Research article

Keratin gene signature expression drives epithelial-mesenchymal transition through enhanced TGF-β signaling pathway activation and correlates with adverse prognosis in lung adenocarcinoma

Gang Li [a,1], Jinbao Guo [b,1], Yunfei Mou [c], Qingsong Luo [a], Xuehai Wang [a], Wei Xue [d], Ting Hou [d], Tianyang Zeng [b,**], Yi Yang [c,*]

[a] Department of Thoracic Surgery, Sichuan Academy of Medical Sciences and Sichuan People's Hospital, University of Electronic Science and Technology of China, Chengdu, 610072, China
[b] Department of Cardiothoracic Surgery, First Affiliated Hospital of Chongqing Medical University, Chongqing, 400016, China
[c] Department of Thoracic Surgery, Chengdu Third People's Hospital, Chengdu, 610082, China
[d] Burning Rock Biotech, Guangzhou, 510300, China

ARTICLE INFO

ABSTRACT

Background: Lung adenocarcinoma (LUAD) stands as the foremost histological subtype of non-small-cell lung cancer, accounting for approximately 40% of all lung cancer diagnoses. However, there remains a critical unmet need to enhance the prediction of clinical outcomes and therapy responses in LUAD patients. Keratins (KRTs), serving as the structural components of the intermediate filament cytoskeleton in epithelial cells, play a crucial role in the advancement of tumor progression. This study investigated the prognostic significance of the KRT family gene and developed a KRT gene signature (KGS) for prognostic assessment and treatment guidance in LUAD.

Methods: Transcriptome profiles and associated clinical details of LUAD patients were meticulously gathered from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. The KGS score was developed based on the expression of five prognostic KRT genes (*KRT7*, *KRT8*, *KRT17*, *KRT18*, and *KRT80*), and the upper quartile of the KGS score was chosen as the cutoff. The Kaplan-Meier method was evaluated to compare survival outcomes between KGS-high and KGS-low groups. The underlying mechanism was further investigated by GSEA, GSVA, and other bioinformatic algorithms.

Results: High expression of the KGS signature exhibited a robust association with poorer overall survival (OS) in the TCGA-LUAD dataset (HR: 1.81; 95% CI: 1.35–2.42, $P = 0.00011$). The association was further corroborated in three external GEO cohorts, including GSE31210 (HR: 3.31; 95% CI: 1.7–6.47, $P = 0.00017$), GSE72094 (HR: 1.95; 95% CI: 1.34–2.85, $P = 0.00057$) and GSE26939 (HR: 3.19; 95% CI: 1.74–5.84, $P < 0.0001$). Interestingly, KGS-high tumors revealed enrichments in TGF-β and WNT-β catenin signaling pathways, exhibited heightened activation of the epithelial-mesenchymal transition (EMT) pathway and proved intensified tumor stemness compared to their KGS-low counterparts. Additionally, KGS-high tumor cells exhibited increased

* Corresponding author.
** Corresponding author.
   E-mail addresses: z_ty1991@163.com (T. Zeng), 191888@qq.com (Y. Yang).
[1] These authors contributed equally to this work.

sensitivity to several targeted agents, including gefitinib, erlotinib, lapatinib, and trametinib, in comparison to KGS-low cells.

Conclusion: This study developed a KGS score that independently predicts the prognosis in LUAD. High expression of KGS score, accompanied by upregulation of TGF-β and WNT-β catenin signaling pathways, confers more aggressive EMT and tumor progression.

1. Introduction

Lung cancer continues to be the leading cause of cancer-related fatalities globally [1], posing a substantial burden on public health and economies [2]. Lung adenocarcinoma (LUAD), as the most prevalent histological form of lung cancer, accounts for approximately 40% of all cases. Typically, lung cancer patients are diagnosed at an advanced stage, with local progression or distant metastases, leading to unfavorable prognoses and clinical outcomes. Precise prognostic predictions can significantly enhance the guidance provided for clinical decision-making and patient management, ultimately improving patients' outcomes. Pathologic parameters, such as TNM staging, remain the major prognostic factors for lung cancer adopted by current clinical guidelines. Finding molecular prognostic markers may bring the opportunity of refining the current prediction system. Furthermore, gaining insights into the molecular mechanisms associated with these prognostic markers will expand the scientific comprehension of lung cancer biology.

Keratins (KRTs) play a crucial role in forming the intermediate filament cytoskeleton within epithelial cells. The Human KRT gene family comprises 54 members categorized into two types: 28 of Type I encoding acidic KRT proteins and 26 of Type II encoding basic KRTs. These KRTs serving as protectors of epithelial structural integrity under stressful conditions are involved in regulating various cellular functions, including motility, signaling, growth and protein synthesis [3]. KRTs have long been utilized as immunohistochemical markers for tumor diagnosis, as epithelial malignancies largely preserve the KRT expression patterns specific to their respective cells of origin [4]. Emerging evidence further points to an active role of KRTs in the invasive and metastatic behavior of cancer cells across various epithelial tumors [3,5]. For instance, *KRT23* has been observed to upregulate ovarian tumor cells migration via epithelial-mesenchymal transition (EMT) by regulating transforming growth factor β (TGF-β)/Smad signaling pathway [6]. The TGF-β pathway integral to normal development and homeostasis [7,8], dysregulation of which could contribute to tumorigenesis by affecting cancer proliferation, progression, EMT and metastasis. *TFAP2A*-induced *KRT16* overexpression promotes tumorigenicity in LUAD via EMT, and *KRT16* expression could serve as an independent prognostic marker [9]. Moreover, a few other KRT genes (KRT17, KRT7, and KRT8, etc.) have also demonstrated prognostic significance in lung cancer [10–12].

However, the prognostic significance and the molecular and biological functions of all KRT family genes remain insufficiently explored. Herein, we performed a systemic investigation on KRT gene family members and developed a KRT-based prognostic signature, namely KGS score, in patients with LUAD. In addition, we delved into mechanisms underlying the differential prognosis by conducting analyses of differentially expressed genes (DEGs), pathway enrichments and drug sensitivities.

2. Materials and methods

2.1. Data collection and preprocessing

RNA sequencing data of a LUAD cohort (n = 592) were downloaded from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/) portal and merged using gdc-rnaseq-tool (https://gdc.cancer.gov/content/gdc-rnaseq-tool). The raw read counts and clinical characteristics, including age, gender, stage, and overall survival (OS) (and vital status), as well as the outcome of LUAD patients, were obtained from UCSC Xena websites (https://xenabrowser.net/datapages/). To further validate the prognostic significance of the candidate gene set, the microarray data and clinical information of GSE31210 (n = 226), GSE72094 (n = 398), and GSE26939 (n = 113) datasets were downloaded from NCBI GEO.

2.2. Differentially expressed genes (DEGs) analysis

Raw read counts of RNA-sequencing of the TCGA-LUAD dataset were normalized into TPM (transcripts per million), which was utilized to recognize DEGs between tumor and normal. Genes that were upregulated were defined using the threshold of a P value < 0.05 and a fold change >1.5. Genes that were upregulated were defined using the cutoff criteria of a P value < 0.05 and a fold change >1.5. The heatmap of candidate gene expression profiles was created following the z-score normalization of their TPM values using the "pheatmap" package within the R computing environment.

2.3. Survival analysis

The Kaplan-Meier method was employed to estimate OS utilizing the "survival" and "survminer" R packages, while the log-rank test ascertained differences in survival curves across groups stratified by varying expressions of KRT genes. The gene expression threshold was set at the upper quartile of TPM values. Multivariate analysis with the Cox proportional hazards model (R packages "survival", "survminer" and "forestplot") was conducted to adjust for potential clinical confounders such as age, gender, and tumor stage. P < 0.05 was defined to be statistically significant.

2.4. Development of a KRT signature score for predicting the prognosis of patients with LUAD

A KRT gene signature (KGS) score was generated by summing the z-scores of five KRT genes as described previously [13–15]. Patients were classified into KGS-high and KGS-low groups using the upper quartile of the KGS score as the cutoff. OS between two distinct groups was compared using Kaplan–Meier curves in TCGA-LUAD dataset and in independent validation datasets of GSE31210, GSE72094, and GSE26939 to evaluate the prognostic significance of the KGS score in LUAD.

2.5. Gene set enrichment analysis and GO analysis

Gene expression profile was compared between KGS-high and KGS-low groups of TCGA-LUAD. DEGs were identified by R package "DEseq2". Furthermore, Gene Set Enrichment Analysis (GSEA) analysis was utilized to identify significant enrichment pathways with hallmark gene sets (c2.cp.v7.4.symbols.gmt) from the Molecular Signatures Database (MSigDB) (http://software.broadinstitute.org/gsea/msigdb/). Finally, GO annotations, including biological process, cellular component, and molecular function analysis were employed by the R package "clusterProfiler" to handle the DEGs and visualize the enriched GO terms between the two groups.

2.6. Gene Set Variation Analysis (GSVA)

In order to quantify the enrichment scores of gene sets in individuals of TCGA-LUAD dataset, ssGSEA algorithm was employed utilizing the R package "GSVA" [16]. Gene sets for GSVA activity pathway analysis of KGS-high and KGS-low groups were selected from the previous study [17]. Specifically, two gene sets, namely the Pan-fibroblast TGF-β response signature (Pan-F-TBRS) and the epithelial-mesenchymal transition (EMT) gene sets, were chosen to validate the potential functional roles of the TGF-β and Wnt signaling pathways.

2.7. EMT score

EMT score based on the 16-gene epithelial-mesenchymal transition (EMT) gene signatures (16 GS) was collected in lung cancer and calculated as previously described [13,15,18,19] for each sample of the TCGA-LUAD dataset. Among them, 16 GS, TJP1, DSP, and CDH1 were epithelial markers. While CDH2, FN1, FOXC2, GSC, ITGB6, MMP2, MMP3, MMP9, SNAI1, SNAI2, SOX10, TWIST1 and VIM were defined as the mesenchymal markers.

2.8. Stemness indices

Two indices of stemness regulated by epigenetic modifications were computed: the RNA expression-based Epigenetically regulated-mRNAsi (EREG-mRNAsi) and the DNA methylation-based Epigenetically regulated-mDNAsi (EREG-mDNAsi), as detailed in previously reported research [18].

2.9. Estimation of drug sensitivity

Drug sensitivity was assessed by determining the concentration at which 50% of cellular growth inhibition occurred, and these estimations were conducted using the R package "pRRophetic" [20].

2.10. Protein-protein interaction analysis

In order to forecast the function and interaction among essential candidate genes and gene sets, we conducted an analysis of the correlations between five KRT family member genes and their interactive genes by GeneMANIA (http://genemania.org/) [21].

2.11. Quantitative real-time polymerase chain reaction (qRT-PCR)

Total cellular RNAs were extracted from cells utilizing Trizol Reagent (Invitrogen, Carlsbad, CA, United States) adhering to the protocol provided by the manufacturer. The reverse transcription was carried out utilizing the reverse transcription kit from Takara (Otsu, Shiga, Japan). Subsequently, real-time polymerase chain reaction (RT-PCR) was conducted utilizing a QuantiTect SYBR Green PCR Kit from Takara, and on an Applied Biosystems QuantStudio 1 (Thermo, Waltham, MA, United States). Relative quantification was determined using the $-2\Delta\Delta Ct$ method. The expression levels of mRNA for each gene were normalized against the expression level of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) mRNA to obtain relative expression values. The primers were synthesized by GenePharma Inc. (Shanghai, China), the sequence of which were listed in Supplementary Table 1. $*P < 0.05$, $**P < 0.01$.

2.12. Cell lines, small interfering RNAs, and cell transfection

Two human lung cancer cells NCI–H1299 and A549, were purchased from National Collection of Authenticated Cell Cultures (Shanghai, China). These cells were cultured in DMEM (Gibco, Carlsbad, CA, United States) supplemented with 10% fetal bovine serum (Gibco, Carlsbad, CA, United States) and 1% penicillin–streptomycin at 37 °C in a 5% CO_2 humidified atmosphere incubator.

The small interfering RNAs (siRNA) designed to target *KRT7* (si-*KRT7*) and *KRT8* (si-*KRT8*), along with non-specific control siRNA (si-Control) were synthesized by RiboBio (Guangzhou, China). Transfection of si-*KRT7*, si-*KRT8* and si-Control into NCI–H1299 and A549 cells was performed utilizing Lipofectamine RNAi Max (Thermo Fisher, Waltham, MA, United States) following the manufacturer's instructions. ns, not significant, **$P < 0.05$, **$P < 0.01$ and ***$P < 0.001$.

2.13. Cell proliferation assays

NCI–H1299 and A549 cells transfected with si-*KRT7*, si-*KRT8* or si-Control were seeded into a 96-well plate, continuously cultured for 120 h, and the number of living cells was measured every 24 h with CCK8 kit (Biosharp, Hefei, Anhui, China). The optical density at 450 nm for each well was quantified with a microplate reader (Allsheng, Hangzou, China). All data were expressed as the means ± standard deviation (SD) which were obtained from three independent experiments. **$P < 0.01$ and ***$P < 0.001$.
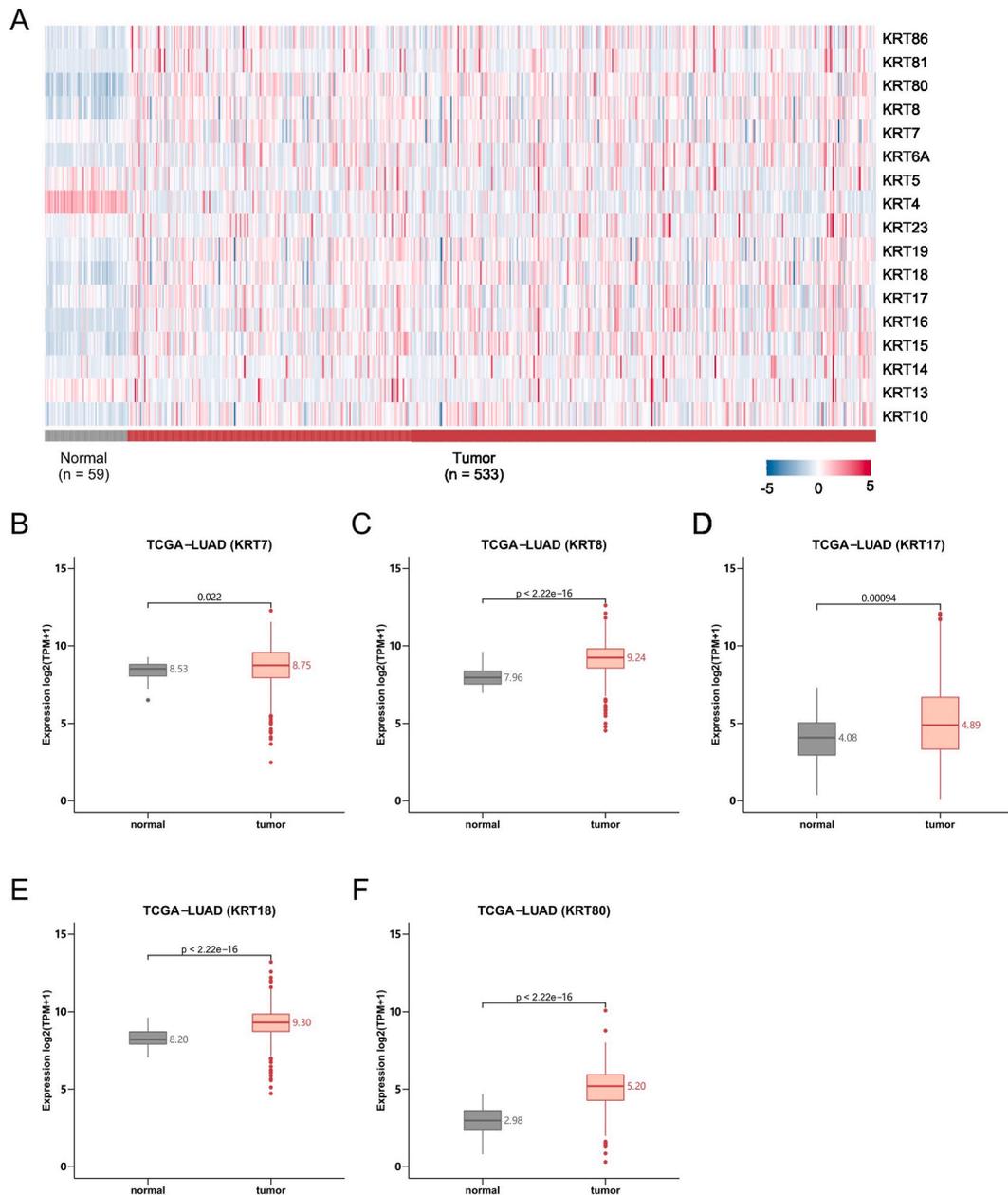
Fig. 1. Differential expression of keratin gene family members in the TCGA-LUAD cohort. A, Heatmap of 17 keratin family genes in normal and tumor using the TCGA-LUAD database. B–F, High mRNA expression levels of five prognostic keratin genes (*KRT7, KRT8, KRT17, KRT18* and *KRT80*) in LUAD vs. normal tissues (TCGA database).

2.14. Colony formation assay

The specified quantity of cells were seeded for transfection. Once colonies became clearly visible even without the need for microscopic examination, we stained them with crystal violet and then captured photographs.

2.15. Statistics analysis

Statistically significant differences were determined utilizing a two-tailed Student's t-test with the R platform (R v4.0.3). To assess the correlation between the two datasets, the Pearson correlation coefficient was calculated using the same R platform. The False Discovery Rate (FDR) was applied to adjust the p-values. Significance levels were defined as P-values or FDR <0.05.An adjusted P < 0.05 was established as the threshold for significance. All statistical analyses were conducted using the R platform (R v4.0.3).
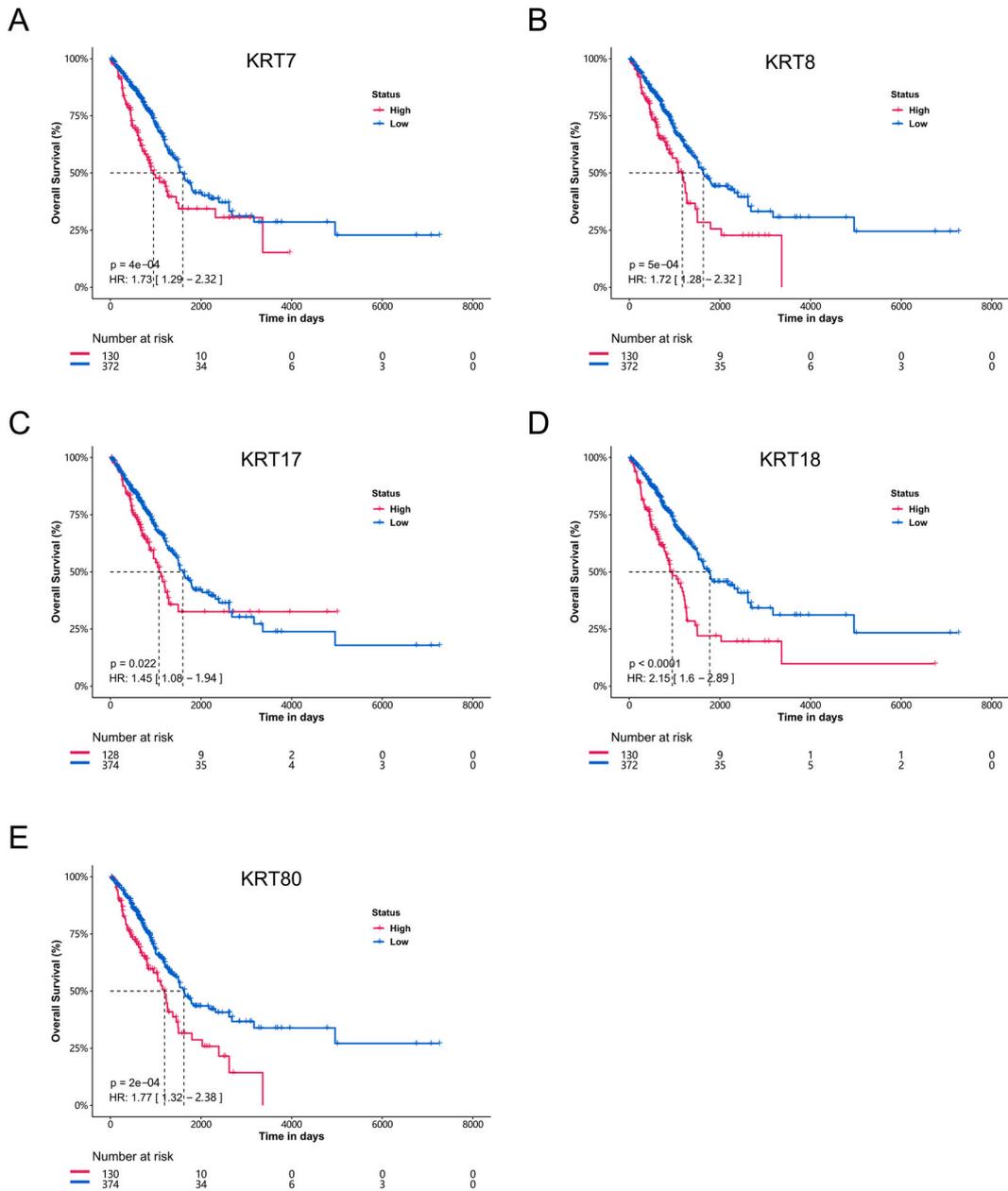
Fig. 2. Prognostic value of mRNA expression of KRT family members in the TCGA-LUAD cohort. Survival curves comparing the high and low expression of *KRT7* (A), *KRT8* (B), *KRT17* (C), *KRT18* (D) and *KRT80* (E) in TCGA-LUAD.

3. Results

3.1. Identification of KRT genes associated with prognosis in LUAD

In order to estimate the mRNA expression of the 54 different KRT family genes in LUAD patients, the TPM of each gene was calculated for the TCGA-LUAD dataset. Among the 54 KRT genes, 37 with extremely low expression (TMP <0.1) were excluded from subsequent analyses. Expression levels of the remained 17 KRT genes were compared between 533 tumor and 59 normal samples (Fig. 1A). Subsequently, the univariate Cox analysis was performed with the 14 DEGs using the TCGA-LUAD dataset and identified five genes (*KRT8, KRT17, KRT7, KRT18,* and *KRT80*) that were significantly upregulated in tumors vs. normal lung tissues (Fig. 1B–F). Furthermore, using the upper quartile value of single gene expression level as the cutoff, patients were stratified into two groups with significantly differential OS (Fig. 2A–E). High-expression groups exhibited inferior survivals compared to low-expression groups. The flow chart of the overall study design is shown in Figure S1.
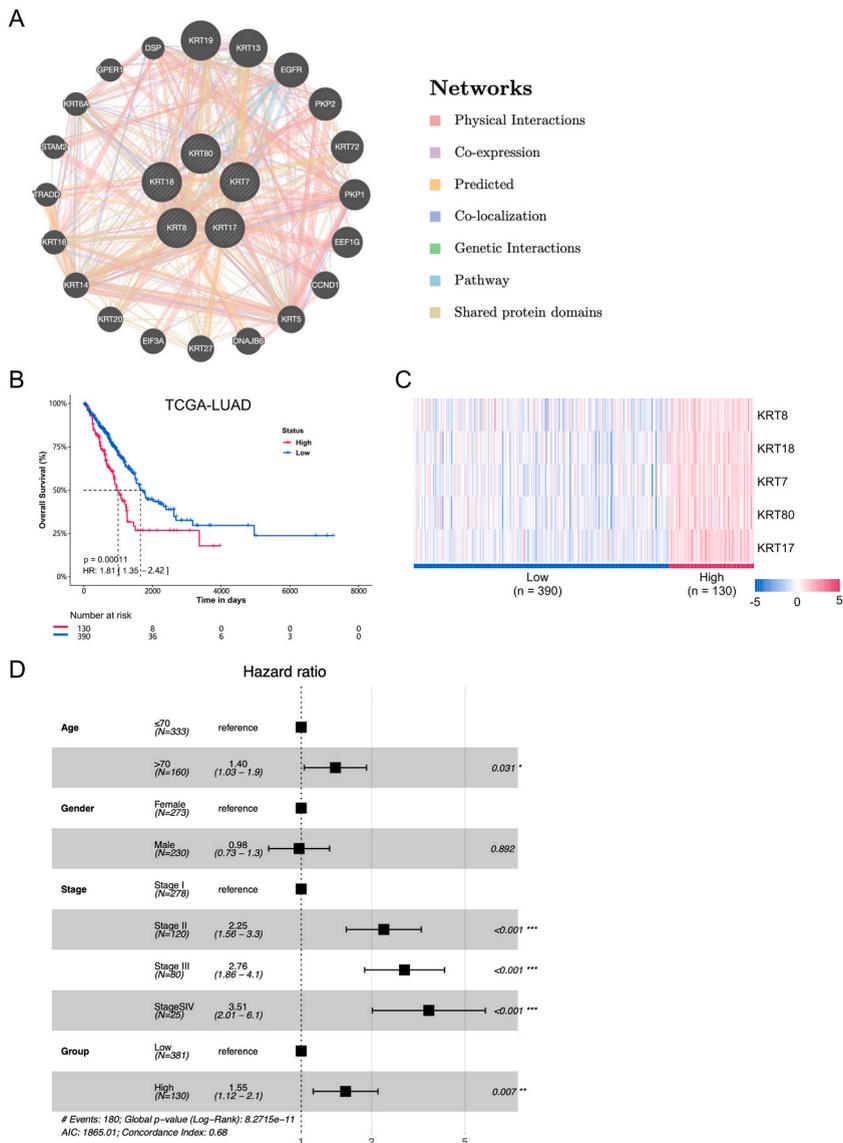
Fig. 3. Development of a KRT signature (KGS) score based on the expression of the five prognostic KRT genes. A, Network of the five keratin genes and their 20 related genes analyzed by GeneMANIA. B, Overall survivals between the two subgroups of TCGA-LUAD stratified by the KGS score (top 25% vs. the remaining). C, Heatmap illustrating the expression levels of the five KRT genes between KGS-high and KGS-low TCGA-LUAD subgroups. D, Multivariable analysis of the KRT signature in the TCGA-LUAD cohort.

3.2. Co-expression network analysis of the five prognostic KRT genes

To investigate the potential functions of the five KRT genes we identified in LUAD, the protein-protein interaction network between KRT genes and their similar function genes was established using GeneMANIA. As a result, the top 20 KRT-related genes were enriched and shown (Fig. 3A), including *KRT5*, *KRT6A*, *KRT13*, *KRT14*, *KRT16*, *KRT19*, *KRT20*, *KRT27*, *KRT72*, *EGFR*, *PKP2*, *PKP1*, *EEF1G*, *CCND1*, *DNAJB6*, *EIF3A*, *TRADD*, *STAM2*, *GPER1,* and *DSP*. As previously reported, *KRT5* expression was inhibited by miRNA let-7a-5p, which plays an essential role in regulating the development of LUAD(22). Furthermore, *KRT16* can be induced by *TFAP2A* and
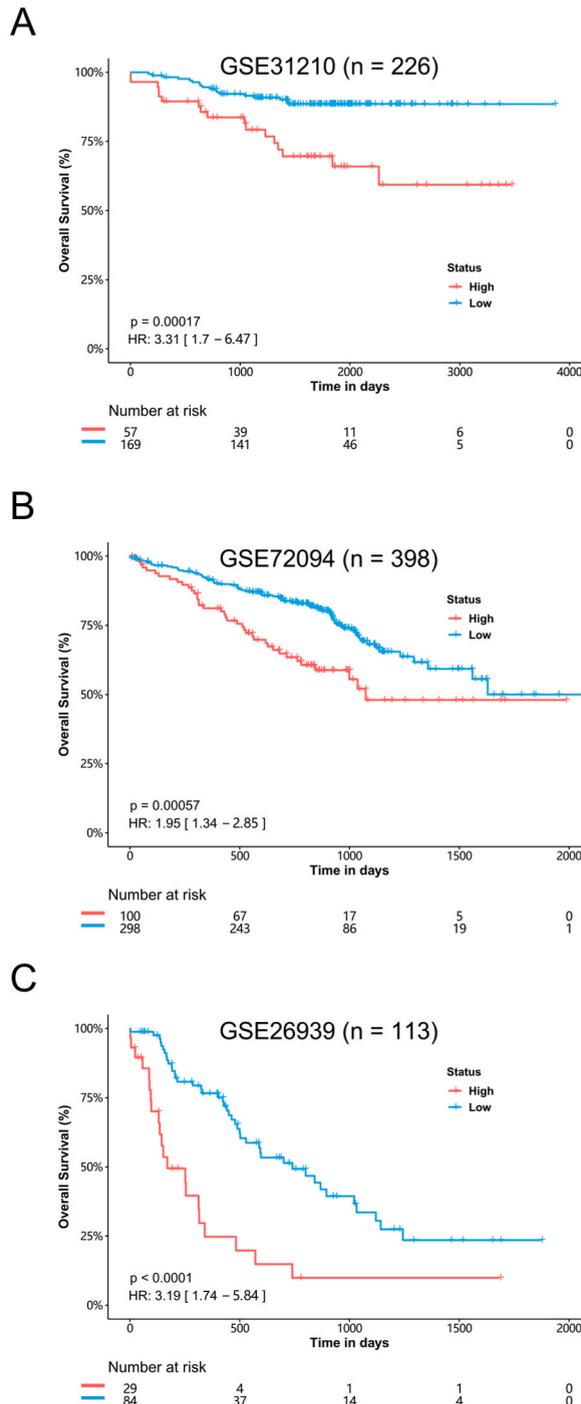
Fig. 4. Validation of the prognostic value of KRT signature in GEO cohorts. A, GSE31210 cohort. B, GSE72094 cohort. C, GSE26939 cohort.

function as an oncogene via the EMT pathway in LUAD [9]. Importantly, we identified *EGFR*, a major oncogenic driver gene in lung cancers, co-expressed with the five KRT genes. Collectively, these results suggested that the five prognostic KRT genes interact with their associated genes, mainly regulating tumorigenesis in LUAD.

3.3. Construction of a KRT gene-based prognostic signature (KGS) in the TCGA-LUAD cohort

Next, we developed a KRT gene signature (KGS) score based on the expression of the five prognostic-associated KRT genes. Patients were classified into two groups using the upper quartile of the KGS score as its threshold. The group with a high KGS (KGS-high) score had shorter OS than the group with a low KGS (KGS-low) score (HR = 1.81, p = 0.00011) (Fig. 3B). Higher expression levels of the five KRT genes were observed in the KGS-high group than KGS-low group (Fig. 3C). Subsequently, multivariate Cox analysis was employed to adjust for potential clinical confounders and found KGS score was an independent prognostic factor (HR = 1.55, p = 0.007) (Fig. 3D). The KGS-high and KGS-low groups revealed comparable age, gender, and statuses of node and metastasis (Figure S2).

3.4. Validation of the prognostic significance of KGS score in external cohorts

We proceeded to corroborate the accuracy and universality of the prognostic significance in external GEO cohorts of LUAD. Patients in these cohorts were also categorized into KGS-high and KGS-low groups using the previously mentioned cutoff (Fig. 4). The data from GSE31210 (n = 226) [23,24] (HR = 3.31, *P* = 0.00017, Fig. 4A), GSE72094 (n = 398) [25] (HR = 1.95, *P* = 0.00057, Fig. 4B) and GSE26939(n = 113) [26] (HR = 3.19, *P* < 0.0001, Fig. 4C) further supported that patients with a high KGS score exhibited poor OS than those with a low KGS score. These results collectively indicate that the KGS score serves as an independent factor associated with OS in LUAD, as observed in both the TCGA-LUAD dataset and the external GEO cohorts (Figs. 3D and 4).

3.5. Association of the KRT gene signature with the EMT process

Subsequently, we compared the gene expression profiles between the KGS-high and KGS-low groups of TCGA-LUAD tumor samples
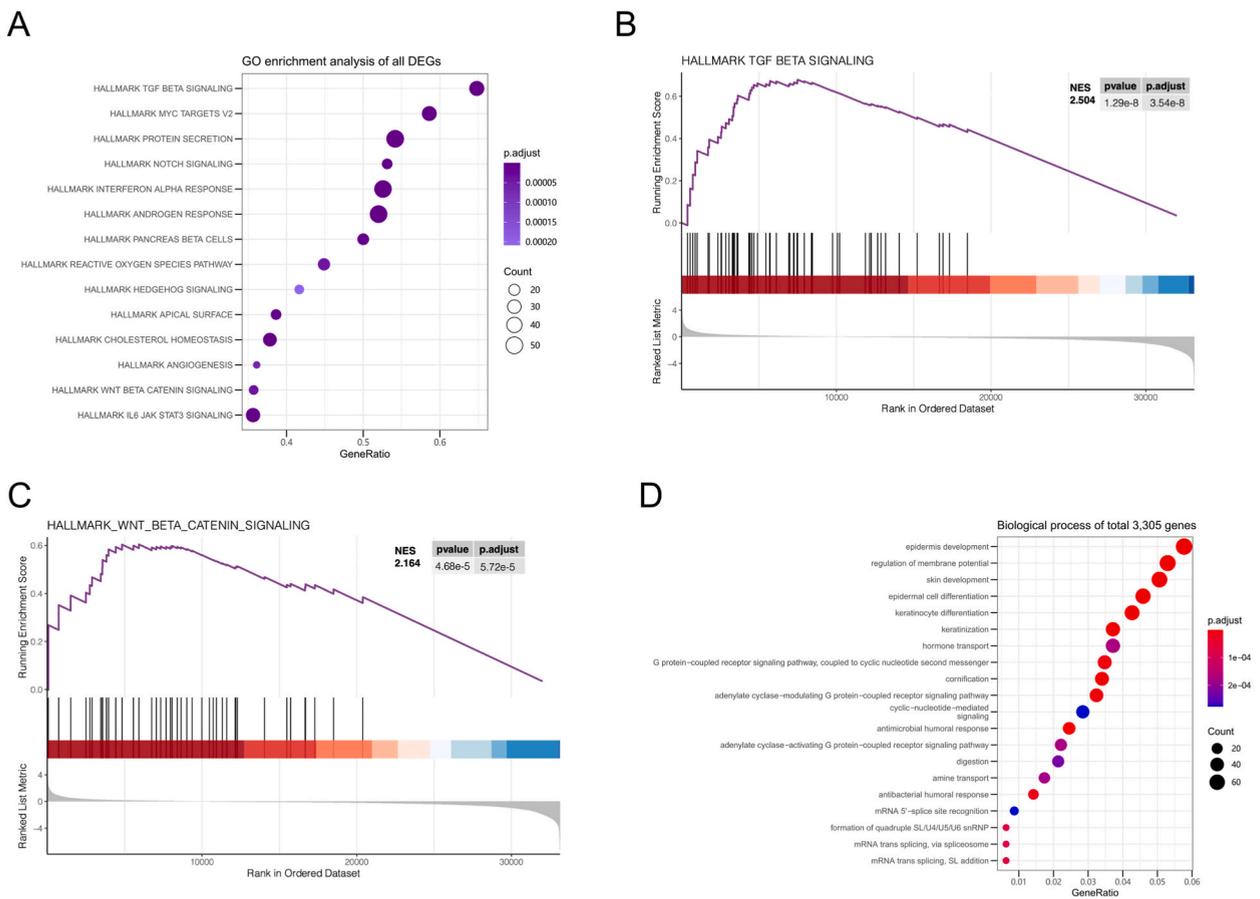
Fig. 5. Function enrichment in the KGS-high vs. KGS-low TCGA-LUAD. A, Pathway analysis of hallmark gene sets. B, GSEA analysis of TGF-β signaling pathway. C, GSEA analysis of WNT-β catenin signaling pathway. D, Biological processes analysis.

and performed GSEA for the functional annotation of DEGs. Notably, GSEA analysis revealed hallmark gene sets of TGF-β signaling pathways (Fig. 5A and B, NES = 2.504, p = 1.29e-8, adjusted p = 3.54e-8) and WNT-β catenin signaling pathways (Fig. 5A and C, NES = 2.164, p = 4.68e-5, adjusted p = 5.72e-5) were significantly more enriched in the KGS-high group as compared to the KGS-low group.

The GO enrichment analysis revealed that KGS-high LUAD expressed higher levels of genes related to epidermis development, epidermal cell differentiation, keratinocyte differentiation, and keratinization (Fig. 5D). Studies have reported that overexpression of KRT genes promotes epithelial-mesenchymal transition (EMT) in malignancies [27–29]. Our result also suggests the activation of epithelial-mesenchymal transition (EMT) in LUAD with a high KGS score, the molecular mechanism underlying which was
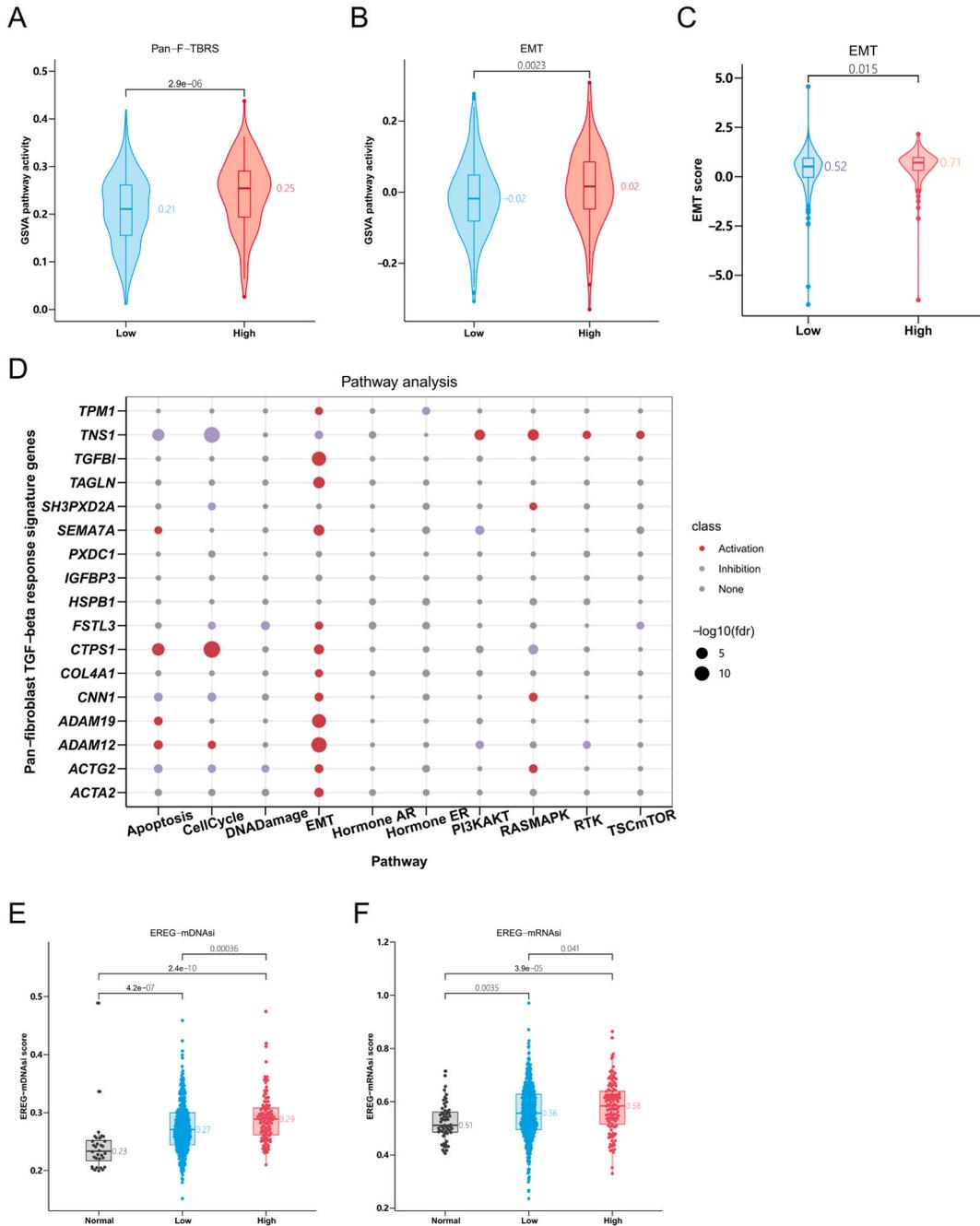
Fig. 6. GSVA analysis of Pan-fibroblast TGF-β response signature (Pan-F-TBRS) and EMT pathway genes. A, GSVA analysis of Pan-F-TBRS signature genes. B, GSVA analysis of EMT signature genes. C, EMT score of KGS-high and KGS-low groups. D, Pathway analysis of Pan-F-TBRS signature genes, which shows activation in EMT process. Stemness indices EREG-mDNAsi (E) and EREG-mRNAsi (F) of normal, KGS-low and KGS-high tumors.

subsequently investigated.

Additionally, Gene Set Variation Analysis (GSVA) was conducted to assess the expression of specific gene signatures between the KGS-high and KGS-low groups in TCGA-LUAD dataset. The gene sets of Pan-F-TBRS and EMT as previously reported [17] were employed to estimate the activity of the TGF-β pathway. The results showed that the Pan-F-TBRS signature exhibited higher expression in the KGS-high than the KGS-low group (Fig. 6A, $P = 2.9e\text{-}06$). The expression levels of individual Pan-F-TBRS genes were also higher in the KGS-high group (Figures S3A–S3C). Furthermore, GSVA indicated that the EMT pathway activity was also significantly enriched in KGS-high tumors (Fig. 6B, $P = 0.0023$). Additionally, we computed an EMT score using 16 EMT genes relevant to lung cancer, including three epithelial genes and 13 mesenchymal markers described in previous study [13]. Consistent with the GSVA score of the EMT signature, the EMT score was also higher in the KGS-high group compared to the KGS-low group (Fig. 6C, $P = 0.015$). The expression levels of individual EMT signature genes were also higher in the KGS-high than in the KGS-low group (Figures S3D–S3F).

Moreover, we investigated the co-expression patterns of the 17 Pan-F-TBRS signature genes with several essential pathways in tumors using GSCA (http://bioinfo.life.hust.edu.cn/GSCA/#/). Notably, twelve out of the seventeen genes displayed strong co-expression with the EMT pathway, indicating a significant activation of the TGF-β pathway in the process of EMT compared with other processes (Fig. 6D). We also calculated two stemness indexes: EREG-mDNAsi and EREG-mRNAsi [18], and conducted a comparative analysis among different samples. Both KGS-low and KGS-high tumors exhibited significantly elevated EREG-mDNAsi and EREG-mRNAsi compared to normal samples (Fig. 6E and F). And KGS-high tumors possessed higher stemness indices than KGS-low tumors. These findings collectively demonstrated that KGS-high tumors maintained a more efficient stemness and exhibited an
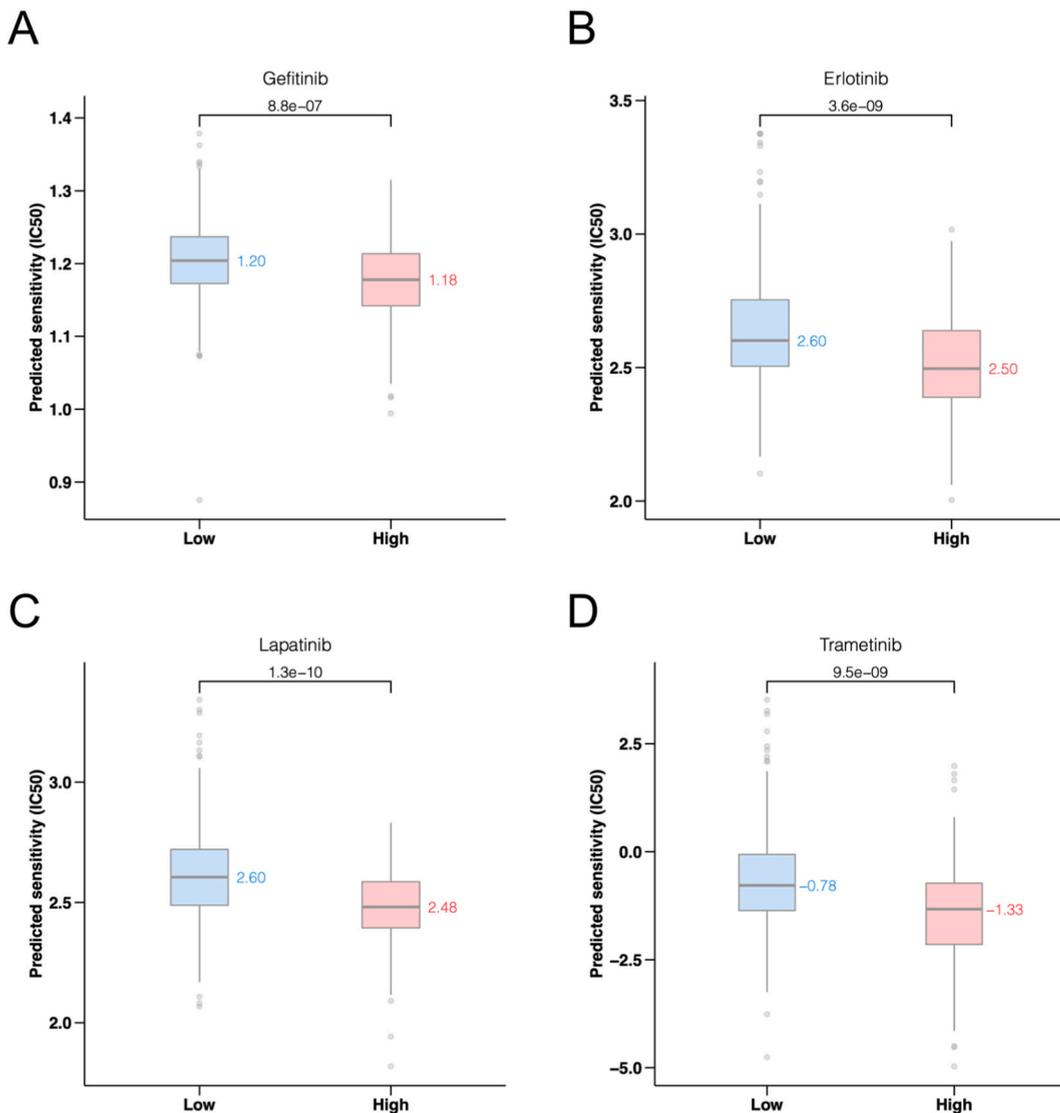
Fig. 7. Drug sensitivities comparison between KGS-high and KGS-low groups. IC50 value of Gefitinib (A), Erlotinib (B), Lapatinib (C) and Trametinib (D).

activated EMT progression.

3.6. Comparison of drug sensitivity between KGS-high and KGS-low groups

Knowing the tumor drug sensitivity may better guide the therapeutic decision-making in the clinic. We compared the sensitivity to various drugs between KGS-high and KGS-low groups. KGS-high tumors revealed higher sensitivities, as indicated by lower IC50 values, to both gefitinib (Fig. 7A, $P = 8.8e\text{-}07$) and erlotinib (Fig. 7B, $P = 3.6e\text{-}09$) compared to KGS-low tumors, indicating that
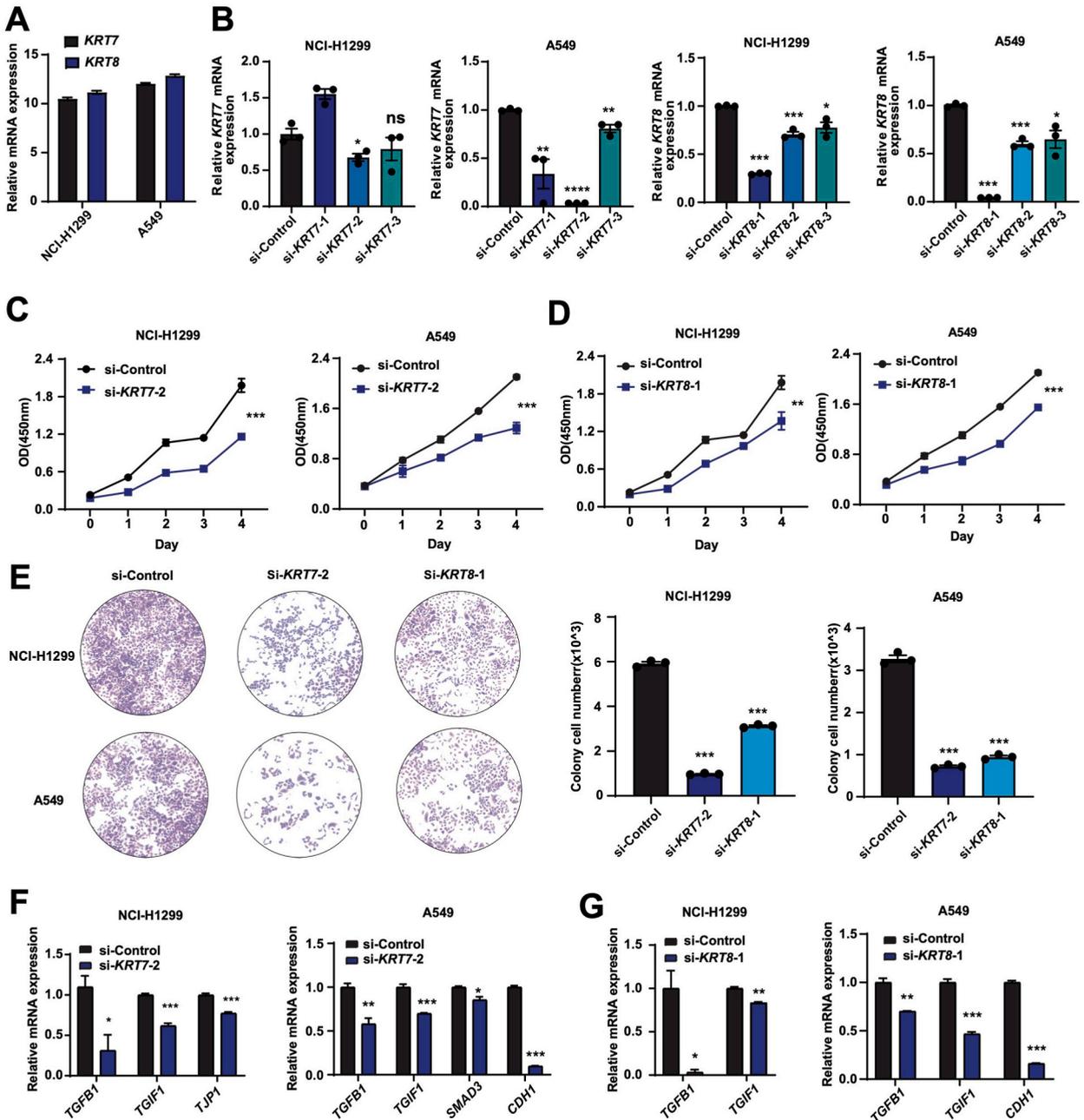
Fig. 8. *In vitro* studies on KGS signature genes in LUAD tumorigenesis. A, The mRNA expression level of *KRT7* and *KRT8* in human lung cells (NCI–H1299 and A549 cell lines). B, The mRNA expression level of *KRT7* and *KRT8* in NCI–H1299 and A549 cells transfected with si-*KRT7*, si-*KRT8*, and si-Control. C, Proliferation curves assessed by CCK8 assay during 120 h for *KRT7* knockdown cell model of NCI–H1299/A549 cells. D, Proliferation curves assessed by CCK8 assay during 120 h for *KRT8* knockdown cell model of NCI–H1299 and A549 cell lines. E, The cloning ability for *KRT7* and *KRT8* knockdown cell models of NCI–H1299 and A549 cell lines. F, The expression level of TGF-β pathway-related genes in NCI–H1299 and A549 cell lines with si-*KRT7*. G, The expression level of TGF-β pathway-related genes in NCI–H1299 and A549 cell lines with si-*KRT8*.

patients with KGS-high LUAD may benefit more from 1st generation epidermal growth factor receptor tyrosine kinase inhibitor (EGFR-TKI). Besides, the KGS-high group was also more sensitive to lapatinib (an ErbB-2/EGFR TKI) (Fig. 7C, $P = 1.3e-10$) and trametinib (a drug targeting RAS-RAF-MEK-ERK-MARP pathway) (Fig. 7D, $P = 9.5e-09$). These findings indicated that the KGS score might be associated with the treatment efficacies of targeted drugs in lung cancer patients.

3.7. Exploration of biological roles of KGS-related signature genes with in vitro studies

In this study, we investigated the expression status of KGS-related signature genes in various human lung cell lines, including *KRT7* and *KRT8* (Fig. 8A). Notably, *KRT7* and *KRT8* exhibited high mRNA expression levels in NCI–H1299 and A549 cell lines (Fig. 8B). To assess the efficacy of siRNA knockdown targeting *KRT7* and *KRT8* in NCI–H1299 and A549 cell lines, we employed qRT-PCR three days post-transfection (Fig. 8C and D). Remarkably, all siRNA sequences resulted in a significant reduction in *KRT7* and *KRT8* mRNA expression. To probe the potential biological roles of *KRT7* and *KRT8* in LUAD, we carried out *in vitro* experiments, encompassing cell proliferation assays and colony formation analyses, to evaluate the effects of *KRT7* and *KRT8* knockdown on LUAD cell proliferation (Fig. 8E) . Strikingly, we observed a substantial decrease in the proliferation ability of LUAD cell lines NCI–H1299 ($P < 0.001$) and A549 ($P < 0.001$) upon *KRT7* and *KRT8* knockdown. Additionally, we performed a cloning ability assay to determine whether the knockdown of *KRT7* and *KRT8* influenced lung cell proliferation *in vitro*. Cloning ability of NCI–H1299 ($P < 0.0001$) and A549 ($P < 0.0001$) cells was also obviously restrained. We also investigated the expression levels of genes associated with the TGF-β signaling pathway (*TGFB1*, *TGIF1*, *TJP1*, *SMAD3* and *CDH1*) in lung cell lines following *KRT7* and *KRT8* knockdown (Fig. 8F and G). These findings suggest that *KRT7* and *KRT8* might enhance cell proliferation and growth via TGF-β signaling pathway in lung cancer.

4. Discussion

Identifying reliable biomarkers that can accurately predict tumor prognosis holds significant promise for tailoring optimal treatment strategies for patients and gaining insights into the molecular mechanisms driving tumor progression and metastasis. This study shed light on the prognostic significance and potential functional roles of five keratin gene family members (*KRT7*, *KRT8*, *KRT17*, *KRT18,* and *KRT80*) in LUAD patients. Regarding the role of these KRT genes in regulating the development of LUAD, it's important to note the complication of lung cancer, involving a interplay of genetic and environmental determinants. While keratins are primarily associated with epithelial tissues, and LUAD originates in the lung epithelium, the exact mechanisms through which these specific keratins and their related genes (*KRT16*, *KRT5*, etc.) may be involved in LUAD development. We found that expression levels of KRT genes were generally higher in LUAD than normal tissues, and high KRT expression was linked to unfavorable prognosis and clinical outcomes, as demonstrated across both the TCGA-LUAD dataset and the validation GEO cohorts.

Previous studies have highlighted the potential of individual KRT genes as independent prognostic biomarkers in various malignancies. *KRT17*, for instance, exhibits high expression levels in numerous tumor tissues and has been associated with poor OS in endometrial carcinomas [30], LUAD [22], LIHC, and KIRC [31]. Interestingly, *KRT17*, known to affect cell migration, proliferation, and invasion, was found to be positively associated with *KRT7* (R = 0.44, $P = 1.1e-24$), *KRT8* (R = 0.38, $P = 1.8e-18$), *KRT18* (R = 0.29, $P = 4.8e-11$) and *KRT80* (R = 0.24, $P = 1.1e-07$) by Gene Expression Profiling Interactive Analysis, version 2 (GEPIA2, http://gepia2.cancer-pku.cn/#index) in TCGA-LUAD tumors. However, the co-expression network among these KRT family genes has not been explored and reported.

While *KRT7* mRNA expression was identified as a sensitive method for detecting *KRT7*-positive circulating tumor cell-resembling A549 cells in peripheral whole blood, its correlation with chemotherapy response and patient survival in advanced LUAD remains uncertain and requires further investigation. Notably, the study highlighted several key findings related to other KRT genes, such as *KRT8*, *KRT17*, and *KRT18*, as independent prognostic factors in LUAD. High *KRT8* expression was linked to poorer OS and recurrence-free survival in LUAD patients. *KRT17* was correlated with advanced disease stage and poor overall survival, with cellular studies demonstrating its role in promoting tumor progression. *KRT18* emerged as a prognostic factor for OS and disease-free survival, and its knockdown enhanced the sensitivity of lung cancer cells to paclitaxel. Furthermore, the study highlighted the prognostic value of 11 genes in LUAD, including *KRT80*, whose downregulation restrained cell growth, migration, and invasion in LUAD cells. Collectively, these findings advance our comprehension of potential molecular markers and therapeutic targets in LUAD, shedding light on its molecular mechanisms and offering avenues for further research and clinical applications.

The EMT process is one of the malignant foundations of LUAD, but the role of KRT genes in this process has not been completely understood. *KRT7* was reported as an immune panel-based signature that could predict the prognosis of LUAD patients and was associated with the infiltration of neutrophils [32]. Elevated KRT7 expression has been associated with increased proliferation, migration, and EMT progression in ovarian carcinoma cells [33]. Interestingly, circKRT7, a circular RNA derived from the back-spliced exon of the linear KRT7 gene locus, has been shown to promote EMT-related cell progression by sponging miR-29a-3p in ovarian cancer [34]. Another keratin gene, *KRT8,* regulates lung carcinogenesis and is associated with EMT. High *KRT8* expression in LUAD is correlated with an unfavorable prognosis [29]. Another study has reported that *KRT8* expression was upregulated along with prognosis and metastasis in LUAD. Functional assays, such as the Cell-Counting Kit-8 (CCK-8) and colony formation assays, established that *KRT8* knockdown could inhibit the proliferation of LUAD cells and significantly impede NF-κB signaling [35]. These findings underscored the potential of *KRT8* as a therapeutic target for anticancer treatments.Furthermore, *KRT18* has been found to upregulate epithelial cell adhesion molecule (EPCAM) by activating the Wnt/β-catenin signaling in breast cancer MCF-7 cells [36]. Consistently, the stemness indices and metastatic properties may be attenuated by the knockdown of EPCAM in *KRT18*-depleted MCF-7 cells. These results imply that *KRT8* could be a promising treatment marker for metastatic and advanced breast cancer [36,37]. In gastric cancer (GC),

CircPIP5K1A/miR-671-5p/*KRT80* axis may be a potential therapeutic target [38]. Meanwhile, in colorectal carcinoma (CRC), *KRT80* has emerged as an independent prognostic biomarker and has been certified to interact with *PRKDC* to activate the AKT pathway, thereby promoting CRC migration and invasion [39]. Collectively, these results suggest that keratin family genes could affect tumor cell expression, proliferation, and migration in a spectrum of cancer types. Particularly in LUAD, KRT genes are involved in the EMT process and impact prognosis.

The clinical significance of KRT (keratin) genes is underscored by their potential role in the response to targeted cancer therapies like Gefitinib, Erlotinib, Lapatinib, and Trametinib. These medications are commonly used in the treatment of various cancers, particularly those with aberrant signaling pathways such as EGFR mutations or MAPK pathway dysregulation. KRT genes, typically associated with epithelial tissues, may be believed to contribute to the processes of cancer cell proliferation, migration, and invasion. Exploring the interplay between KRT gene expression and the response to these targeted therapies holds significant clinical implications. It may lead to the identification of biomarkers that can help predict patient responses to these drugs, thereby enabling more personalized treatment strategies and potentially improving outcomes for individuals with cancer. Further research is needed to elucidate the precise mechanisms linking KRT genes and these targeted therapies and to validate their clinical utility.

Although the expression and prognostic significance of each individual KRT gene (*KRT7*, *KRT8*, *KRT17*, *KRT18*, and *KRT80*) in our KGS score model have been studied in a variety of cancers, our study focused on integrating these five KRT genes by summing their z-scores and investigating mechanisms underlying the prognostic value of the KGS score in LUAD. In the present study, we established that these five genes were significantly upregulated in LUAD tissues compared to normal tissues. Using the upper quartile of KGS score as a cutoff, we categorized the tumor patients into two groups: KGS-high and KGS-low. The KGS-high and KGS-low groups revealed differential prognoses, confirmed by TCGA-LUAD and three independent GEO cohorts. Specifically, the KGS-high group, characterized by higher expression of KRT genes, exhibited inferior OS. Subsequent analyses suggested the inferior prognosis in the KGS-high group could be attributable to the upregulated TGF-β signaling pathway, which promoted the EMT process in LUAD. Additionally, our study revealed that tumor cells with high KGS scores were more sensitive to EGFR-TKI inhibitors (gefitinib, erlotinib, and lapatinib) as well as MAPK inhibitors (Trametinib) compared to the KGS-low group. These findings shed light on the potential therapeutic implications of the KGS score in LUAD.

In summary, our study developed a KRT gene signature that may serve as an independent prognostic marker in LAUD. Our results also indicate that high expression of KRT signature, accompanied by upregulation of TGF-β and WNT-β catenin signaling pathways, confer more aggressive EMT and tumor progression, ultimately leading to a poorer prognosis. However, it's important to note that a limitation of this study is that the findings ideally require validation in a broader cohort of LUAD patients encompassing a wider range of underlying disease etiologies. While *KRT7* and *KRT8* have been validated, the validation of additional genes, such as *KRT17*, *KRT18*, and *KRT80*, would necessitate more extensive resources and a larger patient population for a comprehensive assessment.

Data availability statement

The data underlying this study are freely available from the TCGA dataset (https://portal.gdc.cancer.gov/projects/TCGA-LUAD) and the GEO dataset (http://www.ncbi.nlm.nih.gov/geo/).

Fundings

Not applicable.

CRediT authorship contribution statement

Gang Li: Writing – review & editing, Writing – original draft, Validation, Formal analysis, Data curation. **Jinbao Guo:** Validation, Formal analysis, Data curation. **Yunfei Mou:** Formal analysis, Data curation. **Qingsong Luo:** Investigation, Formal analysis. **Xuehai Wang:** Methodology, Formal analysis. **Wei Xue:** Formal analysis, Methodology. **Ting Hou:** Writing – review & editing. **Tianyang Zeng:** Investigation, Conceptualization. **Yi Yang:** Writing – review & editing, Writing – original draft, Conceptualization.

Declaration of competing interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e24549.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA A Cancer J. Clin. 71 (3) (2021) 209–249.

[2] S. Saab, H. Zalzale, Z. Rahal, Y. Khalifeh, A. Sinjab, H. Kadara, Insights into lung cancer immune-based biology, prevention, and treatment, Front. Immunol. 11 (2020) 159.

[3] V. Karantza, Keratins in health and cancer: more than mere epithelial cell markers, Oncogene 30 (2) (2011) 127–138.

[4] R. Moll, M. Divo, L. Langbein, The human keratins: biology and pathology, Histochem. Cell Biol. 129 (6) (2008) 705–733.

[5] H. Zhang, Y. Zhang, T. Xia, L. Lu, M. Luo, Y. Chen, et al., The role of Keratin17 in human tumours, Front. Cell Dev. Biol. 10 (2022) 818416.

[6] M. Ren, Y. Gao, Q. Chen, H. Zhao, X. Zhao, W. Yue, The overexpression of keratin 23 promotes migration of ovarian cancer via epithelial-mesenchymal transition, BioMed Res. Int. 2020 (2020) 8218735.

[7] V. Syed, TGF-Beta signaling in cancer, J. Cell. Biochem. 117 (6) (2016) 1279–1287.

[8] A. Vander Ark, J. Cao, X. Li, TGF-beta receptors: in and beyond TGF-beta signaling, Cell. Signal. 52 (2018) 112–120.

[9] L. Yuanhua, Q. Pudong, Z. Wei, W. Yuan, L. Delin, Z. Yan, et al., TFAP2A induced KRT16 as an oncogene in lung adenocarcinoma via EMT, Int. J. Biol. Sci. 15 (7) (2019) 1419–1428.

[10] J. Liu, L. Liu, L. Cao, Q. Wen, Keratin 17 promotes lung adenocarcinoma progression by enhancing cell proliferation and invasion, Med. Sci. Mon. Int. Med. J. Exp. Clin. Res. 24 (2018) 4782–4790.

[11] A. Koren, E. Sodja, M. Rijavec, M. Jez, V. Kovac, P. Korosec, et al., Prognostic value of cytokeratin-7 mRNA expression in peripheral whole blood of advanced lung adenocarcinoma patients, Cell. Oncol. 38 (5) (2015) 387–395.

[12] L. Xie, Y. Dang, J. Guo, X. Sun, T. Xie, L. Zhang, et al., High KRT8 expression independently predicts poor prognosis for lung adenocarcinoma patients, Genes 10 (1) (2019).

[13] Y.K. Chae, S. Chang, T. Ko, J. Anker, S. Agte, W. Iams, et al., Epithelial-mesenchymal transition (EMT) signature is inversely associated with T-cell infiltration in non-small cell lung cancer (NSCLC), Sci. Rep. 8 (1) (2018) 2918.

[14] J.C. Smith, J.M. Sheltzer, Genome-wide identification and analysis of prognostic features in human cancers, Cell Rep. 38 (13) (2022) 110569.

[15] J.C. Thompson, W.T. Hwang, C. Davis, C. Deshpande, S. Jeffries, Y. Rajpurohit, et al., Gene signatures of tumor inflammation and epithelial-to-mesenchymal transition (EMT) predict responses to immune checkpoint blockade in lung cancer with high accuracy, Lung Cancer 139 (2020) 1–8.

[16] S. Hanzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, BMC Bioinf. 14 (2013) 7.

[17] S. Mariathasan, S.J. Turley, D. Nickles, A. Castiglioni, K. Kuen, Y. Wang, et al., TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells, Nature 554 (7693) (2018) 544–548.

[18] T.M. Malta, A. Sokolov, A.J. Gentles, T. Burzykowski, L. Poisson, J.N. Weinstein, et al., Machine learning identifies stemness features associated with oncogenic dedifferentiation, Cell 173 (2) (2018) 338–354 e15.

[19] L.A. Byers, L. Diao, J. Wang, P. Saintigny, L. Girard, M. Peyton, et al., An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance, Clin. Cancer Res. 19 (1) (2013) 279–290.

[20] P. Geeleher, N. Cox, R.S. Huang, pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels, PLoS One 9 (9) (2014) e107468.

[21] D. Warde-Farley, S.L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, et al., The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function, Nucleic Acids Res. 38 (Web Server issue) (2010) W214–W220.

[22] H. Yu, Z. Pang, G. Li, T. Gu, Bioinformatics analysis of differentially expressed miRNAs in non-small cell lung cancer, J. Clin. Lab. Anal. 35 (2) (2021) e23588.

[23] M. Yamauchi, R. Yamaguchi, A. Nakata, T. Kohno, M. Nagasaki, T. Shimamura, et al., Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma, PLoS One 7 (9) (2012) e43923.

[24] H. Okayama, T. Kohno, Y. Ishii, Y. Shimada, K. Shiraishi, R. Iwakawa, et al., Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas, Cancer Res. 72 (1) (2012) 100–111.

[25] M.B. Schabath, E.A. Welsh, W.J. Fulp, L. Chen, J.K. Teer, Z.J. Thompson, et al., Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma, Oncogene 35 (24) (2016) 3209–3216.

[26] M.D. Wilkerson, X. Yin, V. Walter, N. Zhao, C.R. Cabanski, M.C. Hayward, et al., Differential pathogenesis of lung adenocarcinoma subtypes involving sequence mutations, copy number, chromosomal instability, and methylation, PLoS One 7 (5) (2012) e36530.

[27] Z. Liu, S. Yu, S. Ye, Z. Shen, L. Gao, Z. Han, et al., Keratin 17 activates AKT signalling and induces epithelial-mesenchymal transition in oesophageal squamous cell carcinoma, J. Proteonomics 211 (2020) 103557.

[28] Z. Wang, M.Q. Yang, L. Lei, L.R. Fei, Y.W. Zheng, W.J. Huang, et al., Overexpression of KRT17 promotes proliferation and invasion of non-small cell lung cancer and indicates poor prognosis, Cancer Manag. Res. 11 (2019) 7485–7497.

[29] W. Wang, J. He, H. Lu, Q. Kong, S. Lin, KRT8 and KRT19, associated with EMT, are hypomethylated and overexpressed in lung adenocarcinoma and link to unfavorable prognosis, Biosci. Rep. 40 (7) (2020).

[30] J.D.K. Bai, S. Babu, L. Roa-Pena, W. Hou, A. Akalin, L.F. Escobar-Hoyos, et al., Keratin 17 is a negative prognostic biomarker in high-grade endometrial carcinomas, Hum. Pathol. 94 (2019) 40–50.

[31] C. Li, Y. Teng, J. Wu, F. Yan, R. Deng, Y. Zhu, et al., A pan-cancer analysis of the oncogenic role of Keratin 17 (KRT17) in human tumors, Transl. Cancer Res. 10 (10) (2021) 4489–4501.

[32] Y. Zhou, L. Tang, Y. Chen, Y. Zhang, W. Zhuang, An immune panel signature predicts prognosis of lung adenocarcinoma patients and correlates with immune microenvironment, Front. Cell Dev. Biol. 9 (2021) 797984.

[33] Q. An, T. Liu, M.Y. Wang, Y.J. Yang, Z.D. Zhang, Z.J. Liu, et al., KRT7 promotes epithelialmesenchymal transition in ovarian cancer via the TGFbeta/Smad2/3 signaling pathway, Oncol. Rep. 45 (2) (2021) 481–492.

[34] Q. An, T. Liu, M.Y. Wang, Y.J. Yang, Z.D. Zhang, Z.J. Lin, et al., circKRT7-miR-29a-3p-COL1A1 Axis promotes ovarian cancer cell progression, OncoTargets Ther. 13 (2020) 8963–8976.

[35] H. Chen, X. Chen, B. Pan, C. Zheng, L. Hong, W. Han, KRT8 serves as a novel biomarker for LUAD and promotes metastasis and EMT via NF-kappaB signaling, Front. Oncol. 12 (2022) 875146.

[36] R. Shi, L. Liu, F. Wang, Y. He, Y. Niu, C. Wang, et al., Downregulation of cytokeratin 18 induces cellular partial EMT and stemness through increasing EpCAM expression in breast cancer, Cell. Signal. 76 (2020) 109810.

[37] Y. Na, S.C. Kaul, J. Ryu, J.S. Lee, H.M. Ahn, Z. Kaul, et al., Stress chaperone mortalin contributes to epithelial-mesenchymal transition and cancer metastasis, Cancer Res. 76 (9) (2016) 2754–2765.

[38] H. Song, Y. Xu, T. Xu, R. Fan, T. Jiang, M. Cao, et al., CircPIP5K1A activates KRT80 and PI3K/AKT pathway to promote gastric cancer development through sponging miR-671-5p, Biomed. Pharmacother. 126 (2020) 109941.

[39] C. Li, X. Liu, Y. Liu, X. Liu, R. Wang, J. Liao, et al., Keratin 80 promotes migration and invasion of colorectal carcinoma by interacting with PRKDC via activating the AKT pathway, Cell Death Dis. 9 (10) (2018) 1009.