



OPEN ACCESS

Edited by:

Ye Wang,
Qingdao University Medical College,
China

Reviewed by:

Ying Wang,
Eastern Hepatobiliary Surgery
Hospital, China
Belamy B. Cheung,
Children's Cancer Institute, Australia

***Correspondence:**

Chenwei Li
NBLCW2008@126.com
Xinjian Li
Dxjs1961@sina.com
Wentao Hu
34836674@qq.com

Specialty section:

This article was submitted to
Cancer Genetics,
a section of the journal
Frontiers in Oncology

Received: 17 June 2021

Accepted: 02 September 2021

Published: 21 October 2021

Citation:

Shi J, Wang Z, Guo J, Chen Y, Tong C,
Tong J, Hu W, Li C and Li X (2021)
Identification of a Three-Gene
Signature Based on Epithelial-
Mesenchymal Transition of Lung
Adenocarcinoma Through
Construction and Validation of a
Risk-Prediction Model.
Front. Oncol. 11:726834.
doi: 10.3389/fonc.2021.726834

Identification of a Three-Gene Signature Based on Epithelial-Mesenchymal Transition of Lung Adenocarcinoma Through Construction and Validation of a Risk-Prediction Model

Jianguang Shi, Zishan Wang, Jing Guo, Yingqi Chen, Changyong Tong, Jingjie Tong, Wentao Hu*, Chenwei Li* and Xinjian Li*

Thoracic Surgery Department, Ningbo First Hospital, Ningbo, China

Epithelial-mesenchymal transition (EMT) process, which is regulated by genes of inducible factors and transcription factor family of signaling pathways, transforms epithelial cells into mesenchymal cells and is involved in tumor invasion and progression and increases tumor tolerance to clinical interventions. This study constructed a multigene marker for lung predicting the prognosis of lung adenocarcinoma (LUAD) patients by bioinformatic analysis based on EMT-related genes. Gene sets associated with EMT were downloaded from the EMT-gene database, and RNA-seq of LUAD and clinical information of patients were downloaded from the TCGA database. Differentially expressed genes were screened by difference analysis. Survival analysis was performed to identify genes associated with LUAD prognosis, and overlapping genes were taken for all the three. Prognosis-related genes were further determined by combining LASSO regression analysis for establishing a prediction signature, and the risk score equation for the prognostic model was established using multifactorial COX regression analysis to construct a survival prognostic model. The model accuracy was evaluated using subject working characteristic curves. According to the median value of risk score, samples were divided into a high-risk group and low-risk group to observe the correlation with the clinicopathological characteristics of patients. Combined with the results of one-way COX regression analysis, HGF, PTX3, and S100P were considered as independent predictors of LUAD prognosis. In lung cancer tissues, HGF and PTX3

expression was downregulated and S100P expression was upregulated. Kaplan-Meier, COX regression analysis showed that HGF, PTX3, and S100P were prognostic independent predictors of LUAD, and high expressions of all the three were all significantly associated with immune cell infiltration. The present study provided potential prognostic predictive biological markers for LUAD patients, and confirmed EMT as a key mechanism in LUAD progression.

Keywords: epithelial-mesenchymal transition, lung adenocarcinoma, prognosis, LASSO, risk score

INTRODUCTION

Lung adenocarcinoma (LUAD) is the most common subtype of non-small-cell lung cancer (NSCLC), with an increasing incidence and poor patient prognosis (1–3). LUAD is highly heterogeneous and aggressive, which is often associated with genetic mutations (4, 5). In recent years, advances in chemotherapy, radiotherapy, and targeted therapies have reduced LUAD mortality (6–8), but its long-term survival is still dismal when compared with other cancers (9). To the best of our knowledge, the molecular pathogenesis of LUAD remains largely unknown. Therefore, there is an urgent need to investigate the mechanisms of LUAD development in-depth and to search potential key prognostic markers.

In terms of primary tumor, spread of cancer cells resulted from metastasis is a main cause of lung cancer-related death (10). During tumor metastasis, local invasion of primary tumor affects only surrounding tissues, especially stromal cells, and shed tumor cells enter the circulatory system through endocytosis and survive in a fluid circulatory system environment. Tumor cells in circulatory system will in turn penetrate the circulatory system and migrate into distant tissues, forming tiny metastatic clones, which will then form visible metastatic foci through proliferation (11, 12). In such a tumor metastasis process, epithelial-mesenchymal transition (EMT) is a highly critical mechanism, through which epithelial cells lose cell-cell junctions and polarity, resulting in loss of epithelial properties and acquisition of mesenchymal properties with invasive and migratory abilities (13). EMT process is often activated during the development of tumorigenesis and growth, invasion, migration, colonization, and therapeutic resistance (14–16). In recent years, many studies have reported EMT as a marker of poor prognosis in LUAD patients (17, 18). Some studies showed the prognostic significance and biological functions of some EMT-related genes in LUAD (19), but such studies are more at an early stage and lack in-depth analysis.

In this study, we first collected EMT-related genes. Candidate genes were obtained through co-analysis with differentially expressed genes and total survival-related genes in LUAD. Then a gene signature prognostic risk prediction model was constructed by LASSO and COX regression analysis and validated. LUAD patients were divided into two groups of high and low risk according to risk scores, and prognostic significance of these gene signatures in LUAD was assessed through analyzing the correlation between different risk groupings and clinicopathological characteristics of patients.

METHODOLOGY

Data Collection

The Tumor Genetic Atlas (TCGA) database and Genotype-Tissue Expression (GTEx) database were downloaded to obtain human cancer gene expression data and clinical data. The information of LUAD samples and other 32 cancer types in TCGA, including clinically relevant pathological parameters and RNA-seq of tumor tissue samples, was assessed according to data completeness of clinical sample and degree of matching with sequenced samples remove duplicate and censored samples as well as cases without clinical outcomes. EMT-related genes were obtained from the Epithelial-Mesenchymal Transition (EMT) gene database (<http://www.dbemt.bioinfo-minzhao.org/>) download. The independent validation cohorts GSE68465 and GSE68571 were downloaded from the Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo>) using the GEO query R package. Tumor name abbreviations and corresponding meanings are given in **Supplementary Table 1**.

Differentially Expressed Genes

Differentially expressed genes in LUAD were screened using the edgeR package in R software with $|\log_{2}FC| \geq 1.5$ and ADJUST P value < 0.05 . Volcano plots of the differentially expressed genes were drawn using ggplot2.

Prognostic Risk Modeling and Analysis

LASSO regression was applied to identify genes and develop a gene signature. The effect of gene signature on prognosis was validated by one-way COX regression. Multifactor COX regression was employed to detect prognostically independent predictors and construct a prognostic risk model. The risk score was calculated according to the model, and the median risk score was the threshold to divide patients into high-risk and low-risk groups. Kaplan-Meier survival curves were drawn to compare the survival of high-risk and low-risk patients. Also, the working curves of subjects were drawn, and the area under the curve was calculated to assess the predictive efficacy of the model.

Survival Analysis

Univariate cox regression analysis was done, and forest plots through the “forestplot” R package were used to display the P value, HR, and 95% CI of each variable. Use R software v4.0.3 for

statistical analysis. If not otherwise stated, the rank sum test detects two sets of data, and a P value of <0.05 is considered statistically significant.

TIMER Analysis

The TIMER database (<http://timer.comp-genomics.org/>) was used to analyze the expression of genes in LUAD correlated with immune cell infiltration.

Protein Expression Validation

Immunohistochemical staining maps of protein expression in both liver cancer tissues and normal tissues were downloaded from the HPA database (The Human Protein Atlas) for validating the gene signature.

Gene-Set Enrichment Analysis

RNA-seq profiles were uploaded to GSEA to investigate key gene-related signaling pathways in the high-risk group and the low-risk group. The enriched sets were screened based on a FDR < 0.25 and P < 0.05 after 1,000 permutations.

Cell Culture

The human bronchial epithelial cell line BEAS-2B and human LUAD cell lines NCI-H2009 were all purchased from Beyotime Biotechnology. All the cell lines were cultured in RPMI Medium 1640 (Gibco, USA) containing 10% fetal bovine serum (FBS) (Gibco, USA) and were incubated in a constant temperature incubator at 37°C and 5% CO₂ for future use.

Western Blot

The ice-cold lysate buffer was added to extract cell lines proteins. After centrifugation, the supernatant was extracted (12,000 rpm, 4°C, 10 min), and the protein concentration was measured using a BCA Protein Assay Kit (Beyotime, China). A total of 20 µg protein was separated using 10% SDS/PAGE, and then transferred to the membrane by wet transfer. The membrane was blocked with 5% skimmed milk powder. The primary antibodies of HGF (Abcam, 1:200), PTX3 (Abcam, 1:1,000), S100P (Abcam, 1:1,000), GAPDH (Abcam, 1:1,000) were added into membrane and then incubated overnight in shaking bed at 4°C. Tris-buffered saline with Tween 20 was used to wash the membrane for three times, 10 min each time. The secondary antibody (Goat Anti-Rabbit IgG, Abcam, 1:2,000) was subsequently added and incubated for 2 h at room temperature. Finally, protein bands were detected using the ECL Western Blot Detection Kit (Beyotime, China), and β-actin was used as the internal reference protein.

RESULTS

Differentially Expressed Gene Screening

Under the conditions of |logFC|>1 and P<0.05 as the screening conditions, a total of 1,811 differentially expressed genes, including 1,200 downregulated genes and 611 upregulated genes, were

obtained from TCGA-LUAD data. The visualization results are shown in the volcano plot (**Figure 1A**). The prognostic characteristics of the top 20 genes with the most significant single-factor COX analysis in LUAD are shown in **Figure 1B**. The Epithelial-Mesenchymal transition (EMT) gene database (<http://www.dbemt.bioinfo-minzhao.org/>) was used to obtain 1,263 EMT-related genes, and the three categories of genes with overlaps were screened to obtain a total of 47 key genes (**Figure 1C**).

Prognostic Prediction Modeling

LASSO regression analysis was conducted to further narrow the range of prognosis-related genes to ensure the result stability (**Figures 2A, B**). The risk score for each sample was calculated using the risk score formula:

$$\begin{aligned} \text{Risk score} = & (0.0822) * KRT8 + (0.0563) * ADM + (0.0099) * ECT2 \\ & + (0.0869) * CCNA2 + (0.0418) * TYMS + (0.0951) * FSTL3 \\ & + (0.0374) * GOLM1 + (0.0083) * EGLN3 + (0.1054) * LGR4 \\ & + (-0.1263) * HGF + (-0.006) * KL + (0.0367) * S100P \\ & + (-0.013) * AGER + (0.0513) * NT5E + (-0.0113) * NDRG2 \\ & + (-0.0022) * GPC3 + (0.0034) * CA9 + (0.0975) * TWIST2 \\ & + (-0.0198) * FHL1 + (0.0829) * PTX3 \end{aligned}$$

The median risk score served as the threshold value to divide the patients into high-risk and low-risk groups. The survival curves were plotted (**Figure 2C**), and the results showed that the survival prognosis of high-risk patients was significantly worse than that of low-risk patients. The ROC curves of this risk score model in predicting the 1-, 3-, 5-year survival of LUAD patients were plotted (**Figure 2D**), and the AUC areas were 0.712, 0.718, and 0.712, respectively, indicating that this risk model had a high prediction accuracy.

Clinicopathological Characteristics of Different Risk Score Subgroups

Details of the distribution of clinicopathological characteristics of different risk score subgroups (high-risk and low-risk) are shown in a heatmap (**Figure 3A**), and the details are shown in **Table 1**. The results of univariate and multifactorial COX regression analyses displayed that risk score, TNM stage, T stage, S100P, HGF, and PTX3 could be used as independent predictors of prognosis of patients with LUAD (**Figures 3B–D**).

Validating the Prediction of the Risk Model in Independent Cohort

We constructed a prognosis predicting risk model based on HGF, PTX3, and S100P base on TCGA datasets (**Figure 4A**). The risk score for each sample was calculated with the risk score formula:

$$\text{Risk score} = (-0.2793) * HGF + (0.0885) * S100P + (0.3172) * PTX3$$

The optimal cutoff risk score served as the threshold to divide the patients into high-risk and low-risk groups. The survival curves were plotted, and the results showed that the survival prognosis of

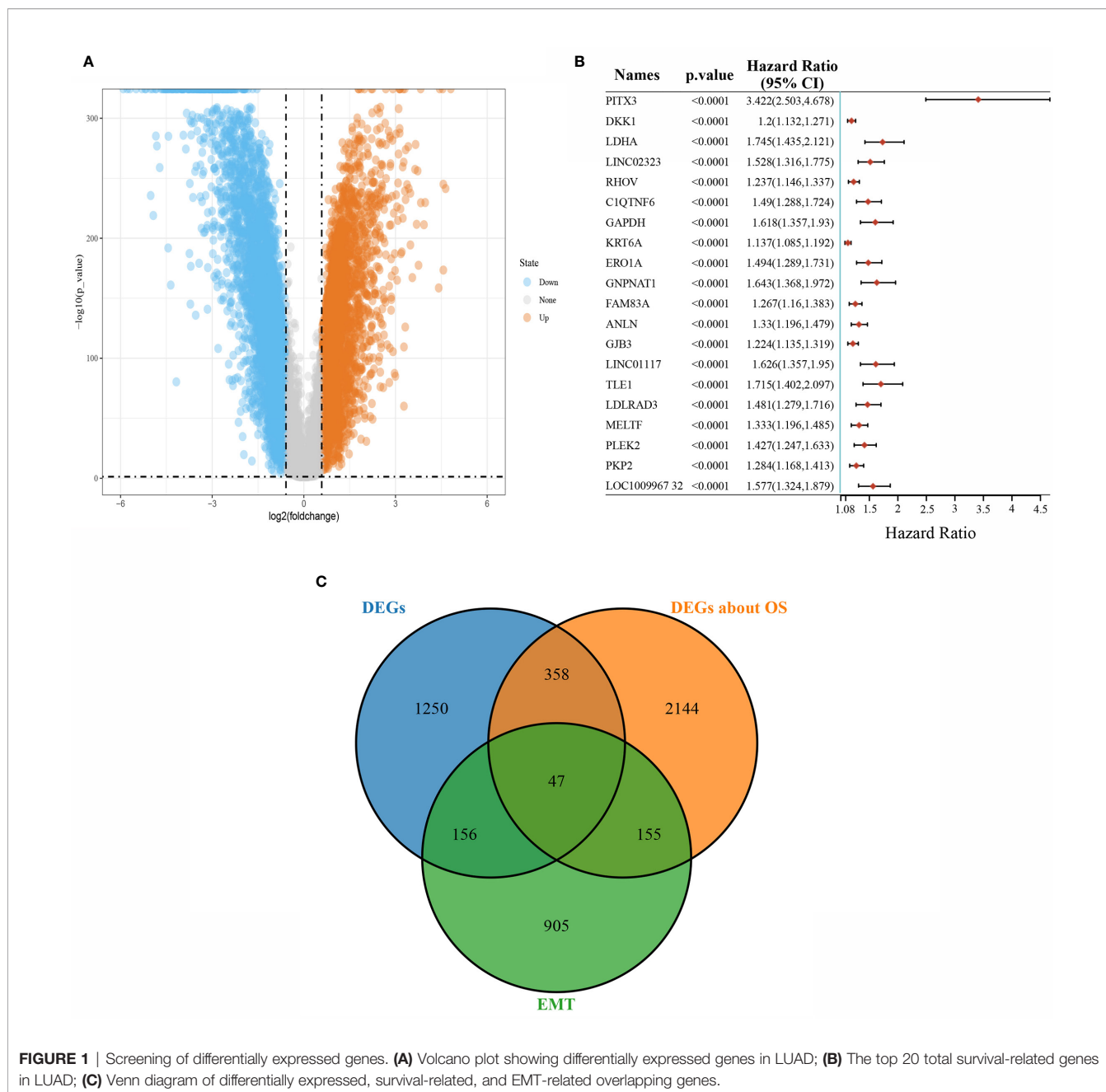


FIGURE 1 | Screening of differentially expressed genes. **(A)** Volcano plot showing differentially expressed genes in LUAD; **(B)** The top 20 total survival-related genes in LUAD; **(C)** Venn diagram of differentially expressed, survival-related, and EMT-related overlapping genes.

high-risk patients was significantly worse than low-risk patients (**Figure 4B**). To further assess the prediction of the risk model, GSE68465 and GSE68571 derived from GEO database were employed as validation cohorts. Cox regression analysis and Kaplan-Meier curve demonstrated that the prognosis of high-risk patients was worse than low-risk patients, which was consistent with the results found in the TCGA-LUAD cohort (**Figures 4C, D**).

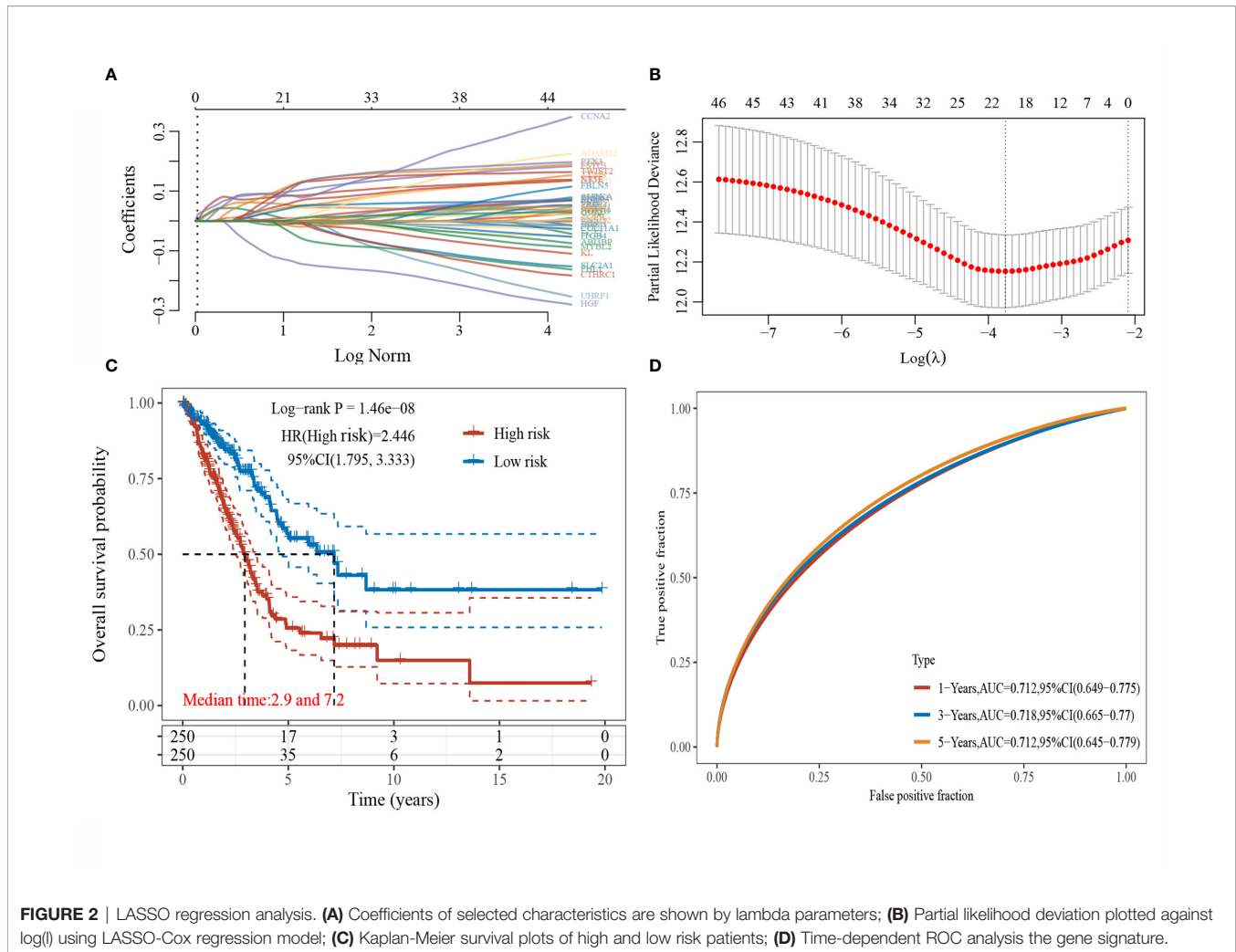
Expression of HGF, PTX3, and S100P in Pan-Cancer and LUAD

We analyzed the expression of HGF, PTX3, and S100P in tumors by integrating data from TCGA and GTEx database samples, and it

was found that HGF and PTX3 expression was downregulated and S100P expression was upregulated in most tumors as compared with normal tissues (**Figures 5A, C, E**). TCGA database samples and integration of GTEx database normal samples displayed that in LUAD, HGF and PTX3 expressions were both downregulated and S100P expression was upregulated (**Figures 5B, D, F**).

Survival Analysis of HGF, PTX3, and S100P

The results of Kaplan-Meier analysis showed (**Figures 6A, C, E**) that high and low expressions of HGF, PTX3, and S100P were significantly correlated with patient prognosis. Specifically, high expression of HGF, PTX3, and S100P were all considered as



poor prognostic factors in LUAD patients; however, ROC curves showed that all the three were less accurate when predicting the prognosis (Figures 6B, D, F).

We further investigated the association of HGF, PTX3, and S100P expression with overall survival in 33 tumors *via* univariate survival analysis. As shown in Figure 7, HGF could significantly affect the overall survival of BLCA, ESCA, KIRC, LGG, LIHC, STAD, THCA (Figure 7A); PTX3 could noticeably affect the overall survival of ACC, BLCA, CESC, GBM, HNSC, KIRC, LGG, LIHC (Figure 7B); S100P could significantly affect the overall survival of CESC, LUAD, THCA, THYM (Figure 7C).

Correlation of HGF, PTX3, and S100P With Immune Cell Infiltration

The expression levels of HGF in LUAD were significantly positively correlated with tumor purity, B cells, CD8+ T cells, CD4+ T cells, neutrophils, and dendritic cells (Figure 8A). However, the expression levels of PTX3 in LUAD were significantly negatively correlated with tumor purity and

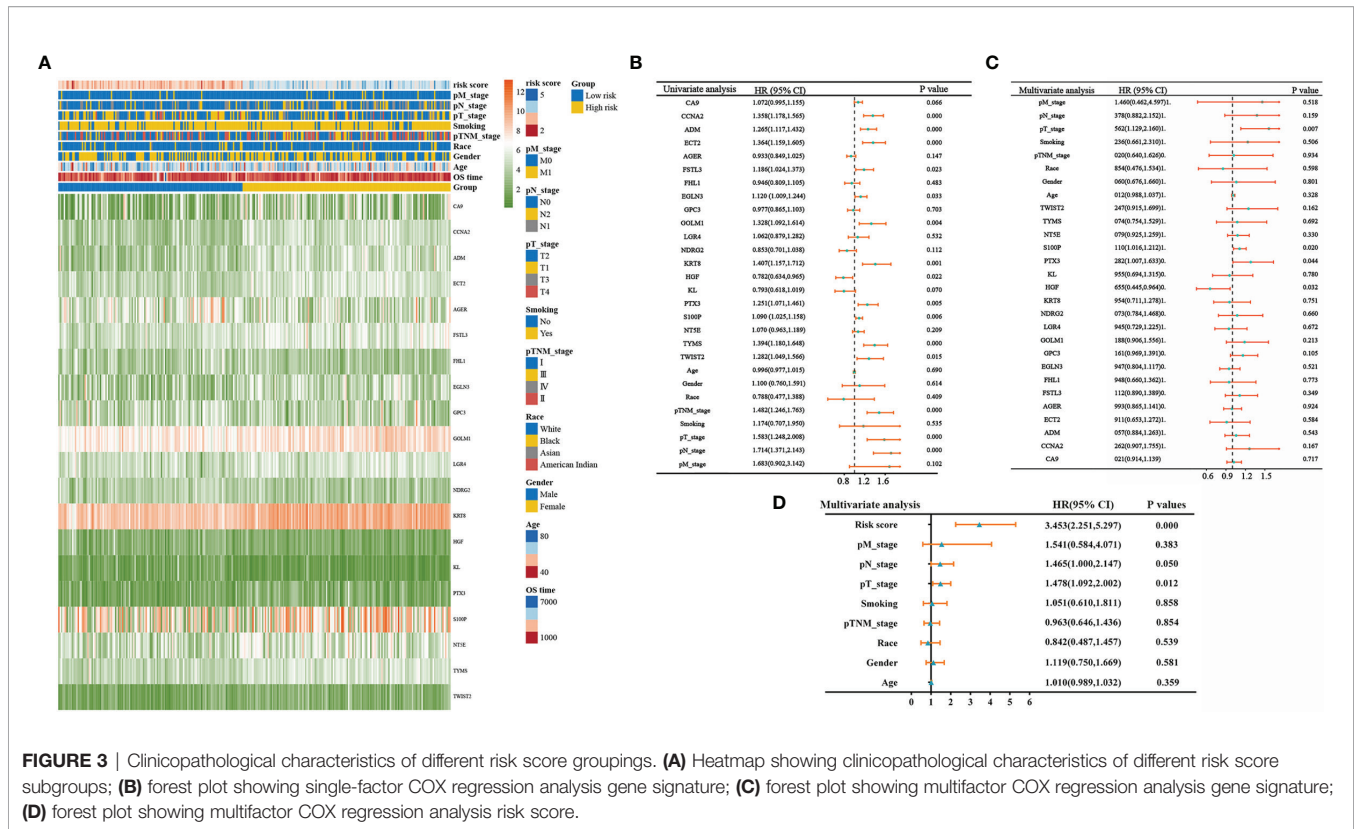
positively correlated with B cells, CD8+ T cells, CD4+ T cells, neutrophils, and dendritic cells (Figure 8B). The expression level of S100P was negatively correlated with B cells, CD8+ T cells, CD4+ T cells, neutrophils, and dendritic cells but positively correlated with tumor purity in LUAD (Figure 8C).

Validation of HGF, PTX3, and S100P Protein Expression

The protein expression of HGF, PTX3, and S100P in lung cancer tissues and normal lung tissues was validated using the HPA online database. The results demonstrated that HGF was expressed in normal lung tissues but it was not detected in lung cancer tissues (Figure 9A). PTX3 was mildly expressed in lung cancer tissues but was not detected in normal alveolar cells (Figure 9B). S100P was moderately expressed in lung cancer tissues but was not detected in normal lung tissues (Figure 9C).

Gene Set Enrichment Analysis

To examine the effect of gene expression on tumors, we divided the human pan-cancer samples into two groups with high and



low expression, according to the expression of HGF, PTX3, S100P, and the enrichment of signaling pathways in KEGG and HALLMARK in high- and low-expression groups was analyzed by GSEA. The top three signaling pathways most significantly enriched in both databases have been listed. GSEA verified that HGF was mainly enriched in hematopoietic cell lineage and inflammatory response (**Figure 10A**), and that PTX3 was mainly enriched in hematopoietic cell lineage and epithelial mesenchymal transition (**Figure 10B**), and that S100P was mainly enriched in ribosome and MYC target (**Figure 10C**).

In Vitro Experiments

We further verified the protein expression level of HGF, PTX3, S100P *in vitro*. Western blotting results showed that the protein expression level of HGF and PTX3 were downregulated in NCI-H2009 *versus* BEAS-2B, while S100P was upregulated (**Figure 11**).

DISCUSSION

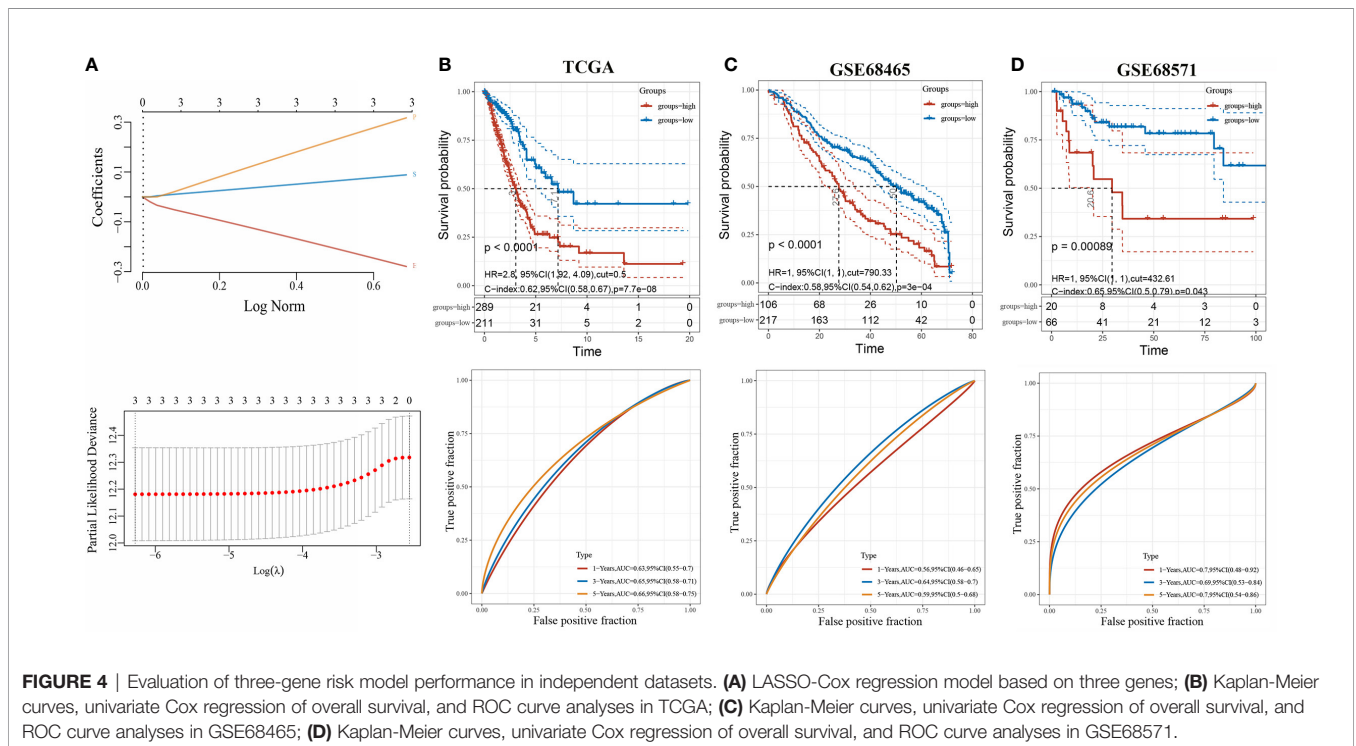
Epithelial-mesenchymal transition (EMT) is a fundamental and critical cell biological process during embryonic development (20). EMT is a reversible process, with the reversal being referred to as mesenchymal-epithelial transition (MET). EMT

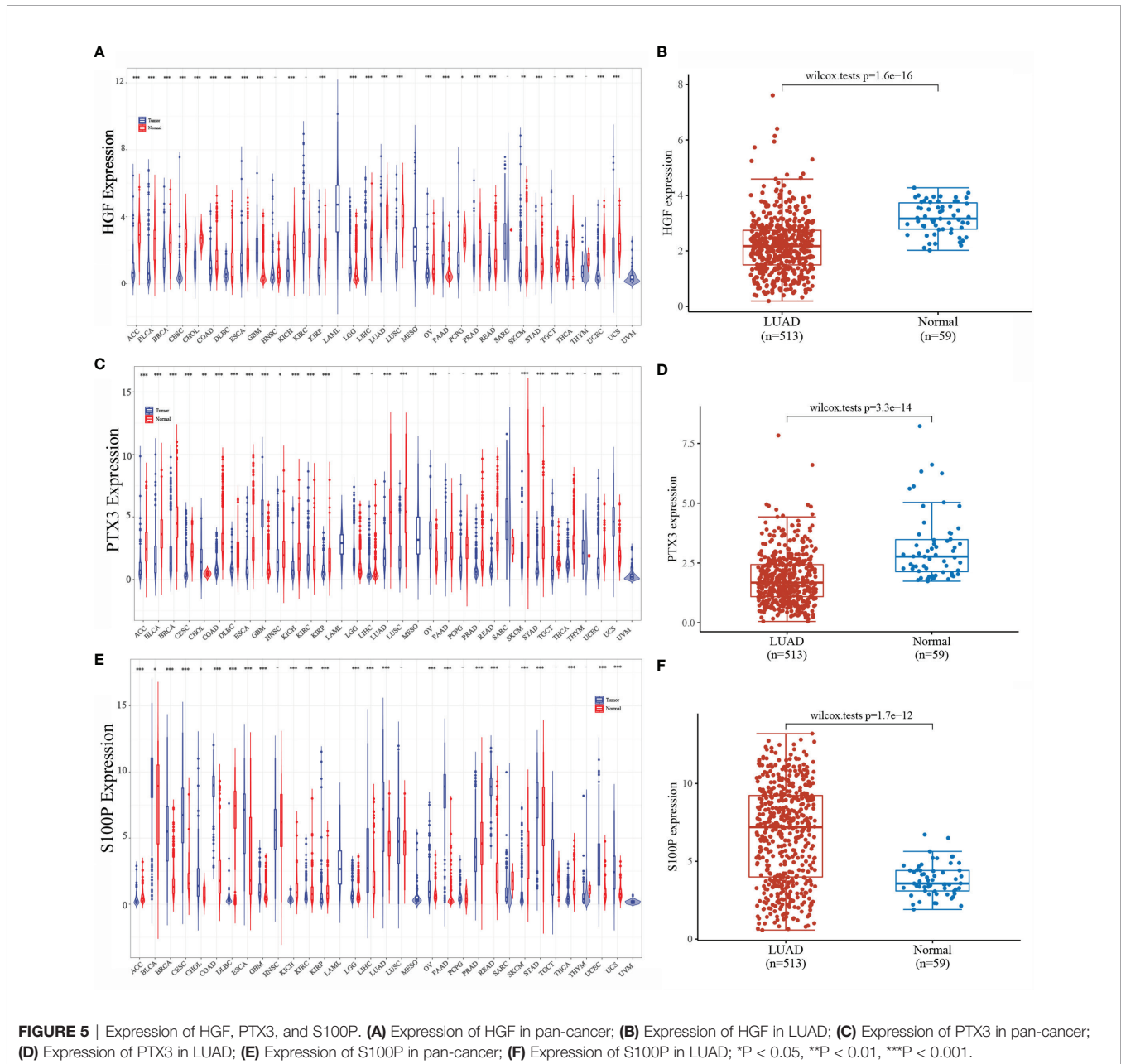
and MET are essential for embryonic development, tissue regeneration, and wound healing, and its role in contributing to organ fibrosis, cancer progression, and metastasis requires a full understanding (21, 22). EMT confers tumor cells the ability to detach from the primary tumor mass and invade lymphatic vessels and blood vessels through extracellular matrix, allowing them to reach secondary tissues or distal organs and reactivate epithelial cell properties to form metastatic foci at secondary sites *via* the MET pathway (23). Assessing EMT state in tumors is challenging as the process is transient and reversible (24). Previous studies have focused on exploring the biological functions and molecular mechanisms of EMT-related genes, such as the expression of E-cadherin, claudins, ocludins, and cytokeratins, and considered these genes as common markers of epithelial state, and Vimentin (VIM), fibronectin, and α -SMA as the markers of mesenchymal state (25–27). However, for some circulating tumor cells (CTCs) underwent EMT, the expression of these (25–27) epithelial cell adhesion-based molecular markers is difficult to be detected (28). Immunicon, an FDA-approved company, has demonstrated that the number of CTCs is indicative of patient prognosis, and Tang et al. (19) first developed a new EMT-related gene signature and constructed a nomogram to predict prognosis of LAUD patients. In addition, the construction of predictive models based on EMT-related genes in LAUD has not been reported. Therefore, searching prognostic predictive biomarkers for LAUD patients based on EMT-related genes may be a promising approach.

TABLE 1 | Clinical characteristics of patients in high- and low-risk subgroups (TCGA, n=283).

Characteristics	Number (High/Low-risk group)	Proportion of patient
Age (years)		
≥65	74/76	53.0%
<65	76/57	47.0%
Sex		
Female	76/74	53.0%
Male	74/59	47.0%
Race		
White	136/114	88.3%
Black	11/16	9.5%
Asian	2/3	1.8%
American Indian	1/0	0.4%
Smoking		
Yes	126/111	83.7%
No	24/22	16.3%
Stage		
I	67/82	52.6%
II	42/27	24.4%
III	31/16	16.6%
IV	10/8	6.4%
T stage		
T1	42/50	32.5%
T2	89/68	55.5%
T3	15/8	8.1%
T4	4/7	3.9%
N stage		
N0	84/101	65.4%
N1	37/19	19.8%
N2	29/13	14.8%
M stage		
M0	140/125	93.6%
M1	10/8	6.4%

In this study, we identified three prognosis-related gene signatures, namely, HGF, PTX3, and S100P. We first constructed a prognostic prediction risk model based on 20 EMT-related gene signatures and verified that the model had strong prediction accuracy through KM curve and ROC curve. Patients were grouped according to their risk score differences. HGF, PTX3, S100P, and risk score were all identified as independent predictors of prognosis of LUAD patients through performing univariate and multifactorial COX regression analyses on these 20 genes and risk scores. The results showed that hGF and PTX3 expression was downregulated and S100P expression was upregulated in LUAD, and that the expression of all the three was correlated with immune cell infiltration, suggesting that all the three gene could promote tumor progression. HGF, which is a cytokine produced by mesenchymal fibroblasts (29), could stimulate the migration, proliferation, migration, cell survival, morphogenesis, and angiogenesis of epithelial cells (30). It has been found that the HGF/c-Met signaling pathway may influence multiple aspects of tumor development by activating specific pathways that induce interactions between cancer cells and the tumor microenvironment in which they reside (29). For example, in EGFR-mutant lung cancer, HGF may be involved in endogenous and acquired resistance to acid kinase inhibitors (31). GSEA results also demonstrated that HGF was mainly enriched in inflammatory response. PTX3 in the pentraxin family is an immunomodulatory factor involved in angiogenesis, proliferation, and immune escape in cancer (32). It has been reported that PTX3 tends to be expressed in the stroma rather than in tumor cell components, suggesting





that the PTX3 gene is epigenetically modified and silenced in cancer cells (33). This is similar to our findings, as we found that PTX3 was low-expressed in tumor cells by protein expression. S100P is a member of the S100 calcium-binding protein. S100P in the S100 calcium-binding protein family is originally isolated from human placenta (34), and S100P expression is elevated in a variety of tumor cell lines and tumor tissues, including in lung cancer (35), pancreatic cancer (36), and breast cancer (37). Moreover, it is also a tumor microenvironment-associated gene (38). Our western blotting results also demonstrated that the protein expression level of HGF and PTX3 were upregulated in NCI-H2009 cell

line compared to BEAS-2B cell line, while the protein expression level of S100P was downregulated.

There are some limitations to this study. Firstly, LUAD is a highly heterogeneous tumor, but certain key clinical variables were not available in public databases, which limited the comprehensiveness when developing the prognostic model. Furthermore, the predictive ability of the EMT-related gene signature using these three genes was only based on bioinformatics analysis, and more basic experiments and clinical evidence are needed to validate the signature.

In summary, this study constructed a prognostic model based on 20 EMT-related gene signatures and validated the

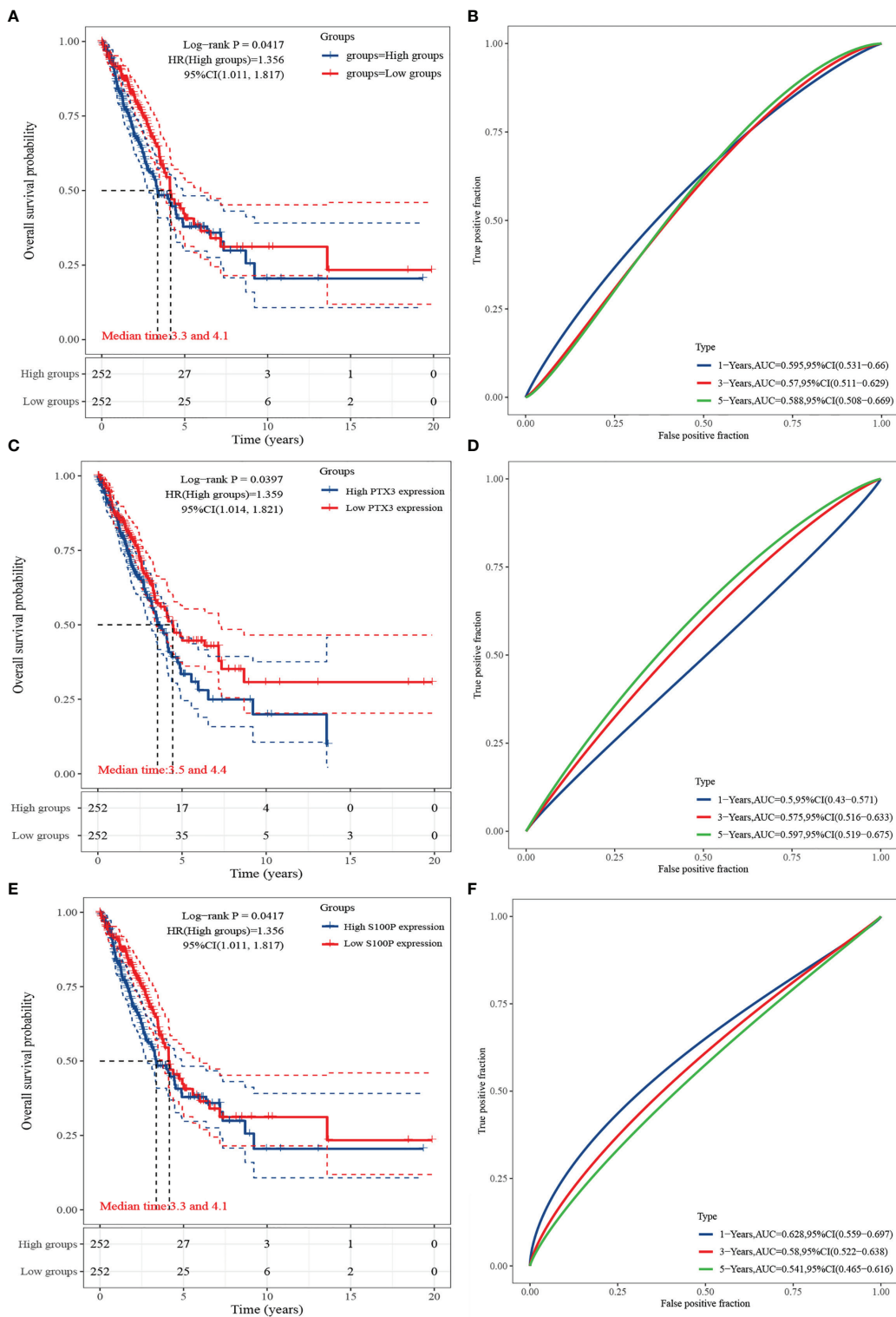


FIGURE 6 | Survival analysis of HGF, PTX3, and S100P. **(A)** KM curve of HGF; **(B)** ROC curve of HGF; **(C)** KM curve of PTX3; **(D)** ROC curve of PTX3; **(E)** KM curve of S100P; **(F)** ROC curve of S100P.

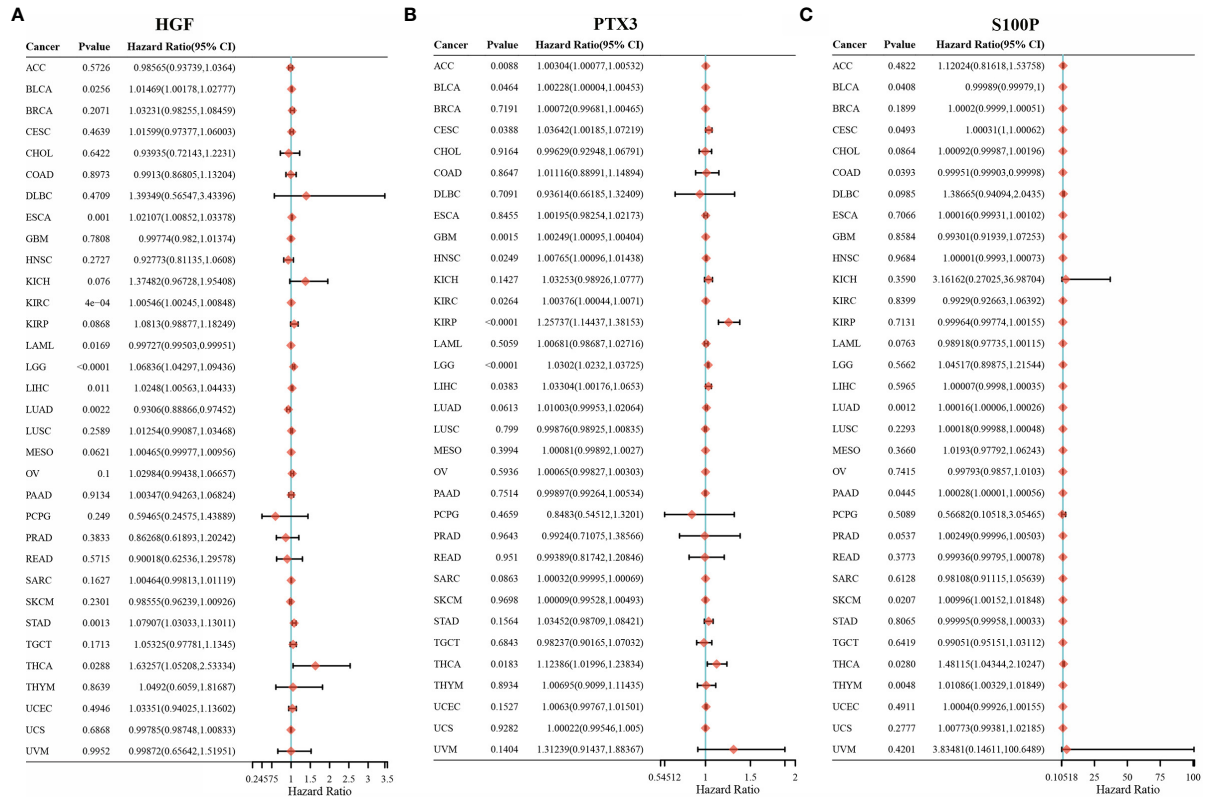


FIGURE 7 | HGF, PTX3, S100P on the prognosis of other cancers. **(A)** Forest plot of HGF on the prognosis in 33 cancer types; **(B)** Forest plot of PTX3 on the prognosis in 33 cancer types; **(C)** Forest plot of S100P on the prognosis in 33 cancer types.

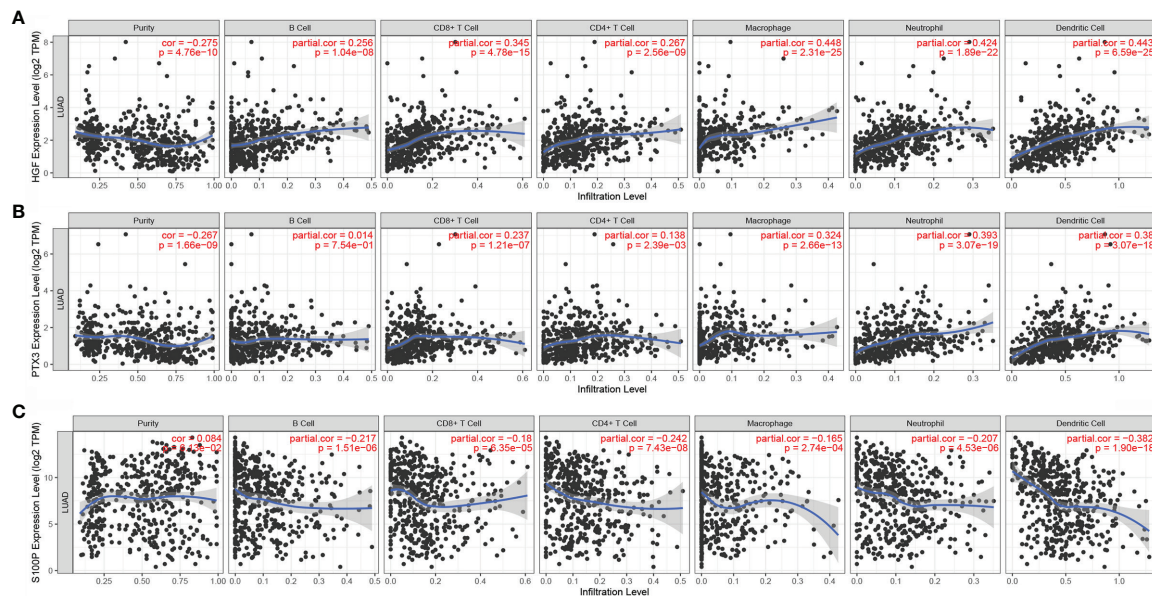


FIGURE 8 | Correlation of HGF, PTX3, and S100P with immune cell infiltration. **(A)** Correlation between HGF and immune cell infiltration; **(B)** Correlation between PTX3 and immune cell infiltration; **(C)** Correlation between S100P and immune cell infiltration.

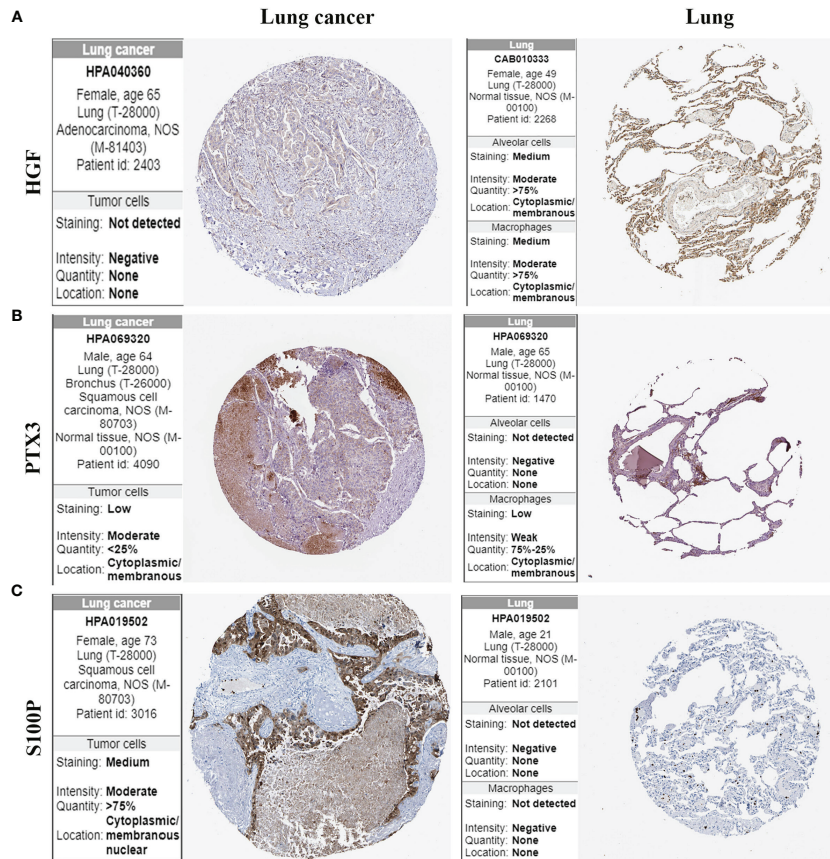


FIGURE 9 | Protein expression of HGF, PTX3, and S100P in lung cancer tissues and normal lung tissues. **(A)** HGF protein expression; **(B)** PTX3 protein expression; **(C)** S100P protein expression.

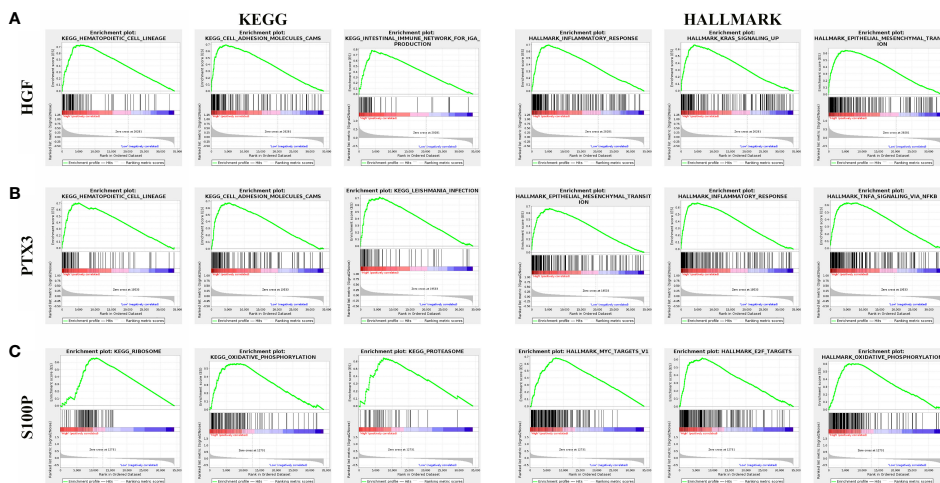


FIGURE 10 | Gene set enrichment analysis of three genes associated with signaling pathways in KEGG and hallmark datasets. **(A)** Results of GSEA of HGF ranked in the top three for its correlation with signaling pathways in KEGG and HALLMARK database. **(B)** Results of GSEA of PTX3 ranked in the top three for its correlation with signaling pathways in KEGG and HALLMARK database. **(C)** Results of GSEA of S100P ranked in the top three for its correlation with signaling pathways in KEGG and HALLMARK database.

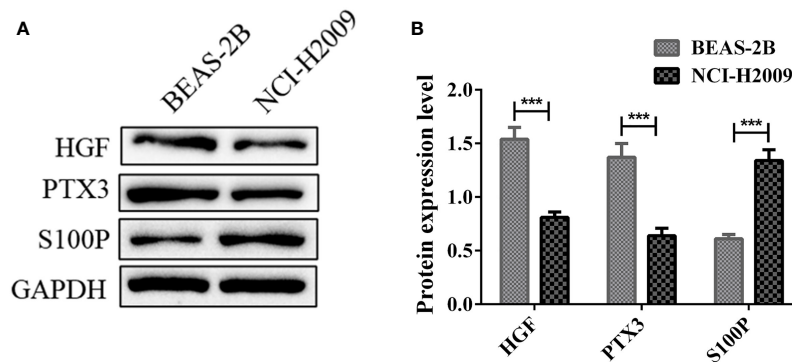


FIGURE 11 | Protein expression levels of HGF, PTX3, S100P in BEAS-2B and NCI-H2009. **(A)** Western blot bands representing HGF, PTX3, S100P protein expression level in BEAS-2B and NCI-H2009; **(B)** Protein expression level of HGF, PTX3, S100P. *** $P < 0.001$.

performance the model. The final three prognostic predictors of LUAD patients were analyzed by COX regression analysis to obtain three independent factors. These findings may provide a new direction for prognosis prediction and individualized treatment for patients with LUAD.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

REFERENCES

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2018) 68(6):394–424. doi: 10.3322/caac.21492
- Herbst RS, Morgensztern D, Boshoff C. The Biology and Management of non-Small Cell Lung Cancer. *Nature* (2018) 553(7689):446–54. doi: 10.1038/nature25183
- Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2019. *CA Cancer J Clin* (2019) 69(1):7–34. doi: 10.3322/caac.21551
- Devarakonda S, Morgensztern D, Govindan R. Genomic Alterations in Lung Adenocarcinoma. *Lancet Oncol* (2015) 16(7):e342–51. doi: 10.1016/S1470-2045(15)00077-7
- Kadota K, Sima CS, Arcila ME, Hedvat C, Kris MG, Jones DR, et al. KRAS Mutation Is a Significant Prognostic Factor in Early Stage Lung Adenocarcinoma. *Am J Surg Pathol* (2016) 40(12):1579. doi: 10.1097/PAS.0000000000000744
- Guo S, Chen Y, Shi S, Wang X, Zhang H, Zhan Y, et al. Arctigenin, a Novel TMEM16A Inhibitor for Lung Adenocarcinoma Therapy. *Pharmacol Res* (2020) 155:104721. doi: 10.1016/j.phrs.2020.104721
- Huang W, Lin A, Luo P, Liu Y, Xu W, Zhu W, et al. EPHA5 Mutation Predicts the Durable Clinical Benefit of Immune Checkpoint Inhibitors in Patients With Lung Adenocarcinoma. *Cancer Gene Ther* (2020) 28(7):864–74. doi: 10.1038/s41417-020-0207-6
- Niu Y, Lin A, Luo P, Zhu W, Wei T, Tang R, et al. Prognosis of Lung Adenocarcinoma Patients With NTRK3 Mutations to Immune Checkpoint Inhibitors. *Front Pharmacol* (2020) 11:1213. doi: 10.3389/fphar.2020.01213
- Wang Y, Zhang Q, Gao Z, Xin S, Zhao Y, Zhang K, et al. A Novel 4-Gene Signature for Overall Survival Prediction in Lung Adenocarcinoma Patients With Lymph Node Metastasis. *Cancer Cell Int* (2019) 19(1):1–9. doi: 10.1186/s12935-019-0822-1
- Wan L, Pantel K, Kang Y. Tumor Metastasis: Moving New Biological Insights Into the Clinic. *Nat Med* (2013) 19(11):1450–64. doi: 10.1038/nm.3391
- Chen Y, Wang D-D, Wu Y-P, Su D, Zhou T-Y, Gai R-H, et al. MDM2 Promotes Epithelial–Mesenchymal Transition and Metastasis of Ovarian Cancer SKOV3 Cells. *Br J Cancer* (2017) 117(8):1192–201. doi: 10.1038/bjc.2017.265
- Liao TT, Yang MH. Revisiting Epithelial–Mesenchymal Transition in Cancer Metastasis: The Connection Between Epithelial Plasticity and Stemness. *Mol Oncol* (2017) 11(7):792–804. doi: 10.1002/1878-0261.12096
- Williams ED, Gao D, Redfern A, Thompson EW. Controversies Around Epithelial–Mesenchymal Plasticity in Cancer Metastasis. *Nat Rev Cancer* (2019) 19(12):716–32. doi: 10.1038/s41568-019-0213-x
- Brabletz T, Kalluri R, Nieto MA, Weinberg RA. EMT in Cancer. *Nat Rev Cancer* (2018) 18(2):128–34. doi: 10.1038/nrc.2017.118
- Nieto MA. Context-Specific Roles of EMT Programmes in Cancer Cell Dissemination. *Nat Cell Biol* (2017) 19(5):416–8. doi: 10.1038/ncb3520
- Ribatti D, Tamma R, Annese T. Epithelial–Mesenchymal Transition in Cancer: A Historical Overview. *Trans Oncol* (2020) 13(6):100773. doi: 10.1016/j.tranon.2020.100773
- Miao Y, Li A-L, Wang L, Fan C-F, Zhang X-P, Xu H-T, et al. Overexpression of NEDD9 Is Associated With Altered Expression of E-Cadherin, β -Catenin and N-Cadherin and Predictive of Poor Prognosis in Non-Small Cell Lung Cancer. *Pathol Oncol Res* (2013) 19(2):281–6. doi: 10.1007/s12253-012-9580-2
- Shi Y, Wu H, Zhang M, Ding L, Meng F, Fan X. Expression of the Epithelial–Mesenchymal Transition-Related Proteins and Their Clinical Significance in Lung Adenocarcinoma. *Diagn Pathol* (2013) 8:89. doi: 10.1186/1746-1596-8-89
- Tang Y, Jiang Y, Qing C, Wang J, Zeng Z. Systematic Construction and Validation of an Epithelial–Mesenchymal Transition Risk Model to Predict

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.726834/full#supplementary-material>

Supplementary Table 1 | Tumor name abbreviations and corresponding meanings.

- Prognosis of Lung Adenocarcinoma. *Aging (Albany NY)* (2021) 13(1):794. doi: 10.18632/aging.202186
20. Nieto MA, Huang RY-J, Jackson RA, Thiery JP. EMT: 2016. *Cell* (2016) 166(1):21–45. doi: 10.1016/j.cell.2016.06.028
 21. Chaffer CL, Weinberg RA. A Perspective on Cancer Cell Metastasis. *Science* (2011) 331(6024):1559–64. doi: 10.1126/science.1203543
 22. Pei D, Shu X, Gassama-Diagne A, Thiery JP. Mesenchymal–Epithelial Transition in Development and Reprogramming. *Nat Cell Biol* (2019) 21(1):44–53. doi: 10.1038/s41556-018-0195-z
 23. Yoshida T, Ozawa Y, Kimura T, Sato Y, Kuznetsov G, Xu S, et al. Eribulin Mesilate Suppresses Experimental Metastasis of Breast Cancer Cells by Reversing Phenotype From Epithelial–Mesenchymal Transition (EMT) to Mesenchymal–Epithelial Transition (MET) States. *Br J Cancer* (2014) 110(6):1497–505. doi: 10.1038/bjc.2014.80
 24. Vasaiakar SV, Deshmukh AP, den Hollander P, Addanki S, Kuburich NA, Kudaravalli S, et al. EMTome: A Resource for Pan-Cancer Analysis of Epithelial–Mesenchymal Transition Genes and Signatures. *Br J Cancer* (2021) 124(1):259–69. doi: 10.1038/s41416-020-01178-9
 25. Chen C, Wei Y, Hummel M, Hoffmann TK, Gross M, Kaufmann AM, et al. Evidence for Epithelial–Mesenchymal Transition in Cancer Stem Cells of Head and Neck Squamous Cell Carcinoma. *PLoS One* (2011) 6(1):e16466. doi: 10.1371/journal.pone.0016466
 26. Wu C, Zeng M-H, Liao G, Qian K, Li H. Neuropilin-1 Interacts With Fibronectin-1 to Promote Epithelial–Mesenchymal Transition Progress in Gastric Cancer. *Onco Targets Ther* (2020) 13:10677. doi: 10.2147/OTT.S275327
 27. Zhitnyak IY, Rubtsova SN, Litovka NI, Gloushankova NA. Early Events in Actin Cytoskeleton Dynamics and E-Cadherin-Mediated Cell–Cell Adhesion During Epithelial–Mesenchymal Transition. *Cells* (2020) 9(3):578. doi: 10.3390/cells9030578
 28. Zheng Y, Zhang J, Huang M, Wang T, Qu X, Wu L, et al. Selection of Aptamers Against Vimentin for Isolation and Release of Circulating Tumor Cells Undergoing Epithelial Mesenchymal Transition. *Anal Chem* (2020) 92(7):5178–84. doi: 10.1021/acs.analchem.9b05690
 29. Zambelli A, Biamonti G, Amato A. HGF/c-Met Signalling in the Tumor Microenvironment. In: *Tumor Microenvironment*. United States: Springer (2021). p. 31–44.
 30. Yano S, Wang W, Li Q, Matsumoto K, Sakurama H, Nakamura T, et al. Hepatocyte Growth Factor Induces Gefitinib Resistance of Lung Adenocarcinoma With Epidermal Growth Factor Receptor–Activating Mutations. *Cancer Res* (2008) 68(22):9479–87. doi: 10.1158/0008-5472.CAN-08-1643
 31. Turke AB, Zejnullahu K, Wu Y-L, Song Y, Dias-Santagata D, Lifshits E, et al. Preexistence and Clonal Selection of MET Amplification in EGFR Mutant NSCLC. *Cancer Cell* (2010) 17(1):77–88. doi: 10.1016/j.ccr.2009.11.022
 32. Netti GS, Lucarelli G, Spadaccino F, Castellano G, Gigante M, Divella C, et al. PTX3 Modulates the Immunoflogosis in Tumor Microenvironment and is a Prognostic Factor for Patients With Clear Cell Renal Cell Carcinoma. *Aging (Albany NY)* (2020) 12(8):7585. doi: 10.18632/aging.103169
 33. Infante M, Allavena P, Garlanda C, Nebuloni M, Morengi E, Rahal D, et al. Prognostic and Diagnostic Potential of Local and Circulating Levels of Pentraxin 3 in Lung Cancer Patients. *Int J Cancer* (2016) 138(4):983–91. doi: 10.1002/ijc.29822
 34. Becker T, Gerke V, Kube E, Weber K. S100P, a Novel Ca²⁺-Binding Protein From Human Placenta: cDNA Cloning, Recombinant Protein Expression and Ca²⁺ Binding Properties. *Eur J Biochem* (1992) 207(2):541–7. doi: 10.1111/j.1432-1033.1992.tb17080.x
 35. Hsu Y-L, Hung J-Y, Liang Y-Y, Lin Y-S, Tsai M-J, Chou S-H, et al. S100P Interacts With Integrin $\alpha 7$ and Increases Cancer Cell Migration and Invasion in Lung Cancer. *Oncotarget* (2015) 6(30):29585. doi: 10.18632/oncotarget.4987
 36. Arumugam T, Simeone DM, Van Golen K, Logsdon CD. S100P Promotes Pancreatic Cancer Growth, Survival, and Invasion. *Clin Cancer Res* (2005) 11(15):5356–64. doi: 10.1158/1078-0432.CCR-05-0092
 37. Kikuchi K, McNamara KM, Miki Y, Iwabuchi E, Kanai A, Miyashita M, et al. S100P and Ezrin Promote Trans-Endothelial Migration of Triple Negative Breast Cancer Cells. *Cell Oncol* (2019) 42(1):67–80. doi: 10.1007/s13402-018-0408-2
 38. Gao X, Yamazaki Y, Pecori A, Tezuka Y, Ono Y, Omata K, et al. Histopathological Analysis of Tumor Microenvironment and Angiogenesis in Pheochromocytoma. *Front Endocrinol* (2020) 11. doi: 10.3389/fendo.2020.587779

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Shi, Wang, Guo, Chen, Tong, Tong, Hu, Li and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.