



Single-cell data analysis of malignant epithelial cell heterogeneity in lung adenocarcinoma for patient classification and prognosis prediction

Xu Ran^{a,b}, Lu Tong^{a,b}, Wang Chenghao^{a,b}, Li Qi^c, Peng Bo^{a,b}, Zhao Jiaying^{a,b}, Wang Jun^d, Zhang Linyou^{a,*}

^a Department of Thoracic Surgery, The Second Affiliated Hospital of Harbin Medical University, Harbin, China

^b The Second Clinical Medical College, Harbin Medical University, Harbin, China

^c Department of Child and Adolescent Health, School of Public Health, Harbin Medical University, Harbin, China

^d Department of Thoracic Surgery, Baoji Central Hospital, Baoji, China

ARTICLE INFO

Keywords:

Lung adenocarcinoma
Single-cell RNA-Seq
Prognosis prediction
Lung cancer

ABSTRACT

Lung cancer is one of the leading causes of cancer-related death. Most advanced lung adenocarcinoma (LUAD) patients have poor survival because of drug resistance and relapse. Neglecting intratumoral heterogeneity might be one of the reasons for treatment insensitivity, while single-cell RNA sequencing (scrRNA-seq) technologies can provide transcriptome information at the single-cell level. Herein, we combined scrRNA-seq and bulk RNA-seq data of LUAD and identified a novel cluster of malignant epithelial cells - KRT81⁺ malignant epithelial cells - associated with worse prognoses. Further analysis revealed that the hypoxia and EMT pathways of these cells were activated to predispose them to differentiate into metastatic lung adenocarcinoma cells. Finally, we also studied the role of these tumor cells in the immune microenvironment and their role in the classification and prognosis prediction of lung adenocarcinoma patients.

1. Introduction

Lung cancer is the leading cause of cancer-related death [1]. About 80–85% of lung cancer cases are non-small cell lung cancer and include two main histological subtypes: lung squamous cell carcinoma and lung adenocarcinoma (LUAD). Besides, LUAD accounts for most new cases. Advanced lung cancer is the leading cause of death in lung cancer cases. Although the overall mortality rate of lung cancer has rapidly decreased with the clinical application of PD-1/PD-L1 immunotherapy and EGFR-TKIs targeted therapy in advanced lung cancer, many cases are still insensitive to these treatments [2]. Therefore, finding new prognostic monitoring biomarkers and therapeutic targets remains necessary.

As RNA sequencing (RNA-seq) matures, its high-throughput and efficient advantages help identify tumor molecular features to mine new therapeutic targets and clinical biomarkers [3]. RNA-seq also has a great role in promoting the current concept of precision treatment. However, bulk RNA-seq is based on whole tissue transcriptomes, ignoring intratumoral heterogeneity. At the same time, although many algorithms [4–6] use bulk RNA-seq data to infer the composition of immune cells in the tumor microenvironment, the inability to obtain transcriptome information of immune cells also weakens the understanding of the tumor microenvironment.

* Corresponding author.

E-mail address: lyzhang@hrbmu.edu.cn (Z. Linyou).

Intratumoral heterogeneity often leads to different responsiveness of patients throughout immunotherapy and targeted therapy cycles [7]. Moreover, tumor cells that are insensitive to treatment are one of the causes of drug resistance and relapse after treatment [8,9]. Therefore, understanding the transcriptional landscape within tumors in a higher dimension is urgent. Meanwhile, single-cell RNA-seq

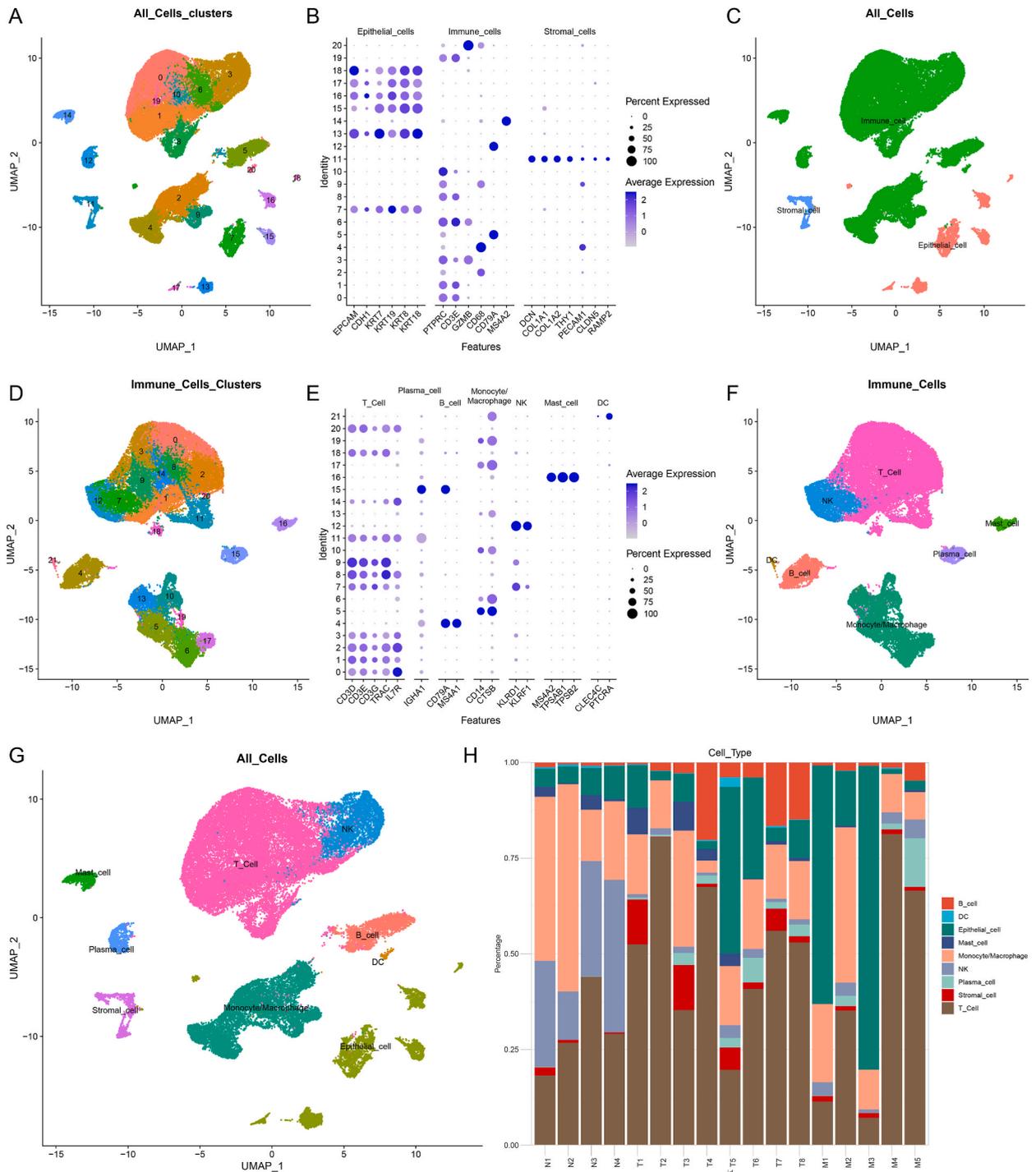


Fig. 1. Single cell atlas of primary and metastatic lung adenocarcinoma. (A) UMAP plot for all cells clustered into 21 clusters. (B) Expression of cell marker genes in each cell cluster. (C) All cells were annotated as three major cell types including immune cells, epithelial cells, and stromal cells. (D) UMAP plot for immune cells clustered into 22 clusters. (E) Expression of immune cell marker genes in each cell cluster. (F) Cell clusters were annotated as seven different types of immune cells. (G) UMAP Plot for all cells after all cell annotations are integrated. (H) Proportion of different cell types in each sample.

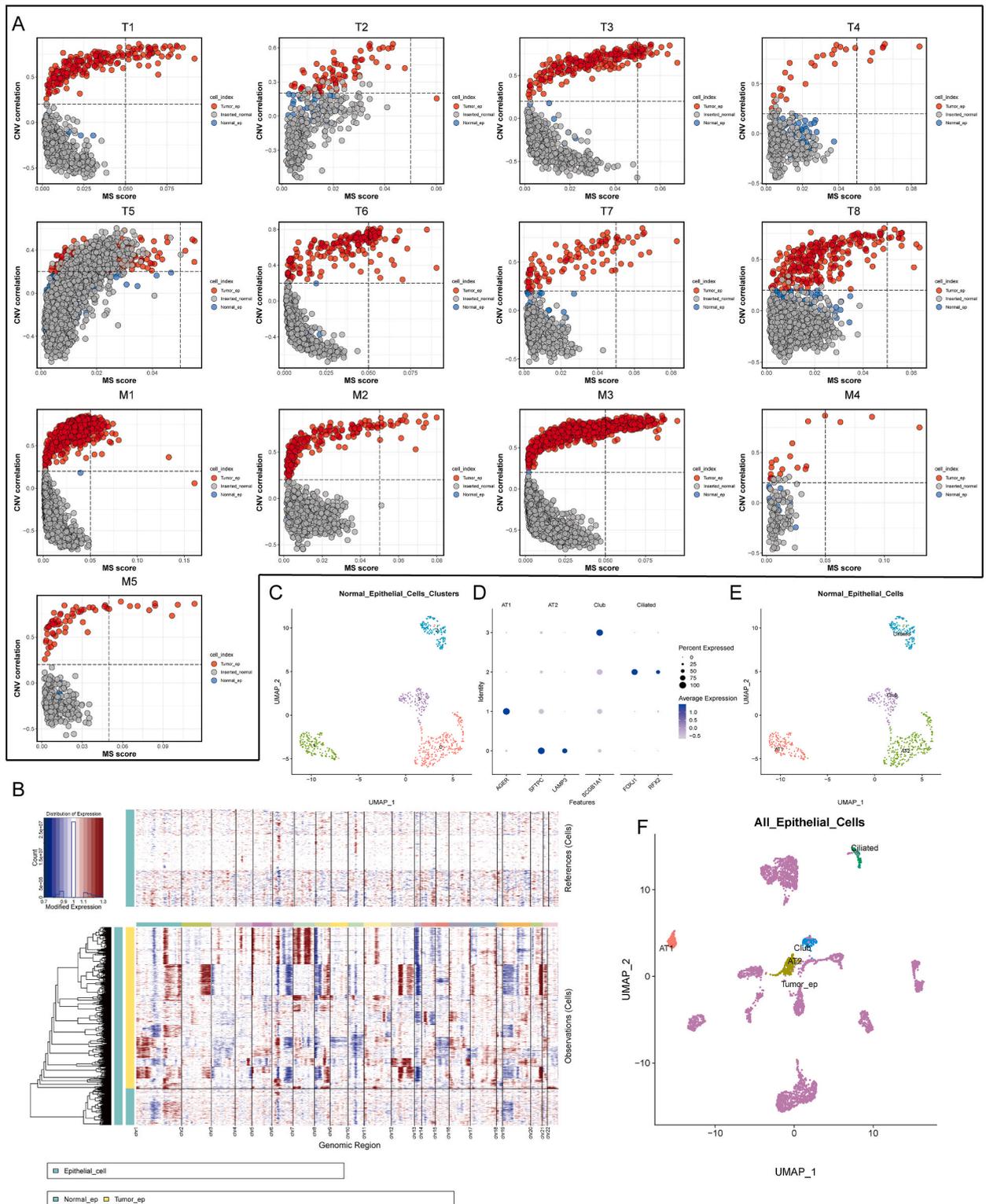


Fig. 2. Copy Number Analysis for All Epithelial Cells. (A) CNV analysis of all tumor tissue-derived epithelial cells with inserted normal tissue-derived normal cells; the gray dots represent inserted normal tissue-derived normal cells, blue dots are defined as normal epithelial cells in tumor tissue, and red dots are defined as tumor cells. (B) Analysis of copy number loss or amplification of each chromosome in all epithelial cells by InferCNV algorithm. (C) UMAP plot for all normal epithelial cells clustered into four clusters. (D) Expression of cell marker genes in each cell cluster. (E) All clusters were annotated as four major cell types. (F) UMAP plot of all normal and malignant epithelial cells.

(scRNA-seq) can solve the lack of understanding of intratumoral heterogeneity and tumor microenvironment in bulk RNA-seq. Additionally, scRNA-seq provides transcriptome sequencing at single-cell resolution, and various associated single-cell algorithms have deepened the understanding of single-cell transcriptional characteristics, cell dynamics, cell-cell interactions, and intratumoral heterogeneity [10,11].

The single-cell transcriptome landscape of LUAD has been reported in several articles [12–19], establishing a preliminary understanding of its intratumoral heterogeneity and providing valuable raw data for subsequent studies. Herein, we combined scRNA-seq and bulk RNA-seq data from LUAD to identify a novel cluster of LUAD tumor cells associated with prognosis. We also profoundly investigated the transcriptome characteristics, the role in the tumor microenvironment, and the unique role of prognostic monitoring. Overall, we revealed an entirely novel cluster of tumor cells that can be used for prognostic prediction and might be potential therapeutic targets for LUAD.

2. Results

2.1. Primary and metastatic LUAD single-cell atlas

The single-cell RNA-seq dataset used in this study contains eight primary cases, five metastatic LUAD, and four normal lung tissues (Table S1). After quality control, dimension reduction, and clustering, we identified 21 unique clusters (Figure S1 and Fig. 1A).

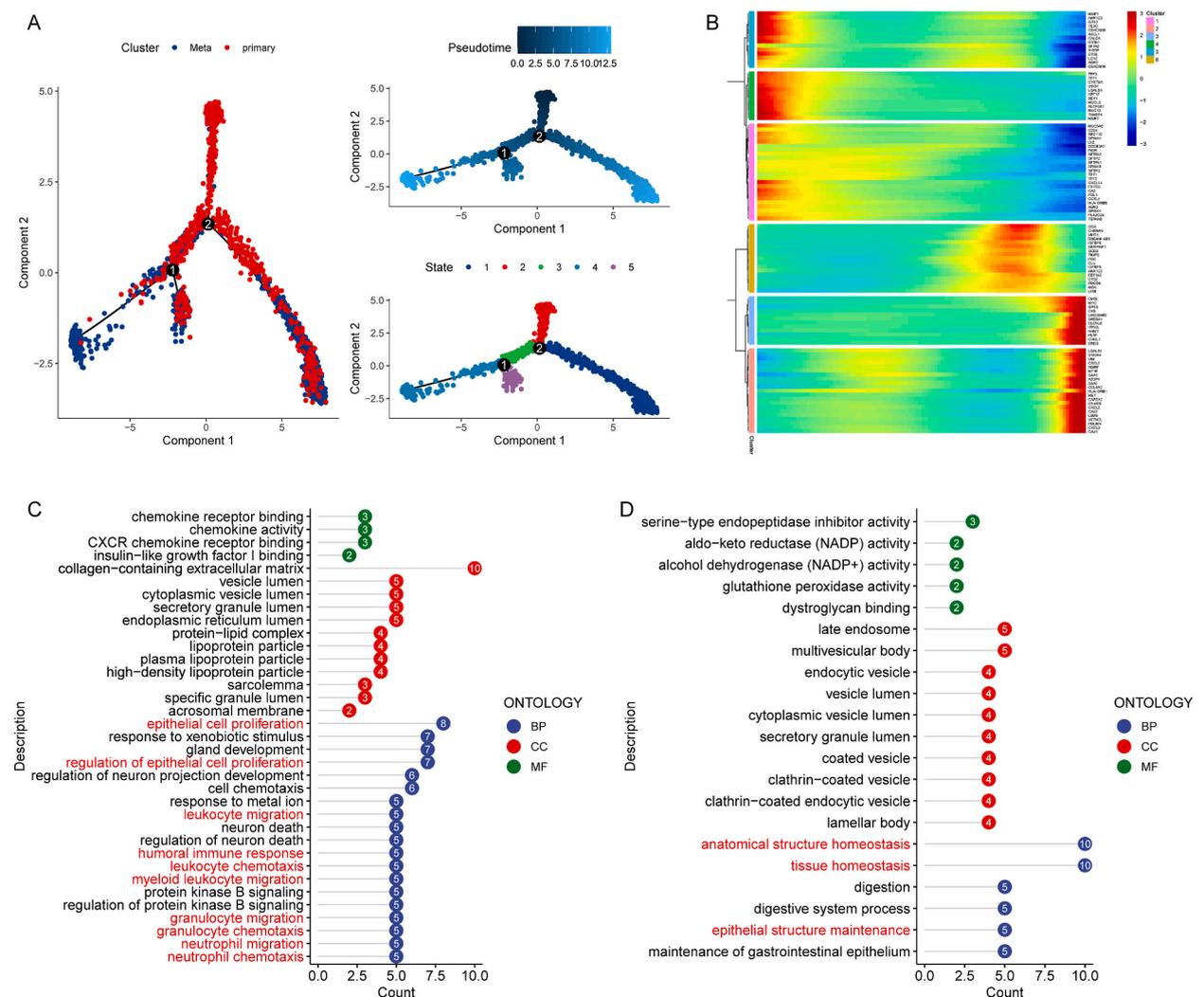


Fig. 3. Cell trajectory analysis of tumor cells. (A) All tumor cells differentiation trajectories, pseudotime distribution, and cell clusters on pseudotime. (B) Top 50 highly expressed genes and top 50 lowly expressed genes in pseudo-time. (C) Gene Ontology Functional Enrichment Analysis of Top 50 highly expressed genes. (D) Gene Ontology Functional Enrichment Analysis of Top 50 lowly expressed genes. *BP:Biological Process, CC:Cell Component, MF:Molecular Function. And the numbers on the x-axis represent the number of genes enriched in each pathway.

According to the marker gene expression in clusters, cells were mainly annotated as epithelial cells (cell marker: EPCAM, CDH1, KRT7, KRT19, KRT8 and KRT18), immune cells (cell marker: PTPRC, CD3E, GZMB, CD68, CD79A, and MS4A2), and stromal cells (cell marker: DCN, COL1A1, COL1A2, THY1, PECAM1, CLDN5, and RAMP2) (Fig. 1B–C). For a more refined annotation of immune cells, we extracted them and re-performed dimensionality reduction and clustering (Fig. 1D). Similarly, we annotated each of the seven major cell types based on their cell-specific gene expression marker (Fig. 1E–F). In the UMAP graph for all cells, cells of the same type were adjacent and showed large genomic differences between different cell types (Fig. 1G). Most immune or stromal cells of the same type were adjacent, while a larger distance was detected between epithelial cell clusters. Since LUAD is derived from epithelial cells [20, 21], this result reflected the heterogeneity between LUAD cancer cells. After counting the proportion of cells in each sample, we found that the cell proportion distribution also had a large difference between samples (Fig. 1H). Altogether, these results suggest that there is significant heterogeneity in LUAD cells.

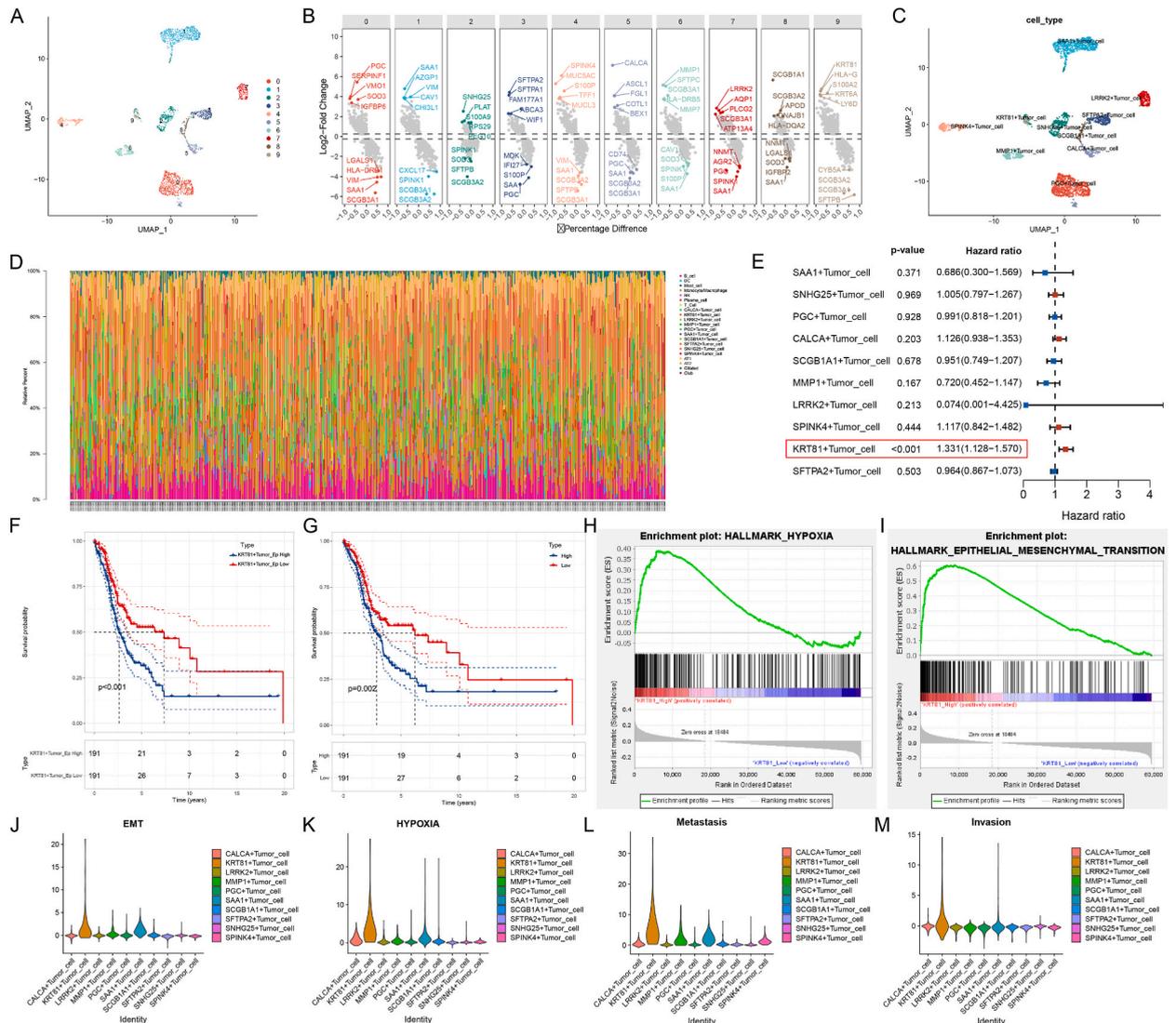


Fig. 4. KRT8⁺ tumor cells proportion are associated with worse prognosis in lung adenocarcinoma. (A) All tumor cells were divided into 10 clusters. (B) Top 5 highly expressed genes and top 5 low expressed genes in each tumor cell cluster. (C) Ten tumor cell clusters were named according to the highest differentially expressed genes. (D) Deconvolute the TCGA-LUAD cohort using the CibersortX algorithm to infer the proportion of each cell type in each sample. (E) Univariate COX analysis of the proportion of each tumor cell cluster in the TCGA-LUAD cohort, p-value <0.05 was considered to be associated with prognosis. (F) Survival analysis of the TCGA-LUAD cohort grouped according to the median proportion of KRT8⁺ tumor cells. (G) Survival analysis of the TCGA-LUAD cohort grouped according to the median KRT81 gene expression. (H–I) GSEA functional enrichment analysis of the TCGA-LUAD cohort grouped according to the median KRT81 gene expression. (J–M) Gene set scoring in single-cell datasets using the AddModuleScore function of the "Seurat" R package.

2.2. Copy number analysis of epithelial cells in LUAD

To exclude the incorporation of normal epithelial cells into tumor samples due to sampling, we used a previously published method [16] to infer the exact malignant cells using copy number variation (CNV). According to previously published articles, epithelial cells with high CNV were considered as malignant cells [22–24]. In our result, almost all tumor tissues, most epithelial cells presented higher CNV compared with artificially inserted epithelial cells derived from normal lung tissues. We defined these cells as tumor cells and identified the remaining fraction as normal lung epithelial cells incorporated during tissue sampling (Fig. 2A). Further, we used inferCNV [25] to characterize each chromosome and showed good agreement with the above methods. Notably, we found significant copy number deletions on chromosome 13 in almost all tumor cells (Fig. 2B). Previous studies have reported frequent loss of heterozygosity of alleles in the chromosome 13 region in lung cancer, leading to the inactivation of multiple tumor suppressor genes [26–28]. Therefore, CNV and loss of heterozygosity on chromosome 13 might be associated with lung cancer development. Next, all normal epithelial cells were re-clustered and cell-type annotated. We identified normal epithelial cells as alveolar type I, alveolar type II, secretory club, and airway ciliated cells (Fig. 2C–E). After integrating all epithelial cells and performing re-clustering, AT2 and Club cells were closer to tumor cells, which might be related to previous studies indicating that AT2 and Club cells are potential LUAD progenitors [21,29] (Fig. 2F).

2.3. Trajectory analysis of LUAD tumor cells reveals heterogeneity in lung adenocarcinoma

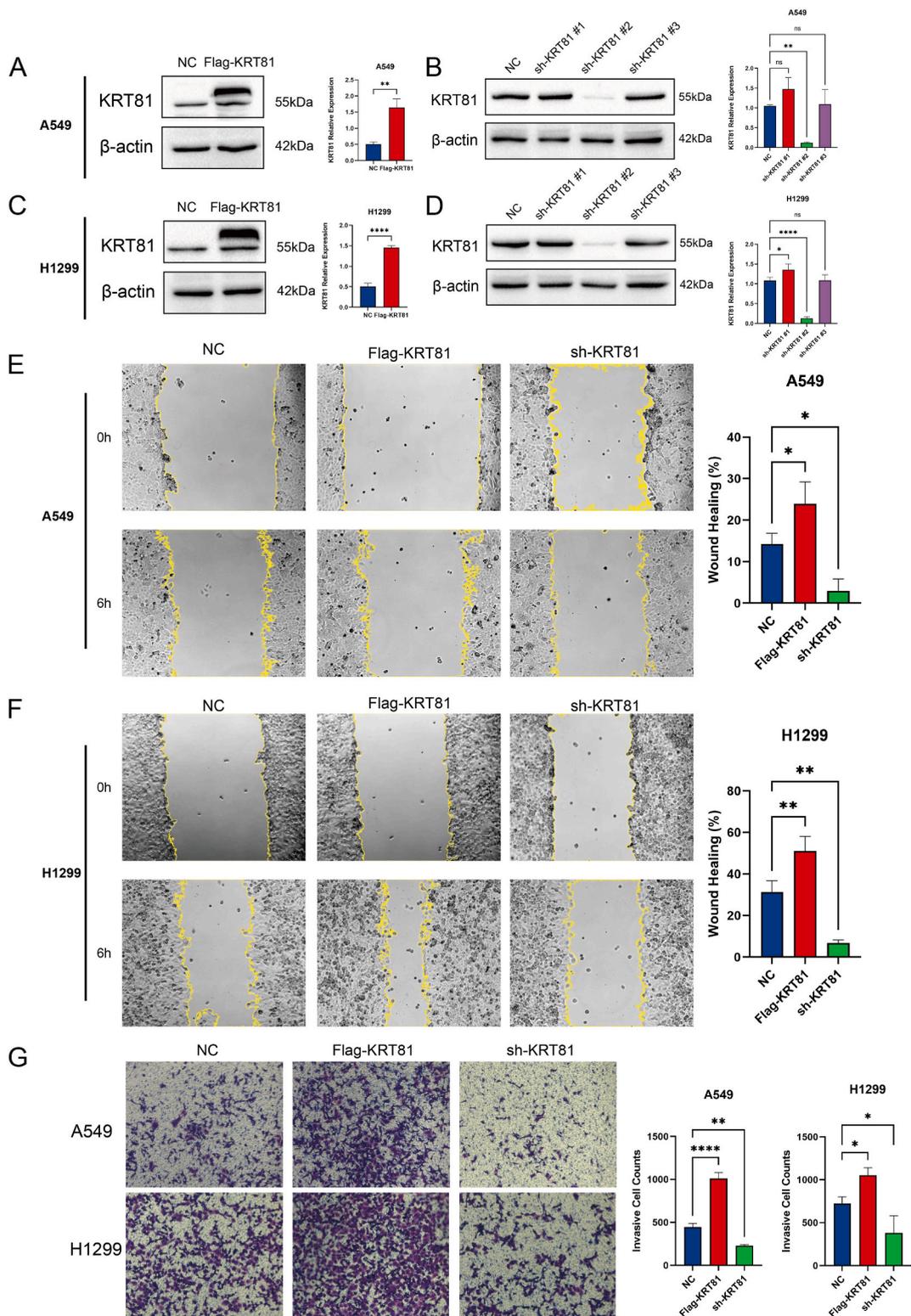
Subsequently, We performed trajectory analysis on all tumor cells using the 'moncel2' R package, an algorithm that can identify cellular transcriptional heterogeneity from single-cell data. We perform trajectory inference on all tumor cells. The two trajectory inflection points identified showed differences in the heterogeneity of lung adenocarcinoma tumor cells. In addition, we also identified developmental pseudotimes and five cell subpopulations on the trajectory (Fig. 3A). In pseudo-time, primary tumor cells showed three different developmental trajectories to metastatic tumor cells, illustrating the heterogeneity presented by the development of primary tumor cells into metastatic LUAD tumor cells (Fig. 3A). In pseudo time, we identified the top 100 hypervariable genes. These genes were clustered into 6 clusters according to their pseudotemporal expression, with a total of 50 upregulated and 50 downregulated (Fig. 3B). After Gene Ontology (GO) functional enrichment analysis, we found that upregulated genes were involved in epithelial cell proliferation, chemotaxis, and migration of various immune cells. However, downregulated genes were enriched in the maintenance and homeostasis of epithelial cells (Fig. 3C–D and Tables S2–S3). These results indicated that uncontrolled cell proliferation and loss of homeostasis maintenance are involved in the heterogeneous metastasis of lung cancer tumor cells and can regulate chemotaxis and migration of various immune cells.

2.4. KRT81⁺ tumor cells proportion are associated with poor prognosis in LUAD patients

To further study the potential functions of these heterogeneous tumor cells. We further clustered all tumor cells and named them according to the characteristically expressed genes of each cluster (Fig. 4A–B). All 10 clusters emerged as unique clusters in the UMAP plot, demonstrating large differences in their transcriptome status (Fig. 4C). Since single-cell data sets are limited by sample size and prognostic follow-up information, we used the CIBERSORT [5] algorithm to deconvolute the TCGA-LUAD dataset to evaluate the impact of the proportion of these cell clusters on prognosis (Fig. 4D). The clinical information of the TCGA-LUAD dataset has been summarized in Table S4. Then, in the TCGA-LUAD cohort, we performed a univariate Cox analysis of the estimated proportion of each tumor cell cluster. The proportion of KRT81⁺ tumor cells was a risk factor for the poor prognosis of LUAD patients (Fig. 4E). The survival curves also confirmed this result for KRT81⁺ tumor cell proportion or KRT81 gene expression grouping (Fig. 4F–G). In addition, we also examined the effect of the expression of another high-expression gene HLA-G and a low-expression gene SFTPB of KRT81⁺ cell cluster on the prognosis in lung adenocarcinoma. The results showed that the expression level of HLA-G was not related to the prognosis, but the patients with high expression of SFTPB had a better prognosis (Figs. S2A–B). However, it is worth noting that when we examined the expression of these three genes in all epithelial cells in single-cell dataset, only KRT81 was uniquely expressed in tumor cells, and SFTPB was more expressed in normal alveolar epithelium (Figs. S2C–E). Therefore, the association between KRT81⁺ cell cluster and prognosis may be more driven by the KRT81 gene. We also examined the association of KRT81 gene expression with prognosis in other cancers using univariate COX, and the results showed that it was also associated with poorer prognosis in seven other cancers from the TCGA database. But it is worth noting that KRT81 was only associated with poorer prognosis in lung adenocarcinoma compared with lung squamous cell carcinoma (Fig. S2F). Therefore, the association of KRT81 with poor prognosis is not specific to lung cancer, but its role in lung adenocarcinoma is relatively specific for other lung cancer pathological subtypes. In the TCGA-LUAD cohort, we detected significant activation of hypoxia and epithelial-mesenchymal transition pathways in the high KRT81 expression group (Fig. 4H–I). This result was consistent with the higher scoring of gene sets for epithelial-mesenchymal transition (EMT), hypoxia, metastasis, and invasion phenotypes for the KRT81⁺ cell cluster in the scRNA-seq data (Fig. 4J–M). Overall, we identified a new cluster of tumor cells associated with poor prognosis in LUAD, KRT81⁺ tumor cells, that might be associated with the activation of EMT and hypoxia pathways.

2.5. Experimental validation of the influence of KRT81 on migration and invasion in lung adenocarcinoma cell lines

Through the analysis above, we conclude that KRT81 may be associated with a poorer prognosis in lung adenocarcinoma. Previous research indicates that KRT81 enhances cell migration and invasion in breast cancer [30] and melanoma [31]. Therefore, to



(caption on next page)

Fig. 5. Experimental validation in lung adenocarcinoma cell lines. (A–D) KRT81 protein expression was detected by Western blot in A549 and H1299 cell lines transfected with overexpression or knockdown plasmids. The relative expression of KRT81 was calculated for statistical analysis. (E) After the overexpression or knockdown of KRT81 in A549 and H1299 cells in the 6-well plate reached 100% growth density, artificial wounds were made and replaced with serum-free medium. Wound pictures were taken at 0h and 6h respectively, and the percentage of wound healing was calculated. Percentage of wound healing = (Area of wound at 0h - Area of wound at 6h)/Area of wound at 0h. (F) Inoculate the same number of control cells and cells with overexpression or knockdown of KRT81 in the upper chamber. After 24 h, stain and photograph the cells in the lower chamber. Count the number of cells that have invaded the lower chamber and perform statistical analysis between groups. In above figures: * represents p-value <0.05, ** represents p-value <0.01, *** represents p-value <0.001, **** represents p-value <0.0001.

investigate its role in lung adenocarcinoma cell lines, we carried out the following experimental validation. We constructed an overexpression plasmid marked with Flag tag for KRT81 and three shRNAs for KRT81 knockdown, and transfected them into A549 and H1299 cell lines, with the transfection effects confirmed by Western blot. The results demonstrated that the overexpression plasmid achieved good overexpression of KRT81 in both cell lines, and among the three shRNAs, the 2nd one yielded good knockdown effect on KRT81 at the protein level (Fig. 5A–D). Hence, sh-KRT81#2 was used for the subsequent experiments. The wound healing experiment

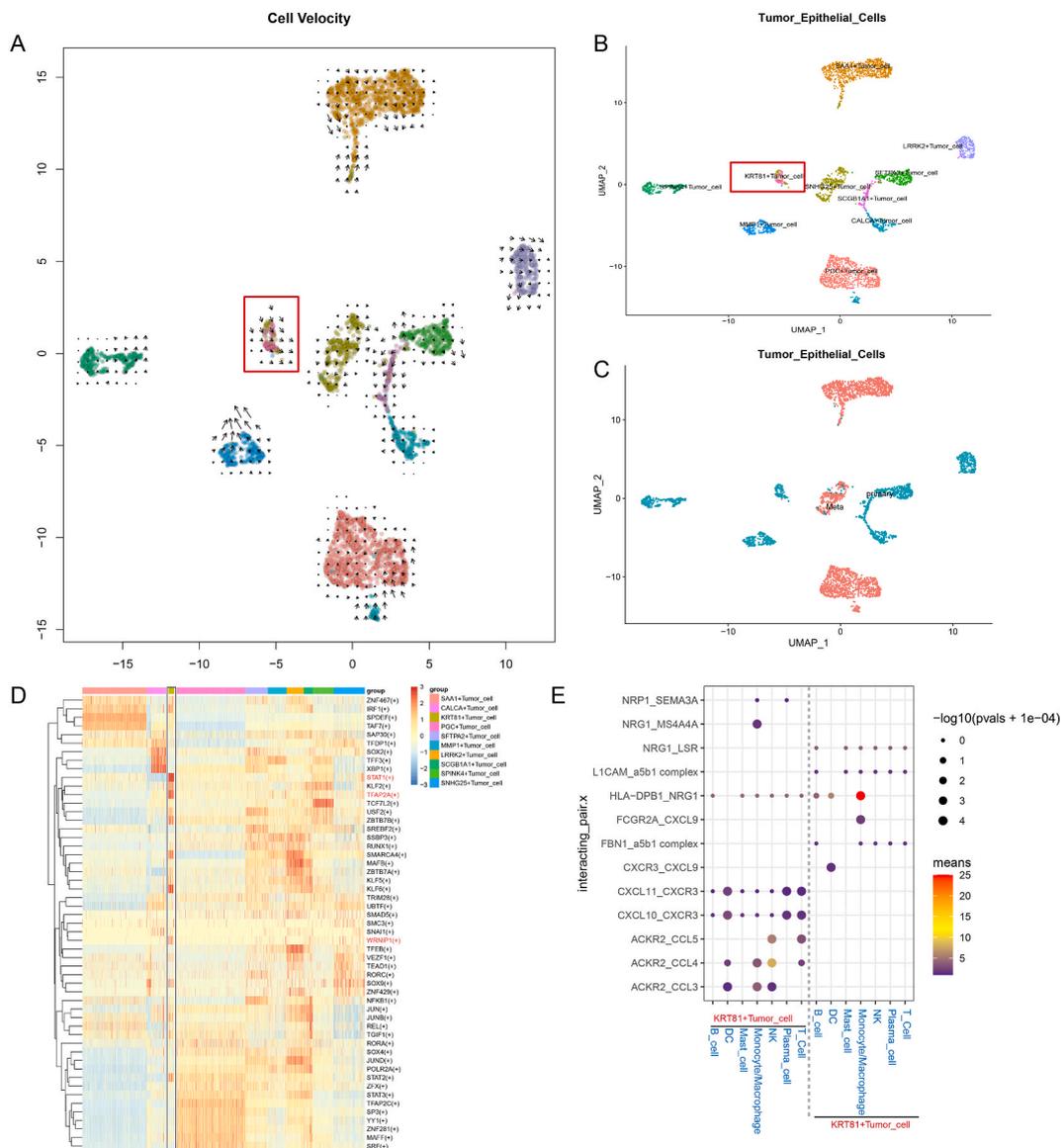


Fig. 6. Transcriptional characterization of KRT81⁺ tumor cells and interactions with immune cells. (A) RNA velocity analysis of malignant epithelial cells. (B) UMAP plot of all malignant epithelial cells. (C) All tumor epithelial cells are annotated according to primary or metastatic origin. (D) Transcriptional factor analysis of all malignant epithelial cells. (E) Analysis of ligand-receptor interaction between KRT81⁺ tumor cells and immune cells using "CellPhoneDB" R package.

was used to measure cell migration ability. The results indicated that compared to the control group, the wound healing ratio in A549 and H1299 cells overexpressing KRT81 significantly increased, while the two cell lines with knockdown of KRT81 demonstrated a decrease in migration ability (Fig. 5E–F). The Transwell assay indicated that cells overexpressing KRT81 invade more into the lower chamber, whereas the number of invasive cells decreased following the knockdown of KRT81 (Fig. 5G). In conclusion, these experimental results indicate that KRT81 promotes the migration and invasion of LUAD cell lines.

2.6. Transcriptional characterization of KRT81⁺ tumor cell cluster and interactions with immune cells

Based on the association of KRT81⁺ tumor cells with worse prognosis in LUAD patients, we further analyzed this tumor cell cluster. To study the transcriptional state of KRT81⁺ tumor cells, we used a python-based "Velocity" algorithm. This algorithm characterizes RNA velocity based on the ratio of spliced and unspliced mRNA and infers developmental direction [32]. In our results, KRT81⁺ tumor cell cluster (a subset of primary tumor cells) had a tendency to develop into metastatic tumor cells. Thus, such clusters of tumor cells may contribute to disease progression (Fig. 6A–C). Moreover, in the transcriptional factor analysis of various tumor cell clusters, we identified two transcription factors of interest, TFAP2A and WRNIP1, in KRT81⁺ tumor cells (Fig. 6D). TFAP2A has been reported to promote LUAD progression by activating the EMT [33–35] and promoting angiogenesis in acquired resistance [36]. WRNIP1 is associated with cell cycle promotion in LUAD [37] and tumor radioresistance [38].

Furthermore, we analyzed the interaction of KRT81⁺ tumor cells with immune cells using the "CellphoneDB" R package [39], an algorithm for analyzing cell and cell-cell interactions in single-cell datasets, to study their role in the tumor microenvironment (Fig. 6E). The NRP1/SEMA3A interacting pair can recruit macrophages into hypoxic tumor areas to promote angiogenesis and suppress antitumor immunity [40]. Additionally, the CXCL9, -10, -11/CXCR3 axis has dual roles. Its autocrine axis can promote tumor cell proliferation and metastasis. In contrast, its paracrine axis can activate the migration and differentiation of various immune cells [41]. Hence, inhibiting its autocrine and activating the paracrine axis might be a potential cancer therapeutic target. This idea was experimentally validated in preclinical models of colon cancer [42], breast cancer [43,44], and osteosarcoma [45] but has not yet been reported in LUAD. Herein, we detected a significant enrichment of the CXCL9, -10, -11/CXCR3 axis in KRT81⁺ tumor cells, indicating that this strategy might equally possess potential therapeutic value in LUAD. In summary, these results demonstrated the potential of the KRT81⁺ tumor cell cluster to differentiate into metastatic LUAD cells, which might be mediated by the characteristic transcription factors TFAP2A and WRNIP1. We also revealed the complex regulatory role of the KRT81⁺ tumor cell cluster in the tumor microenvironment.

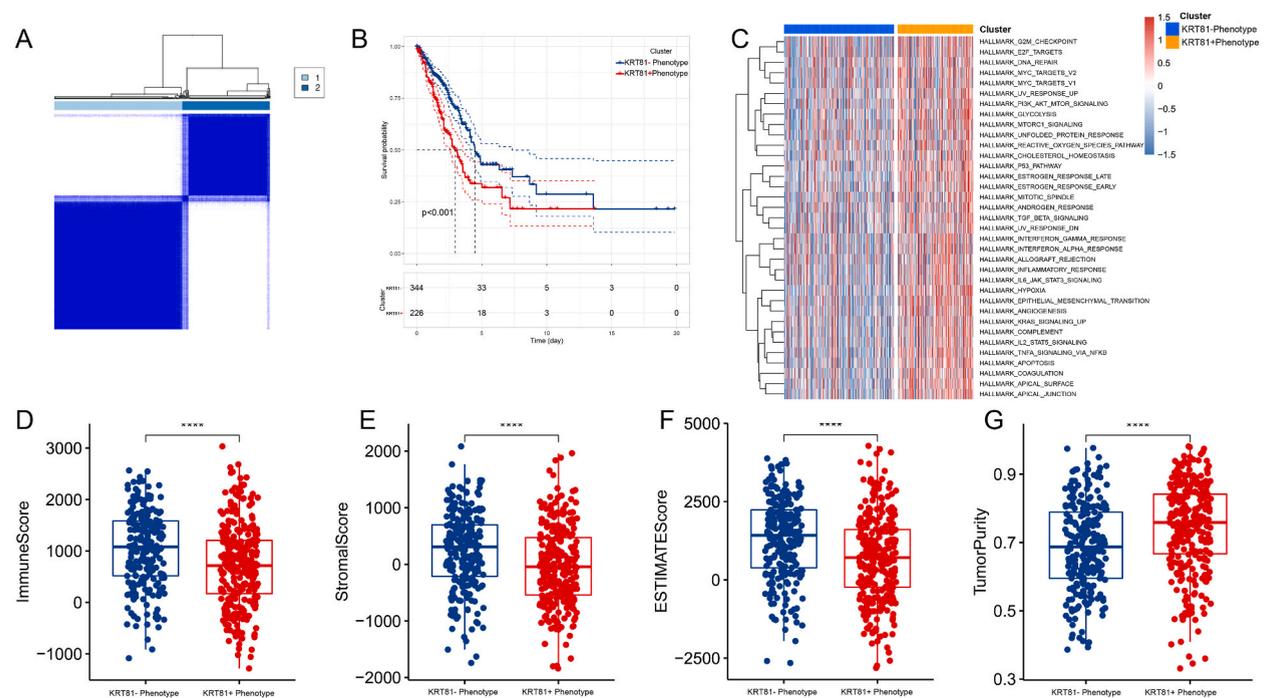


Fig. 7. Identification of KRT81⁺ Phenotypes. (A) Consensus cluster analysis of the TCGA-LUAD cohort using KRT81⁺ tumor cell marker gene expression. (B) Survival analysis of two groups after consensus clustering. (C) GSEA analysis of two groups using the HALLMARK gene set. (D–G) ESTIMATE algorithm estimates immune, stromal and tumor purity score.

2.7. The phenotype of KRT81⁺ tumor cells can be used as a prognostic biomarker in LUAD patients

To further assess the clinical value of KRT81⁺ tumor cells, we performed consensus clustering of all TCGA-LUAD cohorts based on the expression of KRT81⁺ tumor cell markers. We divided them into two clusters according to the expression of KRT81⁺ tumor cell markers: KRT81⁺ and KRT81⁻ phenotypes (Fig. 7A and Fig. S2). The KRT81⁺ phenotype was associated with poor prognosis (Fig. 7B) and was significantly enriched in various hallmark gene sets in the GSVA analysis (Fig. 7C). Besides, We used the ESTIMATE algorithm [4] to infer immune-stromal cell scores in TCGA-LUAD data. The KRT81⁺ phenotype presented lower immune stromal cell infiltration but higher tumor purity (Fig. 7D–G). The poor prognosis for the KRT81⁺ phenotype might be associated with antitumor immunosuppression.

Next, we performed differential gene expression analysis between the two clusters to investigate the power of the two phenotypes

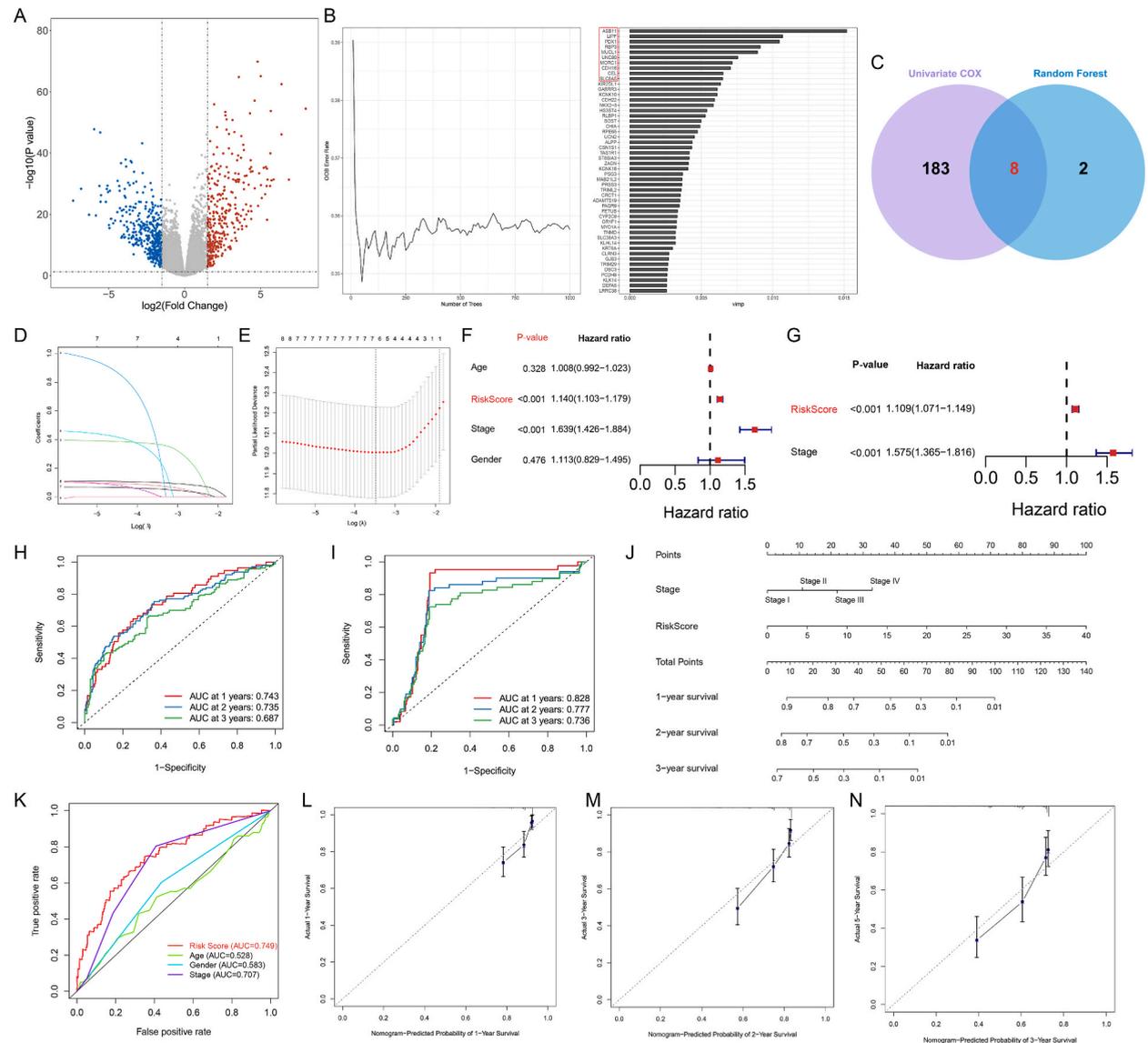


Fig. 8. KRT81⁺ tumor cells phenotype can be used as a prognostic biomarker in patients with lung adenocarcinoma. (A) Identification of differentially expressed genes in two groups. (B) Random Forest algorithm to evaluate the effect of DEGs on survival. (C) The intersection of univariate COX result and the top ten importance in random forest identified robust prognosis-related genes. (D–E) Further screening of prognosis-related genes by LASSO algorithm. (F) Univariate Cox analysis of Risk Score and clinical index in the TCGA-LUAD cohort. (G) Multivariate Cox analysis of Risk Score and clinical index in the TCGA-LUAD cohort. (H) ROC curves of the predictive ability of the prediction model for the 1, 2, and 3-year prognosis in the training dataset (TCGA-LUAD). (I) ROC curves of the predictive ability of the prediction model for the 1, 2, and 3-year prognosis in the testing dataset (GSE30219 and GSE31210). (J) Constructing a nomogram for prognosis prediction in the TCGA-LUAD dataset. (K) ROC curves of risk score and clinical information on prognosis prediction. (L–N) Calibration curves of the nomogram at 1, 2, and 3 years.

in predicting LUAD prognoses (Fig. 8A). We assessed the predictive effect of all differentially expressed genes on prognosis using univariate Cox (Table S5) and Random Forest (Fig. 8B) algorithms. Finally, we obtained eight robust prognosis-related genes (ASB11, MUCL1, LIPF, SLC6A5, UNC80, RBP3, CDH16, PDX1) based on the intersection of genes with $p < 0.05$ in the univariate Cox results and top 10 ranked genes with importance in the Random Forest algorithm (Fig. 8C). To prevent model overfitting, we used the least absolute shrinkage and selection operator (LASSO) algorithm to screen the variables, and the smallest λ was detected when the six variables were retained. Therefore, the prognostic model was constructed for these six genes (ASB11, MUCL1, LIPF, SLC6A5, UNC80, and CDH16) (Fig. 8D–E).

$$\text{RiskScore} = 0.258042 \times \text{ASB11} + 0.157409 \times \text{MUCL1} + 0.157873 \times \text{LIPF} + 0.172779 \times \text{SLC6A5} + 0.129903 \times \text{UNC80} + 0.200211 \times \text{CDH16}$$

According to our constructed model, we calculated the risk score for each TCGA-LUAD patient. The p -value of the risk score was < 0.05 in both univariate and multivariate Cox analyses, indicating that the risk score can be used as an independent prognostic risk factor for LUAD patients (Fig. 8F–G). Next, we used ROC curves to evaluate the ability of the model to predict prognosis. Since the proportion of TCGA-LUAD patients who survived more than 5 years was small, we only assessed the short-term prognosis. The ROC curves at 1, 2, and 3 years demonstrated the model's good predictive ability. Additionally, the AUC values observed at 1, 2, and 3 years were > 0.7 in the external GEO cohort (Fig. 8H–I). Additionally, we compared the predictive performance of our model with previously published prognostic models [46–49] constructed using LUAD scRNA-seq data in testing dataset. Our model demonstrated better predictive ability (Figs. S4A–C). These results indicated that our constructed model has a good short-term survival predictive ability for LUAD patients. After contrasting the prognostic predictive ability with other clinical indexes, we found that the risk score had similar AUC values to the clinical stage (Fig. 8K). Therefore, we constructed a nomogram combining the risk score and clinical stage to enhance the current use of the clinical stage alone and to guide patients requiring further intervention (Fig. 8J). The nomogram calibration curves showed good accuracy for short-term prognosis (Fig. 8L–N). Overall, our current findings suggested a novel subtype of LUAD patients with the KRT81⁺ tumor cell phenotype and proposed a prognostic model associated with this phenotype to assist the currently used clinical stage and to enhance the prediction of short-term prognosis and clinical treatment guidance.

3. Discussion

Despite the widespread use of targeted and immunotherapy for advanced LUAD, relapse and drug resistance remain major concerns [50,51]. Thus, deepening the understanding of LUAD and finding new therapeutic targets and prognostic monitoring biomarkers is urgent. RNA-seq has already provided strong evidence on disease diagnosis, prognostic monitoring, and selection of therapeutic targets. However, this strategy uses whole tissue transcriptome sequencing and ignores intratumoral heterogeneity [3,52]. Notably, intratumoral heterogeneity is one of the main causes of recurrence and drug resistance after treatment [53–57]. The scRNA-seq provides transcriptome data in a single cell dimension to precisely address this problem, allowing a deeper understanding of heterogeneity at the cell level [58]. For LUAD, several articles related to scRNA-seq have delineated the landscape in an atlas approach and provided valuable raw data for subsequent studies [15–17,19]. Therefore, searching for new prognostic biomarkers or therapeutic targets in LUAD from the scRNA-seq perspective has advantages over bulk RNA-seq that ignore intratumoral heterogeneity. Herein, we investigated the heterogeneity of malignant epithelial cells in LUAD by combining scRNA-seq and bulk RNA-seq. We found a new malignant epithelial cell cluster associated with prognoses and conducted an in-depth study of its characteristics and clinical significance.

We collected scRNA-seq data from four normal lung tissues, eight primary LUAD, and five distant metastatic LUAD tissues. After dimensionality reduction, clustering, and cell type annotation, heterogeneous tumor purity and immune cell infiltration were reflected in each sample, indicating the importance of individualized treatment for LUAD patients. Regarding the CNV, most tumor cells have higher CNV than normal lung epithelium. Previous studies have found extensive amplification of chromosome 3q in lung squamous cell carcinoma [59]. However, we did not observe it in our LUAD analysis. Additionally, on chromosome 13, LUAD tumor cells presented significant copy number loss. Nevertheless, due to the limited sample size and clinical information investigated, how to correlate this CNV signature with potential clinical predictive value remains to be further determined. In the pseudo-time analysis of all primary and metastatic LUAD tumor cells, we observed a heterogeneous developmental orientation rooted in the primary tumor cells, with malignant epithelial cell proliferation activated and structural homeostasis disrupted during the development of heterogeneous metastatic tumor cells. We also observed the activation of neutrophil chemotaxis and migration pathways. The notion that neutrophils can promote LUAD cell proliferation and metastasis [60] and are associated with worse prognosis [61] has already been validated in previous studies. To further investigate the potential association of these heterogeneous tumor cells with clinical outcomes, we divided all tumor cells into ten clusters and named them by highly expressed genes. Based on highly expressed genes from each cluster to deconvolute bulk RNA-seq data from TCGA-LUAD, we obtained more clinical information and found that KRT81⁺ tumor cells were associated with a worse LUAD prognosis. Besides, the high KRT81 expression group was enriched for EMT and hypoxia pathways and were validated in the scRNA-seq data. The EMT can lead to tumor cell metastasis by enhancing the migration, invasion, and resistance to apoptotic stimuli of tumor cells. EMT-derived tumor cells can also acquire stemness and significant therapeutic resistance [62,63]. The activation of the hypoxia pathway can promote tumor angiogenesis and cell proliferation [64,65]. Hypoxia can also induce the activation of the EMT pathway [66]. The synergistic effect of hypoxia and EMT might be responsible for the association between KRT81⁺ tumor cells and poor prognosis. Subsequently, we revealed that KRT81⁺ tumor cells tend to differentiate and metastasize, and that their characteristic transcription factor TFAP2A might be responsible for the activation of the EMT pathway. Moreover, activation of the NRP1/SEMA3A ligand-receptor pair by KRT81⁺ tumor cells in the tumor microenvironment recruits macrophages to

hypoxic tumor areas, promoting angiogenesis and suppressing antitumor immunity [40]. Hence, targeting TFAP2A and NRP1 might potentially treat KRT81⁺ tumor cells. Strategies targeting NRP1 have been reported in previous studies [67], but our findings of TFAP2A might comprehend an entirely novel potential therapeutic target. Previous studies have reported that KRT81 promotes the migration and invasion of breast cancer [30]. Besides, KRT81 knockdown inhibits the proliferation, invasion, and migration of melanoma and promotes apoptosis by downregulating the expression of IL-8 [31]. Although previous studies have reported that KRT81 is a risk factor for the prognosis of lung adenocarcinoma [68], no further research has been done on the role of cell cluster with high expression of KRT81. Therefore, this study focused on revealing the prognostic impact and potential functions of KRT81 high-expression tumor cell cluster in single-cell data sets. Further analysis showed that the new LUAD subtypes divided according to the marker genes of KRT81⁺ tumor cells had significant prognostic differences and had a strong predictive ability for the short-term prognosis of LUAD patients. In contrast to previous studies, our research not only involved the construction of a prognostic prediction model but also delved into in-depth investigations. These investigations encompassed exploring the activation of their transcription factors and their interactions with immune cells. Moreover, we identified several potential therapeutic targets specific to this cell subpopulation.

In conclusion, we combined scRNA-seq and bulk RNA-seq data from LUAD to profile the heterogeneity of LUAD tumor cells. We identified a novel cluster of KRT81⁺ tumor cells that might tend to metastasize through the activation of the EMT and hypoxia pathways. Furthermore, we showed that the KRT81⁺ tumor cell phenotype has a good predictive ability for the short-term prognosis of LUAD patients. Our study provides new ideas for predicting the prognosis and treatment strategies of LUAD. But for the current situation, single-cell data sets often lack large sample size and prognosis follow-up data due to high cost. We have to try our best to combine bulk RNA-seq datasets with single-cell datasets to make up for each other's shortcomings, so developing robust integration algorithms is a matter of concern.

4. Materials and methods

4.1. Cell culture

Human A549 and NCI-H1299 cell lines were purchased from Procell Life Science & Technology Co. Ltd., located in Wuhan, China. Both cell lines were cultured using RPMI-1640 medium (Gibco, USA) supplemented with 10% fetal bovine serum (VivaCell Biosciences, Shanghai XP Biomed Ltd.) and incubated at 37 °C with 5% CO₂.

4.2. Datasets and processing methods

The raw human lung adenocarcinoma single-cell RNA-seq data were obtained from The European Nucleotide Archive (ENA) database (Accession ID: PRJNA510251 [69]). Clinical information for all samples is available from the original publication [69]. Raw sequencing data were matched to the genome and quantified for gene expression following Cell Ranger version 7.0 software Official running pipelines. The R package "Seurat" [70] was used for merge and quality control of multiple sample data. Cells with nFeature between 300 and 7500, nCount between 300 and 100000 and Mitochondrial genes expression less than 20% were retained for subsequent analysis. Data were normalized using the "SCTransform" method followed by dimensionality reduction clustering and visualization. In addition, the annotation of cell types is a characteristic marker of various cells summarized according to previous published articles and CellMarker database (<http://xteam.xbio.top/CellMarker/>). The Bulk RNA-seq and microarray data used in this study were obtained from the TCGA and GEO databases (TCGA-LUAD, GSE30219 and GSE31210).

4.3. Copy Number Analysis

For inferring normal epithelium mixed in tumor samples, we used methods published in a previous article [16]. Briefly, epithelial cells from each tumor sample were first extracted and incorporated into a certain number of normal epithelial cells from normal samples so that the number of epithelial cells from the tumor samples were below 20%. Genes expressed in fewer cells were eliminated as background genes. Following conversion of expression to z-score, copy number signals were inferred using expression. Two measures were used to determine cell types, which were the estimated mean squared CNV signal and the correlation between each cell and the mean copy number of the top 5% of cells (Code and detailed instructions are available at <https://github.com/SGL-LungCancer/SingleCell>). We used inferCNV [71] algorithm to revalidate the above results and characterized the copy number landscape of each chromosome in detail. Eventually all normal epithelial cells underwent dimensionality reduction clustering, visualization and cell-type annotation using the Seurat R package.

4.4. Cell trajectory analysis

Cell trajectory analysis was performed using the Monocle [72] R package. The parameters used in the analysis were all the default parameters, and the cell trajectories were visualized using the "DDRTree" method, setting the primary tumor cells as the pseudo-time root, and searching for differential expressed genes in the pseudo-time sequence. Eventually, we extracted the top 50 genes that were up- or down-regulated according to the pseudo-time and performed GO functional enrichment analysis using the clusterProfiler [73, 74] R package.

4.5. Identification of tumor cell cluster associated with poor prognosis

All tumor cells were subjected to dimensionality reduction, clustering and visualization using the Seurat R package and identification of cluster signature genes using the FindallMarker function. The clusters were annotation according to highest expressed genes of each cluster. Similarly, we used the FindallMarker function in all other cell types to search for signature genes for each cell type. For all cell clusters, we calculated the average expression profile of the top 50 highly expressed genes in each cell and combined them as a reference data set. Finally, the TCGA-LUAD bulk RNA-seq data was deconvoluted with our constructed reference dataset using the CibersortX [75] algorithm to estimate the proportion of various cell clusters in TCGA lung adenocarcinoma samples. Univariate Cox analysis was used to identify risk factors associated with poor prognosis. Finally, we grouped patients by median cell proportion or median gene expression, and plotted survival curves using the "survival" R package. The EMT and hypoxia pathway gene sets used for functional enrichment are from the "MsigDB" database, and the metastasis and invasion pathway gene sets are from the "cancerSEA" database.

4.6. Plasmid transfection

The KRT81 gene overexpression and shRNA plasmids were purchased from Shanghai GeneChem Co.,Ltd. The sequence for KRT81 shRNA is as follows:

KRT81-shRNA#1: 5'-GATCCCGCGGCCAATTGAACACCACCTCTCGAGAGGTGGTGTTC AATTGGCCGCTTTTGGAT-3'; KRT81-shRNA#2: 5'-GATCCCGCTCACATTTCTCTGTGTGATCTCGAGATCACACAGAGAAATGTGAGGTTTTGGAT-3'; KRT81-shRNA#3: 5'-GATCCCAGGCATTGGGGCTGTGAATGTCTCGAGACATTCACAGCCCCAATGCCTTTTGGAT-3'.

A549 and H1299 cell lines were seeded in six-well plates and cultured overnight until reaching a density of 50%–80%. Transfection was performed using Lipofectamine 2000 transfection reagent (Invitrogen; CA, USA) according to the manufacturer's instructions. Furthermore, the cells were cultured in the presence of G418 (Beyotime Biotechnology, Shanghai, China) at a concentration of 500 µg/ml for one week to perform resistance selection.

4.7. Western blot

After removing the culture medium, the cells were washed three times with pre-chilled PBS buffer. Total cellular proteins were extracted using RIPA lysis buffer (Beyotime Biotechnology, Shanghai, China) supplemented with a protease inhibitor cocktail (Beyotime Biotechnology, Shanghai, China) and PMSF (Beyotime Biotechnology, Shanghai, China). The protein concentration was determined using the BCA protein assay kit (Thermo Fisher Scientific; CA, USA). Appropriate SDS-PAGE loading buffer was added to prepare protein samples for Western blot analysis. The samples were resolved by 10% SDS-PAGE gel electrophoresis, with a protein ladder (Thermo Fisher Scientific; CA, USA) serving as the molecular weight marker. After electrophoresis, the proteins were transferred onto a PVDF membrane (Millipore, USA) for subsequent antibody incubation. The primary antibody used in this study were anti-KRT81 (Proteintech, Wuhan, China, 1:1000) and anti-βactin (Proteintech, Wuhan, China, 1:10000), and the secondary antibody used was Goat anti-Rabbit HRP (Proteintech, Wuhan, China, 1:10000). The visualization of protein bands was achieved using the ECL detection kit (Beyotime Biotechnology, Shanghai, China).

4.8. Wound healing assay

The cells were seeded in six-well plates and allowed to reach 100% confluency. Scratch wounds were created using a 200 µl pipette tip, carefully scraping the cell monolayer. Subsequently, the cells were cultured in serum-free 1640 medium. At designated time intervals, photographs were captured to monitor the healing process of the scratch wounds at the identical locations.

4.9. Transwell assay

The Transwell assay was conducted using 24-well, 8 µm transwell chambers (NEST Biotechnology Co.LTD. Wuxi, China). The cells were dissociated into a single-cell suspension, and a seeding density of 1×10^5 cells was used for the upper chamber. The upper chamber was supplemented with serum-free medium, while the lower chamber contained medium with 10% FBS. After 24 h of incubation, the cells were fixed with 4% paraformaldehyde (Beyotime Biotechnology, Shanghai, China) and stained using 1% crystal violet (Beyotime Biotechnology, Shanghai, China). After gently removing the cells on the upper surface of the chamber, images were captured using an inverted microscope in at least three random fields of view.

4.10. RNA velocity, transcription factor and cellular interaction analysis

RNA Velocity were analyzed using the python-based "velocity [32]" package. BAM files for all samples were used in this pipeline to assess lineage developmental orientation in an unsupervised manner by the proportion of spliced mRNA and unspliced mRNA and mapped onto UMAP for visualization. Transcription factor analysis was performed using the python-based SCENIC [76] package to mine the characteristic transcription factors of single cell clusters by cis-regulatory analysis to provide biological insights into the heterogeneous mechanisms of cell clusters. Cell-to-cell communication analysis was performed using the python-based CellphoneDB [39] package. In processing of the results, we extracted ligand-receptor pairs between clusters of KRT81+ tumor cells and other cells

and screened them using means > 1.

4.11. Construction of prognostic model associated with KRT81⁺ Phenotype

In the TPM expression data from the TCGA-LUAD cohort, univariate COX analysis of the KRT81⁺ tumor cell marker genes obtained prognostic KRT81⁺ tumor cell markers. Expression of these genes in the TCGA-LUAD cohort was then consensus clustered using the "ConsensusClusterPlus [77]" R package. The cluster algorithm was set to k-means, and sampling proportion set to eighty percents. Eventually we identified two clusters and defined them as KRT81⁺ phenotype cluster and KRT81⁻ phenotype cluster based on the KRT81⁺ marker genes expression. Survival curves for the two clusters were plotted using the "survival" R package. Next, we performed functional enrichment analysis of the two clusters using the "GSVA" R package, and the HALLMARK gene set was obtained from MSigDB database. Differentially expressed genes from clusters were identified using the "edgeR" R package with |FoldChange| > 1.5 and p-value < 0.05 as criterion. To identify robust prognosis related DEGs, we used two algorithms for screening, univariate COX analysis and random forest. By intersection of genes with p-value < 0.05 in univariate COX analysis and top 10 genes of importance in random forest algorithm, we obtained eight robust prognosis-related genes. Random forest analysis used the "randomForestSRC" R package with the number of trees set to 1000 and node size set to 15. The LASSO algorithm used the "glmnet" R package, which retains seven genes for subsequent prognostic model construction when λ is a minimum. Prognostic models were constructed using multivariate COX analysis. The "timeROC" R package was used to plot ROC curves for prognostic models to assess the predictive effect of the model. For external validation of the model, we combined two GEO datasets (GSE30219 and GSE31210) that both used the GPL570 platform after removing batch effects using the "ComBat" algorithm and then converted to z-scores. The nomogram is constructed using the "rms" R package and the calibrate curve is plotted using "caret" R package.

4.12. Statistical analysis

The ImageJ software [78] was used for quantitative analysis of Western blot, wound healing, and Transwell assays, with all experiments conducted in triplicate for statistical analysis. The T-test was used for comparisons between two groups, while one-way ANOVA was employed for analysis of three or more groups.

Author contribution statement

Xu Ran: Performed the experiments; Analyzed and interpreted the data; Wrote the paper. </p>

Lu Tong; Wang Chenghao; Li Qi; Peng Bo; Wang Jun: Performed the experiments; Analyzed and interpreted the data. </p>

Zhao Jiaying: Contributed reagents, materials, analysis tools or data. </p>

Zhang Linyou: Conceived and designed the experiments. </p>

Data availability statement

Data included in article/supplementary material/referenced in article.

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.

Acknowledgments

We thank Dr. Jianming Zeng (University of Macau), and all the members of his bioinformatics team, biotrainee, for generously sharing their experience and codes.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2023.e20164>.

References

- [1] R.L. Siegel, K.D. Miller, H.E. Fuchs, A. Jemal, Cancer statistics, 2022, *CA Cancer J Clin* 72 (1) (2022) 7–33.
- [2] R.S. Herbst, D. Morgensztern, C. Boshoff, The biology and management of non-small cell lung cancer, *Nature* 553 (7689) (2018) 446–454.
- [3] C. Costa, A. Gimenez-Capitan, N. Karachaliou, R. Rosell, Comprehensive molecular screening: from the RT-PCR to the RNA-seq, *Transl. Lung Cancer Res.* 2 (2) (2013) 87–91.
- [4] K. Yoshihara, M. Shahmoradgol, E. Martinez, R. Vegesna, H. Kim, W. Torres-Garcia, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, *Nat. Commun.* 4 (2013) 2612.

- [5] A.M. Newman, C.L. Liu, M.R. Green, A.J. Gentles, W. Feng, Y. Xu, et al., Robust enumeration of cell subsets from tissue expression profiles, *Nat. Methods* 12 (5) (2015) 453–457.
- [6] E. Becht, N.A. Giraldo, L. Lacroix, B. Buttard, N. Elarouci, F. Petitprez, et al., Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression, *Genome Biol.* 17 (1) (2016) 218.
- [7] I. Dagogo-Jack, A.T. Shaw, Tumour heterogeneity and resistance to cancer therapies, *Nat. Rev. Clin. Oncol.* 15 (2) (2018) 81–94.
- [8] I. Vitale, E. Shema, S. Loi, L. Galluzzi, Intratumoral heterogeneity in cancer progression and response to immunotherapy, *Nat Med* 27 (2) (2021) 212–224.
- [9] P.R. Prasetyanti, J.P. Medema, Intra-tumor heterogeneity from a cancer stem cell perspective, *Mol. Cancer* 16 (1) (2017) 41.
- [10] Y. Lei, R. Tang, J. Xu, W. Wang, B. Zhang, J. Liu, et al., Applications of single-cell sequencing in cancer research: progress and perspectives, *J. Hematol. Oncol.* 14 (1) (2021) 91.
- [11] S.H. Gohil, J.B. Iorgulescu, D.A. Braun, D.B. Keskin, K.J. Livak, Applying high-dimensional single-cell technologies to the analysis of cancer immunotherapy, *Nat. Rev. Clin. Oncol.* 18 (4) (2021) 244–256.
- [12] R. Zilionis, C. Engblom, C. Pfirschke, V. Savova, D. Zemmour, H.D. Saaticioglu, et al., Single-cell Transcriptomics of human and mouse lung cancers reveals conserved myeloid populations across individuals and species, *Immunity* 50 (5) (2019) 1317–1334 e10.
- [13] F. Wu, J. Fan, Y. He, A. Xiong, J. Yu, Y. Li, et al., Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer, *Nat. Commun.* 12 (1) (2021) 2540.
- [14] A. Maynard, C.E. McCoach, J.K. Rotow, L. Harris, F. Haderk, D.L. Kerr, et al., Therapy-induced evolution of human lung cancer revealed by single-cell RNA sequencing, *Cell* 182 (5) (2020) 1232–1251 e22.
- [15] A.M. Leader, J.A. Grout, B.B. Maier, B.Y. Nabet, M.D. Park, A. Tabachnikova, et al., Single-cell analysis of human non-small cell lung cancer lesions refines tumor classification and patient stratification, *Cancer Cell* 39 (12) (2021) 1594–1609 e12.
- [16] N. Kim, H.K. Kim, K. Lee, Y. Hong, J.H. Cho, J.W. Choi, et al., Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma, *Nat. Commun.* 11 (1) (2020) 2285.
- [17] Y. Kashima, D. Shibahara, A. Suzuki, K. Muto, I.S. Kobayashi, D. Plotnick, et al., Single-cell analyses reveal diverse mechanisms of resistance to EGFR Tyrosine kinase inhibitors in lung cancer, *Cancer Res.* 81 (18) (2021) 4835–4848.
- [18] X. Guo, Y. Zhang, L. Zheng, C. Zheng, J. Song, Q. Zhang, et al., Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing, *Nat Med* 24 (7) (2018) 978–985.
- [19] J. Chen, Y. Tan, F. Sun, L. Hou, C. Zhang, T. Ge, et al., Single-cell transcriptome and antigen-immunoglobulin analysis reveals the diversity of B cells in non-small cell lung cancer, *Genome Biol.* 21 (1) (2020) 152.
- [20] S.P. Rowbotham, C.F. Kim, Diverse cells at the origin of lung adenocarcinoma, *Proc Natl Acad Sci U S A* 111 (13) (2014) 4745–4746.
- [21] P. Sarode, S. Mansouri, A. Karger, M.B. Schaefer, F. Griminger, W. Seeger, et al., Epithelial cell plasticity defines heterogeneity in lung cancer, *Cell. Signal.* 65 (2020), 109463.
- [22] J. Xu, S. Qin, Y. Yi, H. Gao, X. Liu, F. Ma, et al., Delving into the heterogeneity of different breast cancer subtypes and the prognostic models utilizing scRNA-seq and bulk RNA-seq, *Int. J. Mol. Sci.* 23 (17) (2022).
- [23] K. Chen, Y. Wang, Y. Hou, Q. Wang, D. Long, X. Liu, et al., Single cell RNA-seq reveals the CCL5/SDC1 receptor-ligand interaction between T cells and tumor cells in pancreatic cancer, *Cancer Lett.* 545 (2022), 215834.
- [24] K. Chen, X. Liu, W. Liu, F. Wang, X. Tian, Y. Yang, Development and validation of prognostic and diagnostic model for pancreatic ductal adenocarcinoma based on scRNA-seq and bulk-seq datasets, *Hum. Mol. Genet.* 31 (10) (2022) 1705–1719.
- [25] A.P. Patel, I. Tirosh, J.J. Trombetta, A.K. Shalek, S.M. Gillespie, H. Wakimoto, et al., Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma, *Science* 344 (6190) (2014) 1396–1401.
- [26] J. Yokota, T. Sugimura, M. Terada, Chromosomal localization of putative tumor-suppressor genes in several human cancers, *Environ. Health Perspect.* 93 (1991) 121–123.
- [27] K. Tamura, X. Zhang, Y. Murakami, S. Hirohashi, H.J. Xu, S.X. Hu, et al., Deletion of three distinct regions on chromosome 13q in human non-small-cell lung cancer, *Int. J. Cancer* 74 (1) (1997) 45–49.
- [28] F.M. Kwong, P.S. Wong, M.L. Lung, Genetic alterations detected on chromosomes 13 and 14 in Chinese non-small cell lung carcinomas, *Cancer Lett.* 192 (2) (2003) 189–198.
- [29] W.K. Cheung, D.X. Nguyen, Lineage factors and differentiation states in lung cancer progression, *Oncogene* 34 (47) (2015) 5771–5780.
- [30] N. Nanashima, K. Horie, T. Yamada, T. Shimizu, S. Tsuchida, Hair keratin KRT81 is expressed in normal and breast cancer cells and contributes to their invasiveness, *Oncol. Rep.* 37 (5) (2017) 2964–2970.
- [31] K. Zhang, Y. Liang, W. Zhang, N. Zeng, S. Tang, R. Tian, KRT81 knockdown inhibits malignant progression of melanoma through regulating interleukin-8, *DNA Cell Biol.* 40 (10) (2021) 1290–1297.
- [32] G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, et al., RNA velocity of single cells, *Nature* 560 (7719) (2018) 494–498.
- [33] 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.
- [34] Y. Xiong, Y. Feng, J. Zhao, J. Lei, T. Qiao, Y. Zhou, et al., TFAP2A potentiates lung adenocarcinoma metastasis by a novel miR-16 family/TFAP2A/PSG9/TGF-beta signaling pathway, *Cell Death Dis.* 12 (4) (2021) 352.
- [35] Z. Guoren, F. Zhaohui, Z. Wei, W. Mei, W. Yuan, S. Lin, et al., TFAP2A induced ITPKA serves as an oncogene and interacts with DBN1 in lung adenocarcinoma, *Int. J. Biol. Sci.* 16 (3) (2020) 504–514.
- [36] L.L. Zhang, J. Lu, R.Q. Liu, M.J. Hu, Y.M. Zhao, S. Tan, et al., Chromatin accessibility analysis reveals that TFAP2A promotes angiogenesis in acquired resistance to anlotinib in lung cancer cells, *Acta Pharmacol. Sin.* 41 (10) (2020) 1357–1365.
- [37] Y. Mano, K. Takahashi, N. Ishikawa, A. Takano, W. Yasui, K. Inai, et al., Fibroblast growth factor receptor 1 oncogene partner as a novel prognostic biomarker and therapeutic target for lung cancer, *Cancer Sci.* 98 (12) (2007) 1902–1913.
- [38] W. Jiang, X. Han, J. Wang, L. Wang, Z. Xu, Q. Wei, et al., miR-22 enhances the radiosensitivity of small-cell lung cancer by targeting the WRNIP1, *J. Cell. Biochem.* 120 (10) (2019) 17650–17661.
- [39] M. Efreanova, M. Vento-Tormo, S.A. Teichmann, R. Vento-Tormo, CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes, *Nat. Protoc.* 15 (4) (2020) 1484–1506.
- [40] A. Casazza, D. Laoui, M. Wenes, S. Rizzolio, N. Bassani, M. Mambretti, et al., Impeding macrophage entry into hypoxic tumor areas by Sema3A/Nrp1 signaling blockade inhibits angiogenesis and restores antitumor immunity, *Cancer Cell* 24 (6) (2013) 695–709.
- [41] R. Tokunaga, W. Zhang, M. Naseem, A. Puccini, M.D. Berger, S. Soni, et al., CXCL9, CXCL10, CXCL11/CXCR3 axis for immune activation - a target for novel cancer therapy, *Cancer Treat Rev.* 63 (2018) 40–47.
- [42] B. Cambien, B.F. Karimjee, P. Richard-Fiardo, H. Bziouech, R. Barthel, M.A. Millet, et al., Organ-specific inhibition of metastatic colon carcinoma by CXCR3 antagonism, *Br. J. Cancer* 100 (11) (2009) 1755–1764.
- [43] G. Zhu, H.H. Yan, Y. Pang, J. Jian, B.R. Achyut, X. Liang, et al., CXCR3 as a molecular target in breast cancer metastasis: inhibition of tumor cell migration and promotion of host anti-tumor immunity, *Oncotarget* 6 (41) (2015) 43408–43419.
- [44] T.C. Walser, S. Rifat, X. Ma, N. Kundu, C. Ward, O. Goloubeva, et al., Antagonism of CXCR3 inhibits lung metastasis in a murine model of metastatic breast cancer, *Cancer Res.* 66 (15) (2006) 7701–7707.
- [45] E. Pradelli, B. Karimjee-Soilhi, J.F. Michiels, J.E. Ricci, M.A. Millet, F. Vandenbos, et al., Antagonism of chemokine receptor CXCR3 inhibits osteosarcoma metastasis to lungs, *Int. J. Cancer* 125 (11) (2009) 2586–2594.
- [46] Z. Zhang, H. Zhu, X. Wang, S. Lin, C. Ruan, Q. Wang, A novel basement membrane-related gene signature for prognosis of lung adenocarcinomas, *Comput. Biol. Med.* 154 (2023), 106597.

- [47] P. Zhang, J. Liu, S. Pei, D. Wu, J. Xie, J. Liu, et al., Mast cell marker gene signature: prognosis and immunotherapy response prediction in lung adenocarcinoma through integrated scRNA-seq and bulk RNA-seq, *Front. Immunol.* 14 (2023), 1189520.
- [48] J. Zhang, X. Liu, Z. Huang, C. Wu, F. Zhang, A. Han, et al., T cell-related prognostic risk model and tumor immune environment modulation in lung adenocarcinoma based on single-cell and bulk RNA sequencing, *Comput. Biol. Med.* 152 (2023), 106460.
- [49] X. Huang, H. Xiao, Y. Shi, S. Ben, Integrating single-cell and bulk RNA sequencing to develop a cancer-associated fibroblast-related signature for immune infiltration prediction and prognosis in lung adenocarcinoma, *J. Thorac. Dis.* 15 (3) (2023) 1406–1425.
- [50] H. Kurahara, K. Maemura, Y. Mataka, K. Tanoue, S. Iino, Y. Kawasaki, et al., Lung recurrence and its therapeutic strategy in patients with pancreatic cancer, *Pancreatol.* 20 (1) (2020) 89–94.
- [51] S. Kobayashi, T.J. Boggon, T. Dayaram, P.A. Janne, O. Kocher, M. Meyerson, et al., EGFR mutation and resistance of non-small-cell lung cancer to gefitinib, *N. Engl. J. Med.* 352 (8) (2005) 786–792.
- [52] S. Shukla, J.R. Evans, R. Malik, F.Y. Feng, S.M. Dhanasekaran, X. Cao, et al., Development of a RNA-seq based prognostic signature in lung adenocarcinoma, *J Natl Cancer Inst* 109 (1) (2017).
- [53] N.D. Marjanovic, M. Hofree, J.E. Chan, D. Canner, K. Wu, M. Trakala, et al., Emergence of a high-plasticity cell state during lung cancer evolution, *Cancer Cell* 38 (2) (2020) 229–246 e13.
- [54] Z.F. Lim, P.C. Ma, Emerging insights of tumor heterogeneity and drug resistance mechanisms in lung cancer targeted therapy, *J. Hematol. Oncol.* 12 (1) (2019) 134.
- [55] M. Jamal-Hanjani, G.A. Wilson, N. McGranahan, N.J. Birkbak, T.B.K. Watkins, S. Veeriah, et al., Tracking the evolution of non-small-cell lung cancer, *N. Engl. J. Med.* 376 (22) (2017) 2109–2121.
- [56] V.M.L. de Sousa, L. Carvalho, Heterogeneity in lung cancer, *Pathobiology* 85 (1–2) (2018) 96–107.
- [57] W. Fang, H. Jin, H. Zhou, S. Hong, Y. Ma, Y. Zhang, et al., Intratumoral heterogeneity as a predictive biomarker in anti-PD-(L)1 therapies for non-small cell lung cancer, *Mol. Cancer* 20 (1) (2021) 37.
- [58] Z.X. Chong, W.Y. Ho, S.K. Yeap, M.L. Wang, Y. Chien, N.D. Verusingam, et al., Single-cell RNA sequencing in human lung cancer: applications, challenges, and pathway towards personalized therapy, *J. Chin. Med. Assoc.* 84 (6) (2021) 563–576.
- [59] P. Mendez, J.L. Ramirez, Copy number gains of FGFR1 and 3q chromosome in squamous cell carcinoma of the lung, *Transl. Lung Cancer Res.* 2 (2) (2013) 101–111.
- [60] S. Sangaletti, R. Ferrara, C. Tripodo, M.C. Garassino, M.P. Colombo, Myeloid cell heterogeneity in lung cancer: implication for immunotherapy, *Cancer Immunol. Immunother.* 70 (9) (2021) 2429–2438.
- [61] P.O. Azevedo, A.E. Paiva, G.S.P. Santos, L. Lousado, J.P. Andreotti, I.F.G. Sena, et al., Cross-talk between lung cancer and bones results in neutrophils that promote tumor progression, *Cancer Metastasis Rev.* 37 (4) (2018) 779–790.
- [62] G. Pan, Y. Liu, L. Shang, F. Zhou, S. Yang, EMT-associated microRNAs and their roles in cancer stemness and drug resistance, *Cancer Commun.* 41 (3) (2021) 199–217.
- [63] V. Mittal, Epithelial mesenchymal transition in tumor metastasis, *Annu. Rev. Pathol.* 13 (2018) 395–412.
- [64] W.J. Wang, C. Ouyang, B. Yu, C. Chen, X.F. Xu, X.Q. Ye, Role of hypoxia-inducible factor2alpha in lung cancer (Review). *Oncol Rep* 45 (5) (2021).
- [65] J. Ancel, J.M. Perotin, M. Dewolf, C. Launois, P. Mulette, B. Nawrocki-Raby, et al., Hypoxia in lung cancer management: a Translational approach, *Cancers* 13 (14) (2021).
- [66] A.A. Tirpe, D. Gulei, S.M. Ciortea, C. Crivii, I. Berindan-Neagoe, Hypoxia: overview on hypoxia-mediated mechanisms with a focus on the role of HIF genes, *Int. J. Mol. Sci.* 20 (24) (2019).
- [67] A. Dumond, G. Pages, Neuropilins, as relevant oncology target: their role in the tumoral microenvironment, *Front. Cell Dev. Biol.* 8 (2020) 662.
- [68] C. Liu, X. Li, H. Shao, D. Li, Identification and validation of two lung adenocarcinoma-development characteristic gene sets for diagnosing lung adenocarcinoma and predicting prognosis, *Front. Genet.* 11 (2020), 565206.
- [69] A.M. Laughney, J. Hu, N.R. Campbell, S.F. Bakhoun, M. Setty, V.P. Lavallee, et al., Regenerative lineages and immune-mediated pruning in lung cancer metastasis, *Nat Med* 26 (2) (2020) 259–269.
- [70] Y. Hao, S. Hao, E. Andersen-Nissen, W.M. Mauck 3rd, S. Zheng, A. Butler, et al., Integrated analysis of multimodal single-cell data, *Cell* 184 (13) (2021) 3573–3587 e29.
- [71] S.V. Puram, I. Tirosh, A.S. Parikh, A.P. Patel, K. Yizhak, S. Gillespie, et al., Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer, *Cell* 171 (7) (2017) 1611–1624 e24.
- [72] C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, et al., The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells, *Nat. Biotechnol.* 32 (4) (2014) 381–386.
- [73] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, *OMICS* 16 (5) (2012) 284–287.
- [74] T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, et al., clusterProfiler 4.0: a universal enrichment tool for interpreting omics data, *Innovation* 2 (3) (2021), 100141.
- [75] A.M. Newman, C.B. Steen, C.L. Liu, A.J. Gentles, A.A. Chaudhuri, F. Scherer, et al., Determining cell type abundance and expression from bulk tissues with digital cytometry, *Nat. Biotechnol.* 37 (7) (2019) 773–782.
- [76] S. Aibar, C.B. Gonzalez-Blas, T. Moerman, V.A. Huynh-Thu, H. Imrichova, G. Hulselmans, et al., SCENIC: single-cell regulatory network inference and clustering, *Nat. Methods* 14 (11) (2017) 1083–1086.
- [77] M.D. Wilkerson, D.N. Hayes, ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking, *Bioinformatics* 26 (12) (2010) 1572–1573.
- [78] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis, *Nat. Methods* 9 (7) (2012) 671–675.