BMC Cancer

RESEARCH

Open Access

Highly proliferating cancer cells function as novel prognostic biomarkers for lung adenocarcinoma with particular usefulness for stage IA risk stratification

Yanlu Xiong[1,2,3†], Yongfu Ma[1†], Jie Lei[2†], Jianfei Zhu[2,4], Nianlin Xie[2], Feng Tian[2], Qiang Lu[2], Miaomiao Wen[2], Qian Zheng[1], Yong Han[2,5], Tao Jiang[2*] and Yang Liu[1*]

Abstract

Background The refinement of risk stratification in lung adenocarcinoma (LUAD) plays a pivotal role in advancing precision medicine; however, the current staging classification falls short of comprehensiveness, particularly in the case of stage IA patients. We aimed to molecularly stratify LUAD patients especially for stage IA.

Methods We analysed tumour heterogeneity and identified highly proliferating cancer cells (HPCs) in LUAD by performing single-cell RNA sequencing (scRNA-seq) analysis, immunohistochemical (IHC) staining using a tissue microarray, flow cytometry and biological experiments. Then, we quantified the content of HPCs in nine LUAD datasets by single-sample gene set enrichment analysis and evaluated the relationship between the percentage of HPCs and overall survival (OS). Next, we analysed the OS predictive effect of HPCs at different LUAD stages, especially for stage I risk stratification. Furthermore, we established a prognostic prediction model based on HPC-associated genes for clinical application. The above findings were validated in another five LUAD datasets. Finally, we explored the relationship between HPCs and the progressive pathological evolution of early-stage LUAD and the driving mutations by scRNA-seq, bulk RNA-seq and IHC staining.

Results LUAD tissues carry a small proportion of HPCs, which show potential for malignant proliferation and intense interactions with the microenvironment. A high HPC content is an independent risk factor for OS in LUAD patients, even in stage IA patients. HPCs can be used to establish a cut-off point for the prognosis of stage IA disease, with patients with a higher risk showing a prognosis similar to that of patients with stage IB disease. We built an R package (HSurADs) based on HPC-associated genes, which exhibited good efficacy for the prognostic prediction of LUAD. HPCs gradually increase with the pathological evolution of early-stage LUAD, which may be affected by TP53 mutations.

†Yanlu Xiong, Yongfu Ma and Jie Lei contributed equally to this work.

*Correspondence:
Tao Jiang
jiangtaochest@163.com
Yang Liu
sunny301x@sina.com

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

Conclusion The HPC content can be used as a novel prognostic factor for LUAD, especially for stage IA risk stratification.

Keywords Lung adenocarcinoma, Highly proliferating cancer cells, Stage IA, Prognosis

Introduction

Lung cancer is an important disease that endangers human health. According to global statistics, in 2020, there were approximately 2.21 million new cases of lung cancer worldwide and approximately 1.8 million deaths [1]. Approximately 40% of lung cancer cases are lung adenocarcinoma (LUAD), which is the main pathological subtype of lung cancer, and thus, LUAD is the focus of lung cancer prevention and treatment approaches [2]. LUAD is highly heterogeneous, and accurate classification of subtypes with different biological and clinical characteristics is the key to the precise treatment of LUAD [3]. TNM staging, in which tumour size, lymph node invasion and distant metastasis are the main criteria, is the cornerstone for LUAD classification and determination of the subsequent treatment (surgery, systematic treatment, etc.) [4]. Histological typing (such as atypical adenomatous hyperplasia (AAH), adenocarcinoma in situ (AIS), minimally invasive adenocarcinoma (MIA), invasive adenocarcinoma (IAC) and its various subtypes) and radiological characteristics (such as ground glass/solid components) are also important for the risk stratification of LUAD and accurate treatment, especially in very early-stage LUAD (stage IA LUAD characterized by small nodules) [5, 6].However, LUADs with the same TNM stage and the same pathological subtype or radiological characteristics still exhibit high heterogeneity, and new typing parameters need to be followed up [6–10].

Tumour biological characteristics have strong implications for risk stratification and clinical management for LUAD patients. For example, driver mutation status and immune characteristics provide indications for screening targeted and immunotherapy-sensitive populations, and minimal residual disease (MRD) with circulating tumour DNA (ctDNA) as its main content effectively predicts relapse and chemotherapy sensitivity [11–13]. Moreover, the infiltration of macrophages or fibroblasts also reflects the ability to predict the prognosis of LUAD patients [14, 15]. In fact, for the third phase of the International Association for the Study of Lung Cancer (IASLC) Staging Project for the forthcoming (Ninth) Edition of the TNM Classification of Lung Cancer, molecular information has been collected to evaluate both its prognostic value and the feasibility of incorporating this information into the TNM classification [16, 17]. However, at the level of stage IA LUAD, the above classification is still inadequate, and large gaps remain in the biological classification of IA LUAD. By performing single-cell RNA sequencing (scRNA-seq), we identified a subset of highly proliferating cancer cells (HPCs) in LUAD tissues, and we demonstrated the strong prognostic significance of HPCs in multi-institutional LUAD datasets, especially for risk stratification of stage IA patients.

Materials and methods

Study design

The design process of this study is presented in Fig. 1. First, we identified a subset of HPCs from four LUAD cases from the scRNA-seq datasets (E-MTAB-6149: patient 1 and patient 2; GSE171145: patient 3 and patient 4) and explored their function by bioinformatics analysis. Then, we screened the characteristic gene sets and identified marker genes of HPCs (differentially expressed gene (DEG) screening: TCGA-LUAD, GSE10072, GS30219, GSE31210, GSE32863, GSE43458, and GSE63459; survival-related gene screening: TCGA-LUAD, GS30219, GSE31210, GSE41217, GSE42127, GSE50081, and GSE72094). Then, HPCs were detected by immunohistochemistry (IHC) and flow cytometry (FCM) in a tissue microarray (TMA) and in fresh LUAD tissues, respectively. In addition, gene intervention technology targeting marker genes in cell lines and subsequent in vitro experiments were used to investigate the function of HPCs. Furthermore, we used single-sample gene set enrichment analysis (ssGSEA) to quantify the relative content of HPCs in nine LUAD datasets of the bulk transcriptome (GSE13213, GS30219, GSE31210, GSE41217, GSE42127, GSE50081, GSE68465, GSE72094, and TCGA-LUAD) (n=2194) and explored the relationships between HPCs and prognosis and clinical stage. Subsequently, we examined the risk stratification effect of HPC content on stage I LUAD, especially stage IA LUAD. In addition, we screened the key genes from the characteristic gene set of HPCs through machine learning algorithms (random forest (RF) and extreme gradient boosting (XGBoost)) to establish risk scores and corresponding R packages (HSurADs) for clinical application. Furthermore, we validated these results on four independent LUAD bulk transcriptome datasets (GSE11969, GSE14814, GSE81089, and our dataset: GSE282774) (n=327) and a LUAD clinical dataset from our centre (n=729). The SEER-LUAD dataset (n=18620) was used to analyse the clinical significance of existing IA classifications. Finally, we explored the relationship between HPCs and the progression of early-stage LUAD through scRNA-seq, bulk transcriptome sequencing and IHC staining of samples from different

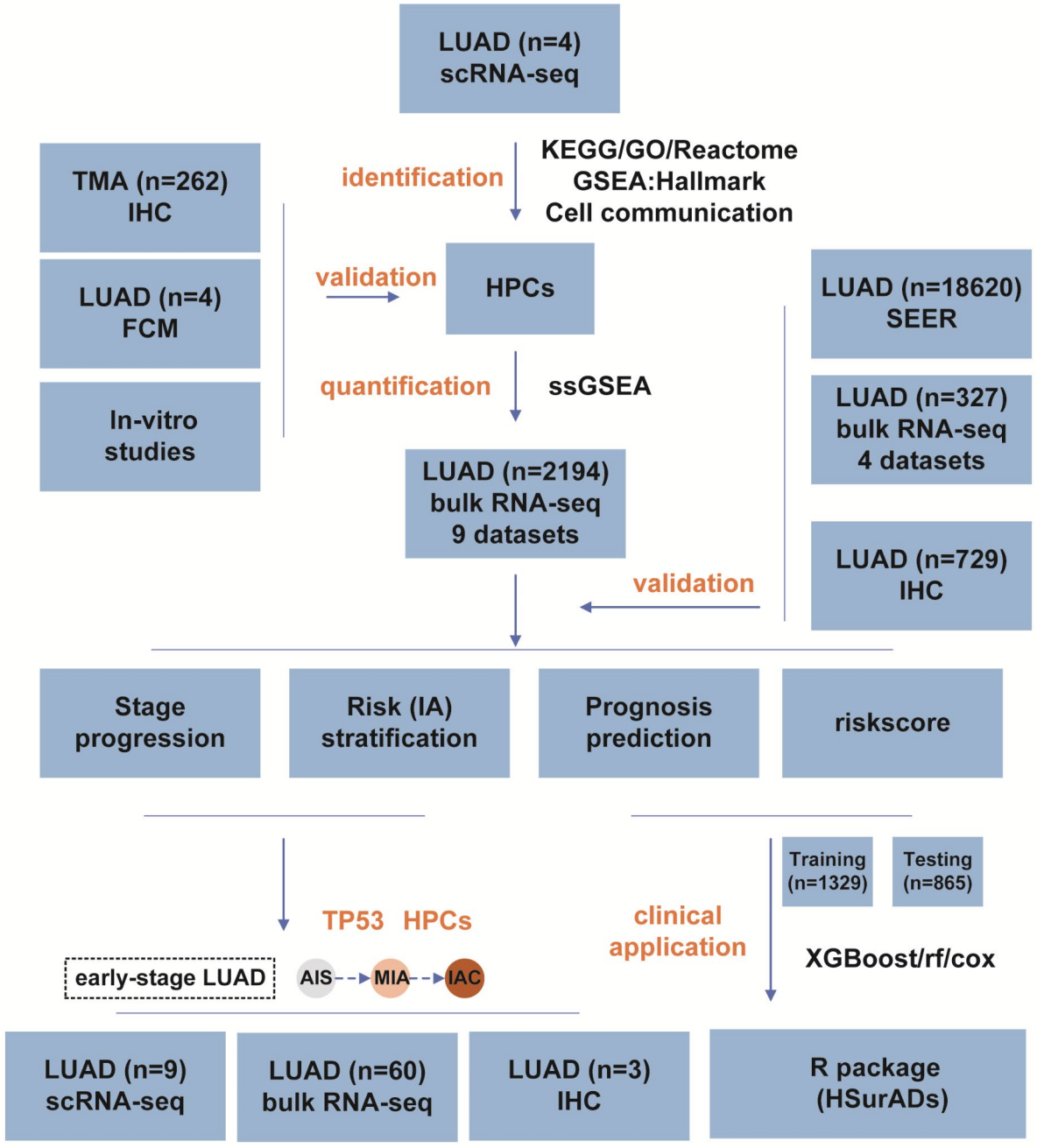

Fig. 1 Study design. Workflow of this study. LUAD, lung adenocarcinoma; scRNA-seq, single-cell RNA sequencing; ssGSEA, single-sample gene set enrichment analysis; SEER, Surveillance Epidemiology and End Results; KEGG, Encyclopedia of Genes and Genomes; GO, Gene Ontology; GSEA, Gene Set Enrichment Analysis; TMA, Tissue Microarray; IHC, Immunohistochemistry; FCM, flow cytometry; HPCs, highly proliferating cancer cells; rf, Random Forest; XGBoost, eXtreme Gradient Boosting

pathological stages of early-stage LUAD (AIS, MIA and IAC (IA)).

Datasets

Public datasets, such as bulk transcriptome datasets (TCGA-LUAD, GSE10072, GSE11969, GSE13213, GSE14814, GS30219, GSE31210, GSE32863, GSE41217, GSE42127, GSE43458, GSE50081, GSE63459, GSE68465, GSE72094, and GSE81089), scRNA-seq datasets (E-MTAB-6149 and GSE171145), and LUAD clinical datasets (SEER-LUAD), were downloaded from the TCGA [18–20], ArrayExpress [21], GEO [22] and SEER

databases [23]. Datasets and LUAD specimens from our centre, including scRNA-seq (GSE189357), bulk transcriptome data (GSE282617 and GSE282774), TMA, and a clinical dataset (Ki-67 index), were collected from Tangdu Hospital, the Fourth Military Medical University (Xi'an, China), in accordance with ethics authority approval. Detailed information on the public datasets and procedures used for IHC, FCM, scRNA-seq and bulk RNA-seq are provided in the Supplementary Material 1.

Biological experiment

Human LUAD cell lines (PC-9) were purchased from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China). For transient transfection, siRNAs (Sangon Biotech, Shanghai, China) were transfected into cells using Lipofectamine 2000 (Invitrogen, CA, USA) according to the manufacturer's instructions. Quantitative reverse transcription–polymerase chain reaction (qRT–PCR) and Western blotting (WB) were used to measure mRNA and protein expression, respectively. CCK8 proliferation assays were used to examine the proliferation rate of cells. Detailed information is provided in the Supplementary Material 1.

Bioinformatics and statistical analysis

R language was used for bioinformatics and statistical analysis, and a two-sided p value <0.05 was considered to indicate statistical significance. Detailed information is provided in the Supplementary Material 1. The research is being reported in line with the relevant guideline of Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK) [24].

Results

LUAD possesses HPCs in a minor proportion

We analysed the scRNA-seq data of four LUAD tissue samples (E-MTAB-6149: Patients 1 and 2; GSE171145: Patients 3 and 4) and identified immune cells (T/NK cells: TRBC1/GNLY; myeloid cells: AIF1/LYZ; B cells: CD79A/MS4A1), stromal cells (endothelial cells: PLVAP/VWF; fibroblasts: COL1A1/DCN), and epithelial cells (EPCAM/KRT7)) (Supplemental Fig. 1A, 1B). Cancer cells were further identified by subdividing epithelial cells and performing CNV analysis (all stroma and immune cells were used as references) (Supplemental Fig. 1C, 1D). Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis revealed a typical subpopulation: HPCs (Fig. 2A), and Gene Ontology (GO) and Reactome analyses also demonstrated a high level of division and proliferation processes in this subpopulation (Supplemental Fig. 2A, 2B). Gene set enrichment analysis (GSEA) based on the Hallmark database also confirmed that the subpopulation had high malignant proliferation ability (Supplemental Fig. 2C). Cell communication analysis demonstrated close communication between HPCs and other cells in the microenvironment, especially through the MDK pathway and ANXA1 pathway (Supplemental Fig. 2D 2E). These pathways can not only promote the self-proliferation of tumour cells, resistance to apoptosis, and metastasis but also regulate other cells in the microenvironment, promote immunosuppression and angiogenesis, and support other microenvironments conducive to tumour occurrence and progression [25–30]. Through intersection analysis, we determined the characteristic gene sets of HPCs (Fig. 2B, Supplemental Table 2). We considered the abnormal spindle-like microcephaly associated (ASPM) gene as a possible marker gene for identifying HPCs and characterized its biological function by DEG screening (cancer-normal tissue) in 7 LUAD datasets and prognostic gene screening (Surv) in 7 LUAD datasets (Fig. 2C-E). ASPM was first recognized as a spindle tubulin-related protein, and an increasing number of studies have suggested that it plays an important regulatory role in malignant tumours by regulating the functional activities of various proteins through its macromolecular structure [31]. Some databases have also shown that ASPM is distributed in the nucleus, cytoplasm and membrane (https://www.genecards.org/) [32]. We also selected Ki-67, a classic proliferative marker, as a marker of HPCs. Ki-67 is commonly expressed in proliferating cells, except for cells in the G0 phase, and is often used as a marker to evaluate the proliferative activity of cancer cells [33, 34]. We demonstrated a strong correlation between HPCs and Ki-67 and ASPM in multiple datasets (Fig. 2F). We demonstrated the presence of a hyperproliferative cell subpopulation in the TMA via IHC staining of Ki-67 and ASPM (Fig. 2H). In addition, in the LUAD cell line PC-9, after small interfering RNA (siRNA) was used to silence ASPM, the proliferation rate significantly decreased, and the content of cyclic-related molecules also decreased (Supplemental Fig. 2F-H). We also demonstrated the presence of HPCs in fresh LUAD tissues by flow cytometric measurement of carboxyfluorescein diacetate succinimidyl ester (CFSE) dye dilution (Supplemental Fig. 2I). Both scRNA-seq, TMA IHC and FCM revealed a small proportion of hyperproliferative cell subsets in LUAD (Fig. 2G and I; Supplemental Fig. 2I).

The HPC content is clearly correlated with the prognosis of LUAD

We evaluated the clinical significance of the HPC content in LUAD. First, based on the characteristic gene set (Supplemental Table 2) of the gene expression signatures of HPC, we used ssGSEA to quantify the relative content of the HPCs in the LUAD samples using 9 different LUAD datasets. A higher HPC content predicted shorter overall survival (OS) in all 9 LUAD datasets, and a meta-analysis

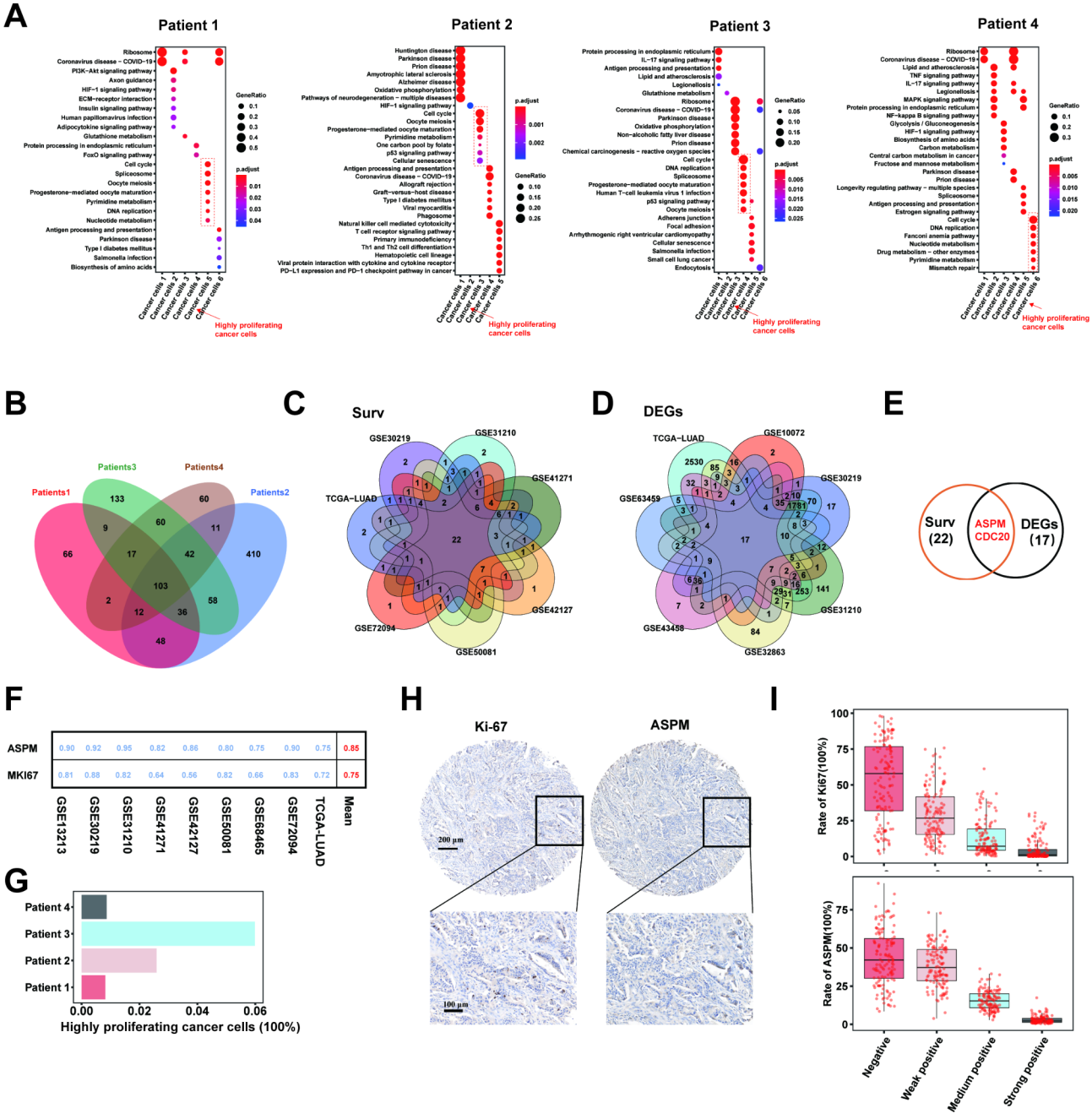

Fig. 2 HPCs in LUAD. (A) Biological pathway enrichment (KEGG) of malignant epithelial cell subtypes in four LUAD tissue samples. (B) Common DEGs of the HPCs in four LUAD tissue samples. (C) Survival prediction ability of 103 genes comprising the HPC gene signature in 7 LUAD datasets was evaluated (Surv). (D) Analysis of the common DEGs between cancer tissue and normal tissue in 7 LUAD datasets. (E) Intersection analysis of survival predictor genes (Surv) and DEGs to identify possible functional marker genes of HPCs: ASPM. (F) High correlation between APSM/MKI67 and HPCs across multiple datasets. (G) Proportion of HPCs in four LUAD tissue samples. (H) IHC staining for Ki-67 (left) and ASPM (right) protein levels in LUAD tissues (up, low resolution; down, high resolution). (I) Expression intensity of Ki-67 (up) and ASPM (down) protein in LUAD-TMA. In the LUAD tissue samples used for scRNA-seq, Patient 1 and Patient 2 were from E-MTAB-6149, and Patient 3 and Patient 4 were from GSE171145. KEGG, Encyclopedia of Genes and Genomes; LUAD, lung adenocarcinoma; scRNA-seq, single-cell RNA sequencing; DEGs, differentially expressed genes; HPCs, highly proliferating cancer cells; TMA, Tissue Microarray; IHC, Immunohistochemistry; ASPM, abnormal spindle-like microcephaly associated

showed an overall risk prediction function (common effect: HR = 1.42, 95% CI = 1.32–1.52; random effect: HR = 1.45, 95% CI = 1.33–1.59) (Fig. 3A). Using four datasets with complete clinical information, we demonstrated the independent predictive effect of HPC content on OS (Supplemental Fig. 3A). We integrated the gene expression data of 9 datasets through z-transformation. The datasets that were not integrated showed an obvious

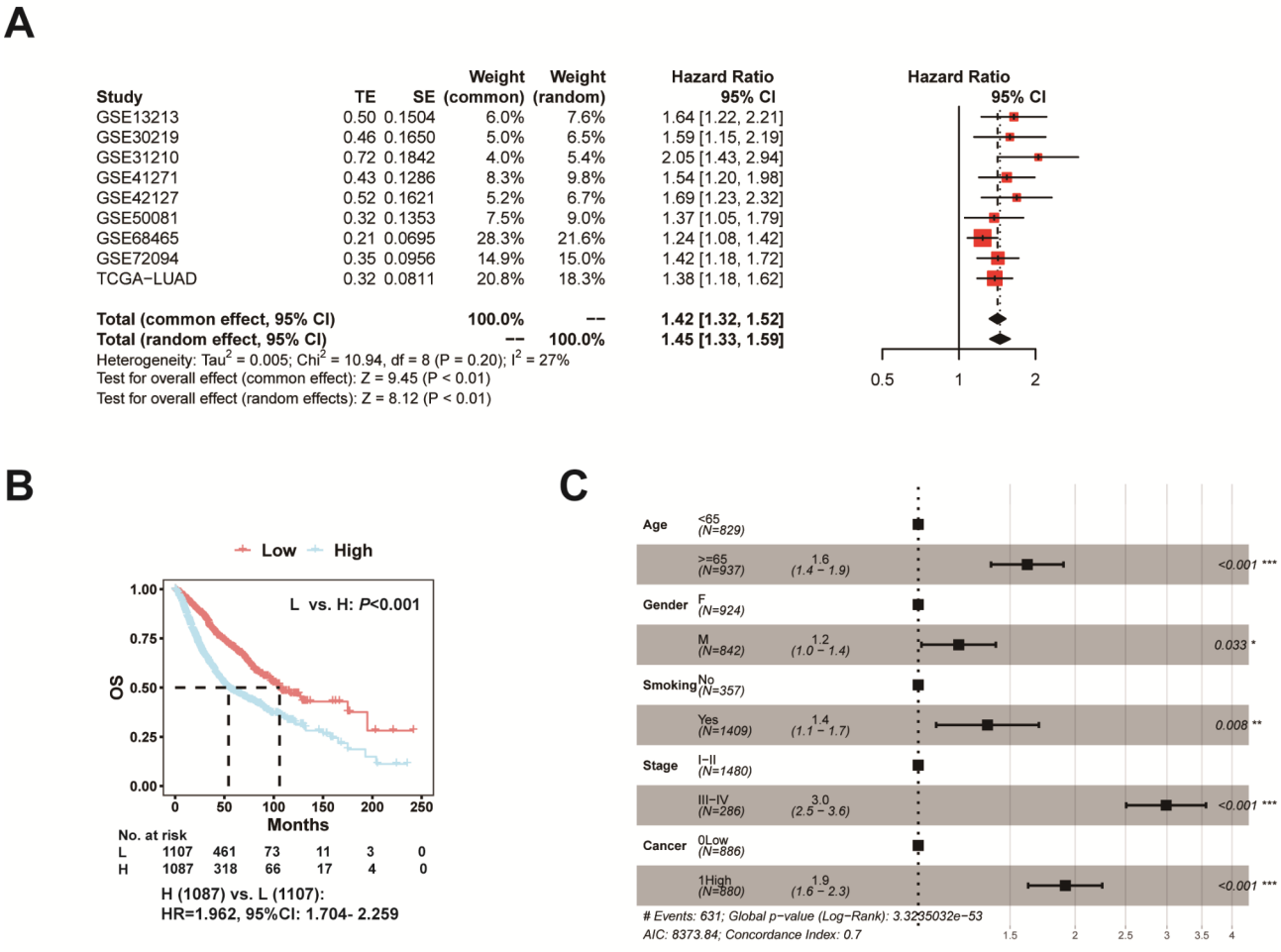

Fig. 3 Relationships between HPCs and the prognosis of LUAD. (**A**) OS predictive efficacy of the level of HPCs in 9 different LUAD datasets by Cox analysis. (**B**) OS curves for two categories of LUAD patients according to the level of HPCs (median value as the cut-off) in the integrated cohort of 9 LUAD datasets. (**C**) Independent OS predictive efficacy curves of HPCs in the integrated cohort of 9 LUAD datasets. HPCs, highly proliferating cancer cells; LUAD, lung adenocarcinoma; OS, overall survival

batch effect (Supplemental Fig. 3B), but the batch effect after integration was not obvious (Supplemental Fig. 3C). We assessed the HPC content in the integrated population via ssGSEA. We found that a higher HPC content was closely associated with worse prognosis (Fig. 3B) and may be an independent prognostic factor after accounting for age, sex, smoking history, and TNM stage (Fig. 3C). Through decision curve analysis, we found that HPCs exerted a stronger prognostic effect than age, sex, or smoking history but were still inferior to TNM stage (Supplemental Fig. 3D).

The relationship between the content of HPCs and the clinical stage of LUAD

We found a significant correlation between HPCs content and the clinical stage of LUAD and a significant increasing relationship in stages I and II, especially for LUAD IA/IB (Fig. 4A). By analysing the integrated data, we found that the HPC content negatively predicted OS in all stages (stage I, II, and III-IV), with greater significance in stage I (Supplemental Fig. 4A). Moreover, multivariate analysis revealed that a higher HPC content was an independent prognostic risk factor for stage I, II and III-IV LUAD (Supplemental Fig. 4B). By further dividing stage I tumours into stage IA and IB tumours, we found that a higher HPC content in both stage IA and IB tumours was still clearly associated with prognosis (Fig. 4B). A multivariate analysis revealed that a higher HPC content was an independent risk factor for stage IA and IB LUAD (Fig. 4C). For each dataset, the meta-analysis also demonstrated that the content of HPCs was positively correlated with prognosis at each stage, especially

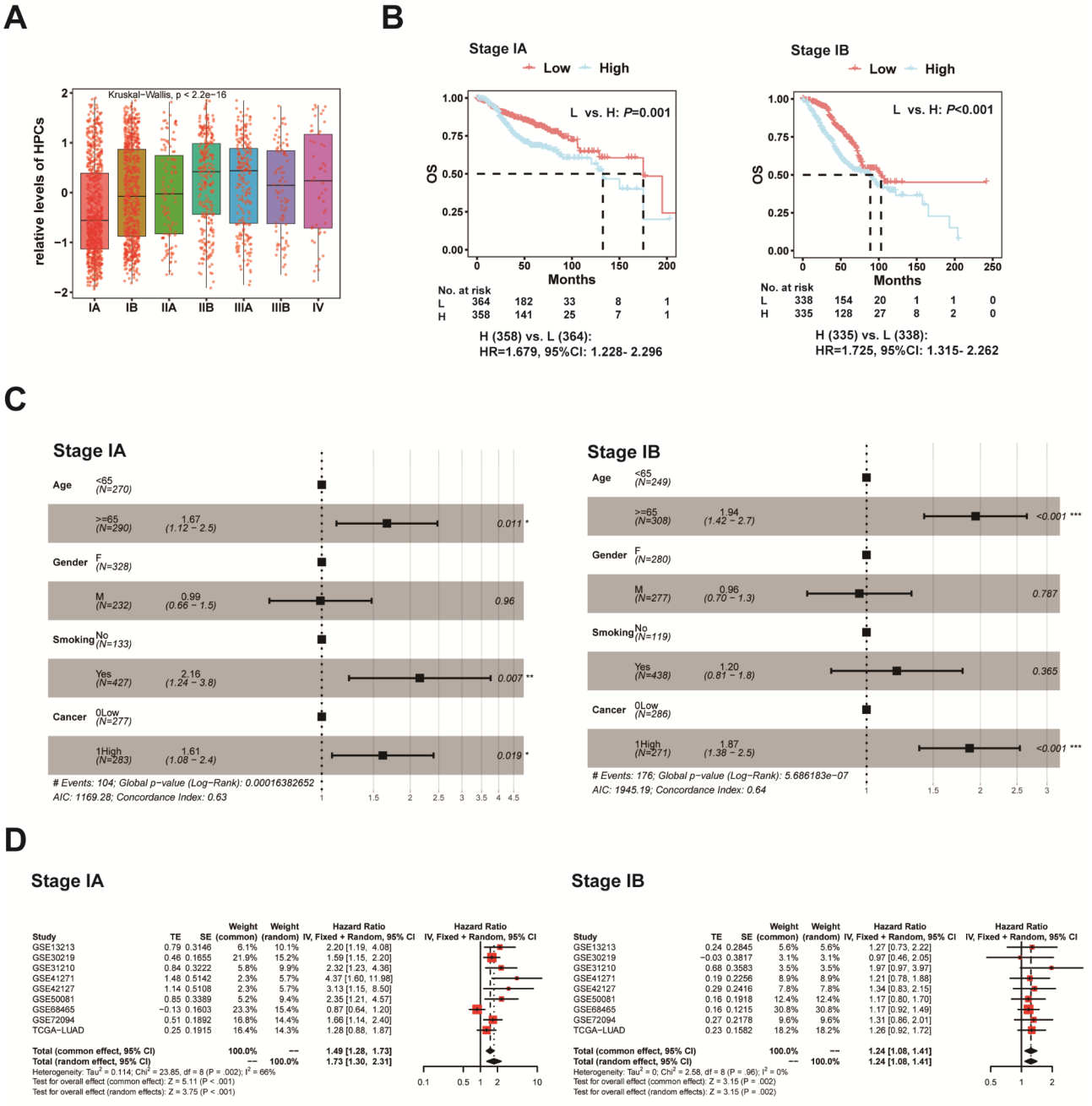

Fig. 4 Relationships between HPCs and the clinical stage of LUAD. (**A**) Content of HPCs in different clinical stages of LUAD in the integrated cohort of 9 LUAD datasets. (**B**) OS curves for two categories of LUAD patients according to the level of HPC (median value as the cut-off) in stage IA (left) and IB (right) in the integrated cohort of 9 LUAD datasets. (**C**) Independent OS predictive efficacy curves of HPCs in stage IA (left) and IB (right) in the integrated cohort of 9 LUAD datasets. (**D**) OS predictive efficacy of HPCs in stage IA (left) and IB (right) in 9 LUAD datasets. HPCs, highly proliferating cancer cells; LUAD, lung adenocarcinoma; OS, overall survival

at stage IA, where 7 of the 9 datasets demonstrated a significant prognostic impact (Fig. 4D) (Supplemental Fig. 4C).

The level of HPCs stratified the risk for stage IA LUAD

Based on the median HPCs, we divided stage IA LUAD patients into a high HPC content group (IA-H) and a low HPC content group (IA-L) according to the integrated data. The prognosis of IA-H patients was significantly worse than that of IA-L patients but similar to that of stage IB patients, suggesting that HPCs can be used for prognostic stratification of stage IA LUAD patients (Fig. 5A). However, the currently used eighth edition of the TNM staging recommendations for IAs, which are based on tumour size, is not sufficient, and the prognosis of the highest risk type (IA3) is still better than that of stage IB (Fig. 5B). We further classified stage IB LUAD by HPC content, and the IB-L subsets showed a lower degree of malignancy, similar to that of stage IA LUAD (Fig. 5C). However, the higher risk type of IB (IB-H) still did not overlap with that of IIA (Fig. 5C). Furthermore, we evaluated the effect of HPC content on the risk stratification of stage IA and IB LUAD patients in each of the 9 datasets. The meta-analysis revealed that the prognosis of low-risk IA (IA-L) was significantly lower than that of IB (common effect: HR=0.54, 95% CI=0.40–0.73; random effect: HR=0.46, 95% CI=0.30–0.70) (Fig. 5D), while the prognosis of higher-risk IA (IA-H) was similar to that of IB (common effect: HR=0.91, 95% CI=0.71–1.17; random effect: HR=0.91, 95% CI=0.71–1.17) (Fig. 5D). The prognosis of IB with a lower risk (IB-L) was similar to that of IA (common effect: HR=1.15, 95% CI=0.87–1.51; random effect: HR=1.15, 95% CI=0.87–1.51) (Supplemental Fig. 5A). The prognosis of IB with a higher risk (IB-H) was much greater than that of IA (common effect: HR=1.84, 95% CI=1.44–2.35; random effect: HR=1.84, 95% CI=1.44–2.35) (Supplemental Fig. 5B). However, the prognosis of IB with a lower risk (IB-L) or higher risk (IB-H) was still significantly better than that of stage IIA (Supplemental Fig. 5C, 5D). These results suggest that the HPC content can be used for risk stratification of stage IA LUAD for screening high-risk subtypes.

LUAD prognostic prediction and stage IA risk stratification based on HPC-associated genes

We aimed to establish a risk score based on HPC-associated genes to clinically evaluate the prognosis of LUAD. After integrating the nine datasets, we randomly divided them into a training set (n=1329) and a test set (n=865) (6:4). The random forest (rf) and eXtreme Gradient Boosting (XGBoost) algorithms were used to screen the feature genes from the characteristic gene set of HPCs closely related to prognosis (Fig. 6A). Through intersection analysis of the most important genes (top 20), we identified 7 characteristic genes (Fig. 6B). Subsequently, the prognostic risk score was established by Cox regression (risk score=ECT2 *(0.1456)+PRC1*(0.2635)+CCNB2*(-0.0683)+PTTG1*(0.0381)+UBE2C*(0.1426)+ASF1B*(-0.1772)+ASPM*(0.0323)). In both the training set and test sets, the risk score consistently predicted the prognosis of LUAD (Fig. 6C) and stage IA LUAD (Fig. 6D). We established a risk score-based prognostic prediction model in the training set (Fig. 6E). Finally, based on the risk score, we developed the R package (HSurADs) and function (HpSurADs) for predicting the survival probability of LUAD patients for clinical application (Fig. 6F).

Validation of the clinical significance of HPCs

To further clarify the clinical significance of HPCs, we validated the above findings. First, we used the Ki-67 index to represent HPCs and evaluated the relationship between Ki-67 expression and the prognosis of LUAD and risk stratification of stage IA LUAD in 729 patients from our centre. We found that the Ki-67 index was associated with a worse prognosis of LUAD (HR=2.46, 95% CI=1.27–4.76) and stage IA LUAD (HR=3.83, 95% CI=1.34–10.99) (Fig. 7A and B). Based on the median Ki-67, we divided stage IA LUAD patients into a high Ki-67 content group (IA-H) and a low Ki-67 content group (IA-L). The prognosis of IA-H was significantly worse than that of IA-L but similar to that of stage IB, suggesting that Ki-67 can be used for prognostic stratification of stage IA LUAD (Fig. 7C). We then collected four independent bulk transcriptome datasets of LUAD (GSE11969, GSE14814, GSE81089, and our dataset: GSE282774) (n=327). We integrated the gene expression data of 4 datasets through z-transformation. The datasets that were not integrated showed an obvious batch effect, but the batch effect after integration was not obvious (Supplemental Fig. 6A). We assessed the content of HPCs in the integrated population via ssGSEA based on the gene expression signatures of HPCs. We found that a higher HPC content was clearly associated with worse prognosis (Fig. 7D) and may be an independent prognostic factor after accounting for age, sex, smoking history, and TNM stage (Supplemental Fig. 6B). Moreover, a higher HPC content was also closely related to the prognosis of stage IA LUAD (Fig. 7E) and was an independent risk factor (Supplemental Fig. 6C). Based on the median HPCs, we divided stage IA LUAD patients into a high-HPC group (IA-H) and a low-HPC group (IA-L) according to the integrated data. The prognosis of IA-H was significantly worse than that of IA-L but similar to that of stage IB, suggesting that HPCs can be used for prognostic stratification of stage IA LUAD patients (Fig. 7F). Finally, we further evaluated the value of the risk score in the validation datasets. Similarly, the above risk score

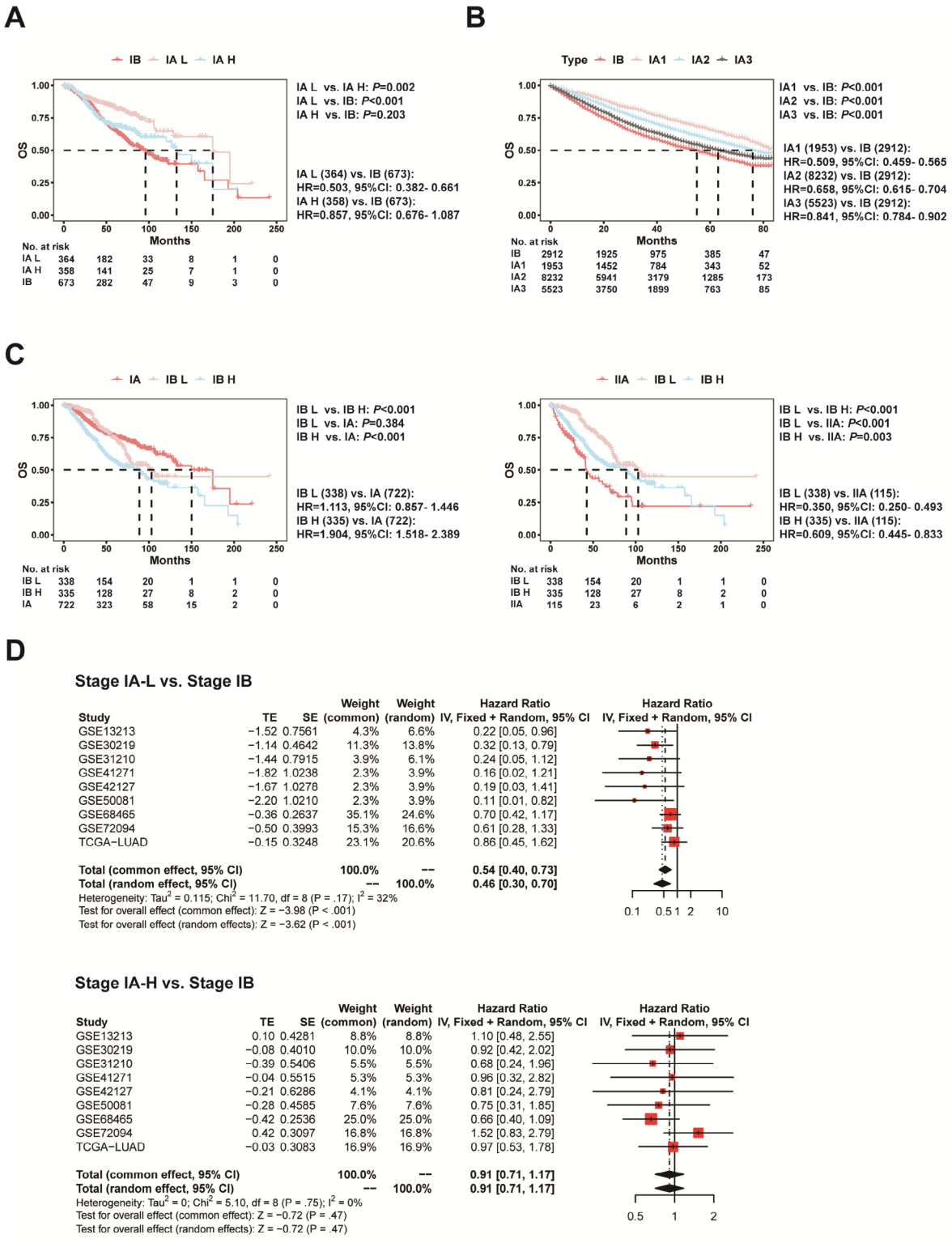

Fig. 5 Significance of HPCs in risk stratification of stage I LUAD. (**A**) OS curves of two risk types of stage IA LUAD patients divided by the median level of HPCs (IA H and IA L) and stage IB LUAD patients in the integrated cohort of 9 LUAD datasets. (**B**) OS curves of IA1, IA2, IA3 and IB LUAD patients in the SEER database. (**C**) OS curves of two risk types of stage IB LUAD patients divided by the median level of HPCs (IB H and IB L) and stage IA LUAD patients (left) and stage IIA LUAD patients (right) in the integrated cohort of 9 LUAD datasets. (**D**) OS analysis between IA L (up) or IA H (down) and stage IB LUAD patients in 9 LUAD datasets. HPCs, highly proliferating cancer cells; LUAD, lung adenocarcinoma; OS, overall survival; SEER, Surveillance Epidemiology and End Results

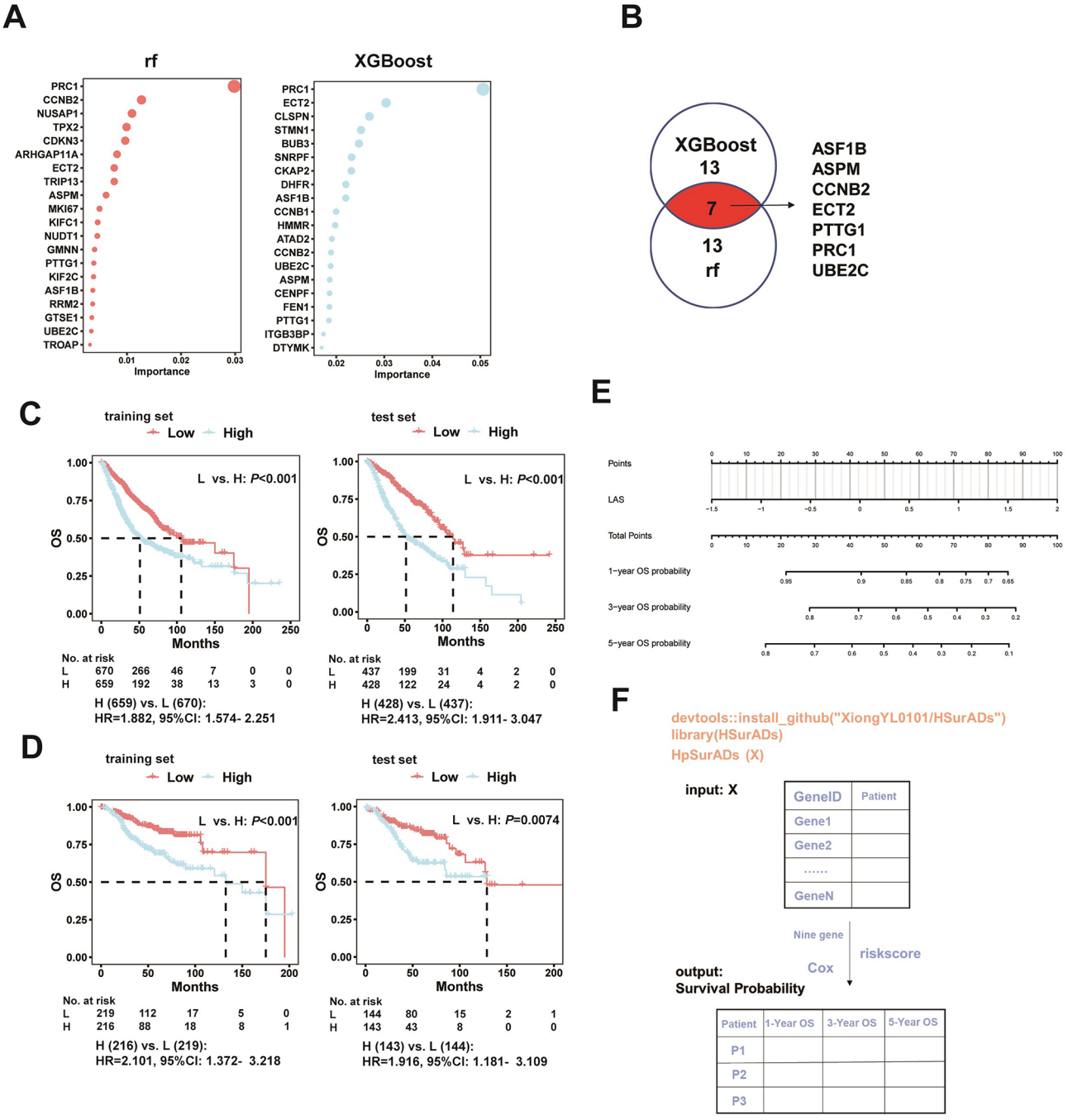

Fig. 6 HPC-based risk score for prognostic prediction of LUAD. (**A**) Importance ranking of HPC-associated genes (top 20) in LUAD prognosis prediction by RF (left) and XGBoost (right). (**B**) Intersection of HPC-associated genes identified by RF and XGBoost. (**C**) Prognostic effect of the risk score in the training set (left) and testing set (right) of LUAD patients. (**D**) Prognostic effect of the risk score in the training set (left) and testing set (right) of stage IA LUAD patients. (**E**) Risk score-based OS prediction model in the training set. (**F**) Prediction of survival probability prediction of LUAD by the R package HSurADs and function HpSurADs. HPCs, highly proliferating cancer cells; LUAD, lung adenocarcinoma; OS, overall survival; rf, Random Forest; XGBoost, eXtreme Gradient Boosting

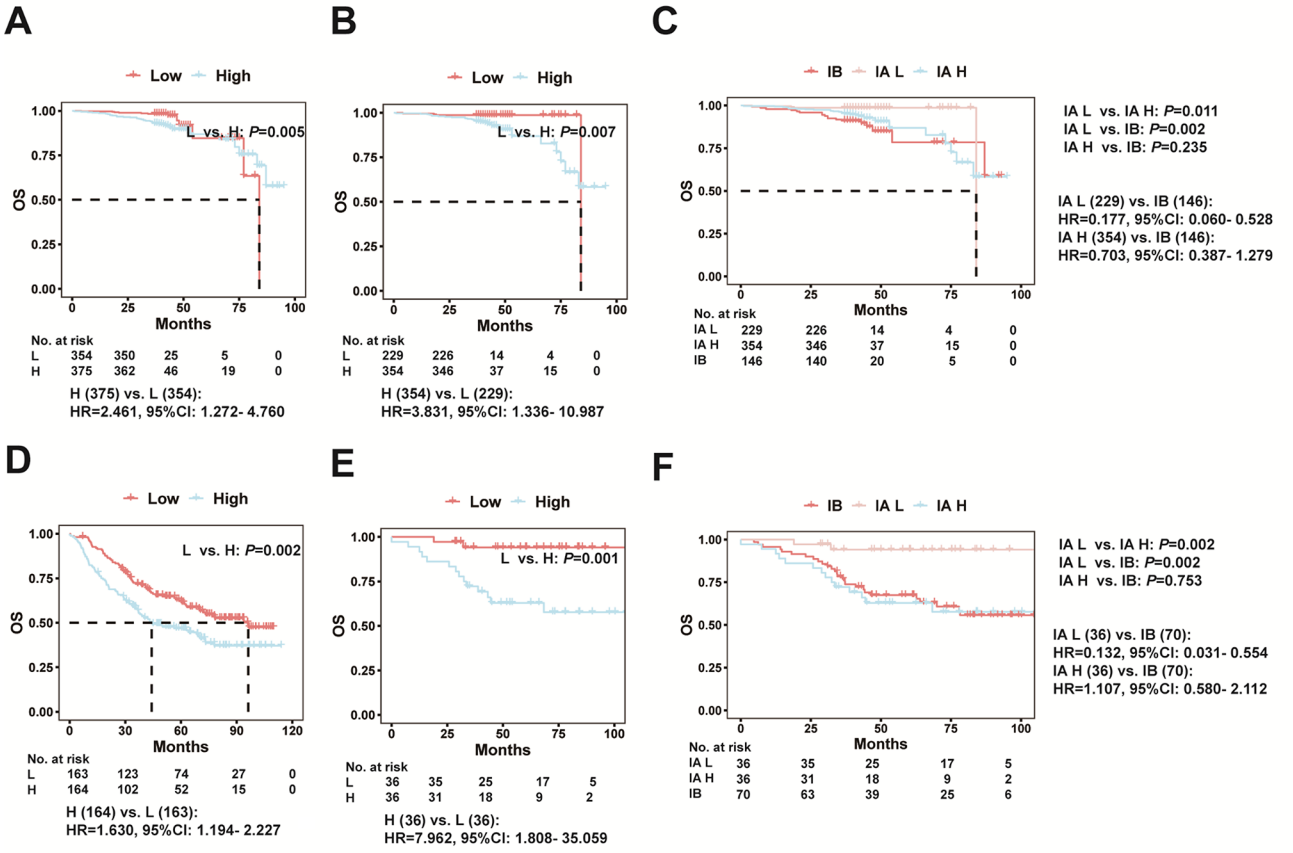

Fig. 7 Validation of the clinical significance of HPCs in LUAD. (**A**) OS curves for two categories of LUAD patients according to the level of Ki-67 (median value as the cut-off). (**B**) OS curves for two categories of stage IA LUAD patients according to the level of Ki-67 (median value as the cut-off). (**C**) OS curves of two risk types of stage IA LUAD patients divided by the median of Ki-67 (IA H and IA L) and stage IB LUAD patients. (**D**) OS curves for two categories of LUAD patients according to the level of HPCs (median value as the cut-off). (**E**) OS curves for two categories of stage IA LUAD patients according to the level of HPCs (median value as the cut-off). (**F**) OS curves of two risk types of stage IA LUAD patients divided by the median of HPCs (IA H and IA L) and stage IB LUAD patients. HPCs, highly proliferating cancer cells; LUAD, lung adenocarcinoma; OS, overall survival

had a significant negative prognostic effect on LUAD and stage IA LUAD (Supplemental Fig. 6D).

The HPC content significantly correlates with the pathological evolution of stage IA LUAD

Why do HPCs have a profound prognostic effect on IA LUAD? We analysed the relationship between the HPC content and the degree of malignancy of stage IA LUAD, which gradually increased during the gradual evolution of AIS-MIA-IAC. We collected early-stage LUAD samples (AIS: 20, MIA: 17, IAC (IA): 23) and conducted bulk RNA-seq. We found that the HPC content increased gradually through the AIS-MIA-IAC progression, as determined by ssGSEA (Fig. 8A). IHC staining of Ki-67 indicated that the HPC content increased gradually through the AIS-MIA-IAC progression (Fig. 8B). Furthermore, we collected early-stage LUAD samples (AIS: 3, MIA: 3, IAC (IA): 3) from our department and performed scRNA-seq. Immune cells (T/NK cells: TRBC1/GNLY, myeloid cells: AIF1/LYZ, B cells: CD79A/MS4A1), stromal cells (endothelial cells: PLVAP/VWF, fibroblasts: COL1A1/DCN) and epithelial cells (EPCAM/KRT7) were identified according to typical marker genes (Fig. 8C) (Supplemental Fig. 7A). The cancer cells were further identified by subdividing epithelial cells and performing CNV analysis (all the stroma and immune cells were the reference cells) (Fig. 8D) (Supplemental Fig. 7B). HPCs were found in early-stage LUAD by KEGG analysis, GO analysis, Reactome analysis and GSEA based on hallmarks (Supplemental Fig. 7C-F). We found that the proportion of HPCs gradually increased with the evolution of AIS-MIA-IAC (Fig. 8E). We further analysed the reasons for the high abundance of HPCs in LUAD. First, the TCGA-LUAD cohort containing mutation records was divided into two groups based on HPC content (Supplemental Fig. 7G). We found that TP53 mutations were the most significant genomic changes in the high-HPC group compared to the low-HPC group (Fig. 8F). A

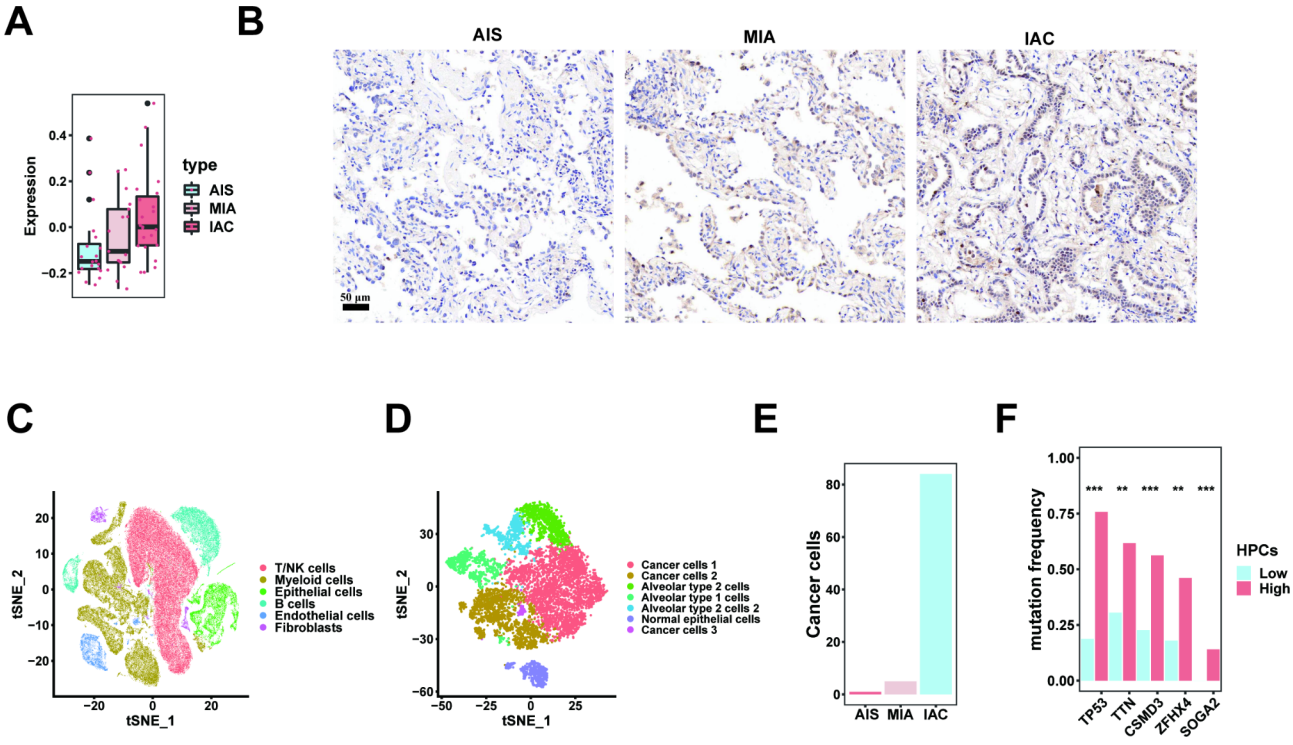

Fig. 8 Relationships between HPCs and the malignant evolution of stage IA LUAD. (**A**) Levels of HPCs in AIS, MIA and IAC (IA) tissues (RNA-seq) quantified by ssGSEA. (**B**) IHC staining for Ki-67 protein levels in AIS, MIA and IAC (IA) tissues. (**C**) tSNE was used to determine the distribution of total microenvironmental cells, including immune cells (T/NK cells, myeloid cells and B cells), stromal cells (endothelial cells and fibroblasts) and epithelial cells, in 9 early-stage LUAD tissues (scRNA-seq). (**D**) tSNE was used to determine the distribution of all epithelial cell subtypes, including normal and malignant epithelial cells, in 9 early-stage LUAD tissues (scRNA-seq). (**E**) Levels of HPCs in AIS, MIA and IAC in 9 early-stage LUAD tissues (scRNA-seq). (**F**) Driver mutation comparison analysis of patients with high and low expression of HPCs in the TCGA-LUAD database. HPCs, highly proliferating cancer cells; LUAD, lung adenocarcinoma; ssGSEA, single-sample gene set enrichment analysis; IHC, Immunohistochemistry; AIS, adenocarcinoma in situ; MIA, microinvasive carcinoma; IAC, invasive adenocarcinoma; scRNA-seq, single-cell RNA sequencing

greater HPC content in the TP53 mutant group was further demonstrated in three LUAD transcription datasets (GSE72094, GSE26939 and GSE13213) (Supplemental Fig. 7H-J). In conclusion, we believe that TP53 mutations promote the accumulation of HPCs and thus promote the progressive progression of early-stage LUAD.

Discussion

Previous studies have focused on the risk stratification of stage IA patients based on radiological and pathological features. For patients with stage IA LUAD, the composition and proportion of ground glass opacities, small tumour size and lepidic growth in pathological manifestations are gradually becoming stratification markers of low-risk subtypes [6, 10, 35, 36]. The prognosis of these tumours is excellent, and reducing the scope of surgery can also result in survival benefits. However, stage IA patients with pathological manifestations such as micropapilla, solid components, lymphovascular invasion, and spread through air spaces often have a worsened prognosis, and active adjuvant therapy may be needed after radical lobectomy [8, 35, 37, 38]. However, there are still defects in the previous classification of stage IA LUAD patients. The first problem is that the screening of high-risk subtypes is relatively insufficient, especially for patients with stage IA disease. The JCOG series of clinical trials has gradually identified low-risk patients who can benefit from sublobular treatment [36, 39]. The National Comprehensive Cancer Network (NCCN) guidelines and the World Health Organization (WHO) classification confirmed the pathological characteristics of low-risk subtypes, such as AIS and MIA [40]. However, the 5-year survival rate of stage IA LUAD patients is only approximately 80% (a considerable number of patients are still at risk of postoperative recurrence), and screening of these patients remains largely unsuccessful [41]. Stage IA3, the most severe subtype in the current staging system, is still associated with a better prognosis than stage IB; therefore, current staging methods cannot provide definite criteria for postoperative adjuvant chemotherapy. Moreover, clinical application based on current staging or risk stratification methods remains controversial and is not sufficiently convenient. For example, for part-solid tumours, subsolid nodules in TNM staging (8th) can be subdivided according to the size of the solid components, but some researchers believe that ground

glass composition can be classified as a good prognostic sign, regardless of the composition [6, 9, 42]. Although the pathological subtype criteria in the 5th WHO classification play an important role in determining prognosis, the extent of infiltration and the relationships between different subtypes in terms of composition proportions and clinical characteristics still lack rigorous evidence. Furthermore, the accurate pathological diagnosis of stage IA LUAD is extremely challenging due to both specimen quality and the ability of pathologists, which to some extent reduces the convenience of clinical applications.

The mechanism of carcinogenesis shows good potential for the risk classification of cancer. The biological characteristics of tumours have also become a potential basis for the clinical evaluation and treatment of LUAD. Molecular typing, represented by driver mutations, enables molecular targeted therapy, especially for epidermal growth factor receptor (EGFR) mutations and anaplastic lymphoma kinase (ALK) rearrangements. These targeted therapies have shown good results in patients with specific genetic backgrounds, as they significantly improved the survival status of patients compared with traditional chemotherapy [43]. Immunotyping based on the tumour immune background has also led to breakthroughs in immunotherapy. For example, immune checkpoint therapy based on the tumour mutation load and expression level of programmed cell death 1 ligand 1 (PD-L1) also shows superior effects to traditional chemotherapy. Moreover, its combination with platinum-containing dual drugs can significantly improve the prognosis of LUAD and even increase the likelihood of long-term survival [11]. Liquid biopsy, such as ctDNA detection, can be used to effectively evaluate the survival, recurrence and treatment sensitivity of LUAD patients [12, 13]. Malignant tumours are complex ecosystems in which a variety of noncancer cells and cancer cells communicate with each other and promote the occurrence and development of tumours. In particular, the rapid development of single-cell sequencing, spatial omics and other technologies has effectively evaluated the biological and clinical significance of different cells in the tumour microenvironment. For example, the levels of different types of fibroblasts can effectively predict the prognosis of LUAD and the risk of LUAD can be classified by the distribution of different types of fibroblasts [15, 44]. The infiltration of different types of macrophages is closely related to the prognosis of LUAD [14]. Risk scores based on tumour neutrophil differentiation-related genes can predict prognosis and immunotherapy outcomes in NSCLC [45]. In fact, molecular information has been collected to evaluate both its prognostic value and the feasibility of incorporating this information into the forthcoming (Ninth) Edition of the TNM Classification of Lung Cancer [16, 17]. However, clinical evaluation methods based on tumour biological characteristics are inadequate for stage IA LUAD.

Through scRNA-seq technology, we found that a small number of malignant cell populations with high proliferation-related gene expression were indeed present in LUAD (multiple biological enrichment methods revealed that this group of cells was a highly proliferative subpopulation, and cell communication analysis revealed that this subpopulation was fully in communication with the remaining microenvironment), and they were also found in LUAD tissues according to proliferation-specific molecular staining and FCM. We also validated the malignant proliferation characteristics of these cells in vitro. These cells are likely to function as stem cells or tumor-initiating cells (possessing higher proliferation rates, higher metastasis rates and other highly malignant features), which are critical to malignant initiation and progression [46, 47]. What is the relationship between HPCs and the malignant features of LUAD? We determined the gene expression characteristics of HPCs in LUAD via intersection analysis and quantified their content in multi-institution-acquired LUAD cohort transcriptome data via ssGSEA, which is a good method for estimating the whole-gene expression profile or functional status of cells. Regardless of the training set, test set, or validation set, HPCs showed significant prognostic value for LUAD. We found that the relative HPC content was an independent prognostic factor for LUAD. However, although the prognostic effect of the HPC content was significantly better than that of classic clinical factors such as age, sex, and smoking history, it was weaker than that of tumour TNM stage, suggesting that it may be used as a supplement to the existing stage.

So can HPCs be used as a basis for further classification of TNM stages? We first demonstrated a significant relationship between the level of HPCs and cancer stage, with a significant increasing relationship at stage I. The level of HPCs was an independent prognostic risk factor at different stages. However, it was more evident in stage I, even in stage IA. Since HPCs can be used as prognostic markers for stage IA LUAD, can they be used as a basis for high- to low-risk classification of stage IA LUAD? We classified stage IA via the HPC content. IA-H (with a higher level of HPCs) led to a significantly worse prognosis than IA-L (with a lower level of HPCs) but led to a prognosis similar to that of stage IB. When we performed the same stratification at the IB stage, we found that the prognosis of IB-L (with a lower level of HPCs), similar to that of stage IA, was better than that of IB-H (with a higher level of HPCs), but the prognosis of IB-H was better than that of stage IIA. These results suggest that the relative HPC content can be used as a basis for the classification of high-risk patients with stage IA disease. To facilitate clinical application, we used machine learning

algorithms to screen characteristic genes (HPC-associated genes) and established an R package (HSurADs) for predicting LUAD OS, which exhibited good efficacy. All the above results were verified in the validation set.

Why are HPCs capable of risk typing for stage IA LUAD? Moreover, what is the role of HPCs in the malignant progression of stage IA LUAD? Understanding the law of transformation from precursor glandular lesions to invasive lesions (AAH-AIS-MIA-IAC) is the key to exploring the malignant mechanisms of stage IA LUAD. In 2019, Hu et al. performed whole-exon sequencing of pathological specimens of early-stage LUAD. The authors found that AIS/MIA/IAC had more single nucleotide variations, copy number variations and subclonal variations than did AAH, and a gradient increase in driver gene variation existed during the evolution of AAH-AIS-MIA-IAC [48]. In 2019, through whole-exon sequencing of tissue samples, Zhang et al. reported that the tumour mutation load increased during the progression of AIS-MIA-IAC, and IAC showed more driver mutations, especially mutations in TP53 and other genes related to tumour DNA repair [49]. In 2019, Chen et al., through whole-exon sequencing and transcriptome sequencing of tissue samples, reported that TP53 mutations, arm-level chromosome copy number changes and HLA heterozygosity deletions increased during the progression of early-stage LUAD, and IAC also exhibited more genetic variation than did AIS/MIA [50]. In 2020, Li et al. confirmed the accumulation of TP53 and other driver genes during the evolution of AAH-AIS-MIA-IAC by organizing multiregional whole-exome sequencing [51]. In 2020, Lu et al. compared single-cell differences between ground glass nodules (MIAs) and solid nodules (IACs). The authors reported that both tumour cells and stromal cells differed and that tumour cell proliferation was relatively weaker in MIA at earlier stages [52]. Xing et al. revealed changes in the cell microenvironment, pathway molecules and other events among subsolid nodules (corresponding to MIA and IAC in early-stage LUAD), normal lung tissue and advanced LUAD and reported that NK cells and the toxic immune fraction in subsolid nodules were decreased (but still higher than those in advanced LUAD) [53]. In 2021, Wang et al. also demonstrated that increased energy metabolism, ribosome synthesis and stem-like tumour cells might promote the pathological evolution of early-stage LUAD [54]. In general, the literature suggests that with the progression of early-stage LUAD, tumour cells show more genetic variation, tumour biological behaviour becomes more malignant, and immunosuppression becomes more obvious. Through sc-RNA-seq, IHC staining and bulk RNA-seq of early-stage LUAD samples, we found that as stage IA LUAD progressed, the content of HPCs continuously increased. Considering the strong malignant proliferation potential of HPCs and their ability to communicate strongly with the microenvironment, we believe that HPCs may be involved in the malignant evolution of stage IA LUAD, which may partially explain the biological reasons for the classification of stage IA risk. In addition, we found that TP53 mutations promote the accumulation of HPCs in LUAD, and TP53 mutations are not only high-frequency mutations in LUAD but also key molecular events in the progression of glandular precursor disease to invasive adenocarcinoma [50, 51, 55, 56].

This study has several limitations. First, a retrospective cohort design (analysis of the retrospective database) was adopted in this study, and some bias is difficult to avoid. Second, RNA-seq, which is relatively expensive, is needed to determine the HPC content and risk score, and corresponding economic methods must be developed to increase the affordability of clinical application. Third, whether HPCs and the risk score can be combined with current IA risk classification parameters (radiological and pathological features) needs further verification. Fourth, whether HPCs can guide the treatment of stage IA LUAD, such as surgical methods and adjuvant therapy, needs to be rigorously explored in clinical trials. Fifth, the biological significance of HPCs needs to be studied in vitro and in vivo.

Conclusion

In conclusion, we provide a basis for HPC-based prognostic prediction and risk classification of early-stage LUAD, which may aid in accurate diagnosis and treatment in clinical practice.

Abbreviations

LUAD	Lung adenocarcinoma
HPCs	Highly proliferating cancer cells
scRNA-seq	Single-cell RNA sequencing
IHC	Immunohistochemical
OS	Overall survival
AAH	Atypical adenomatous hyperplasia
AIS	Adenocarcinoma in situ
MIA	Minimally invasive adenocarcinoma
IAC	Invasive adenocarcinoma
MRD	Minimal residual disease
ctDNA	Circulating tumor DNA
IASLC	International Association for the Study of Lung Cancer
TMA	Tissue Microarray
ASPM	Abnormal spindle-like microcephaly associated
SEER	Surveillance Epidemiology and End Results
TCGA	The Cancer Genome Atlas
CNV	Copy number variation
DEGs	Differentially expressed genes
KEGG	Kyoto Encyclopedia of Genes and Genomes
GO	Gene Ontology
ssGSEA	Single-sample gene set enrichment analysis
BH	Benjamini-Hochberg
rf	Random Forest
XGBoost	eXtreme Gradient Boosting
FCM	Flow cytometry
NCCN	National Comprehensive Cancer Network
WHO	The World Health Organization
EGFR	Epidermal growth factor receptor

ALK	Anaplastic lymphoma kinase
PD-L1	Programmed Cell Death 1 Ligand 1

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-024-13308-0.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Supplementary Material 7

Supplementary Material 8

Supplementary Material 9

Supplementary Material 10

Supplementary Material 11

Acknowledgements
Not applicable.

Author contributions
YLX, TJ, and YL had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis, including and especially any adverse effects; YLX contributed to the conceptualization, data curation and software; YFM and JL contributed to the conception, analysis data and formal analysis; YH, QZ, MMW, and JFZ contributed to the acquisition and interpretation of data for the work. NLX, QL and FT contributed to the design of the work. All authors contributed to the manuscript writing and revising, gave final approval of the manuscript as submitted and agreed to be accountable for all aspects of the work.

Funding
This study was funded by the National Natural Science Foundation of China. (82002421), Health Care Special Research Project of Military Logistics Research Program (23BJZ23,19BJZ20), Key Projects in Military Logistics (BKJ21J009), Youth Talent Lifting Program of Shaanxi Association for Science and Technology (20230311), Key research and development plan of Shaanxi Province (2023-YBSF-318), Health Research Fund of Shaanxi Province (2021B004), Young Talent Program of Tangdu Hospital, Clinical Research Program of Air Force Medical University (2023LC2337), Partner Laboratory Cooperation Exchange Program of Air Force Medical University (2022HB010), New Clinical Technology Project of Tangdu Hospital (XJSXYW-2023041), Discipline Innovation Development Plan Project of Tangdu Hospital (2021LCYJ005), the "Open Competition Mechanism to Select the Best Candidates" Project of Tangdu Hospital (2022TDGS002), and Tangdu Hospital's 2024 Discipline Boosting Program (2024JSYX008).

Data availability
The raw sequence data reported in this paper have been deposited at GEO (https://www.ncbi.nlm.nih.gov/geo/) under accession GSE189357, GSE282774 and GSE282617.

Declarations

Ethics approval and consent to participate
For data obtained from public database, all ethics approvals and informed consents to participate were already obtained by the original studies. Besides, all procedures performed in studies involving human participants were conducted in accordance with the Declaration of Helsinki and its later amendments or comparable ethical standards, and approved by the ethics committee of Tangdu Hospital (approval No: K202107-19, No. K202003-018, and No. TDLL-2017016), registration details of this clinical trial (registry: Chinese Clinical Trial Registry, ChiCTR; trial registration number: ChiCTR2200059416; data of registration: 2022-04-29) and all patients signed informed consent forms.

Consent for publication
Not applicable.

Generative AI in scientific writing
We declare that we have not used artificial intelligence (AI)-assisted technologies in the production of the submitted work.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Thoracic Surgery, The First Medical Center, Chinese PLA General Hospital, Beijing, China
[2]Department of Thoracic Surgery, Tangdu Hospital, Fourth Military Medical University, Xi'an, China
[3]Innovation Center for Advanced Medicine, Tangdu Hospital, Fourth Military Medical University, Xi'an, China
[4]Department of Thoracic Surgery, Shaanxi Provincial People's Hospital, Xi'an, China
[5]Department of Thoracic Surgery, Air Force Medical Center, Fourth Military Medical University, Beijing, China

Received: 19 July 2024 / Accepted: 6 December 2024
Published online: 07 January 2025

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and Mortality Worldwide for 36 cancers in 185 countries. Cancer J Clin. 2021;71(3):209–49.
2. Zappa C, Mousa SA. Non-small cell lung cancer: current treatment and future advances. Translational lung cancer Res. 2016;5(3):288–300.
3. Thomas A, Liu SV, Subramaniam DS, Giaccone G. Refining the treatment of NSCLC according to histological and molecular subtypes. Nat Reviews Clin Oncol. 2015;12(9):511–26.
4. Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman JR, Bharat A, Bruno DS, Chang JY, Chirieac LR, D'Amico TA, et al. Non-small Cell Lung Cancer, Version 3.2022, NCCN Clinical Practice guidelines in Oncology. J Natl Compr Cancer Network: JNCCN. 2022;20(5):497–530.
5. Nicholson AG, Tsao MS, Beasley MB, Borczuk AC, Brambilla E, Cooper WA, Dacic S, Jain D, Kerr KM, Lantuejoul S, et al. The 2021 WHO classification of lung tumors: impact of advances since 2015. J Thorac Oncology: Official Publication Int Association Study Lung Cancer. 2022;17(3):362–87.
6. Travis WD, Asamura H, Bankier AA, Beasley MB, Detterbeck F, Flieder DB, Goo JM, MacMahon H, Naidich D, Nicholson AG, et al. The IASLC Lung Cancer Staging Project: proposals for Coding T Categories for Subsolid Nodules and Assessment of Tumor Size in Part-Solid tumors in the Forthcoming Eighth Edition of the TNM classification of Lung Cancer. J Thorac Oncology: Official Publication Int Association Study Lung Cancer. 2016;11(8):1204–23.
7. Qian F, Yang W, Wang R, Xu J, Wang S, Zhang Y, Jin B, Yu K, Han B. Prognostic significance and adjuvant chemotherapy survival benefits of a solid or micropapillary pattern in patients with resected stage IB lung adenocarcinoma. J Thorac Cardiovasc Surg. 2018;155(3):1227–e12351222.
8. Wang C, Yang J, Lu M. Micropapillary Predominant Lung Adenocarcinoma in Stage IA benefits from Adjuvant Chemotherapy. Ann Surg Oncol. 2020;27(6):2051–60.
9. Hattori A, Matsunaga T, Takamochi K, Oh S, Suzuki K. Neither maximum tumor size nor solid component size is Prognostic in Part-Solid Lung Cancer: impact of tumor size should be Applied exclusively to solid Lung Cancer. Ann Thorac Surg. 2016;102(2):407–15.
10. Hattori A, Suzuki K, Takamochi K, Wakabayashi M, Aokage K, Saji H, Watanabe SI. Prognostic impact of a ground-glass opacity component in clinical stage IA non-small cell lung cancer. J Thorac Cardiovasc Surg. 2021;161(4):1469–80.

11. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Nature. 2018;553(7689):446–54.
12. Rolfo C, Mack P, Scagliotti GV, Aggarwal C, Arcila ME, Barlesi F, Bivona T, Diehn M, Dive C, Dziadziuszko R, et al. Liquid Biopsy for Advanced NSCLC: a Consensus Statement from the International Association for the study of Lung Cancer. J Thorac Oncology: Official Publication Int Association Study Lung Cancer. 2021;16(10):1647–62.
13. Nigro MC, Marchese PV, Deiana C, Casadio C, Galvani L, Di Federico A, De Giglio A. Clinical utility and application of Liquid Biopsy genotyping in Lung Cancer: a Comprehensive Review. Lung Cancer (Auckland NZ). 2023;14:11–25.
14. Hu Z, Jin X, Hong W, Sui Q, Zhao M, Huang Y, Li M, Wang Q, Zhan C, Chen Z. Dissecting the single-cell transcriptome network of macrophage and identifies a signature to predict prognosis in lung adenocarcinoma. Cell Oncol (Dordrecht). 2023;46(5):1351–68.
15. Cords L, Engler S, Haberecker M, Rüschoff JH, Moch H, de Souza N, Bodenmiller B. Cancer-associated fibroblast phenotypes are associated with patient outcome in non-small cell lung cancer. Cancer Cell 2024.
16. Asamura H, Nishimura KK, Giroux DJ, Chansky K, Hoering A, Rusch V, Rami-Porta R. IASLC Lung Cancer Staging Project: the New Database to inform revisions in the Ninth Edition of the TNM classification of Lung Cancer. J Thorac Oncology: Official Publication Int Association Study Lung Cancer. 2023;18(5):564–75.
17. Rami-Porta R, Nishimura KK, Giroux DJ, Detterbeck F, Cardillo G, Edwards JG, Fong KM, Giuliani M, Huang J, Kernstine KH et al. Sr. : The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groups in the Forthcoming (Ninth) Edition of the TNM Classification for Lung Cancer. *Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer* 2024.
18. Wei L, Jin Z, Yang S, Xu Y, Zhu Y, Ji Y. TCGA-assembler 2: software pipeline for retrieval and processing of TCGA/CPTAC data. Bioinf (Oxford England). 2018;34(9):1615–7.
19. Zhu Y, Qiu P, Ji Y. TCGA-assembler: open-source software for retrieving and processing TCGA data. Nat Methods. 2014;11(6):599–600.
20. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee AV, et al. An Integrated TCGA Pan-cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell. 2018;173(2):400–e416411.
21. Athar A, Füllgrabe A, George N, Iqbal H, Huerta L, Ali A, Snow C, Fonseca NA, Petryszak R, Papatheodorou I, et al. ArrayExpress update - from bulk to single-cell expression data. Nucleic Acids Res. 2019;47(D1):D711–5.
22. Davis S, Meltzer PS. GEOquery: a bridge between the Gene expression Omnibus (GEO) and BioConductor. Bioinf (Oxford England). 2007;23(14):1846–7.
23. Doll KM, Rademaker A, Sosa JA. Practical guide to Surgical Data sets: Surveillance, Epidemiology, and end results (SEER) database. JAMA Surg. 2018;153(6):588–9.
24. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. PLoS Med. 2012;9(5):e1001216.
25. Saikia M, Cheung N, Singh AK, Kapoor V. Role of Midkine in Cancer Drug Resistance: regulators of its expression and its molecular targeting. Int J Mol Sci 2023, 24(10).
26. Filippou PS, Karagiannis GS, Constantinidou A. Midkine (MDK) growth factor: a key player in cancer progression and a promising therapeutic target. Oncogene. 2020;39(10):2040–54.
27. Foo SL, Yap G, Cui J, Lim LHK. Annexin-A1 - A blessing or a curse in Cancer? Trends Mol Med. 2019;25(4):315–27.
28. Boudhraa Z, Bouchon B, Viallard C, D'Incan M, Degoul F. Annexin A1 localization and its relevance to cancer. Clin Sci (London England: 1979). 2016;130(4):205–20.
29. Sheikh MH, Solito E. Annexin A1: uncovering the many talents of an old protein. Int J Mol Sci 2018, 19(4).
30. Araújo TG, Mota STS, Ferreira HSV, Ribeiro MA, Goulart LR, Vecchi L. Annexin A1 as a Regulator of Immune Response in Cancer. Cells 2021, 10(9).
31. Wu X, Li Z, Wang ZQ, Xu X. The neurological and non-neurological roles of the primary microcephaly-associated protein ASPM. Front NeuroSci. 2023;17:1242448.
32. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Stein TI, Nudel R, Lieder I, Mazor Y et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Current protocols in bioinformatics* 2016, 54:1.30.31–31.30.33.
33. Warth A, Cortis J, Soltermann A, Meister M, Budczies J, Stenzinger A, Goeppert B, Thomas M, Herth FJ, Schirmacher P, et al. Tumour cell proliferation (Ki-67) in non-small cell lung cancer: a critical reappraisal of its prognostic role. Br J Cancer. 2014;111(6):1222–9.
34. Sobecki M, Mrouj K, Colinge J, Gerbe F, Jay P, Krasinska L, Dulic V, Fisher D. Cell-cycle Regulation accounts for Variability in Ki-67 expression levels. Cancer Res. 2017;77(10):2722–34.
35. Liu C, Wang LC, Chen HS, Yeh YC, Hsu PK, Huang CS, Hsieh CC, Hsu HS. Outcomes of patients with different lepidic percentage and tumor size of stage I lung adenocarcinoma. Thorac cancer 2022.
36. Saji H, Okada M, Tsuboi M, Nakajima R, Suzuki K, Aokage K, Aoki T, Okami J, Yoshino I, Ito H, et al. Segmentectomy versus lobectomy in small-sized peripheral non-small-cell lung cancer (JCOG0802/WJOG4607L): a multicentre, open-label, phase 3, randomised, controlled, non-inferiority trial. Lancet (London England). 2022;399(10335):1607–17.
37. Tsuchiya T, Akamine S, Muraoka M, Kamohara R, Tsuji K, Urabe S, Honda S, Yamasaki N. Stage IA non-small cell lung cancer: vessel invasion is a poor prognostic factor and a new target of adjuvant chemotherapy. Lung cancer (Amsterdam Netherlands). 2007;56(3):341–8.
38. Chen D, Wang X, Zhang F, Han R, Ding Q, Xu X, Shu J, Ye F, Shi L, Mao Y, et al. Could tumor spread through air spaces benefit from adjuvant chemotherapy in stage I lung adenocarcinoma? A multi-institutional study. Therapeutic Adv Med Oncol. 2020;12:1758835920978147.
39. Suzuki K, Watanabe SI, Wakabayashi M, Saji H, Aokage K, Moriya Y, Yoshino I, Tsuboi M, Nakamura S, Nakamura K, et al. A single-arm study of sublobar resection for ground-glass opacity dominant peripheral lung cancer. J Thorac Cardiovasc Surg. 2022;163(1):289–e301282.
40. Nicholson AG, Tsao MS, Beasley MB, Borczuk AC, Brambilla E, Cooper WA, Dacic S, Jain D, Kerr KM, Lantuejoul S et al. The 2021 WHO classification of lung tumors: impact of advances since 2015. J Thorac Oncology: Official Publication Int Association Study Lung Cancer 2021.
41. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, Nicholson AG, Groome P, Mitchell A, Bolejack V. The IASLC Lung Cancer Staging Project: proposals for revision of the TNM Stage groupings in the Forthcoming (Eighth) Edition of the TNM classification for Lung Cancer. J Thorac Oncology: Official Publication Int Association Study Lung Cancer. 2016;11(1):39–51.
42. Ye T, Deng L, Wang S, Xiang J, Zhang Y, Hu H, Sun Y, Li Y, Shen L, Xie L, et al. Lung adenocarcinomas manifesting as Radiological Part-Solid Nodules define a special clinical subtype. J Thorac Oncology: Official Publication Int Association Study Lung Cancer. 2019;14(4):617–27.
43. Chalela R, Curull V, Enríquez C, Pijuan L, Bellosillo B, Gea J. Lung adenocarcinoma: from molecular basis to genome-guided therapy and immunotherapy. J Thorac Dis. 2017;9(7):2142–58.
44. Hanley CJ, Waise S, Ellis MJ, Lopez MA, Pun WY, Taylor J, Parker R, Kimbley LM, Chee SJ, Shaw EC, et al. Single-cell analysis reveals prognostic fibroblast subpopulations linked to molecular and immunological subtypes of lung cancer. Nat Commun. 2023;14(1):387.
45. Pang J, Yu Q, Chen Y, Yuan H, Sheng M, Tang W. Integrating single-cell RNA-seq to construct a Neutrophil prognostic model for predicting immune responses in non-small cell lung cancer. J Translational Med. 2022;20(1):531.
46. Clarke MF. Clinical and therapeutic implications of Cancer Stem cells. N Engl J Med. 2019;380(23):2237–45.
47. Bayik D, Lathia JD. Cancer stem cell-immune cell crosstalk in tumour progression. Nat Rev Cancer. 2021;21(8):526–36.
48. Hu X, Fujimoto J, Ying L, Fukuoka J, Ashizawa K, Sun W, Reuben A, Chow CW, McGranahan N, Chen R, et al. Multi-region exome sequencing reveals genomic evolution from preneoplasia to lung adenocarcinoma. Nat Commun. 2019;10(1):2978.
49. Zhang C, Zhang J, Xu FP, Wang YG, Xie Z, Su J, Dong S, Nie Q, Shao Y, Zhou Q, et al. Genomic Landscape and Immune Microenvironment Features of Preinvasive and Early Invasive Lung Adenocarcinoma. J Thorac Oncology: Official Publication Int Association Study Lung Cancer. 2019;14(11):1912–23.
50. Chen H, Carrot-Zhang J, Zhao Y, Hu H, Freeman SS, Yu S, Ha G, Taylor AM, Berger AC, Westlake L, et al. Genomic and immune profiling of pre-invasive lung adenocarcinoma. Nat Commun. 2019;10(1):5472.
51. Li Y, Li X, Li H, Zhao Y, Liu Z, Sun K, Zhu X, Qi Q, An B, Shen D et al. Genomic characterisation of pulmonary subsolid nodules: mutational landscape and radiological features. Eur Respir J 2020, 55(2).
52. Lu T, Yang X, Shi Y, Zhao M, Bi G, Liang J, Chen Z, Huang Y, Jiang W, Lin Z, et al. Single-cell transcriptome atlas of lung adenocarcinoma featured with ground glass nodules. Cell Discovery. 2020;6:69.

53. Xing X, Yang F, Huang Q, Guo H, Li J, Qiu M, Bai F, Wang J. Decoding the multicellular ecosystem of lung adenocarcinoma manifested as pulmonary subsolid nodules by single-cell RNA sequencing. Sci Adv 2021, 7(5).
54. Wang Z, Li Z, Zhou K, Wang C, Jiang L, Zhang L, Yang Y, Luo W, Qiao W, Wang G, et al. Deciphering cell lineage specification of human lung adenocarcinoma with single-cell RNA sequencing. Nat Commun. 2021;12(1):6500.
55. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511(7511):543–50.
56. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, Colaprico A, Wendl MC, Kim J, Reardon B, et al. Comprehensive characterization of Cancer driver genes and mutations. Cell. 2018;173(2):371–e385318.

Publisher's note