



# PSATluca is a Nomogram Predictor of Survival and Treatment of Lung Cancer Patients Based on Serum and Clinical Characters

Longxiang Xie,<sup>1,#</sup> Shengnan Wu,<sup>1,3,#</sup> Qiang Wang,<sup>2,#</sup> Jing Wang,<sup>3</sup> Yuxuan He<sup>4</sup> and Xiangqian Guo<sup>1,\*</sup>

## Abstract

To date, albeit a great number of serum biomarkers for early cancer diagnosis have been well characterized and widely used in clinics, the prognostic value of these serum biomarkers remains to be comprehensively analyzed. Here, we applied our previously built prognosis analysis database Long-term Outcome and Gene Expression Profiling Database of pan-cancers (LOGpc) to screen the commonly used serum biomarkers with prognostic values. Univariate and multivariate Cox analysis were conducted to evaluate the prognostic value of serum biomarkers and clinical factors in outcome measurement. By analyzing the LOGpc database, high cytokeratin 19 (KRT19) messenger RNA (mRNA) expression was found to be significantly correlated with poor overall survival (OS) in non-small cell lung cancer (NSCLC). A total of 232 serum samples were collected for validation, which reaffirmed that increased CYFRA21-1 (fragmented protein form of KRT19) was independently associated with poor OS in NSCLC. A nomogram integrating serum CYFRA21-1 level and clinico-pathological factors was constructed. This nomogram showed good accuracy in predicting OS, and could recommend treatment options based on CYFRA21-1 abundances and patient characteristics. A web implementation of this nomogram called PSATluca (Predictor of Survival And Treatment of lung cancer Patient) is provided and accessible to users at <https://bioinfo.henu.edu.cn/PSATluca/index.jsp>. This tool can assist clinicians in evaluating prognosis of NSCLC patients and selecting appropriate treatment options.

**Keywords:** CYFRA21-1; Non-small cell lung cancer; Prognosis; Serous biomarker; LOGpc; PASTluca.

Received: 14 April 2025; Revised: 05 June 2025; Accepted: 17 July 2025.

Article type: Research article.

## 1. Introduction

Cancer is a leading cause of mortality. There were, nearly 9.7 million cancer-related deaths and 20 million new cancer cases worldwide in 2022 (Global cancer statistics 2022). In China, there were approximately 3 million cancer deaths and 5 million new cancer cases in 2022.<sup>[1]</sup> Despite substantial improvements in diagnosis and treatment, the prognosis, especially among those with advanced cancers, remains poor. Currently, the AJCC/UICC TNM staging system is the most commonly used tool to predict cancer prognosis.<sup>[2]</sup> However, a growing number of researches had suggested that the TNM staging system alone is insufficient to reliably predict cancer

prognosis. Identifying highly effective and easily accessible molecular biomarkers could assist the prognosis assessment.

Scientists have devoted enormous efforts to discovering and characterizing serous biomarkers for early cancer detection and diagnosis. To date, a great number of biomarkers, including carbohydrate antigen 19-9 (CA19-9), carbohydrate antigen 125 (CA125), carcinoembryonic antigen (CEA), prostate-specific antigen (PSA), alpha-fetoprotein (AFP), cytokeratin 19 fragment 21-1 (CYFRA 21-1), have been established as Tumor early Diagnostic Biomarkers (TDB). For example, CA19-9 is one of the most widely used serous biomarkers for the diagnosis of symptomatic pancreatic cancer and the monitoring of treatment.<sup>[3]</sup> Recently, some serous markers were found to have prognostic values, including CA19-9 as a prognostic biomarker for advanced pancreatic cancer patients receiving gemcitabine chemotherapy,<sup>[4]</sup> and CA125 as an independent unfavorable prognostic biomarker for patients with colorectal cancer (CRC),<sup>[5]</sup> etc. These findings suggested the feasibility of identifying biomarkers with prognostic values from the TDB pool. The identification and validation of prognostic value of the serous TDBs requires more systematic studies.

<sup>1</sup> Cell Signal Transduction Laboratory, Henan Provincial Engineering Center for Tumor Molecular Medicine, Zhongyuan Intelligent Medical Laboratory, School of Basic Medical Sciences, Henan University, Kaifeng, 475004, China

<sup>2</sup> School of Software, Henan University, Kaifeng, 475004, China

<sup>3</sup> Kaifeng Central Hospital, Kaifeng, 475001, China

<sup>4</sup> School of Life Sciences, Henan University, Kaifeng, 475004, China

<sup>#</sup> These authors contributed equally to this work.

\*Email: [xqguo@henu.edu.cn](mailto:xqguo@henu.edu.cn) (X. Guo)

In this study, we aimed to assess the correlation between the serum TDB expression and prognosis by comprehensive bioinformatics analysis and retrospective study, and to construct a nomogram incorporating various clinicopathological factors to support the outcome measurement and therapeutic decision-making in tumor patients.

## 2. Materials and methods

### 2.1 Overall study design

Our study contains four main phases: data collection, potential serous prognostic biomarker discovery, clinical validation, and nomogram construction (Fig. 1).

### 2.2 Collection of tumor diagnostic serous biomarkers (TDB) in clinic

We collected 29 serum TDBs (including NSE, CEA, CA125, CYFRA21-1, *etc.*) that are commonly used in local hospitals. For example, tumors related to the origin of neuroendocrine tissue, especially small cell lung cancer (SCLC), overexpress NSE, resulting in a significantly elevated serous NSE level.<sup>[6,7]</sup> The diagnostic roles of 29 serous biomarkers in human tumors are shown in Fig. 2. ACP: Acid Phosphatase; ACTH: Adrenocorticotrophic Hormone; AFP: Alpha Fetoprotein; ALP: ALkaline Phosphatase; CA125: Carbohydrate Antigen 125; CA15-3: Carbohydrate Antigen 15-3; CA19-9: Carbohydrate

Antigen 19-9; CA24-2: Carbohydrate Antigen 24-2; CA50: Carbohydrate Antigen 50; CA724: Carbohydrate Antigen 724; CEA: Carcinoembryonic Antigen; CYFRA21-1: Cytokeratin 19 Fragment Antigen 21-1; G-17: Gastrin; GP73: Golgi Protein 73; HE4: Human Epididymis Protein 4 ; NMP-22: Nuclear Matrix Protein 22; NSE: Neuron-specific Enolase; PAP: Prostatic Acid Phosphatase; PG: Pepsin; PRL: Prolactin; ProGRP: Pro-Gastrin-Releasing Peptide; PSA: Prostate Specific Antigen; SCC: Squamous Cell Carcinoma Antigen; SF: Ferritin; Tg: Thyroglobulin; TPA: Tissue Polypeptide Antigen;  $\gamma$ -GT:  $\gamma$ -Gglutamyl Transpeptadase;  $\beta$ 2-MG:  $\beta$ 2-Microglobulin Antibody;  $\beta$ -hCG: Human Chorionic Gonadotrophin- $\beta$ .

### 2.3 Bioinformatics analysis of TDB in LOGpc database

We previously developed LOGpc (Long-term Outcome and Gene Expression Profiling Database of pan-cancers), a database (<https://bioinfo.henu.edu.cn/DatabaseList.jsp>) that encompasses 209 gene expression datasets and provides 13 types of survival terms for 31310 patients of 27 distinct malignancies. In this study, we applied LOGpc to determine the correlation between the serous biomarker abundance and prognosis.  $P < 0.05$  was set as the screening cut off to call the prognostic values.

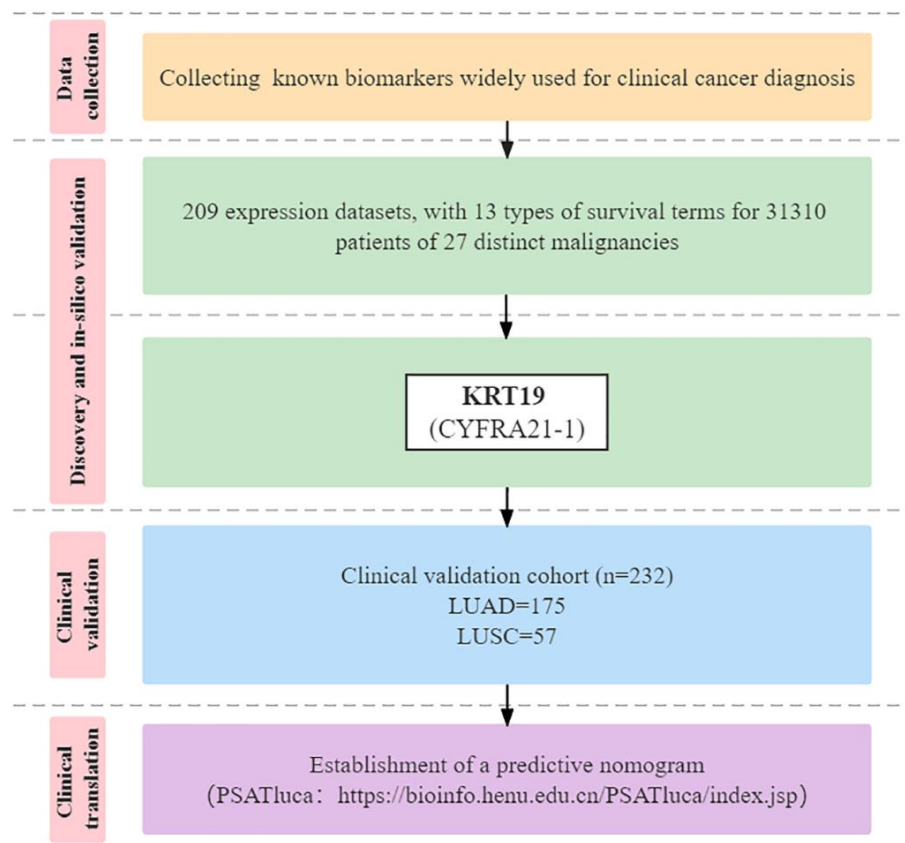
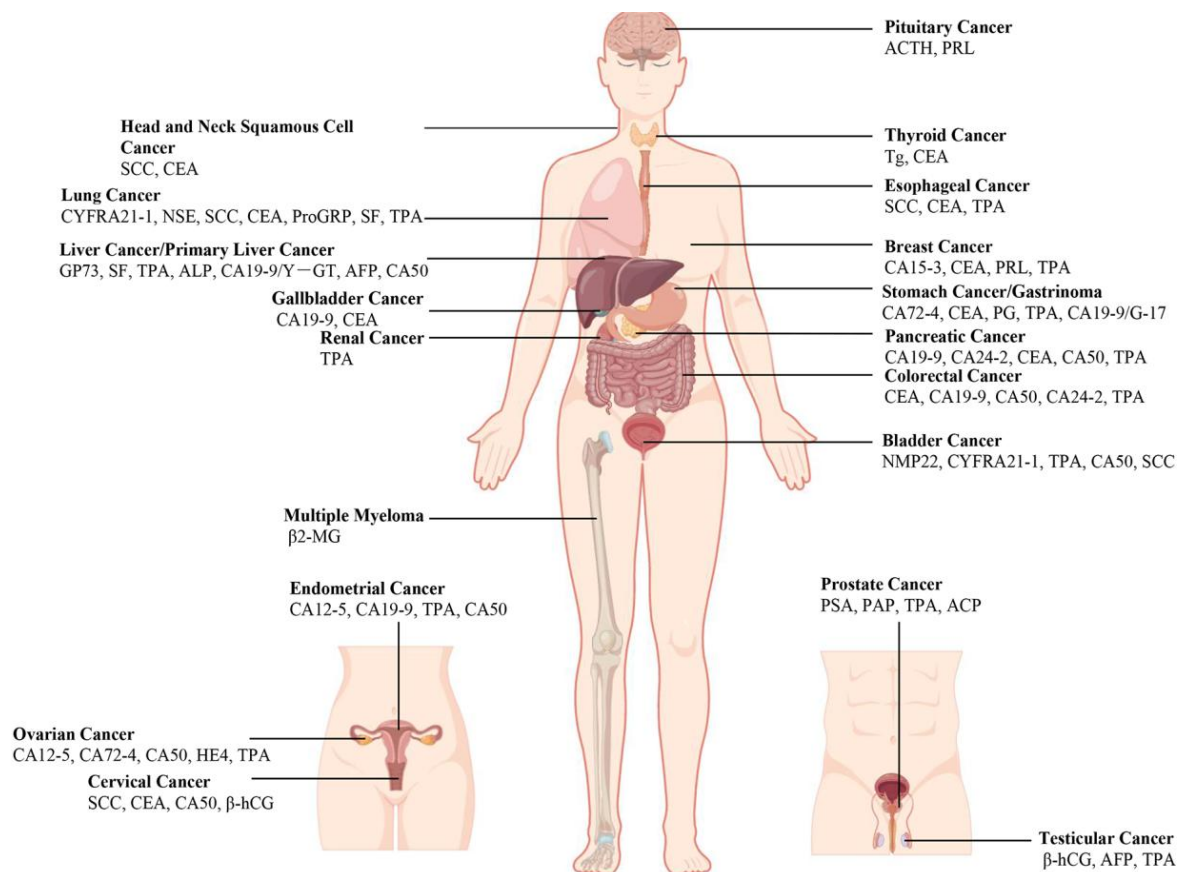


Fig. 1: Diagram of overall experimental design.



**Fig. 2:** The commonly used serous TDBs and their diagnostic roles in different types of tumors, which was made by Figdraw ([www.figdraw.com](http://www.figdraw.com)). The unique ID is AYUSU0b384.

**2.4 Sample collection and testing**

To validate the LOGpc results, 232 serous NSCLC samples were retrospectively collected at Kaifeng Central Hospital, including 175 lung adenocarcinoma (LUAD, 75.4%) and 57 lung squamous cell carcinoma (LUSC, 24.6%) samples. The median age was 70 years (40-90 years), and there were 89 females (38.4%) and 143 males (61.6%). 175 patients (75.4%) had received treatment. The demographic and clinico-pathological information is listed in Table 1.

**2.5 Combination therapy refers to patients who received two or more of these therapies**

To experimentally validate the prognostic value of KRT19 in NSCLC, we collected 232 serous samples of NSCLC patients from Kaifeng central hospital, including 175 LUAD samples and 57 LUSC samples, and analyzed the abundance of CYFRA21-1 (fragmented protein form of KRT19) in these serous samples (Table 1).

The serous samples were prepared from the whole blood of NSCLC patients (without any anticoagulant), placed until coagulated, the cores were separated by centrifugation at 3000 rpm for 10 min within 12 hours. Serous CYFRA21-1 levels were measured by chemiluminescence (by Roche CCM system). After the test, serous samples were stored in a -80 °C freezer. The reference range of serum CYFRA21-1 level in healthy people is 0-3.3 ng/ml. This study was approved by the

Biomedical Research Ethics committee of Henan University and Medical ethics committee of KaiFeng Central Hospital. Because of the retrospective design, informed consents from the participants were waived.

**2.6 Univariate and multivariate Cox regression analysis**

Univariate Cox regression analysis and multivariate Cox regression analysis were used to evaluate the independent prognostic roles of serous biomarkers and clinical factors. The R package “forestplot” was used to visualize the Cox results. Survival analysis was performed by “survival” package. All the aforementioned statistics were performed by R software (version 4.0.3).  $P < 0.05$  was considered as statistically significant.

**2.7 Establishment of the nomogram and Tumor microenvironment (TME) deconvolution estimation**

Constructed by R packages “rms” and “survival”, the nomogram integrates serous biomarkers and clinical attributes and could predict 1-, 3- and 5-year LUCA OS (Overall Survival). The predictive accuracy of the nomogram was evaluated by calibration plot. The web server PSATluca was implemented by Java. The detailed procedure was described as previous study.<sup>[8]</sup> “IOBR” R package was used to calculate the landscape of the TME,<sup>[9]</sup> using three deconvolution methods, namely, CIBERSORT,<sup>[10]</sup> EPIC,<sup>[11]</sup> and xCell.<sup>[12]</sup>

**Table 1:** Clinical features of enrolled NSCLC patients in validation cohort.

Characteristics	Number	%	#High CYFRA21-1	%
Number of Patients	232	100.0	31	13.4
Age				
31-50	11	4.7		
51-70	137	59.1		
71-90	84	36.2		
Sex				
Male	143	61.6	26	11.2
Female	89	38.4	5	2.2
Type				
LUAD	175	75.4	20	8.6
LUSC	57	24.6	11	4.7
Stage				
II	1	0.4	0	0
III	21	9.1	1	0.4
IV	126	54.3	21	9.1
NA	84	36.2	9	3.9
Metastases				
Yes	121	52.2	22	9.5
No	111	47.8	9	3.9
Status				
Living	171	73.7	17	7.3
Dead	61	26.3	14	6.0
Treatment				
Chemotherapy alone	50	21.5	5	2.2
Radiotherapy alone	10	4.3	4	1.7
Targeted therapy alone	48	20.7	8	3.4
*Combination therapy	67	28.9	9	3.9
No treatment	57	24.6	5	2.2

Note: # patients with high serous CYFRA21-1 level above the optimal cut off value (Fig. S1); \* Current treatment regimens include chemotherapy, radiotherapy and targeted therapy.

**2.8 Statistical analysis**

OS was defined from the date of serous CYFRA21-1 collection to the date of death or last follow-up. All statistical analyses were performed using R V.4.1.1, SPSS 21.0, GraphPad Prism 8. To best stratify the cancer patients with distinct outcomes, X-tile V.3.6.1 was used to optimize the CYFRA21-1 abundance cutoff. In present study,  $P < 0.05$  was considered statistically significant.

**3. Results and discussion**

**3.1 Identification of serous prognostic biomarkers in pancreaticancers**

We used previously built prognosis analysis database called

LOGpc, which integrates 209 TCGA and GEO datasets of 31310 tumor cases of 27 distinct malignancies, to determine the prognostic values of the 29 commonly used TDBs. The results showed that *KRT19* mRNA expression (corresponding to serous CYFRA21-1) was significantly and consistently correlated with prognosis of NSCLC patients in multiple cohorts (Table 2), whereas there were only poor correlations between the expression of other TDBs and prognosis across different cancer cohorts (results not shown).

Hence, in following sections of this study, we mainly focused on assessing the relationship between the expression of *KRT19* and prognosis. In NSCLC patients, the high expression of *KRT19* was significantly correlated with poor

**Table 2:** Significant correlation of *KRT19* mRNA expression with prognosis in LUAD cohorts.

Data Source	Sample Size	Country	Cut off	<i>P</i>	HR (95% CI)	Prognosis
GSE67639	1635	USA	Upper 25%	<0.0001	1.572 (1.2814-1.9282)	Poor
GSE13213	363	Japan	Upper 25%	2e-04	2.992 (1.6898-5.2978)	Poor
GSE68571	178	USA	Upper 25%	0.0026	3.452 (1.5397-7.7369)	Poor
GSE31210	633	Japan	Upper 25%	0.003	2.746 (1.4079-5.3546)	Poor
GSE11969	303	Japan	Upper 25%	0.0314	1.995 (1.0636-3.7437)	Poor
GSE29066	77	Sweden	Upper 25%	0.0052	3.174 (1.4126-7.1302)	Poor
GSE30219	178	France	Upper 25%	<0.0001	4.461 (2.3281-8.5492)	Poor
GSE41271	349	USA	Upper 25%	0.0376	1.683 (1.0304-2.7498)	Poor
TCGA	597	NA	Upper 50%	0.0286	1.388 (1.0349-1.8604)	Poor
GSE67639	1635	USA	Upper 50%	2e-04	1.433 (1.1826-1.7335)	Poor
GSE13213	363	Japan	Upper 50%	0.0342	1.864 (1.0473-3.3174)	Poor
GSE68571	178	USA	Upper 50%	0.0188	2.877 (1.1913-6.9498)	Poor
GSE31210	633	Japan	Upper 50%	0.0019	3.495 (1.5848-7.7071)	Poor
GSE11969	303	Japan	Upper 50%	0.0444	1.875 (1.0159-3.4598)	Poor
GSE29066	77	Sweden	Upper 50%	0.0228	2.459 (1.1335-5.3331)	Poor
GSE11117	30	Switzerland	Upper 50%	0.0204	13.066 (1.4898-114.6017)	Poor
GSE30219	178	France	Upper 50%	0.0012	2.962 (1.5358-5.7121)	Poor
GSE26939	158	USA	Upper 50%	0.0076	1.968 (1.1972-3.2359)	Poor

\*Corresponding authors' country.

OS. Subgroup analysis results showed a more pronounced correlation between the high expression of *KRT19* and poor OS in LUAD patients (Fig. S1 and Table S1).

### 3.2 NSCLC patients in Kaifeng cohort benefited from chemotherapy alone or the combination of chemotherapy and targeted therapy

Patients' survival extensions by their respective treatments were analyzed to determine the treatment benefits. Kaplan-Meier survival plots demonstrated that chemotherapy alone (Fig. S2A) or the combination of chemotherapy and targeted therapy (Fig. S2B) resulted in significantly extended OS in NSCLC patients. When NSCLC patients were subdivided into stage II-III and stage IV according to TNM staging system, albeit the sample size was reduced to small, the treatments resulted in significantly improved survival in stage II-III patients (Fig. S3). In contrast, stage IV patients did not benefit from the treatments (Fig. S4).

To study the responses of NSCLC patients in different subtypes to the treatments of chemotherapy and targeted therapy, we performed KM survival analysis in LUAD and LUSC patients respectively. The results showed that monotherapy or combination of chemotherapy and targeted therapy significantly improved the survival in LUSC patients (Figs. S5 and S6), but not in LUAD patients (Fig. S7).

### 3.3 High CYFRA21-1 abundance was significantly correlated with low survival rate in NSCLC patients

To explore the prognostic value of fragmented protein form of *KRT19* (CYFRA21-1) in NSCLC, we examined its prognostic role in the independent Kaifeng NSCLC cohort of 232 samples, and found a strong correlation between CYFRA21-1 level and patient outcomes. To assist the use of CYFRA21-1 as a prognosis biomarker, we used X-tile software to calculate the optimal cut-off value, and found that serous CYFRA21-1 level at 31.1ng/ml is most suitable for NSCLC or LUAD, whereas serous CYFRA21-1 level at 5.5ng/ml is optimal for LUSC prognosis evaluation (Fig. S8). KM curves showed that higher serous CYFRA21-1 levels were associated with significantly

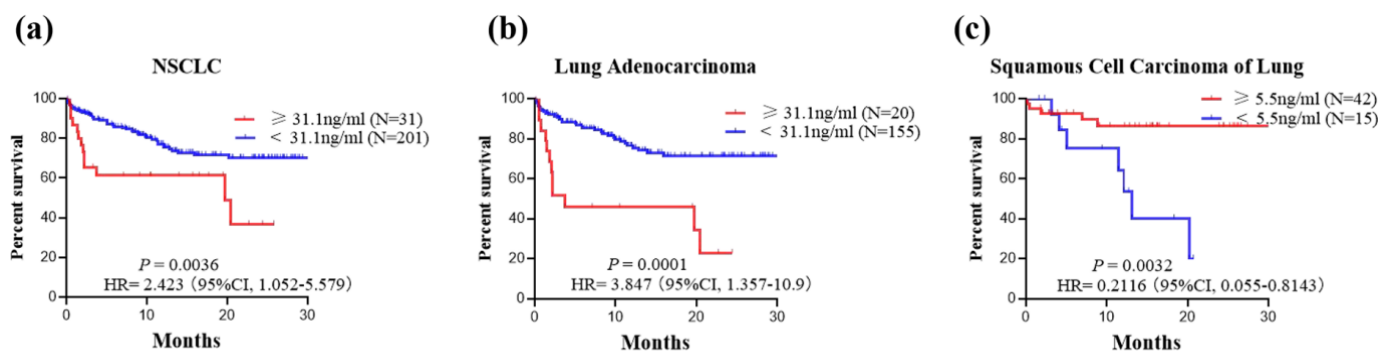
poorer OS in NSCLC patients (Fig. 3a) and in LUAD patients (Fig. 3b). Interestingly, the higher serous CYFRA21-1 levels were associated with better outcomes in LUSC patients (Fig. 3c). To investigate the possibility that treatment in LUSC patients with higher CYFRA21-1 contributed to the better outcome, we compared OS between LUSC patients with low (< 5.5ng/ml) and high ( $\geq$  5.5ng/ml) CYFRA21-1 expressions in treatment or not treatment subgroups, respectively (Fig. S9). In LUSC patients without any treatment, high CYFRA21-1 expression group had a longer OS than low CYFRA21-1 expression group, although the difference was not statistically significant (Fig. S9). The finding was similar among LUSC patients who received treatment (Fig. S9).

### 3.4 CYFRA21-1 could predict treatment outcomes for lung cancer patients

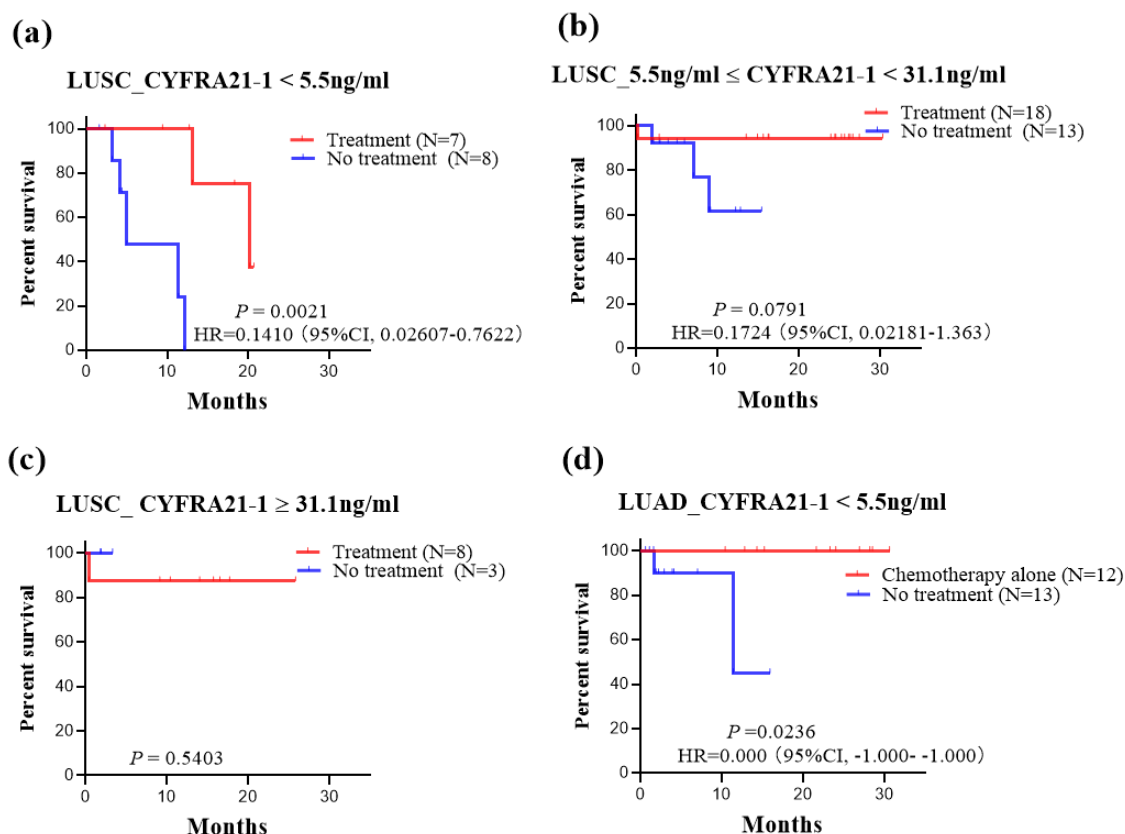
To determine whether serous CYFRA21-1 level could predict treatment outcomes for lung cancer patients, we divided LUSC patients into three subgroups based on CYFRA21-1 concentrations: < 5.5 ng/ml, 5.5 ng/ml-31.1 ng/ml, and  $\geq$  31.1 ng/ml. As shown in Fig. 4a, LUSC patients with serous CYFRA21-1 level at < 5.5 ng/mL were sensitive to treatment (monotherapy or multiple therapies) and had longer OS. LUSC patients with serous CYFRA21-1 level at 5.5 ng/ml-31.1 ng/ml also benefited from treatment and showed a trend of longer OS, although the treatment effect did not reach statistical significance (Fig. 4b). Remarkably, albeit the small sample size, LUSC patients with serous CYFRA21-1 level at  $\geq$  31.1 ng/ml seemed to have better prognosis, regardless of treatment or no treatment (Fig. 4c). As a comparison, chemotherapy could significantly prolong the survival of LUAD patients with serous CYFRA21-1 < 5.5 ng/mL, but not with CYFRA21-1  $\geq$  5.5 ng/mL (Fig. 4d and Figs. S10-S14).

### 3.5 CYFRA21-1 as an independent prognostic indicator can guide NSCLC therapy

By univariate analysis, age, metastasis, or CYFRA21-1 was found to be significantly correlated with poor OS in NSCLC patients, respectively ( $P = 0.029$ ,  $P = 0.044$  and  $P = 0.003$ )



**Fig. 3:** The serum CYFRA21-1 could predict the OS outcomes of NSCLC or LUAD. (a) Increased serous CYFRA21-1 was associated with poor OS in NSCLC patients. (b) Increased serous CYFRA21-1 was related with poor OS in LUAD patients. (c) Increased serous CYFRA21-1 was related with longer OS in LUSC patients. (The optimal cut off value of CYFRA21-1 concentration for prognosis was calculated by X-tile. NSCLC cut off: 31.1ng/ml; LUAD cut off: 31.1ng/ml; LUSC cut off: 5.5ng/ml).



**Fig. 4:** Measurement of the treatment response and its influence on the overall survival of LUSC and LUAD patients. (a) The relationship between treatment and OS in LUSC patients with serum CYFRA21-1 expression less than 5.5 ng/ml. (b) The relationship between OS and treatment in LUSC patients with serum CYFRA21-1 expression between 5.5 and 31.1 ng/ml. (c) The relationship between OS and treatment in LUSC patients with serum CYFRA21-1 expression equal to or greater than 31.1 ng/ml. (d) The relationship between treatment and OS in LUAD patients with serum CYFRA21-1 expression less than 5.5 ng/ml.

(Fig. 5a), whereas gender and histologic type did not show a definitive correlation with OS. Results by multivariate analysis suggested that high serum CYFRA21-1 level could serve as an independent prognostic biomarker for poor OS in NSCLC patients (Fig. 5b).

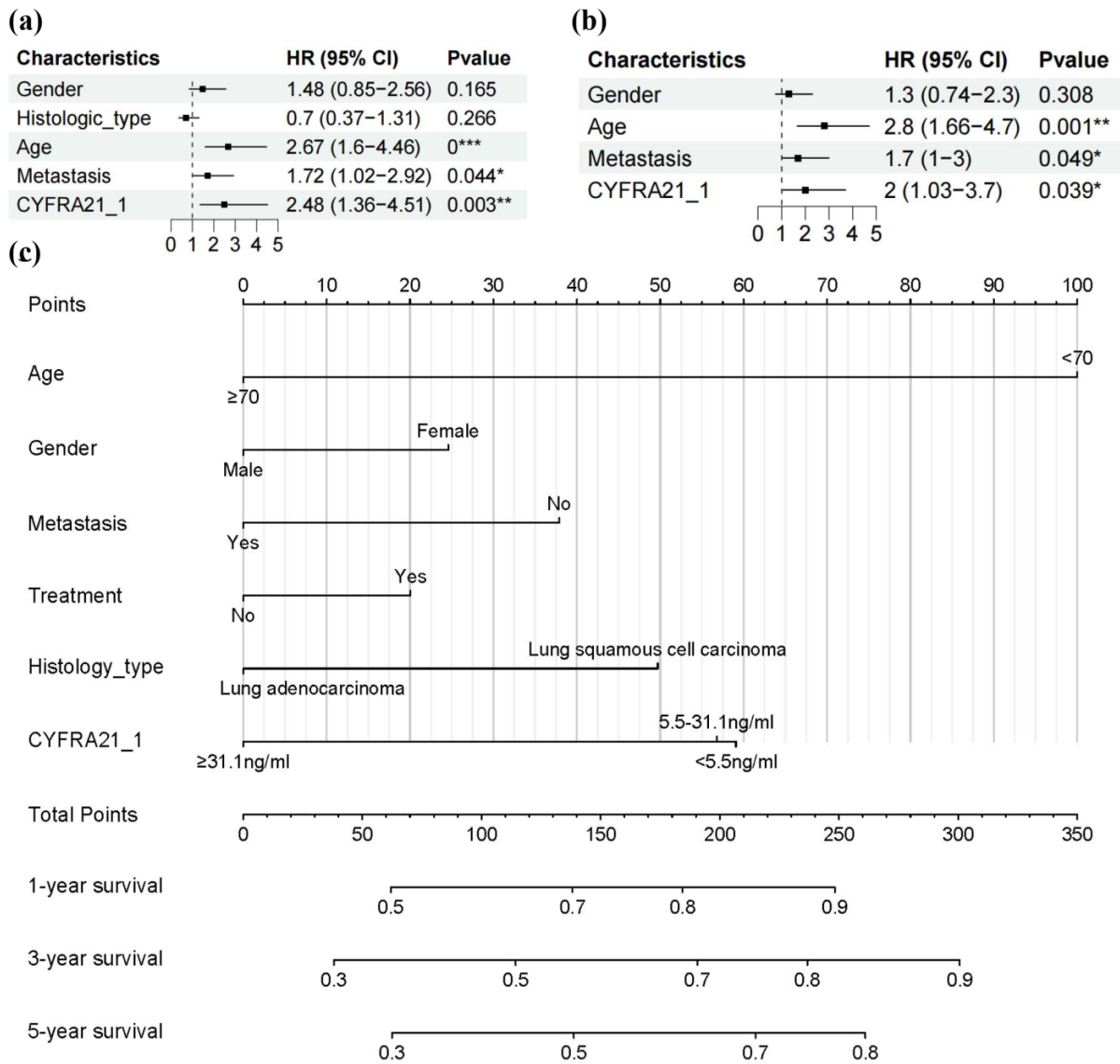
Nomogram has been widely used as a powerful predictive tool for cancer prognosis. To build a model to predict the OS outcome and to guide the clinical intervention for NSCLC patients, a prognostic nomogram was constructed using various clinical factors (*i.e.*, gender, age, metastasis, treatment, and histological type) and CYFRA21-1 based on Cox regression analysis results. Each factor in the nomogram was assigned a weighted number of points. For each individual NSCLC patient, summing up points of all risk factors in the upper panel of nomogram returns a total point that predicts the probability of 1-, 3- and 5- year OS (Fig. 5c). For internal validation purpose, we conducted the bootstrapped calibration plot of nomogram to predict 1-year OS, which worked well with this model (Fig. S15).

To facilitate clinical use of this nomogram, we have implemented an online tool named Predictor of Survival And Treatment of Lung Cancer Patient (PSATluca, <https://bioinfo.henu.edu.cn/PSATluca/index.jsp>). PSATluca contains two functional modules: “Predictor of survival” and

“Predictor of treatment”. In module “Predictor of survival”, users can directly obtain an individualized estimate of overall survival by inputting histological subtype, gender, age, metastasis, treatment and CYFRA21-1 concentration (Fig. 6a). In module “Predictor of treatment”, users can obtain treatment recommendation by entering histological subtype and CYFRA21-1 concentration (Fig. 6b). This tool is easy to use and could be very useful for patients and healthcare providers to predict prognosis and to select appropriate treatment options.

### 3.6 The relationship between KRT19 mRNA expression and diverse infiltrated immune cell types

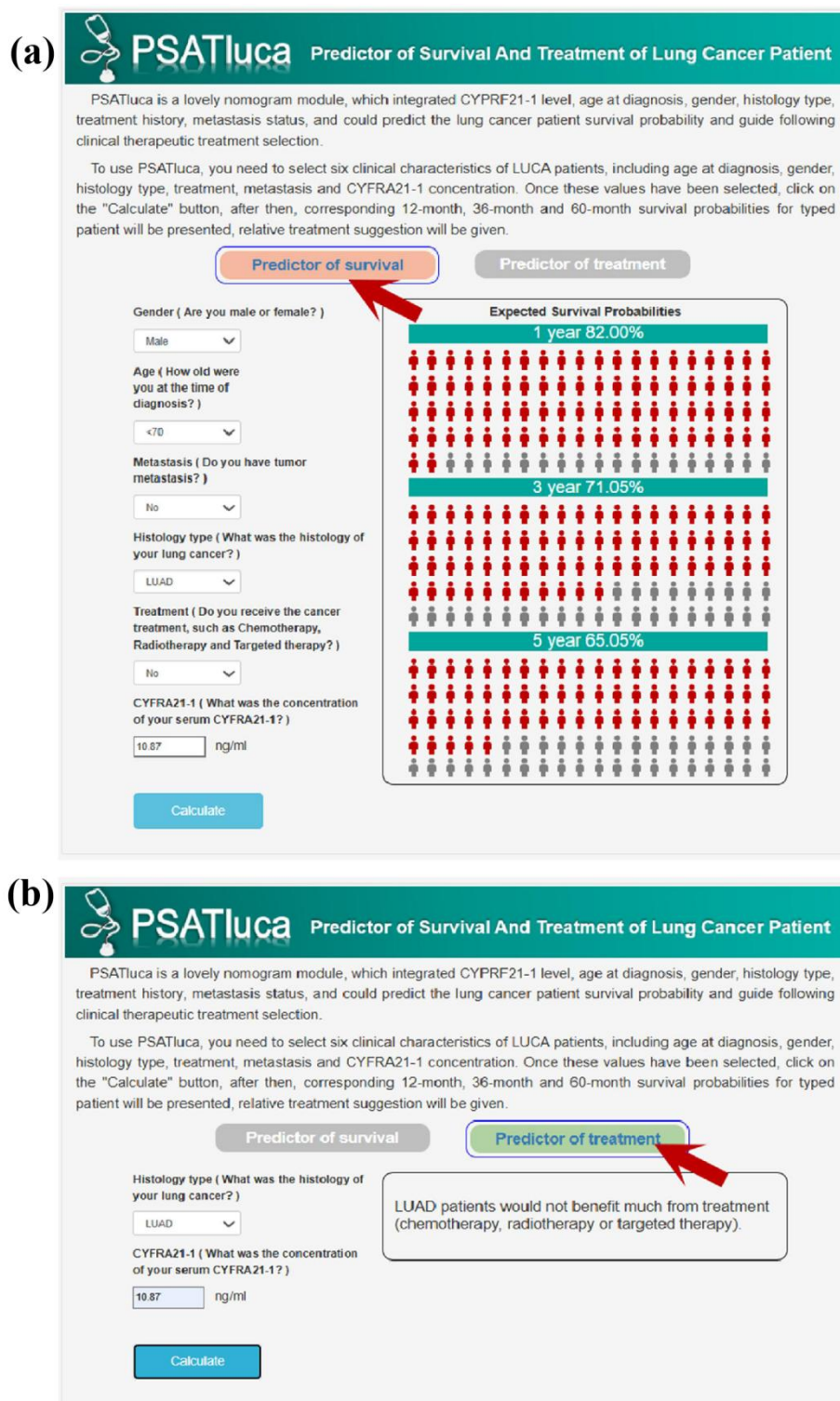
Given the critical role of tumor microenvironment in tumor initiation and progression, we explored the association between *KRT19* mRNA expression and diverse infiltrated immune cell types in LUAD and LUSC from TCGA cohort using CIBERSORT, xCell, and EPIC. The results showed that *KRT19* mRNA expression was positively related to M2 macrophage infiltration, while negatively correlated with T cells, B cells and Plasma cells in LUAD (Fig. 7a). In LUSC, *KRT19* mRNA expression was found to be positively related with activated dendritic cells, while negatively correlated with macrophage cells (Fig. 7b).



**Fig. 5:** Prognostic factors and nomogram for NSCLC patients. Forest plot of univariate (a) and multivariate Cox analysis results (b) for OS in NSCLC patients; (c) Nomogram constructed for NSCLC. Nomogram can be interpreted by adding up the points assigned to each variable in the left, as indicated at the top of the point scale. The total point projected on the bottom scale represents the probability of 1-, 3- and 5-year OS.

Classical prognostic factors such as tumor stage have been widely applied in NSCLC.<sup>[13]</sup> However, NSCLC patients at the same TNM stage might have different treatment responses and resulting outcomes. Therefore, there is an urgent need for prognostic tools that can reliably and objectively predict cancer patients' survival. Integrating molecular biomarkers into tumor staging may help satisfy that need. To now, a wide variety of serous TDBs have been used as diagnostic biomarkers, however, lack of adequate systematic studies about their prognostic values limited their use in cancer prognosis. In 2003, researchers at Memorial Sloan Kettering Cancer Center (MSK) in the United States developed an online tool based on nomogram to predict the prognosis of several common cancers (*e.g.*, breast cancer, bladder cancer, liver cancer, unfortunately without lung cancer) to assist patients and clinicians in making critical treatment decisions

(<https://www.mskcc.org/nomograms>).<sup>[14-16]</sup> However, the prognostic tools developed by MSK were constructed with clinical indicators (*e.g.*, age, sex, TNM stage) and lack molecular indicators. In 2021, based on multidimensional clinical data and 10-year survival follow-up from thousands of esophageal squamous cell carcinoma (ESCC) patients who underwent radical surgery in high-incidence regions of northern China and non-high-incidence regions of southern China, Yang *et al.* developed a postoperative survival prediction tool for ESCC (<https://fxpg.bjcancer.org/ecms/ptools/ec.html>),<sup>[17]</sup> which can be used to assess the survival probability of ESCC patients who underwent esophagectomy as initial treatment, and to evaluate the potential benefits of different postoperative therapies. As we know, tumor has high cellular morphological and molecular heterogeneity, the diagnosis and treatment



**Fig. 6:** Case studies showing the usage of PSATluca online prediction system. The above output webpage presented the prognosis estimation (a) and treatment suggestion (b) for a LUAD patient who is male, less than 70 years old, has no metastasis, didn't receive treatment, and has a 10.87 ng/ml concentration of CYFRA21-1.

techniques of tumors have been advancing, efficacy of therapy has constantly improving, thus the prognosis module needs periodic update accordingly. Unfortunately, only very few nomogram-based prognostic tools of tumors are available to now, there is still lack of molecular risk prediction system or product for NSCLC. Therefore, developing a blood

biomarker-based prognostic tool for Chinese lung cancer patients is highly necessary.

In this study, we first employed in-house built LOGpc system to evaluate the prognostic potentials of all the serum biomarkers used for cancer diagnosis in our local hospital. After discovering the strong correlation of high *KRT19* mRNA

expression with poor survival in multiple NSCLC cohorts, we then used retrospective patients' data from the local hospital to validate that serous level of CYFRA21-1, a protein fragment of KRT19, is an independent prognostic biomarker for NSCLC patients. Finally, we constructed a prognostic nomogram model PSATluca by integrating serous CYFRA21-1 level and various clinical characteristics. PSATluca could allow healthcare providers to quantitatively estimate individual OS probability of NSCLC patients and assist therapeutic decision-making in clinical practice and clinical trials.

*KRT19* (Cytokeratin 19, or *CK19*), a member of the keratin family, is a key molecule in building intermediate fibers in epithelial cells and maintaining the integrity of epithelial cells. *KRT19* also plays key roles in cell cycle, stress response and apoptosis.<sup>[18-20]</sup> Upon tumor necrosis, *KRT19* is degraded and released into the blood, resulting in increased serous protein levels.<sup>[21,22]</sup> Multiple studies have shown that *KRT19* is significantly differentially expressed between normal and cancer tissues, and its expression is often associated with tumor progression. For example, Mohammadi *et al.* found that *KRT19* is a sensitive biomarker which could improve the diagnosis accuracy of early stage colorectal cancer.<sup>[23]</sup> Khair *et al.* confirmed that the expression of *KRT19* and *MUC1* mRNA in peripheral blood of colorectal cancer patients was correlated with Dukes' stage, distant metastasis and differentiation degree.<sup>[24]</sup> Shoichi *et al.* have shown that *KRT19* expression is significantly elevated in moderately and poorly differentiated oral squamous cell carcinoma, and high *KRT19* level is associated with poor prognosis.<sup>[25]</sup> Yang *et al.* demonstrated that compared to the negative group, OS in CK19-positive HCC patients was predicted poor prognosis, and they also developed a five-factor risk (CK19, CA125, Edmondson, BMI,

and tumor number) scoring model with good predictive performance.<sup>[26]</sup> *KRT19* is also involved in migration, invasion, and metastasis of oral squamous cell carcinoma, and may be used as a novel biomarker for metastasis risk of highly aggressive oral squamous cell carcinoma.<sup>[27,28]</sup> Using public databases, Oxford Nanopore Technologies (ONT) sequencing, and bioinformatics methods, Song *et al.* demonstrated that *KRT19* together with ER $\alpha$  formed the ESR1 (ER $\alpha$ )/*KRT19* signaling axis to promote the proliferation, migration, and invasion of papillary thyroid carcinoma (PTC).<sup>[29]</sup> Han and colleagues demonstrated that nuclear *KRT19* functions as a transcriptional corepressor by enhancing the deacetylase activity within the RE-1 silencing transcription factor (CoREST) complex, which comprises histone deacetylase 1 and REST corepressor 1, ultimately leading to dedifferentiation in liver cancer.<sup>[30]</sup> Mi *et al.* discovered that the proteasome-dependent degradation of *KRT19* may suppress the migration and invasion of CRC cells.<sup>[31]</sup> Shi *et al.* demonstrated that oncogenic *KRT19* is negatively regulated by miR-642a-5p, and stimulates pancreatic cancer (PC) progression through the Wnt/ $\beta$ -catenin (WBC) pathway.<sup>[32]</sup>

Recently, Yuan *et al.* found that high *KRT19* expression in lung cancer is correlated with tumor progression and poor prognosis in lung cancer patients, especially LUSC patients.<sup>[33,34]</sup> Although *KRT19* expression and biology in lung cancer has been extensively studied, the potential of using serous *KRT19* level to predict prognosis in NSCLC prognosis remains to be established. In this study, by analyzing clinical data of 232 NSCLC patients with Cox regression, we aimed to identify risk factors for survival and to develop a nomogram prognosis model. Serous level of CYFRA21-1, a protein fragment of *KRT19*, was found to be able to predict the treatment benefit in NSLSC patients. Treatment can

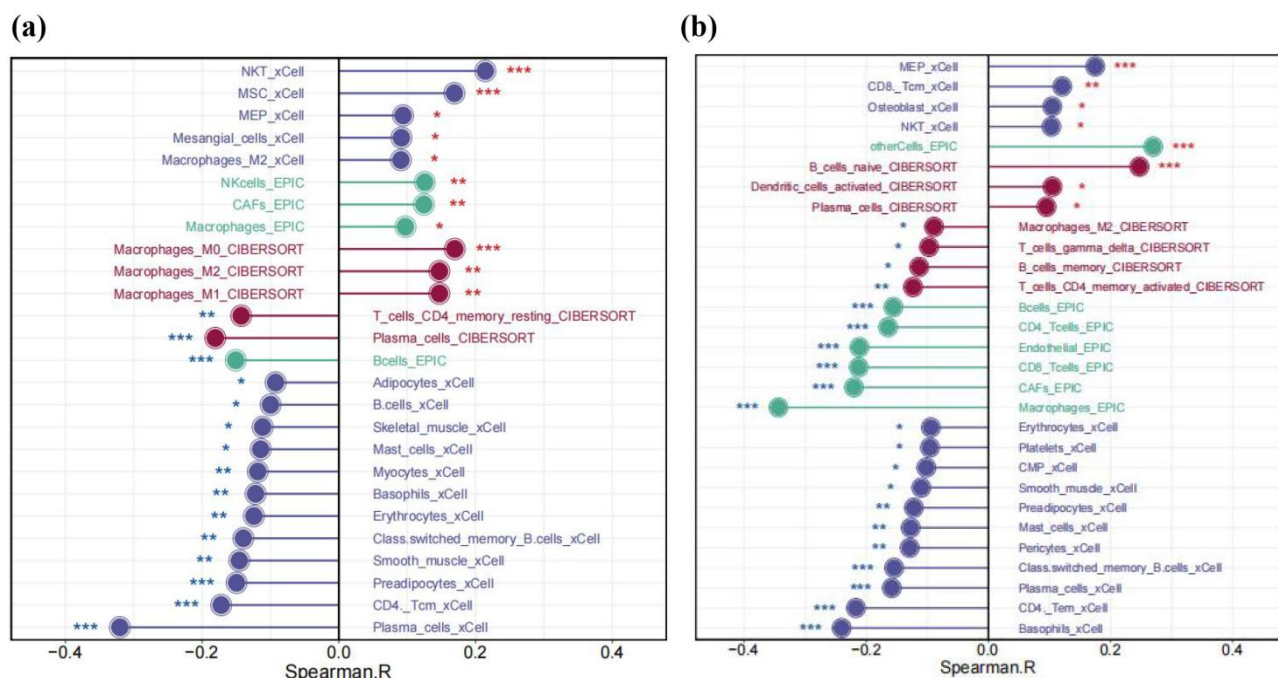


Fig. 7: Correlations between KRT19 mRNA expression and tumor environmental infiltration in LUAD (a) and LUSC (b).

significantly prolong the survival of LUSC and LUAD patients with CYFRA21-1 < 5.5ng/ml regardless of treatment options. Finally, we reaffirmed the clinical significance of serous CYFRA21-1 as an independent prognostic biomarker in NSCLC patients, and developed PSATLuca to predict OS of NSCLC patients.

There are four main advantages in our study. First, we started from a screening of huge population of a couple of hundred tumor cohorts to a retrospective analysis of a local independent validation cohort, aiming to reach a valid conclusion. This is important for translating the findings into future clinical application. Second, we integrated clinical features, serous CYFRA21-1, and therapeutic items into PSATLuca nomogram to predict NSCLC outcomes, which provides a more individualized prognostic model and could be complementary to TNM staging system. Third, all variables included in PSATLuca nomogram are widely used in clinical testing, thus PSATLuca can be easily adopted to assist clinical use. Last but not least, the visual web tool makes PSATLuca easier to use for clinicians and/or patients to predict individualized survival. The limitation of our work is that it is a retrospective analysis. To investigate the potential role of *KRT19* and extend its clinical translation in NSCLC, basic in vitro cell and in vivo animal experiments are needed in future study. By knocking down or rescuing *KRT19* expression in tumor cells, it will help to investigate the molecular roles of *KRT19* by testing cellular functions or mouse bearing tumor growing. The additional small molecule and bioengineering therapy developments targeting *KRT19* are also very meaningful for lung cancer treatment. These treatment methods, combined with prognostic tools such as PSATLuca, have promising prospects for precision treatment of lung cancer in the future. Nevertheless, our study yielded important findings and could also assist the design of some prospective studies.

#### 4. Conclusion

In this work, our comprehensive study showed that serous CYFRA21-1 level is a reliable independent prognostic biomarker for NSCLC. In addition, we established a nomogram named PSATLuca to assess the survival probability for individual NSCLC patients, to identify high-risk NSCLC patients, and to support the appropriate treatment decision-making.

#### Acknowledgments

This work was supported by Key Scientific Research Project of Henan Province (22A310012), Program from Academy for Advanced Interdisciplinary Studies of Henan University (CX3070A0970005), Henan Province Science and Technology Research Project (242102311172), Henan Province Medical Science and Technology Aattack Project (SBCJ202402078), Postdoctoral Research Project of Henan Province (HN2025169), Program for Science & Technology

Innovation Talents in Universities of Henan Province (25HASTIT054).

#### Conflict of Interest

There is no conflict of interest.

#### Supporting Information

Applicable.

#### CRedit Statement

**Xiangqian Guo:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition. **Shengnan Wu, Longxiang Xie:** Writing – Original draft, Methodology, Investigation, Formal analysis, Data curation. **Shengnan Wu, Longxiang Xie, Qiang Wang:** Writing – review & editing, Supervision, Software, Funding acquisition. **Yuxuan He:** Supervision

#### References

- [1] C. Xia, X. Dong, H. Li, M. Cao, D. Sun, S. He, F. Yang, X. Yan, S. Zhang, N. Li, W. Chen, Cancer statistics in China and United States, 2022: profiles, trends, and determinants, *Chinese Medical Journal*, 2022, **135**, 584-590, doi: 10.1097/CM9.0000000000002108.
- [2] Q. Li, Y. Mao, R. Guo, C. Huang, X. Fang, J. Ma, L. Tang, L. Chen, A nomogram based on serum biomarkers and clinical characteristics to predict survival in patients with non-metastatic nasopharyngeal carcinoma, *Frontiers in Oncology*, 2020, **10**, 594363, doi: 10.3389/fonc.2020.594363.
- [3] G. Luo, K. Jin, S. Deng, H. Cheng, Z. Fan, Y. Gong, Y. Qian, Q. Huang, Q. Ni, C. Liu, X. Yu, Roles of CA19-9 in pancreatic cancer: biomarker, predictor and promoter, *Biochimica et Biophysica Acta. Reviews on Cancer*, 2021, **1875**, 188409, doi: 10.1016/j.bbcan.2020.188409.
- [4] T. M. Bauer, B. F. El-Rayes, X. Li, N. Hammad, P. A. Philip, A. F. Shields, M. M. Zalupski, T. Bekaii-Saab, Carbohydrate antigen 19-9 is a prognostic and predictive biomarker in patients with advanced pancreatic cancer who receive gemcitabine-containing chemotherapy, *Cancer*, 2013, **119**, 285-292, doi: 10.1002/ncr.27734.
- [5] X. Yang, Y. Li, C. Chen, C. Peng, S. Liu, Y. Liu, Preoperative serum carbohydrate antigen 125 level is an independent negative prognostic marker for overall survival in colorectal cancer, *Medical Oncology*, 2011, **28**, 789-795, doi: 10.1007/s12032-010-9518-z.
- [6] X. Liu, L. Wang, X. Yin, The diagnostic value of serum CEA, NSE and MMP-9 for on-small cell lung cancer, *Open Medicine*, 2016, **11**, 59-62, doi: 10.1515/med-2016-0012.
- [7] Q. Li, S. Sang, Diagnostic value and clinical significance of combined detection of serum markers CYFRA21-1, SCC Ag, NSE, CEA and ProGRP in non-small cell lung carcinoma, *Clinical Laboratory*, 2020, **66**, 2189, doi: 10.7754/Clin.Lab.2020.191243.

- [8] L. Xie, Q. Wang, X. Ma, G. Zhao, Z. Yan, H. Li, W. Zhu, Z. Zuo, X. Guo, OSTme: Online consensus survival analysis web server for tumor microenvironment components of pan-cancers, *Genes & Diseases*, 2023, **10**, 2256-2259, doi: 10.1016/j.gendis.2023.02.043.
- [9] D. Zeng, Z. Ye, R. Shen, G. Yu, J. Wu, Y. Xiong, R. Zhou, W. Qiu, N. Huang, L. Sun, X. Li, J. Bin, Y. Liao, M. Shi, W. Liao, IOBR: multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures, *Frontiers in Immunology*, 2021, **12**, 687975, doi: 10.3389/fimmu.2021.687975.
- [10] B. Chen, M. S. Khodadoust, C. L. Liu, A. M. Newman, A. A. Alizadeh, Profiling tumor infiltrating immune cells with CIBERSORT, *Cancer Systems Biology*, New York, NY: Springer New York, 2018, 243-259, doi: 10.1007/978-1-4939-7493-1\_12.
- [11] J. Racle, K. de Jonge, P. Baumgaertner, D. E. Speiser, D. Gfeller, Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data, *eLife*, 2017, **6**, e26476, doi: 10.7554/elife.26476.
- [12] D. Aran, Z. Hu, A. J. Butte, xCell: digitally portraying the tissue cellular heterogeneity landscape, *Genome Biology*, 2017, **18**, 220, doi: 10.1186/s13059-017-1349-1.
- [13] C. Bortolotto, A. Lancia, C. Stelitano, M. Montesano, E. Merizzoli, F. Agustoni, G. Stella, L. Preda, A. R. Filippi, Radiomics features as predictive and prognostic biomarkers in NSCLC, *Expert Review of Anticancer Therapy*, 2021, **21**, 257-266, doi: 10.1080/14737140.2021.1852935.
- [14] J. L. B. Bevilacqua, M. W. Kattan, J. V. Fey, H. S. Cody III, P. I. Borgen, K. J. Van Zee, Doctor, what are my chances of having a positive sentinel node? a validated nomogram for risk estimation, *Journal of Clinical Oncology*, 2007, **25**, 3670-3679, doi: 10.1200/jco.2006.08.8013.
- [15] C. S. Sima, W. R. Jarnagin, Y. Fong, E. Elkin, M. Fischer, D. Wuest, M. D'Angelica, R. P. DeMatteo, L. H. Blumgart, M. Gönen, Predicting the risk of perioperative transfusion for patients undergoing elective hepatectomy, *Annals of Surgery*, 2009, **250**, 914-921, doi: 10.1097/sla.0b013e3181b7fad3.
- [16] Bochner BH, Kattan MW, Vora KC, Postoperative nomogram predicting risk of recurrence after radical cystectomy for bladder cancer, *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 2006, **24**, 3967-3972, doi: 10.1200/JCO.2005.05.3884.
- [17] W. Yang, F. Liu, R. Xu, W. Yang, Y. He, Z. Liu, F. Zhou, F. Heng, B. Hou, L. Zhang, L. Chen, F. Zhang, F. Cai, H. Xu, M. Lin, M. Liu, Y. Pan, Y. Liu, Z. Hu, H. Chen, Z. He, Y. Ke, Is adjuvant therapy a better option for esophageal squamous cell carcinoma patients treated with esophagectomy? a prognosis prediction model based on multicenter real-world data, *Annals of Surgery*, 2023, **277**, e61-e69, doi: 10.1097/sla.0000000000004958.
- [18] H. Yu, X. Huang, Z. Zhu, Y. Hu, W. Ou, L. Zhang, N. Zhou, Significance of combined detection of LunX mRNA and tumor markers in diagnosis of lung carcinoma, *Chung-Kuo Yen Cheng Yen Chiu*, 2014, **26**, 89-94, doi: 10.3978/j.issn.1000-9604.2014.02.04.
- [19] H. Jamaati, A. Khosravi, A. Abedini, A. Kiani, P. Tabarsi, H. Dargahi, N. Bahrami, M. Kazempour Dizaji, F. Sadegh Beigee, M. Pourabdollah, A. Gheysouri, N. Ahmadian, M. Abniki, A. Mohamadnia, A. Velayati, Three markers in cancerous and healthy cells of patients with non-small-cell lung carcinoma (NSCLC), *Asian Pacific Journal of Cancer Prevention*, 2019, **20**, 2281-2285, doi: 10.31557/APJCP.2019.20.8.2281.
- [20] S. K. Saha, K. Kim, G. M. Yang, H. Y. Choi, S. G. Cho, Cytokeratin 19 (*KRT19*) has a role in the reprogramming of cancer stem cell-like cells to less aggressive and more drug-sensitive cells, *International Journal of Molecular Sciences*, 2018, **19**, 1423, doi: 10.3390/ijms19051423.
- [21] K. Dohmoto, S. Hojo, J. Fujita, Y. Yang, Y. Ueda, S. Bandoh, Y. Yamaji, Y. Ohtsuki, N. Dobashi, T. Ishida, J. Takahara, The role of caspase 3 in producing cytokeratin 19 fragment (CYFRA21-1) in human lung cancer cell lines, *International Journal of Cancer*, 2001, **91**, 468-473, doi: 10.1002/1097-0215(200002)9999:9999::aid-ijc1082>3.0.co;2-t.
- [22] O. Nordgård, G. Singh, S. Solberg, L. Jørgensen, A. R. Halvorsen, R. Smaaland, O. T. Brustugun, Å. Helland, Novel molecular tumor cell markers in regional lymph nodes and blood samples from patients undergoing surgery for non-small cell lung cancer, *PLoS One*, 2013, **8**, e62153, doi: 10.1371/journal.pone.0062153.
- [23] P. Mohammadi, M. Saidijam, A. Kaki, K. Etemadi, N. Shabab, R. Yadegarazari, A pilot study of CK19, CK20 and GCC mRNA in the peripheral blood as a colorectal cancer biomarker panel, *International Journal of Molecular and Cellular Medicine*, 2016, **5**, 30-36.
- [24] G. Khair, J. R. T. Monson, J. Greenman, Epithelial molecular markers in the peripheral blood of patients with colorectal cancer, *Diseases of the Colon and Rectum*, 2007, **50**, 1188-1203, doi: 10.1007/s10350-006-0875-9.
- [25] A. S. Rathore, S. Katyal, A. Jain, D. C. Shetty, Biochemical analysis of cytokeratin fragment 21-1 concentration and expression of cytokeratin 19 in oral potentially malignant disorders, *Journal of Cancer Research and Therapeutics*, 2020, **16**, 452-457, doi: 10.4103/jrcrt.JCRT\_893\_17.
- [26] C. Yang, W. Xiang, Z. Wu, N. Li, G. Xie, J. Huang, L. Zeng, H. Yu, B. Xiang, CK19 protein expression: the best cutoff value on the prognosis and the prognosis model of hepatocellular carcinoma, *BMC Cancer*, 2025, **25**, 55, doi: 10.1186/s12885-024-13399-9.
- [27] S. Tanaka, S. Kawano, T. Hattori, R. Matsubara, T. Sakamoto, Y. Hashiguchi, N. Kaneko, Y. Mikami, M. Morioka, Y. Maruse, R. Kitamura, E. Hamada, M. Hiwatashi, K. Oobu, T. Kiyoshima, S. Nakamura, Cytokeratin 19 as a biomarker of highly invasive oral squamous cell carcinoma with metastatic potential, *Journal of Oral and Maxillofacial Surgery, Medicine, and Pathology*, 2020, **32**, 1-7, doi: 10.1016/j.ajoms.2019.10.007.
- [28] Z. Zhang, F. Bi, Z. Zhang, W. Tian, W. Guo, Identification of potential biomarkers and available drugs for oral squamous cell carcinoma, *Translational Cancer Research*, 2021, **10**, 141-151, doi: 10.21037/tcr-20-2500.
- [29] Z. M. Song, Y. D. Wang, F. Chai, J. Zhang, S. Lv, J. X. Wang,

Y. Xi, Estrogen enhances the proliferation, migration, and invasion of papillary thyroid carcinoma *via* the ER $\alpha$ /KRT19 signaling axis, *Journal of Endocrinological Investigation*, 2025, **48**, 653-670, doi: 10.1007/s40618-024-02473-5.

[30] S. Han, H. Fan, G. Zhong, L. Ni, W. Shi, Y. Fang, C. Wang, L. Wang, L. Song, J. Zhao, M. Tang, B. Yang, L. Li, X. Bai, Q. Zhang, T. Liang, Y. Xu, X. Feng, C. Ding, D. Fang, B. Zhao, Nuclear KRT19 is a transcriptional corepressor promoting histone deacetylation and liver tumorigenesis, *Hepatology*, 2025, **81**, 808-822, doi: 10.1097/hep.0000000000000875.

[31] S. Mi, J. Hu, W. Chen, J. Chen, Z. Xu, M. Xue, m1A-regulated DIAPH<sub>3</sub> promotes the invasiveness of colorectal cancer *via* stabilization of KRT19, *Clinical & Experimental Metastasis*, 2025, **42**, 10, doi: 10.1007/s10585-024-10323-0.

[32] H. Shi, X. Li, Z. Chen, S. Dong, C. Ye, S. Hou, D. Fan, H. Zhang, W.-C. Zhou, KRT19 is regulated by miR-642a-5p and promotes pancreatic cancer progression through the Wnt/ $\beta$ -catenin pathway, *iScience*, 2024, **27**, 110782, doi: 10.1016/j.isci.2024.110782.

[33] W. Wang, J. He, H. Lu, Q. Kong, S. Lin, KRT8 and KRT19, associated with EMT, are hypomethylated and overexpressed in lung adenocarcinoma and link to unfavorable prognosis, *Bioscience Reports*, 2020, **40**, BSR20193468, doi: 10.1042/BSR20193468.

[34] X. Yuan, M. Yi, B. Dong, Q. Chu, K. Wu, Prognostic significance of KRT19 in lung squamous cancer, *Journal of Cancer*, 2021, **12**, 1240-1248, doi: 10.7150/jca.51179.

**Publisher's Note:** Engineered Science Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits the use, sharing, adaptation, distribution and reproduction in any medium or format, as long as appropriate credit to the original author(s) and the source is given by providing a link to the Creative Commons license and changes need to be indicated if there are any. The images or other third-party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

©The Author(s) 2025