The significance of spread through air spaces in the prognostic assessment model of stage I lung adenocarcinoma and the exploration of its invasion mechanism

YangYang Niu
XinHao Han
Yuan Zeng
Abiyasi Nanding
Qiang Bai
SaiNan Guo
YaLi Hou
Yan Yu
QiuJu Zhang
XiaoMei Li (✉️ 0337@hrbmu.edu.cn)

Research Article

Keywords: lung adenocarcinoma, STAS, nomogram, tumor lymphocyte infiltration, CXCL8

Posted Date: November 11th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2252132/v1

License: ☒️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Journal of Cancer Research and Clinical Oncology on March 7th, 2023. See the published version at https://doi.org/10.1007/s00432-023-04619-z.
Abstract

Purpose
Spread through air spaces (STAS) is a crucial invasive mode of lung cancer and has been shown to be associated with early recurrence and metastasis. We aimed to develop a prognostic risk assessment model for stage I lung adenocarcinoma based on STAS and other pathological features and to explore the potential relationship between CXCL-8, Smad2, Snail, and STAS.

Methods
312 patients who underwent surgery at Harbin Medical University Cancer Hospital with pathologically diagnosed stage I lung adenocarcinoma were reviewed in the study. STAS and other pathological features were identified by H&E staining, and a prognostic risk assessment model was established. The expression levels of CXCL8, Smad2, and Snail were determined by immunohistochemistry.

Results
The nomogram was established based on age, smoking history, STAS, tumor lymphocyte infiltration, tissue subtype, nuclear grade, and tumor size. The C-index for DFS was (training set 0.84 vs validation set 0.77) and for OS was (training set 0.83 vs validation set 0.78). Decision curve analysis showed that the model constructed has a better net benefit than traditional reporting. The prognostic risk score validated the risk stratification value for stage I lung adenocarcinoma. STAS was an important prognostic factor associated with stronger invasiveness and higher expression of CXCL8, Smad2, and Snail. CXCL8 was associated with poorer DFS and OS.

Conclusions
We developed and validated a survival risk assessment model and the prognostic risk score formula for stage I lung adenocarcinoma. Additionally, we found that CXCL8 could be used as a potential biomarker for STAS and poor prognosis, and its mechanism may be related to EMT.

1. Introduction
Lung cancer is the malignant tumor with the highest morbidity and mortality, among which adenocarcinoma accounts for the vast majority (Siegel et al. 2021). With the popularization of physical examination and the development of imaging technology, more and more stage I lung adenocarcinomas have been diagnosed. Although most patients have a favorable prognosis, the cumulative incidence of local recurrence and distant metastasis within 5 years after complete surgical resection is approximately 12%-15% and 17%-22%, respectively (Hung et al. 2012; Varlotto et al. 2015). For patients with stage IA
lung adenocarcinoma, postoperative observation and follow-up should be performed instead of adjuvant chemotherapy, and stage IB patients with high-risk factors may be considered with adjuvant chemotherapy according to National Comprehensive Cancer Network (NCCN) guidelines (National Comprehensive Cancer Network.2021). Exploring and discovering the clinical and pathological factors of poor prognosis in patients with stage I lung adenocarcinoma and establishing a prognosis prediction model are the keys to breakthrough therapy.

At present, the pathological reports of lung cancer mainly focus on tumor size, degree of differentiation, histological type, pleural invasion, tissue margins, and lymph node dissection. However, these clinicopathological factors are insufficient to reflect prognostic differences. Therefore, our study increased the proportion of clinicopathological features, such as tumor tissue content. Therefore, our study added clinicopathological features, such as the proportion of tumor tissue content, the degree of tumor cell differentiation, and the nuclear grade that reflect the characteristics of tumor parenchyma, the proportion of intratumoral stroma content, the proportion of stromal tumor-infiltrating lymphocytes (sTIL), and the proportion of peritumoral tumor-infiltrating lymphocytes (pTIL) that reflected the characteristics of tumor stroma. Most importantly, our study incorporated the observation and analysis of STAS, which proved to be a strong prognostic factor, significantly associated with recurrence-free and poorer overall survival. We demonstrated that STAS was an independent prognostic risk factor for stage I lung adenocarcinoma and contributed significantly to prognostic risk assessment models. However, the relationship between the morphological and molecular properties of STAS remains unclear.

Interleukin-8 (CXCL8, IL-8) is a granulocyte chemokine with multiple roles in the tumor microenvironment, including recruiting immunosuppressive cells to tumors, enhancing tumor angiogenesis, and promoting epithelial-to-mesenchymal transition (EMT) (Fousek et al. 2021; David et al. 2016; Palena et al. 2012). CXCL8 is also a potential biomarker whose overexpression plays an important role in the metastasis and poor prognosis of melanoma (Wu et al. 2012), breast (Fang et al. 2017), and colorectal cancers (Xiao et al. 2015). In our previous study, analyzing the STAS transcriptome in lung adenocarcinoma by high-throughput sequencing found that CXCL8 promotes STAS occurrence and correlates with survival (Zeng et al. 2022). In this study, we investigated the relationship between CXCL8, EMT markers Small mothers against decapentaplegic protein 2 (Smad2), Zinc finger protein transcription factor (Snail) and STAS in clinical samples, which might provide new insights into the mechanism of STAS formation.

2. Methods

2.1 Patients

312 patients with stage I lung adenocarcinoma who underwent surgical resection at Harbin Medical University Cancer Hospital between January 2013 to December 2016 were retrospectively reviewed. The inclusion criteria included: primary stage I (T1/T2aN0M0) lung adenocarcinoma confirmed by pathologists according to the criteria of the American Joint Committee on Cancer (AJCC 8th). Patients with multiple nodules, incomplete clinical and pathological data, other malignancies, positive surgical
margins, neoadjuvant therapy, or who died in the perioperative period were excluded. Clinical information was collected through medical records, including age, smoking history, surgical resection, and stage. Follow-up information was obtained through electronic medical records and telephone interviews. This study was approved by the Ethics Committee of the Harbin Medical University Cancer Hospital.

2.2 Pathological Assessment

Pathological data were independently assessed by two pulmonary pathologists blinded to clinical information, and inconsistent cases were discussed and reached a consensus, including tumor size, tissue subtype, STAS, visceral pleural invasion, tumor stromal ratio (TSR), sTIL, pTIL, and nuclear grades. The tumor area was completely sampled and reassessed per patient.

Histological subtyping was done according to the International Association for the Study of Lung Cancer, the American Thoracic Society, and the European Respiratory Society (IASLC/ATS/ERS) in 2011 classification of lung adenocarcinoma, including low-grade pattern (lepidic), intermediate (acinar or papillary), high-grade pattern (micropapillary or solid). High-grade patterns were classified using a 20% cutoff value. STAS was defined as micropapillary clusters, solid nests, or single cancer cells that invade the lung parenchyma beyond the tumor margins. TSR was assessed as a percentage per 10-fold (10%, 20%, 30%, etc.) for all tumor areas. The cut-off value was 50%, and they were divided into two groups, the stroma-rich group (TSR < 50%) and the stroma-poor group (TSR ≥ 50%). The level of sTIL was calculated according to the proportion of mononuclear cells in the tumor stroma, granulocytes and other polymorphonuclear leukocytes were excluded and were assessed as a percentage per 5-fold (5%, 10%, etc.). The level of pTIL was calculated according to the proportion of mononuclear cells in the peritumoral tumor and was assessed as a percentage per 5-fold (5%, 10%, etc.). The degree of nuclear atypia was assessed as follows. Grade 1 was mild atypia - uniform nuclei in size and shape. Grade 2 was moderate atypia - nuclei in intermediate size with slight irregularity in shape. Grade 3 was severe atypia - enlarged nuclei of varied sizes and irregular contours, with some nuclei at least twice as large as others. Elastic fibers staining was used when pleural invasion cannot be assessed microscopically.

2.3 Immunohistochemistry

The paraffin-embedded tissues of 100 patients with lung adenocarcinoma were selected and cut into 4µm thick sections for immunohistochemical staining. The paraffin sections were dewaxed and hydrated. CXCL8 (1:100, Affinity, USA), Smad2 (1:50, Bioss, China), and Snail (1:500, Affinity, USA) were used as primary antibodies. The HRP Rabbit/Mouse (Zhongshan Goldenbridge, Beijing, China) was used as a secondary antibody, and the immunoreactivity of the tissue sections was visualized using diaminobenzidine (Zhongshan Goldenbridge, Beijing, China). The staining intensity was scored as follows: (0) no staining, (1) weak = light yellow, (2) moderate = yellow-brown, and (3) strong = brown. The percentages of positive cells were scored as follows: (0) < 5%, (1) 5–25%, (2) 25–50%, (3) 50–75%, and (4) > 75%.

2.4 Statistical Analysis
Disease-free survival (DFS) was defined as the date of surgery to the recurrence. Overall survival (OS) was defined as the time of surgery to death. The last follow-up date was July 1, 2021. The training and testing set were generated via random sampling with a ratio of 70/30 to validate the model constructed. Patients in the training set and validation set were compared in terms of baseline variables.

We constructed the prognosis models for both DFS and OS endpoints, respectively. Univariate Cox proportional hazards were performed on each factor first. Then all significant factors on univariate analysis were entered into multivariate Cox regression analysis using a stepwise procedure with an inclusion criterion of 0.10 and exclusion of 0.15. Given the known existence that age would affect the prognosis of cancer patients, the multivariate analysis was adjusted for age as a potential confounder. Finally, the prognosis model for DFS and OS was produced using the training set and validated in the validation set, respectively. DFS and OS nomogram was generated with R package rms. The calibration curves of 3, 5, and 7 years were drawn to display the prediction probability of the constructed model. We used the time-dependent ROC analyses to evaluate our model's discrimination ability using the timeROC package. The model clinical benefit was measured from the decision curve analysis (DCA). And model performances of the DFS model and OS model were validated in the validation set.

The prognostic risk score was constructed by coefficients and their features based on the multivariate Cox regression analysis. To benchmark the prognosis power of the score, the ROC curves were applied compared to other single common characteristics. The patients were divided into the high-score and low-score groups by taking the median score value, and the survival between the two groups was estimated.

All statistical analyses were performed by R version 4.0.3 and GraphPad Prism 9.0. All tests were two-sided and p-value < 0.05 were considered statistically significant. The Chi-square tests were used to compare different groups for categorical variables. For the continuous variables the normality was tested by the Shapiro-Wilk normality test, and the Student's t-test was done if normally distributed or the Wilcoxon rank-sum if nonnormal. Kaplan-Meier survival between the two groups was statistically analyzed by the Log-rank test.

3. Results

3.1 Patients Characteristics

According to the inclusion and exclusion criteria, 312 lung adenocarcinoma patients were enrolled in the study. We split the dataset into a training set (70% of the patients, n = 219) and a validation set (30% of the patients, n = 93) listed in Table 1. Only the visceral pleural invasion differed significantly between the groups. Both the training and validation sets had no statistical differences in STAS, tissue subtype, TSR, nuclear grades, smoking history, surgical resection, stage, age, sTIL, pTIL, and tumor size.
Table 1
Clinicopathological characteristics of the patients.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Training Set (n = 219)</th>
<th>Validation Set (n = 93)</th>
<th>Statistics</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td>58.38 ± 7.76</td>
<td>58.23 ± 8.46</td>
<td>-.1598</td>
<td>0.8732</td>
</tr>
<tr>
<td><strong>Smoking history</strong></td>
<td></td>
<td></td>
<td>0.0355</td>
<td>0.8507</td>
</tr>
<tr>
<td>Yes</td>
<td>92(42.0)</td>
<td>38(40.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>127(58.0)</td>
<td>55(59.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stage</strong></td>
<td></td>
<td></td>
<td>0.0101</td>
<td>0.9198</td>
</tr>
<tr>
<td>IA</td>
<td>187(85.4)</td>
<td>79(84.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IB</td>
<td>32(14.6)</td>
<td>14(15.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Surgical resection</strong></td>
<td></td>
<td></td>
<td>0.9025</td>
<td>0.3421</td>
</tr>
<tr>
<td>Lobar</td>
<td>184(84.0)</td>
<td>74(79.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sublobar</td>
<td>35(16.0)</td>
<td>19(20.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tumor size</strong></td>
<td>2.00(1.50-3.00)</td>
<td>2.20(1.50-3.00)</td>
<td>0.2191</td>
<td>0.8266</td>
</tr>
<tr>
<td><strong>Visceral pleural invasion</strong></td>
<td></td>
<td></td>
<td>4.9353</td>
<td>0.0263</td>
</tr>
<tr>
<td>Present</td>
<td>51(23.3)</td>
<td>33(35.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>168(76.7)</td>
<td>60(64.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>STAS</strong></td>
<td></td>
<td></td>
<td>0.0241</td>
<td>0.8766</td>
</tr>
<tr>
<td>Yes</td>
<td>101(46.1)</td>
<td>42(45.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>118(53.9)</td>
<td>51(54.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tissue subtype</strong></td>
<td></td>
<td></td>
<td>0.0001</td>
<td>0.9904</td>
</tr>
<tr>
<td>High-grade patterns &lt; 20%</td>
<td>127(58.0)</td>
<td>54(58.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-grade patterns ≥ 20%</td>
<td>92(42.0)</td>
<td>39(41.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TSR</strong></td>
<td></td>
<td></td>
<td>0.0764</td>
<td>0.7822</td>
</tr>
<tr>
<td>&lt; 50%</td>
<td>105(47.9)</td>
<td>43(46.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 50%</td>
<td>114(52.1)</td>
<td>50(53.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>sTIL</strong></td>
<td>30(10–60)</td>
<td>40(10–70)</td>
<td>0.7147</td>
<td>0.4748</td>
</tr>
<tr>
<td><strong>pTIL</strong></td>
<td>20(10–40)</td>
<td>20(10–40)</td>
<td>0.6467</td>
<td>0.5178</td>
</tr>
<tr>
<td><strong>Nuclear grades</strong></td>
<td></td>
<td></td>
<td>0.8690</td>
<td>0.6476</td>
</tr>
<tr>
<td>Grade 1</td>
<td>62(28.3)</td>
<td>30(32.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variables</td>
<td>Training Set (n = 219)</td>
<td>Validation Set (n = 93)</td>
<td>Statistics</td>
<td>p</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------</td>
<td>-------------------------</td>
<td>------------</td>
<td>---</td>
</tr>
<tr>
<td>Grade 2</td>
<td>90(41.1)</td>
<td>39(41.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 3</td>
<td>67(30.6)</td>
<td>24(25.8)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

STAS, spread through air space; TSR, Tumor stromal ratio; sTIL, stromal tumor infiltrating lymphocytes; pTIL, peritumoral tumor infiltrating lymphocytes.

### 3.2 Construction and Validation of the DFS Nomogram

Six significant variables associated with DFS by univariate Cox analysis were included in the multivariate analysis. We further adjusted for factors that related to the outcomes of patients, including age, and stepwise regression was used. In the covariate-adjusted model, age, smoking history, STAS, tissue subtype, sTIL, nuclear grades, and tumor size, were significant predictors of DFS (Table 2). Then, the nomogram was constructed to predict DFS at 3, 5, and 7 years. The predictive ability of the model measured by the C-index was 0.84 (95%CI: 0.77–0.90) in the training cohort and 0.77 (95%CI: 0.66–0.88) in the validation cohort. The nomogram was well calibrated, which demonstrated good agreement between nomogram-predicted probability and actual observed probability. The variable with a high AUC above 0.8 was considered to be an effective predictor. In ROC analysis, the AUCs were 0.87, 0.89, and 0.87 in the training set respectively, and higher than 0.8 in the validation set, indicating the good quality and accurate prediction of our model for DFS (Fig. 1). The DCA curves showed that beyond the 7 years, the net benefit of the DFS model was higher than the combined variables included in the traditional pathology report, which demonstrated the clinical utility and value of our DFS model (Fig. 2). Taken together, the calibration curves, ROC analysis, and DCA curves indicated that our model enabled valuables prediction for DFS.
Table 2
Univariate and multivariate cox regression analysis of factors associated with disease-free survival.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Univariate analysis</th>
<th></th>
<th></th>
<th>Multivariate analysis</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>95% CI</td>
<td>p</td>
<td>HR</td>
<td>95% CI</td>
<td>p</td>
</tr>
<tr>
<td>Age</td>
<td>1.034</td>
<td>0.993–1.077</td>
<td>0.1045</td>
<td>1.054</td>
<td>1.009–1.102</td>
<td>0.0189</td>
</tr>
<tr>
<td>Smoking history (Yes VS No)</td>
<td>2.273</td>
<td>1.227–4.211</td>
<td>0.0091</td>
<td>2.674</td>
<td>1.408–5.078</td>
<td>0.0027</td>
</tr>
<tr>
<td>Stage (IB VS IA)</td>
<td>1.887</td>
<td>0.902–3.948</td>
<td>0.0917</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgical resection (Sublobar VS Lobar)</td>
<td>1.295</td>
<td>0.599–2.797</td>
<td>0.5111</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor size</td>
<td>1.640</td>
<td>1.181–2.278</td>
<td>0.0031</td>
<td>1.545</td>
<td>1.079–2.212</td>
<td>0.0176</td>
</tr>
<tr>
<td>Visceral pleural invasion (Absent VS Present)</td>
<td>0.640</td>
<td>0.333–1.232</td>
<td>0.1815</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAS (Yes VS No)</td>
<td>5.457</td>
<td>2.607–11.423</td>
<td>&lt;.0001</td>
<td>2.409</td>
<td>1.033–5.617</td>
<td>0.0417</td>
</tr>
<tr>
<td>Tissue subtype (High-grade patterns ≥ 20% VS &lt; 20%)</td>
<td>4.375</td>
<td>2.235–8.561</td>
<td>&lt;.0001</td>
<td>2.124</td>
<td>0.973–4.634</td>
<td>0.0584</td>
</tr>
<tr>
<td>TSR (≥ 50% VS &lt; 50%)</td>
<td>0.519</td>
<td>0.278–0.968</td>
<td>0.0391</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sTIL</td>
<td>0.978</td>
<td>0.965–0.991</td>
<td>0.0011</td>
<td>0.983</td>
<td>0.970–0.995</td>
<td>0.0061</td>
</tr>
<tr>
<td>pTIL</td>
<td>1.003</td>
<td>0.992–1.015</td>
<td>0.5747</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuclear grades (Grade 2 VS 1)</td>
<td>6.027</td>
<td>1.386–26.215</td>
<td>0.0166</td>
<td>2.891</td>
<td>0.630–13.278</td>
<td>0.0343</td>
</tr>
<tr>
<td>Nuclear grades (Grade 3 VS 1)</td>
<td>14.878</td>
<td>3.513–63.017</td>
<td>0.0002</td>
<td>5.306</td>
<td>1.131–24.892</td>
<td>0.0176</td>
</tr>
</tbody>
</table>

STAS, spread through air space; TSR, Tumor stromal ratio; sTIL, stromal tumor infiltrating lymphocytes; pTIL, peritumoral tumor infiltrating lymphocytes; HR, hazard ratio; CI, confidence interval. Statistically significant (P < 0.05).

3.3 Construction and Validation of the OS Nomogram
The same method was applied to screen significant variables to build the model and validate the model for the prediction of OS. The final OS model contained age, smoking history, STAS, tissue subtype, sTIL, and nuclear grades (Table 3), and was presented as a nomogram. The C-index of the model was 0.83 (95%CI: 0.76–0.91) in the training cohort and 0.78(95%CI: 0.68–0.89) in the validation cohort. The majority of the calibration curves followed the diagonal line, which indicated reliable risk estimates. For the OS model, the AUCs were 0.83, 0.86, and 0.88 at 3-, 5-, and 7-year respectively, which showed a better prediction accuracy (Fig. 3,4).
Table 3
Univariate and multivariate cox regression analysis of factors associated with overall survival.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Univariate analysis</th>
<th></th>
<th></th>
<th></th>
<th>Multivariate analysis</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>95%CI</td>
<td>p</td>
<td>HR</td>
<td>95%CI</td>
<td>p</td>
<td>HR</td>
<td>95%CI</td>
</tr>
<tr>
<td>Age</td>
<td>1.033</td>
<td>0.992–1.076</td>
<td>0.1200</td>
<td>1.047</td>
<td>1.002–1.095</td>
<td>0.0398</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking history (Yes VS No)</td>
<td>2.300</td>
<td>1.240–4.267</td>
<td>0.0082</td>
<td>2.617</td>
<td>1.376–4.975</td>
<td>0.0034</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage (IB VS IA)</td>
<td>1.541</td>
<td>0.735–3.229</td>
<td>0.2522</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgical resection (Sublobar VS Lobar)</td>
<td>1.518</td>
<td>0.697–3.307</td>
<td>0.2938</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor size</td>
<td>1.480</td>
<td>1.063–2.062</td>
<td>0.0204</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visceral pleural invasion (Absent VS Present)</td>
<td>0.715</td>
<td>0.370–1.383</td>
<td>0.3193</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAS (Yes VS No)</td>
<td>5.248</td>
<td>2.507–10.984</td>
<td>&lt; .0001</td>
<td>2.638</td>
<td>1.141–6.100</td>
<td>0.0233</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tissue subtype (High-grade patterns ≥ 20% VS &lt; 20%)</td>
<td>4.210</td>
<td>2.150–8.247</td>
<td>&lt; .0001</td>
<td>2.314</td>
<td>1.059–5.054</td>
<td>0.0353</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSR (≥ 50% VS &lt; 50%)</td>
<td>0.551</td>
<td>0.295–1.030</td>
<td>0.0618</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sTIL</td>
<td>0.980</td>
<td>0.967–0.993</td>
<td>0.0035</td>
<td>0.983</td>
<td>0.971–0.996</td>
<td>0.0087</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pTIL</td>
<td>1.004</td>
<td>0.992–1.016</td>
<td>0.5597</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuclear grades (Grade 2VS 1)</td>
<td>5.625</td>
<td>1.293–24.468</td>
<td>0.0166</td>
<td>2.900</td>
<td>0.633–13.287</td>
<td>0.1704</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuclear grades (Grade 3VS 1)</td>
<td>13.805</td>
<td>3.258–58.500</td>
<td>0.0004</td>
<td>5.081</td>
<td>1.091–23.655</td>
<td>0.0383</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

STAS, spread through air space; TSR, Tumor stromal ratio; sTIL, stromal tumor infiltrating lymphocytes; pTIL, peritumoral tumor infiltrating lymphocytes; HR, hazard ratio; CI, confidence interval. Statistically significant (P < 0.05).

3.4 Clinical significance of the prognostic risk score
Based on the coefficients and the level of variables, the prognostic risk score of each lung adenocarcinoma patient was calculated. The prognostic risk score of each patient was calculated according to the following formula, \( DFS\text{-score} = \text{Age} \times 0.05284 + \text{Smoking history} (\text{yes}) \times 0.98345 + \text{STAS (yes)} \times 0.87929 + sTIL \times 0.01760 + \text{Tissue subtype (high-grade patterns} \geq 20\%) \times 0.75327 + \text{Tumor size} \times 0.43494 + \text{Nuclear grades} (\text{Grade} 2) \times 1.06172 \text{ or (Grade}3) \times 1.66893; \text{OS-score} = \text{Age} \times 0.04627 + \text{Smoking history} (\text{yes}) \times 0.96185 + \text{STAS (yes)} \times 0.97015 + sTIL \times 0.01705 + \text{Tissue subtype} (\text{high-grade patterns} \geq 20\%) \times 0.83893 + \text{Nuclear grades} (\text{Grade} 2) \times 1.06471 \text{ or (Grade}3) \times 1.62548, \) We took the AUC as a measurement of the score's predictive value (Supplement Table 1–2). For DFS, the AUCs of our score in the training set at 3, 5, 7 years were 0.868, 0.886, and 0.869. For OS, it was 0.830, 0.862, and 0.876. ROC analyses showed that the score was better than other clinical variables (Supplement Fig. 1,2). Finally, the patients were divided into high-risk and low-risk subgroups based on the median of the score (DFS = 5.46457; OS = 4.05548). The KM curves showed a significant difference in survival between the high- and low-risk subgroups for DFS and OS in the training cohort and the validation cohort (Fig. 5,6).

### 3.5 Association between STAS and clinicopathological parameters

In this cohort, STAS was observed in 143 patients (45.8%). The relationship between STAS and clinicopathological parameters is summarized in Table 4. STAS is more common in lung adenocarcinomas with more aggressive behavior, such as larger tumor size \((P = 0.0016)\), higher-grade patterns \((P < 0.0001)\), and higher nuclear grade \((P < 0.0001)\). In addition, it was significantly associated with tumor-infiltrating lymphocytes, especially peritumoral tumor-infiltrating lymphocytes \((P < 0.0001)\).
Table 4  
Correlations between clinicopathological parameters and STAS

<table>
<thead>
<tr>
<th>Variables</th>
<th>STAS (-)</th>
<th>STAS (+)</th>
<th>Statistics</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
<td>STAS (-) n = 169</td>
<td>STAS (+) n = 143</td>
<td>Statistics</td>
<td>p</td>
</tr>
<tr>
<td>Age</td>
<td>57.78 ± 8.27</td>
<td>58.99 ± 7.56</td>
<td>-1.341</td>
<td>0.1809</td>
</tr>
<tr>
<td>Smoking history</td>
<td></td>
<td></td>
<td>0.6201</td>
<td>0.4310</td>
</tr>
<tr>
<td>No</td>
<td>102(60.4)</td>
<td>80(55.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>67(39.6)</td>
<td>63(44.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td>1.5756</td>
<td>0.2094</td>
</tr>
<tr>
<td>IA</td>
<td>148(87.6)</td>
<td>118(82.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IB</td>
<td>21(12.4)</td>
<td>25(17.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgical resection</td>
<td></td>
<td></td>
<td>0.6822</td>
<td>0.4088</td>
</tr>
<tr>
<td>Lobar</td>
<td>137(81.1)</td>
<td>121(84.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sublobar</td>
<td>32(18.9)</td>
<td>22(15.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor size</td>
<td>2.00(1.50-3.00)</td>
<td>2.40(2.00–3.00)</td>
<td>3.1597</td>
<td>0.0016</td>
</tr>
<tr>
<td>Visceral pleural invasion</td>
<td></td>
<td></td>
<td>3.6910</td>
<td>0.0547</td>
</tr>
<tr>
<td>Present</td>
<td>38(22.5)</td>
<td>46(32.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>131(77.5)</td>
<td>97(67.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tissue subtype</td>
<td></td>
<td></td>
<td>80.444</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>High-grade patterns &lt; 20%</td>
<td>137(81.1)</td>
<td>44(30.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-grade patterns ≥ 20%</td>
<td>32(18.9)</td>
<td>99(69.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSR</td>
<td></td>
<td></td>
<td>3.4532</td>
<td>0.0631</td>
</tr>
<tr>
<td>&lt; 50%</td>
<td>72(42.6)</td>
<td>76(53.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 50%</td>
<td>97(57.4)</td>
<td>67(46.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sTIL</td>
<td>40(20–60)</td>
<td>30(10–60)</td>
<td>-2.255</td>
<td>0.0241</td>
</tr>
<tr>
<td>pTIL</td>
<td>10(5–30)</td>
<td>30(10–60)</td>
<td>5.2342</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Nuclear grades</td>
<td></td>
<td></td>
<td>59.490</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Grade 1</td>
<td>78(46.2)</td>
<td>14(9.79)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 2</td>
<td>65(38.5)</td>
<td>64(44.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 3</td>
<td>26(15.4)</td>
<td>65(45.5)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.6 STAS and immunohistochemical expression

The association between STAS and immunohistochemical expression is summarized in Fig. 7. The expression of CXCL8, Smad2, and Snail was higher in the STAS-positive group, and the expression of CXCL8 was correlated with the expression of Smad2, and Snail. Survival analysis showed that the STAS-positive group exhibited worse DFS and OS. High expression of CXCL8 predicts poor prognosis (DFS, \( P = 0.0055 \); OS, \( P = 0.0118 \)), high expression of Smad2 is associated with shorter DFS (\( P = 0.0253 \)) and no statistical significance with OS, while snail expression has little effect on survival.

4. Discussion

Although the research on the diagnosis and treatment of stage I lung adenocarcinoma has gradually deepened in recent years, no major breakthrough has been made in effective biomarkers and treatment methods. Surgical resection was the preferred treatment for stage I lung adenocarcinoma, but the prognosis remained unresolved, and some patients experienced recurrence and metastasis within a short period of time (Schuchert et al. 2019). Based on clinical features and content of traditional postoperative pathology reports such as tumor size, tissue subtype, and pleural invasion, it is difficult to detect and alert patients with poor prognosis. This study was the first to comprehensively analyze the pathological characteristics of tumor parenchyma and stroma, and aimed to demonstrate and clarify more clinicopathological factors that could be observed by pathologists in HE sections that indicated poor prognosis, so as to enrich and improve the content of traditional lung adenocarcinoma pathology report, and to establish a prognostic model that could be beneficial to clinical risk stratification of patients with stage I lung adenocarcinoma.

In this research, we initially included factors from traditional pathology reports such as age, smoking history, surgical resection, clinical stage, tumor size, tissue subtype, and pleural invasion, and newly added factors that could more fully reflect the characteristics of tumor parenchyma and stroma, such as STAS, TSR, sTIL, pTIL, and nuclear grade. After univariate and multivariate Cox analysis, we finally identified independent variables affecting prognosis in stage I lung adenocarcinoma, including age, smoking history, tissue subtype, tumor size, STAS, sTIL, and nuclear grade to construct DFS and OS prognostic models, respectively, in which tumor size has less influence on OS and was excluded from the OS nomogram. The results of the C-index for DFS were (training set 0.84 vs validation set 0.77) and for OS was (training set 0.83 vs validation set 0.78) and ROC analysis with the AUC above 0.8 showed that both DFS and OS models we constructed exhibited better survival prediction ability. We also assessed the calibration curve that showed the predicted probabilities of the model are in good agreement with the actual observed probabilities, which also ensured repeatability. Through DCA measurement, the net benefit of the model incorporating STAS, sTIL, and nuclear levels was significantly higher than traditional reports, indicating the clinical application value of the model. In addition, we established the formula for calculating DFS and OS prognostic risk scores according to the different weights of risk factors, and ROC analysis verified that the score performance of this formula was better than that of clinical single variables. By this prognostic risk score formula, we further divided patients into high and low-risk groups with the median risk score as the cutoff point. Kaplan-Meier survival analysis showed a significant
survival difference between the two risk groups. All results were subsequently validated by the validation cohort. The establishment and validation of prognostic models and risk score formulas are of great value for the overall diagnosis and clinical treatment of stage I lung adenocarcinoma. For pathologists, we should pay attention to more factors in the process of diagnosis, so that the routine report can be supplemented and improved. For clinicians, it could more accurately identify intermediate and high-risk groups and be used to guide the need for adjuvant therapy and close clinical follow-up.

The model and formula we established verified that STAS has a larger weight in the prognosis of stage I lung adenocarcinoma. STAS has emerged as a new invasive modality for lung adenocarcinoma and was significantly associated with poorer recurrence-free and overall survival (Kadota et al. 2015; Masai et al. 2017; Gutierrez et al. 2022). Adenocarcinomas with the micropapillary subtype could also be observed in breast and ovarian cancers and the high invasiveness of the clustered cancer nest was prone to vascular and lymphatic vessel invasion causing metastasis to lymph nodes and distant organs (Eren et al. 2022; Fauvet et al. 2012). In lung adenocarcinoma, we could observe that the STAS phenomenon often appeared around the cancer tissue containing micropapillary components and solid components, often as a group of micropapillary or solid cancer cells free in the alveolar cavity or growing in the alveolar wall surface. In our analysis, STAS was associated with micropapillary and solid components, consistent with the previous conclusions (Gutierrez et al. 2022; Kadota et al. 2019). In addition, our study found for the first time that interstitial lymphocytic infiltration and peritumoral lymphocytic infiltrating cells were related to STAS. Tumor-infiltrating lymphocytes are critical for prognosis, Kim et al. reported that high levels of sTIL were associated with better DFS and OS in non-small cell lung cancer (Kim et al. 2019). Understanding the prognostic significance and the mechanisms of TIL recruitment into tumors and its relationship with STAS has important implications for the development of effective immunotherapies, which deserve further investigation. Furthermore, STAS was an independent prognostic risk factor for DFS and OS in this study. So far, many studies have investigated the formation mechanism of STAS, scholars support that EMT was involved in the development of STAS. Liu et al. supported that STAS was more easily observed in lung adenocarcinoma with high metastasis-associated protein 1 (MTA1) expression levels. MTA1 was considered to be an EMT-related protein that promoted tumor aggressiveness, stemness, and cisplatin resistance (Liu et al. 2018). Liu et al. reported that STAS was associated with overexpressed Twist and Slug and showed a worse prognosis (Liu et al. 2020). This might provide a potential mechanism for STAS. Our previous study accomplished a comprehensive analysis of a microarray dataset of STAS that reported high expression of CXCL8 in STAS-positive samples, but its relationship with EMT was unclear.

Previous studies have shown that CXCL8 promotes tumor progression in lung adenocarcinoma and is a poor prognostic factor (Liu et al. 2018). CXCL8 and its receptors CXCR1/CXCR2 play important roles in tumor development, CXCL8-CXCR1/2 Axis can activate multiple signaling pathways, such as the phosphoinositide-3 kinase/Akt (PI3K/Akt), mitogen-activated protein kinase / extracellular signal-regulated kinase (MAPK /ERK), phospholipase C /protein kinase C(PLC/PKC), Janus kinases/activator of transcription protein 3 (JAK/STAT3) signaling to promote tumor proliferation, tumor cell invasion and migration, maintain stemness, and anti-tumor immunity(Waugh et al. 2008; Cheng et al. 2019; Alassaf et
EMT was thought to be one of the mechanisms by which tumor cells acquire invasive and metastatic functions. Previous studies have shown that CXCL8 could interact with Snail to induce EMT to facilitate invasion and metastasis in colon cancer (Hwang et al. 2011). In addition, CXCL8-CXCR2 promoted EMT in oral cancer through the TGF-β signaling pathway (Meng et al. 2020). TGF-β/SMADs signaling was a classical pathway leading to EMT. Smad2 and Snail were EMT-related proteins. The results of this study showed that the CXCL8 expression was correlated with Smad2 and Snail. Furthermore, STAS was more likely to be observed with high expression of CXCL8, Smad2, and Snail. This may provide a relevant mechanism for the formation of STAS, whether CXCL8 interacts with EMT-related pathways to endow STAS with higher invasive and metastatic behavior, which is the direction we need to study further. On the other hand, CXCL8 was also a potential biomarker whose overexpression predicted a poor prognosis (Yang et al. 2020; Jia et al. 2018). Our findings suggested that CXCL8 overexpression was associated with worse DFS and OS. We concluded that CXCL8, Smad2, and Snail were related to the frequency of STAS. CXCL8 might serve as a potential biomarker for STAS formation and poor prognosis in lung adenocarcinoma.

The current study has several limitations. Firstly, the nomogram obtained in this study was only applicable to stage I lung adenocarcinoma. Besides, more indicators could be expected to be added in the future for a comprehensive evaluation, such as molecular indicators, etc.

In conclusion, we proposed and validated the nomogram and the prognostic risk score formulas for predicting the DFS and OS of stage I lung adenocarcinoma mainly based on pathological features. As a pathologist, we try our best to give clinical indicators that reflect the prognosis in H&E-stained sections, which is helpful for the clinical evaluation of the prognosis of stage I lung adenocarcinoma. It also can be easily applied in practice and incorporated into routine pathological diagnosis. We risk-scored factors strongly associated with prognosis, enabling efficient patient stratification. Finally, we found that STAS was significantly associated with stronger invasive behavior. In addition, highly expressed CXCL8, Smad2, and Snail were associated with STAS. Most importantly, CXCL8 could serve as a potential biomarker for STAS formation and poor prognosis and a routine immunohistochemical detection index for stage I lung adenocarcinoma, and the mechanism may be related to EMT.

Declarations

Acknowledgments

The authors would like to thank everyone who participated in this study. This study was supported by the Haiyan Foundation of Harbin Medical University Cancer Hospital (JJZD2020-14).

Authors’ Contributions

All authors contributed to the study’s conception and design. YYN, YZ, and XHH collected and analyzed the data in the study. YYN, YZ, XHH, NA, QB, SNG, YLH, YY, QJZ and XML discussed the results and contributed to the final manuscript. YYN, YZ, and XHH drafted the manuscript. YY, QJZ and XML
conceived of the presented idea and completed the critical revision of the article. All authors have read and approved the final manuscript.

**Conflicts of Interest** The authors declare that they have no potential conflicts of interest.

**References**


Figures
Figure 1

Prognostic nomogram for predicting 3-, 5-, and 7-year disease-free survival probability of stage I lung adenocarcinoma patients (A). The calibration curves for predicting disease-free survival probabilities (B1) at 3-, 5-, and 7 years in the training cohort, at 3-, 5-, and 7 years (B2) in the validation cohort. ROC analysis of nomogram to predict the 3-, 5-, and 7 years DFS rate (C1) in the training cohort, 3-, 5-, and 7 years DFS
rate (C2) in the validation cohort. STAS, spread through air space, HP, high-grade patterns, sTIL, stromal tumor-infiltrating lymphocytes.

Figure 2

Decision curve analysis of the nomogram for 3- (A1), 5- (A2), and 7-year (A3) disease-free survival probability of stage I lung adenocarcinoma patients in the training cohort, and 3-(B1), 5-(B2), and 7 years (B3) in the validation cohort. STAS, spread through air space, sTIL, stromal tumor-infiltrating lymphocytes.
Figure 3

Prognostic nomogram for predicting 3-, 5-, and 7-year overall survival probability of stage I lung adenocarcinoma patients (A). The calibration curves for predicting overall survival probabilities at 3-, 5-, and 7 years (B1) in the training cohort, at 3-, 5-, and 7 years (B2) in the validation cohort. ROC analysis of nomogram to predict the 3-, 5-, and 7 years DFS rate (C1) in the training cohort, 3-, 5-, and 7 years DFS
rate (C2) in the validation cohort. STAS, spread through air space, HP, high-grade patterns, sTIL, stromal tumor-infiltrating lymphocytes.

**Figure 4**

Decision curve analysis of the nomogram for 3- (A1), 5- (A2), and 7-year (A3) overall survival probability of stage I lung adenocarcinoma patients in the training cohort, and 3-(B1), 5-(B2), and 7 years (B3) in the validation cohort. STAS, spread through air space, sTIL, stromal tumor-infiltrating lymphocytes.
Figure 5

Kaplan–Meier curves of high-score and low-score subgroups of stage I adenocarcinoma patients for disease-free survival in the training cohort (A1) and in the validation cohort (A2).

Figure 6

Kaplan–Meier curves of high-score and low-score subgroups of stage I adenocarcinoma patients for overall survival in the training cohort (A1) and in the validation cohort (A2).
Figure 7

STAS (H&E staining, 100×, 400×) and representative images of CXCL8, Smad2, Snail (IHC staining, 100×, 400×) expression in stage I adenocarcinoma(A). KM plotter survival analysis based on STAS and the expression of CXCL8, Smad2, and Snail(B). The expression of CXCL8, Smad2, and Snail was in the STAS-positive group and STAS-negative group, and the relationship between CXCL8 expression and Smad2, Snail expression (C).
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

* SupplementFig.docx
* SupplementTable1.docx
* SupplementTable2.docx