima M, Muroi H, Kikuchi M, Ihara K, Yamaguchi T, Nakagawa M, Morita S, Nakamura T, Tsuchioka T, et al. The prognostic significance of Plakophilin-1 expression in esophageal Cancer. Anticancer Res. 2021;41(7):3401–7.
54. Lee P, Jiang S, Li Y, Yue J, Gou X, Chen SY, Zhao Y, Schober M, Tan M, Wu X. Phosphorylation of Pkp1 by RIPK4 regulates epidermal differentiation and skin tumorigenesis. EMBO J. 2017;36(13):1963–80.
55. Liu H, Li Y, Karsidag M, Tu T, Wang P. Technical and biological biases in bulk transcriptomic data mining for Cancer research. J Cancer. 2025;16(1):34–43.
56. Liu H, Guo Z, Wang P. Genetic expression in cancer research: challenges and complexity. Gene Rep. 2024;37:102042.
57. Spethmann T, Böckelmann LC, Labitzky V, Ahlers AK, Schröder-Schwarz J, Bonk S, Simon R, Sauter G, Huland H, Kypta R, et al. Opposing prognostic relevance of junction Plakoglobin in distinct prostate cancer patient subsets. Mol Oncol. 2021;15(7):1956–69.
58. Winn RA, Bremnes RM, Bemis L, Franklin WA, Miller YE, Cool C, Heasley LE. gamma-Catenin expression is reduced or absent in a subset of human lung cancers and re-expression inhibits transformed cell growth. Oncogene. 2002;21(49):7497–506.
59. Rieger-Christ KM, Ng L, Hanley RS, Durrani O, Ma H, Yee AS, Libertino JA, Summerhayes IC. Restoration of Plakoglobin expression in bladder carcinoma cell lines suppresses cell migration and tumorigenic potential. Br J Cancer. 2005;92(12):2153–9.
60. Chen K, Zeng J, Sun Y, Ouyang W, Yu G, Zhou H, Zhang Y, Yao W, Xiao W, Hu J, et al. Junction Plakoglobin regulates and destabilizes HIF2α to inhibit tumorigenesis of renal cell carcinoma. Cancer Commun (London England). 2021;41(4):316–32.
61. Fang WK, Liao LD, Gu W, Chen B, Wu ZY, Wu JY, Shen J, Xu LY, Li EM. Down-regulated γ-catenin expression is associated with tumor aggressiveness in esophageal cancer. World J Gastroenterol. 2014;20(19):5839–48.
62. Smirnova L, Seregin A, Boksha I, Dmitrieva E, Simutkin G, Kornetova E, Savushkina O, Letova A, Bokhan N, Ivanova S, et al. The difference in serum proteomes in schizophrenia and bipolar disorder. BMC Genomics. 2019;20(Suppl 7):535.
63. Seregin AA, Smirnova LP, Dmitrieva EM, Boksha IS, Savushkina OK, Simutkin GG, Ivanova SA. [Correlations between clinical features of bipolar affective disorder and serum concentrations of ANKRD12 gene product, coagulation factor XIII, and Cadherin 5]. Zhurnal Nevrologii I Psikhiatrii Imeni SS Korsakova. 2022;122(11):137–42.
64. Peter B, Wijsman EM, Nato AQ Jr., Matsushita MM, Chapman KL, Stanaway IB, Wolff J, Oda K, Gabo VB, Raskind WH. Genetic candidate variants in two multigenerational families with childhood apraxia of speech. PLoS ONE. 2016;11(4):e0153864.
65. Chen CY, Tian R, Ge T, Lam M, Sanchez-Andrade G, Singh T, Urpa L, Liu JZ, Sanderson M, Rowley C, et al. The impact of rare protein coding genetic variation on adult cognitive function. Nat Genet. 2023;55(6):927–38.
66. Bai R, Li D, Shi Z, Fang X, Ge W, Zheng S. Clinical significance of Ankyrin repeat domain 12 expression in colorectal cancer. J Experimental Clin cancer Research: CR. 2013;32(1):35.
67. Yang Y, Zhang H, Liu Z, Ma N, Li C, Wang Y, Li Z. Use of exosome transcriptome-based analysis to identify novel biomarkers in patients with locally advanced esophageal squamous cell carcinoma undergoing neoadjuvant chemoradiotherapy. Annals Translational Med. 2023;11(4):182.
68. Wang J, Su W, Zhang T, Zhang S, Lei H, Ma F, Shi M, Shi W, Xie X, Di C. Aberrant Cyclin D1 splicing in cancer: from molecular mechanism to therapeutic modulation. Cell Death Dis. 2023;14(4):244.
69. Karedath T, Ahmed I, Al Ameri W, Al-Dasim FM, Andrews SS, Samuel S, Al-Azwani IK, Mohamoud YA, Rafii A, Malek JA. Silencing of ANKRD12 circrna induces molecular and functional changes associated with invasive phenotypes. BMC Cancer. 2019;19(1):565.
70. Hu J, Huang R, Liang C, Wang Y, Wang M, Chen Y, Wu C, Zhang J, Liu Z, Zhao Q, et al. TRIM50 inhibits gastric Cancer progression by regulating the ubiquitination and nuclear translocation of JUP. Mol cancer Research: MCR. 2023;21(10):1107–19.
71. Ahmad US, Uttagomol J, Wan H. The regulation of the Hippo pathway by intercellular junction proteins. Life (Basel Switzerland) 2022, 12(11).
72. Knights AJ, Funnell AP, Crossley M, Pearson RC. Holding tight: cell junctions and Cancer spread. Trends cancer Res. 2012;8:61–9.
73. Jahangiri L. Updates on liquid biopsies in neuroblastoma for treatment response, relapse and recurrence assessment. Cancer Genet. 2024;288–289:32–9.
74. Li J, Guan X, Fan Z, Ching LM, Li Y, Wang X, Cao WM, Liu DX. Non-Invasive biomarkers for early detection of breast Cancer. Cancers 2020, 12(10).

Publisher's note