Comprehensive molecular characterizations of stage I–III lung adenocarcinoma with tumor spread through air spaces

Ronghao Ye[1], Yongfeng Yu[1], Ruiying Zhao[2], Yuchen Han[2] and Shun Lu[1]*

[1]Shanghai Lung Cancer Center, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China,
[2]Department of Pathology, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China

Purpose: The aim of this study is to investigate integrative genomic spectra of stage I–III lung adenocarcinoma with tumor spread through air spaces (STAS).

Methods: We retrospectively identified 442 surgically resected lung adenocarcinoma patients of pathological stage I–III in Shanghai Chest Hospital from January 2018 to February 2021. Surgically resected tissues were used for next-generation sequencing (NGS) with a panel of 68 lung cancer-related genes to profile comprehensive molecular characterizations.

Results: A total of 442 cases were analyzed, including 221 (50%) STAS-positive (SP) and 221 (50%) STAS-negative (SN) lung adenocarcinoma patients. In total, 440 cases (99.6%) were positive for the overall mutational spectrum, and the higher mutational genes were EGFR, TP53, KRAS, ALK, SMAD4, and ERBB2 (62%, 42%, 14%, 10%, 7%, and 7%, respectively). Compared with the SN population, there was significantly lower EGFR alteration in the single-nucleotide variant (SNV) mutation spectrum (52.5% vs 69.7%, $p < 0.001$) and significantly higher TP53 alteration in the SP population (49.8% vs 34.8%, $p = 0.002$). EGFR L858R missense mutation (19.5% vs 37.6%, $p < 0.001$) and ERBB2 exon 20 indel mutation (1.8% vs 5.9%, $p = 0.045$) were more frequent in the SN population. The detection rate of ALK fusion rearrangements in the SP population was significantly higher than that in the SN population (13.1% vs 2.3%, $p < 0.001$). In the analysis of signaling pathways, no significant difference was discovered between SP and SN patients. No difference in 1-year disease-free survival was observed between SP and SN patients in this study.

Conclusion: Significant differences exist in stage I–III lung adenocarcinoma patients with STAS in molecular characterizations.

Introduction

Lung cancer, a common solid tumor, seriously threatens people's mental and psychological health. It was demonstrated in the 2020 WHO International Agency for Research on Cancer (IARC) global cancer data report that the incidence of lung cancer is approximately 2.2 million per year and mortality is approximately 1.8 million per year (Kordiak et al., 2022). Notably, non-small-cell lung cancer (NSCLC) accounts for 85% of all lung cancers, and one of the major pathological subtypes of NSCLC is lung adenocarcinoma (Yang et al., 2020). Despite the

development of imaging techniques that can detect tumors at an early stage, lung adenocarcinoma is associated with a mounting mortality rate due to the poor prognosis and high recurrence rate, which could be explained by tumor invasion and metastasis. Spread through air spaces (STAS), a recently discovered form of infiltrates raised by the WHO in 2015, has already received much attention in the field of lung cancer (Ma et al., 2019).

STAS is defined as tumor cell propagation through the alveolar space into the lung tissue beyond the margins of the tumor. Tumor cells usually form microbial structures, solid tumor cell islands, or the spread of single tumor cells (Han et al., 2021). It has been estimated that approximately 15%–60% of lung cancer patients are STAS-positive (SP) (Eguchi et al., 2019; Tian et al., 2021), and the SP population is more frequently seen in late-stage lung cancer patients. Moreover, SP lung cancer patients are associated with low recurrence-free survival and overall survival (Toyokawa et al., 2018a; Han et al., 2021; Ikeda et al., 2021). Kadota et al. (2015) discovered that the cumulative recurrence rate of distal and local recurrence in SP patients was significantly higher than that in STAS-negative (SN) patients. Several retrospective studies have found that STAS can occur in nearly all pathological types of lung cancer (Lu et al., 2017; Yokoyama et al., 2018; Aly et al., 2019). Among them, STAS is most frequently observed in lung adenocarcinoma (ADC). It has been demonstrated that STAS is an independent prognostic factor for poor tumor-related outcomes and is associated with early-stage ADC sublobar resection (LR, wedge-shaped pulmonary resection, and segmentectomy) (Kadota et al., 2015; Toyokawa et al., 2018b; Shiono et al., 2018).

Currently, the correlation between STAS and the molecular characteristics of lung cancer is rarely reported, and thus its mechanism is still unclear (Ikeda et al., 2021). A few studies have shown high expression of vimentin (Upton et al., 1986; Karim et al., 2017; Jia et al., 2020) and low expression of E-cadherin (Jia et al., 2020) in SP patients. This indicates a potential relationship between STAS and the biomarkers of epithelial–mesenchymal transition (EMT), which is important in tumor cell migration and invasion. Liu et al. (2018) found that the elevated expression of metastasis-related protein 1 (MTA1) is correlated with STAS, which indicated the important effect of STAS on lung cancer metastasis and poor prognosis. In addition, some studies have shown that STAS occurs more frequently in tumors with ROS1 and ALK rearrangements (Kadota et al., 2019; Tian et al., 2021; Pyo and Kim, 2022). However, interestingly, some high-profile mutations, including BRAF, EGFR, and KRAS, have not been definitively linked to STAS status as they have presented mixed results in several studies (Toyokawa et al., 2018b; Jia et al., 2020). Other genes (HER2, PD-L1, TTF1, Napsin, and CK7) were also investigated, but no statistical significance was presented (Warth et al., 2015; Toyokawa et al., 2018b; Toyokawa et al., 2018c; Hu et al., 2018).

In summary, the pathogenesis of STAS is incompletely investigated, and further molecular biological characterization studies are urgently needed to update current understanding. In this study, next-generation sequencing (NGS) was conducted on the pathological tissues of SP and SN lung adenocarcinoma patients grouped by propensity score matching (PSM), and 68 genes related to lung cancer targeted therapies were analyzed to reveal the mutation spectrum and biological associations with STAS.

Methods

Data collection

The study protocol was evaluated and approved by the Shanghai Chest Hospital Institutional Review Board. Between 1 January 2018 and 28 February 2021, we identified patients who underwent surgical resection and tested for 68 targeted lung cancer therapy-related genes. The clinical characteristics of patients, including age, gender, smoking history, P-stage, and type of surgery, were searched and collected from the original medical records. Based on the eighth edition of the AJCC/UICC staging system, the P-stage was restaged (Jia et al., 2020). Follow-ups ranged from 1 to 12 months. There were 240 SP lung cancer patients, including two large cell carcinomas, five squamous cell carcinomas, six small cell carcinomas, and 227 lung adenocarcinomas. The analysis was narrowed to ADC subjects due to insufficient cases of other pathological types of lung cancer except for ADC. A total of 227 ADC patients consisted of 97 stage I subjects, 43 stage II subjects, 81 stage III subjects, and six stage IV subjects. Owing to the small sample size of stage IV ADC and palliative surgery conducted in these patients, a total of 221 SP ADC patients (stage I: 97, stage II: 43, and stage III: 81) and 2027 SN patients (stage I: 1467, stage II: 288, and stage III: 272) were included. Propensity score matching (PSM) (match ratio 1:1 and caliper value 0.03) was used to exclude other confounding factors, such as STAS status, including tumor stage, surgical mode, gender, age, and smoking history. The SP population was matched with the SN population (stage I: 97, stage II: 43, and stage III: 81). Tumor tissue samples from patients were obtained and reviewed by two qualified pathologists independently. Any disagreement was resolved by discussion. The presence or absence of STAS was reported following the unanimous conclusion of the two pathologists. Cases were classified based on whether STAS was present or not. DNA was extracted for the targeted sequencing of 68 genes. For both SP and SN groups, genetic mutation profiles and biological associations were analyzed.

Targeted tumor next-generation sequencing and analysis

Formalin-fixed paraffin-embedded (FFPE) sections of surgically resected tissues were collected for DNA extraction. The tissue DNA was extracted using a QIAamp DNA FFPE Tissue Kit (Qiagen, Hilden, Germany) and subsequently sheared with Covaris M220 (Covaris, MA, United States) for end-repair, phosphorylation, and adapter connection. DNA quality and size were measured using a Qubit 2.0 Fluorometer and double-stranded DNA high-sensitivity detection kit (Life Technologies, Carlsbad, CA, United States). No less than 50 ng of tissue DNA was applied for NGS library construction and target capture using a panel of 68 lung cancer-related genes (spanning 245 kb of the human genome) (Lung Core, Burning Rock Biotech, Guangzhou, China) (Supplementary Table S1). Indexed samples were paired-end sequenced using a NextSeq500 sequencer (Illumina, Inc., Madison, WI, United States) with 1,000-fold paired-end reads and target sequencing depth.

Sequence data were mapped to the reference human genome (hg19) using the Burrows–Wheeler Aligner v.0.7.10 platform (Li and Durbin, 2009). Genome Analysis Toolkit v.3.2 (McKenna et al., 2010) and VarScan (Koboldt et al., 2009) were applied for

TABLE 1 Clinical characteristic baseline before and after PSM.

Characteristic	Entire cohort (n = 2248)			Matched cohort (n = 442)		
	STAS-negative (%)	STAS-positive (%)	*p*-value	STAS-negative (%)	STAS-positive (%)	*p*-value
Gender			0.009			1
Female	1,209 (59.6)	111 (50.2)		111 (50.2)	111 (50.2)	
Male	818 (40.4)	110 (49.8)		110 (49.8)	110 (49.8)	
Age			0.017			0.846
≤60	1,021 (50.4)	92 (41.6)		88 (39.8)	91 (41.2)	
>60	1,006 (49.6)	129 (58.4)		133 (60.2)	130 (58.8)	
Smoking status			<0.001			0.896
Non-smoker	1851 (91.3)	185 (83.7)		187 (84.6)	185 (83.7)	
Smoker	176 (8.7)	36 (16.3)		34 (15.4)	36 (16.3)	
Resection			<0.001			1
Pneumonectomy/lobectomy	1,158 (57.1)	178 (80.5)		178 (80.5)	178 (80.5)	
Segmental/wedge	869 (42.9)	43 (19.5)		43 (19.5)	43 (19.5)	
Stage			<0.001			1
I	1,467 (72.4)	97 (43.9)		97 (43.9)	97 (43.9)	
II	288 (14.2)	43 (19.5)		43 (19.5)	43 (19.5)	
III	272 (13.4)	81 (36.7)		81 (36.7)	81 (36.7)	

PSM, propensity score matching; STAS, spread through air spaces.

local alignment optimization, variant calling, and annotation. Sites with depths below 100 were excluded. Indels require at least five backward reads and SNV calls require at least eight backward reads. Variants with a population frequency greater than 0.1% were classified as SNPs based on the ExAC 1,000 Genomes, dbSNP, and ESP6500SI-V2 databases and were not further analyzed. ANNOVAR (Wang et al., 2010) and SnpEff v3.6 (Yen et al., 2017) were used for the remaining variants. DNA translocation analyses were conducted using TopHat2 (Falch et al., 2018) and FACTERA 1.4.3 (Newman et al., 2014).

Statistical analysis

All statistical analyses were conducted using R software (version 4.0.3, , Austria). Propensity score matching (PSM) was used to control for the confounding effects of the clinical parameters. Propensity scores for all lung adenocarcinoma patients were calculated using multivariate logistic regression with the following covariates: age, sex, smoking history, TNM stage, and surgery protocol. In the matched cohort, 221 SP patients were matched 1:1 with 221 SN patients. The intergroup clinical features were compared using the chi-squared test.

All mutation detection rate differences and mutually exclusive concomitant analyses between STAS+/STAS- are based on Fisher's exact test. Genes were listed when $p < 0.05$ and $p < 0.01$. Genomic maps were clustered using the R package NMF (version 0.23.0) using a non-negative matrix factorization algorithm.

The primary outcome of clinical follow-ups was designed as disease-free survival (DFS), which was defined as the period from surgery to recurrence/cancer-related death. The DFS curves were drawn using Kaplan–Meier analysis with a log-rank test. Moreover, the hazard ratio (HR) and 95% confidence interval (CI) were calculated using Cox proportional hazard regression analysis for

examining clinical characteristics. Statistical significance was set at $p < 0.05$.

Results

Baseline clinical characteristics

PSM was performed separately for age, gender, smoking history, TNM stage, and surgery type to ensure an exact balance between SP and SN patients in the matched cohort. A total of 442 patients were enrolled, including 221 (50%) SP patients and 221 (50%) SN patients, with 49.8% (220/442) males and 50.2% (222/442) females. In total, 58.4% (258/442) of the subjects were aged over 60 years. The vast majority of patients were non-smokers, accounting for 84.6% (374/442), and smoking patients accounted for 15.4% (68/442). Patients undergoing pneumonectomy or lobectomy accounted for 80.5% (356/442) and segmental or wedge resection accounted for 19.5% (86/442). In the cohort, 43.9% (194/442) were diagnosed with stage I, 19.5% (86/442) with stage II, and 36.7% (162/442) with stage III. The baseline characteristics before and after PSM are shown in Table 1.

Characterization of genetic alterations

A total of 442 surgically resected ADC tissues were analyzed, of which 440 samples (99.6%) were positive for the overall mutational spectrum. The top six highest mutational genes were EGFR, TP53, KRAS, ALK, SMAD4, and ERBB2 (62%, 42%, 14%, 10%, 7%, and 7%, respectively). The most frequently detected mutation type was the missense mutation, with detection rates of 78.3% and 79.6% in SP and SN populations, respectively ($p = 0.82$). The SP population presented with a significantly high detection rate of fusion
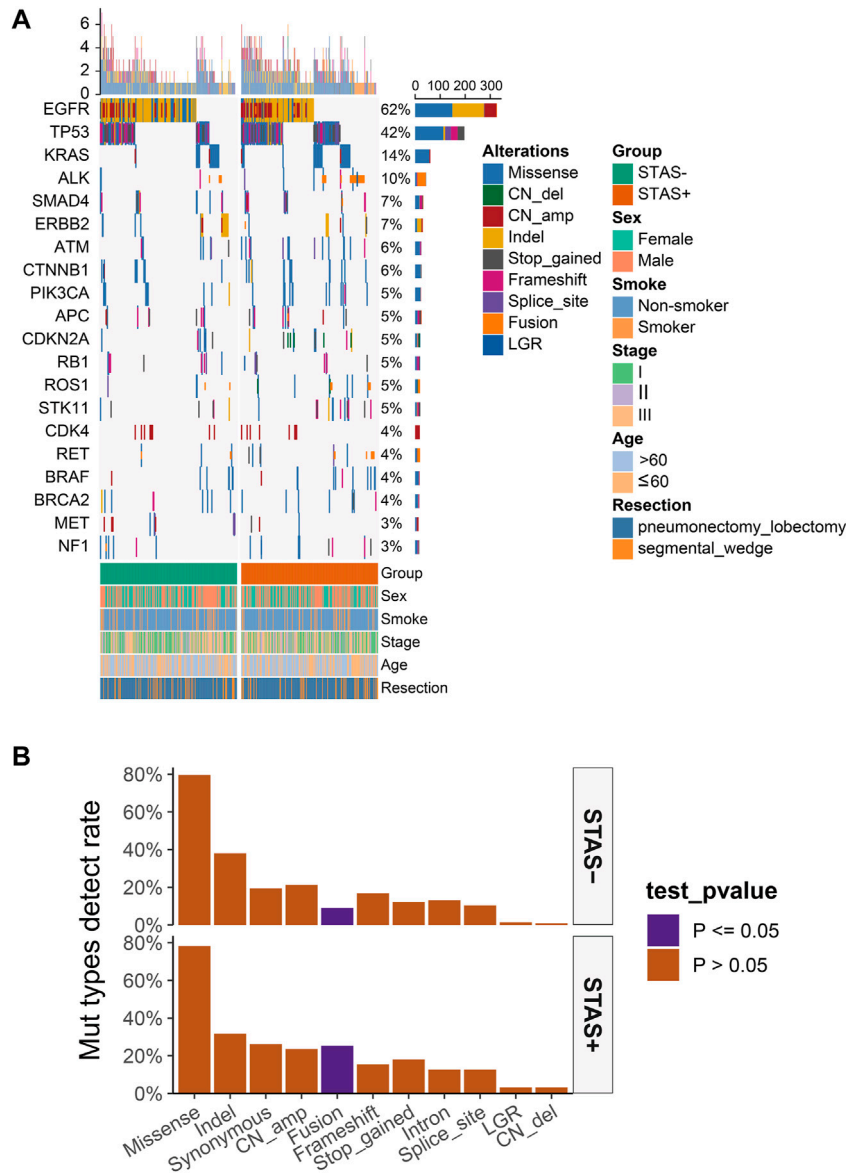
FIGURE 1
(A) Mutational spectrum of lung adenocarcinoma patients grouped according to STAS presence or absence. Gender, smoking status, age, disease stage, age, and type of surgery are also annotated at the bottom of the printout. Each column represents a patient, and each row represents a gene. The left column shows the mutated genes. The right column shows the mutation rate and the number of mutation types for each gene. The top plot represents the total number of mutations carried by the patient. Different colors indicate different types of mutations. **(B)** Overall lung adenocarcinoma patients were grouped according to STAS presence or absence for mutation type detection rates. The X-axis represents the mutation type, and the Y-axis represents the mutation frequency. STAS, propagation through the air space.

rearrangement types compared with the SN population (25.3% vs 9.0%, *p* < 0.05), and the remaining mutation types were not significantly different between the two groups (Figure 1).

Characterization of somatic mutations

The detection rate of EGFR in the SP population in the single-nucleotide variant (SNV) mutation spectrum was significantly lower than that in the SN population (52.5% vs 69.7%, *p* < 0.05). The detection rate of TP53 in the SP population was significantly higher than that in the SN population (49.8% vs 34.8%, *p* < 0.05) (Figures 2A,

C). No significant difference in copy number variation (CNV) of each gene was observed between the SP and SN populations in the CNV mutation profile (Figures 2B, D). The somatic mutations detected above 1% are demonstrated in Supplementary Figure S1.

Characterization of mutational hotspots of driver genes

The detection rate of EGFR L858R missense mutation (19.5% vs 37.6%, *p* < 0.001) and ERBB2 exon 20 indel mutation (1.8% vs 5.9%, *p* < 0.05) was significantly lower in ADC mutational hotspots of driver
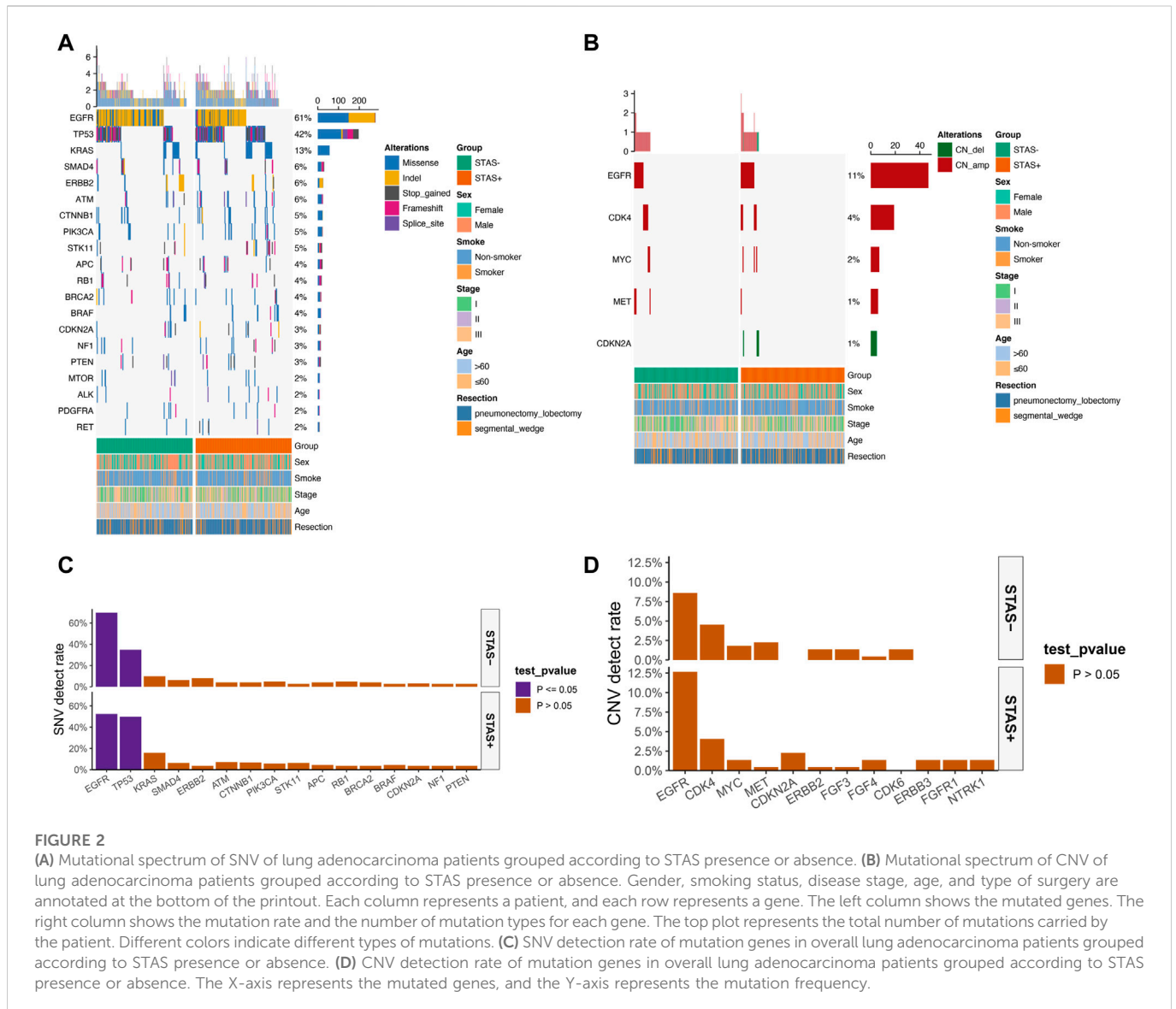
FIGURE 2
(A) Mutational spectrum of SNV of lung adenocarcinoma patients grouped according to STAS presence or absence. **(B)** Mutational spectrum of CNV of lung adenocarcinoma patients grouped according to STAS presence or absence. Gender, smoking status, disease stage, age, and type of surgery are annotated at the bottom of the printout. Each column represents a patient, and each row represents a gene. The left column shows the mutated genes. The right column shows the mutation rate and the number of mutation types for each gene. The top plot represents the total number of mutations carried by the patient. Different colors indicate different types of mutations. **(C)** SNV detection rate of mutation genes in overall lung adenocarcinoma patients grouped according to STAS presence or absence. **(D)** CNV detection rate of mutation genes in overall lung adenocarcinoma patients grouped according to STAS presence or absence. The X-axis represents the mutated genes, and the Y-axis represents the mutation frequency.

genes in the SP population. The detection rate of ALK fusion rearrangements in the SP population was significantly higher than that in the SN population (13.1% vs 2.3%, $p < 0.05$) (Figure 3).

Signaling pathway analysis

To investigate the potential tumor-related signaling pathway differences in SP and SN populations, we calculated the mutation detection rates of genes in some tumor-related signaling pathways. In this part, a significant difference in gene mutation primarily existed in the P53 cell cycle signaling pathway and Wnt and ERBB signaling pathways (Supplementary Figure S2A).

In the P53 cell cycle signaling pathway, we observed significantly more TP53 mutations in SP patients than in SN patients (49.8% vs 34.8%, $p < 0.01$). Moreover, the mutation detection rates of other P53 cell cycle signaling pathway-related genes, including ATM, CDKN2A, and TSC2, were also slightly higher in the SP population than in the SN population (7.2% vs 4.1%, 5.9% vs 3.6%,

and 4.1% vs 1.4%. $p = 0.22$, 0.37, and 0.14, respectively) (Supplementary Figure S2B). For the Wnt signaling pathway, TP53 mutations showed a dominant difference between SP and SN populations (49.8% vs 34.8%, $p < 0.01$). In addition, the mutation detection rates of SMAD4 and CTNNB1 in SP patients were slightly higher than those in SN patients (7.2% vs 6.3% and 6.8% vs 4.5%. $p = 0.85$ and 0.41, respectively) (Supplementary Figure S2C). However, the mutation rate of the ERBB signaling pathway-related genes was lower in the SP population, including EGFR and ERBB2 mutations (52.9% vs 70.1%, $p < 0.01$; 4.5% vs 8.6%, $p = 0.12$) (Supplementary Figure S2D). The detection rates for each gene in the signaling pathway are shown in Supplementary Tables S2A–S2C.

Exclusivity and co-occurrence of mutations and NMF cluster analysis

The results of exclusivity and co-occurrence analyses were examined in this part. In SP subjects, EGFR and driver genes, such
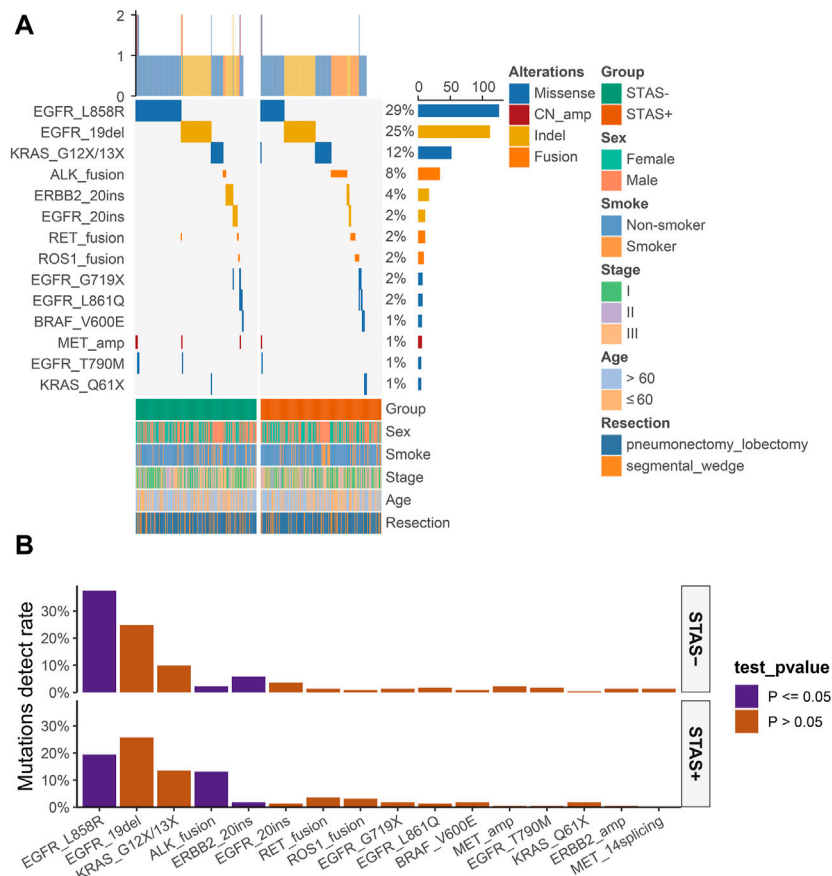
FIGURE 3
(A) Mutational hotspot spectrum of derived genes of lung adenocarcinoma patients grouped according to STAS presence or absence. Gender, smoking status, disease stage, age, and type of surgery are also annotated at the bottom of the printout. Each column represents a patient, and each row represents a gene. The left column shows the mutational hotspots of the derived genes. The right column shows the mutation rate of each gene and the number of mutation types for each gene. The top plot represents the total number of mutations carried by the patient. Different colors indicate different types of mutations. **(B)** Detection rate of mutational hotspots of derived genes in overall lung adenocarcinoma patients grouped according to STAS presence or absence. The X-axis represents the mutation hotspots of the derived genes, and the Y-axis represents the mutation frequency. STAS, transmission through the air space.

as ALK, KRAS, ROS1, and ERBB2, were mutually exclusive; EGFR was previously mutually exclusive with STK11. PIK3CA alterations were substantially correlated with aberrations of CTNNB1 and PTEN; KRAS and STK11 co-alterations occurred frequently (Figure 4A). Among SN subjects, EGFR was mutually exclusive with driver genes, such as ALK, KRAS, ROS1, ERBB2, and BRAF. We detected TP53 and CDKN2A/RB1/APC co-alterations, and SMAD4 alterations correlated with aberrations of CTNNB1. In addition, mutations in KRAS and STK11 co-occurred in the SN population ($p < 0.05$) with less significance than that in SP subjects ($p < 0.001$) (Figure 4B).

The numerical matrix of all samples and their gene mutations was iteratively reduced by non-negative matrix factorization (NMF) and divided into two categories. There were 171 samples in cluster 1, of which 111 (64.9%) were SP. The clustering features showed the occurrence of KRAS, ALK, STK11, ROS1, ATM, and RET mutations. There were 271 samples in cluster 2, of which 161 (59.4%) were SN. The clustering features showed the occurrence of mutations in EGFR, ERBB2, SMAD4, and CTNNB1. Figure 4C presents STAS-positive and -negative subgroups that had similar contributions in the two groups predicted by unsupervised clustering. In addition, a higher

proportion of STAS-positive patients were observed in the two predicted groups. In the NMF results, it was difficult to classify STAS-positive and -negative populations only by mutation spectrum characteristics.

Postoperative follow-up

All patients were regularly followed up every 3 months for a year. One subject died during surgery, and 20 subjects were lost to follow-up. The final follow-up date was 28 February 2022. We tested the correlation between STAS and prognosis in 421 patients with fully resected stage I–III ADC. Additionally, the results showed no significant link between clinical characteristics and prognosis in either group. A total of 421 subjects were included for further prognostic analyses. A total of 41 cases (9.7%) relapsed or died within 1 year, including 22 SP cases (10.4%) and 19 SN cases (9.1%). The association between tumor stage and prognosis in these patients were also investigated [stage I (HR, 0.72, $p = 0.663$), stage II (HR, 2.00, $p = 0.424$), and stage III (HR, 1.14, $p = 0.728$)]. The results are presented in Figure 5A.
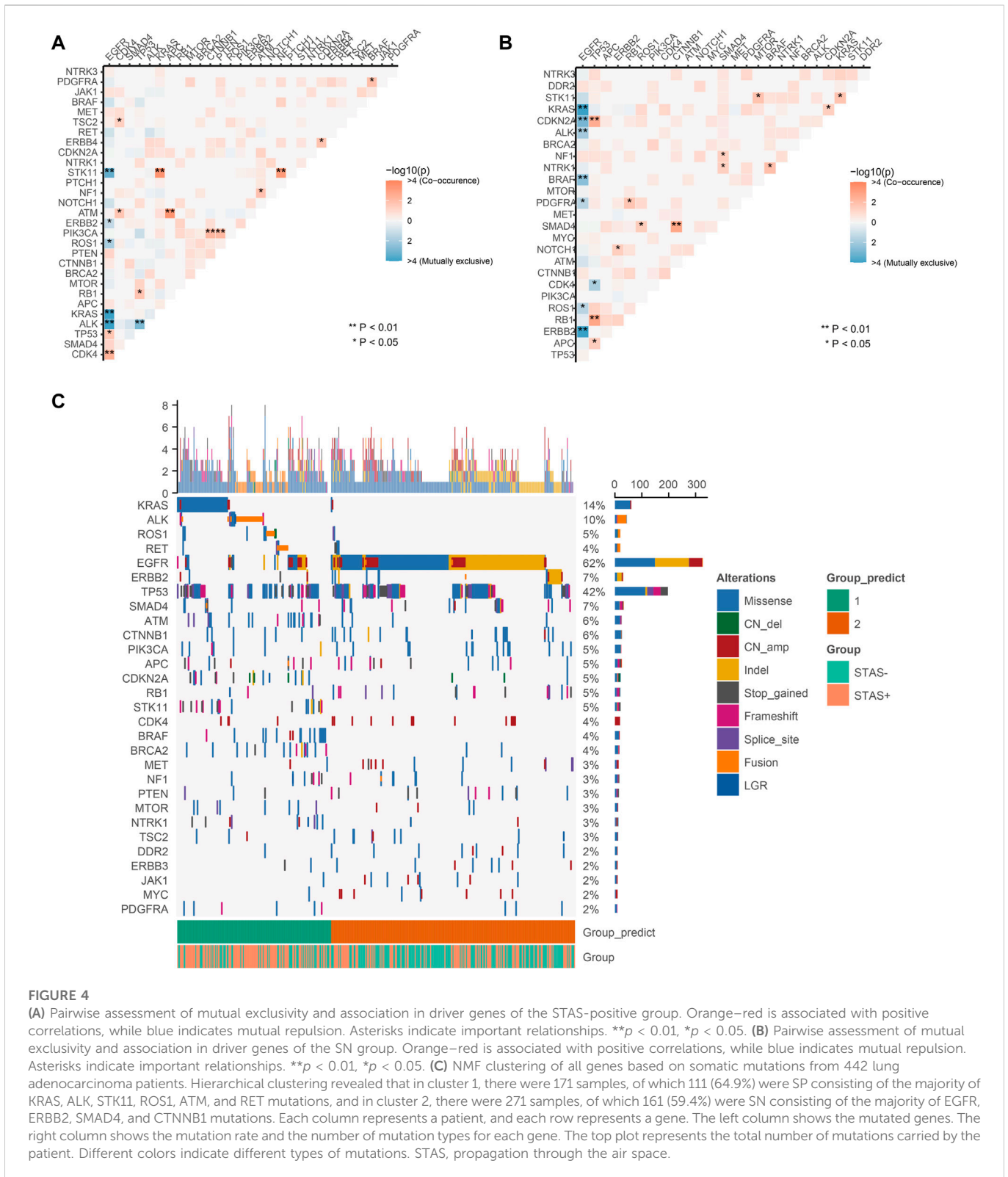
FIGURE 4
(A) Pairwise assessment of mutual exclusivity and association in driver genes of the STAS-positive group. Orange–red is associated with positive correlations, while blue indicates mutual repulsion. Asterisks indicate important relationships. **$p < 0.01$, *$p < 0.05$. **(B)** Pairwise assessment of mutual exclusivity and association in driver genes of the SN group. Orange–red is associated with positive correlations, while blue indicates mutual repulsion. Asterisks indicate important relationships. **$p < 0.01$, *$p < 0.05$. **(C)** NMF clustering of all genes based on somatic mutations from 442 lung adenocarcinoma patients. Hierarchical clustering revealed that in cluster 1, there were 171 samples, of which 111 (64.9%) were SP consisting of the majority of KRAS, ALK, STK11, ROS1, ATM, and RET mutations, and in cluster 2, there were 271 samples, of which 161 (59.4%) were SN consisting of the majority of EGFR, ERBB2, SMAD4, and CTNNB1 mutations. Each column represents a patient, and each row represents a gene. The left column shows the mutated genes. The right column shows the mutation rate and the number of mutation types for each gene. The top plot represents the total number of mutations carried by the patient. Different colors indicate different types of mutations. STAS, propagation through the air space.
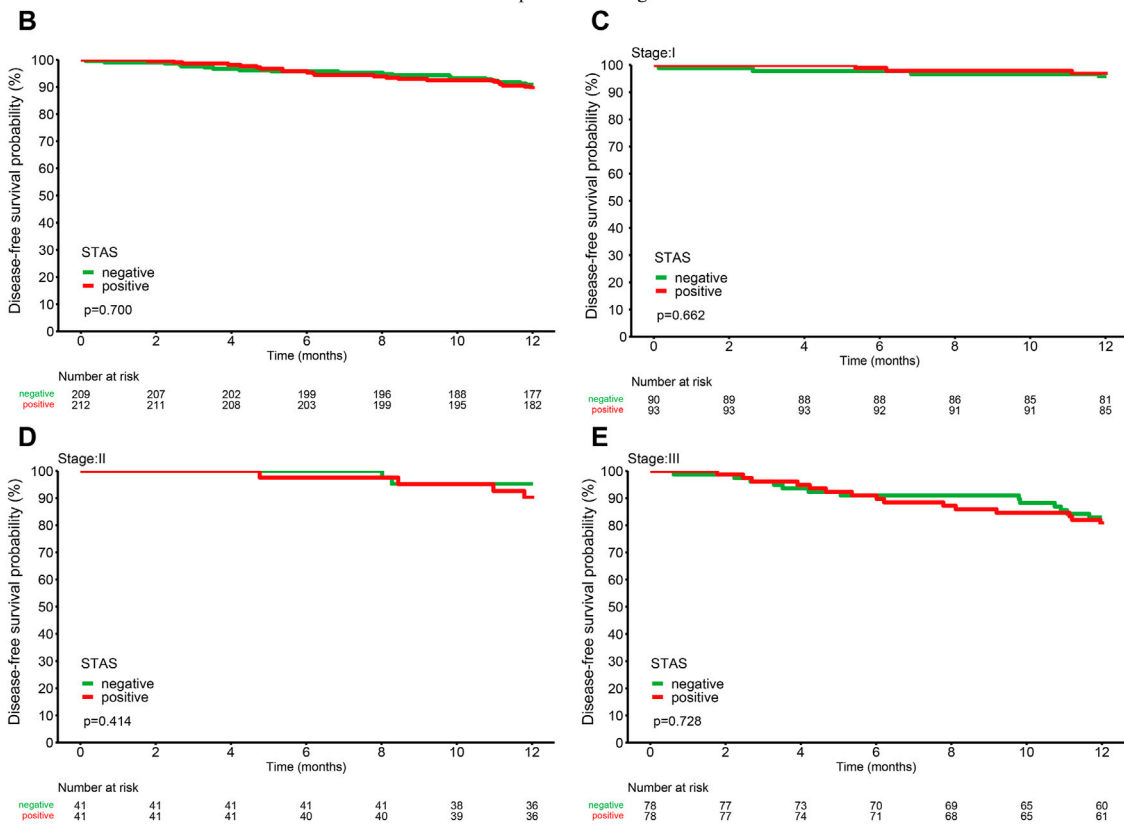
The 1-year DFS in the STAS-positive or -negative group of all ADC patients was 89.6% and 90.7% ($p = 0.7$, Figure 5B), respectively. The 1-year DFS in the STAS$^{+/-}$ subgroup of patients with stage I lung cancer was 96.8% and 95.5% ($p = 0.662$, Figure 5C), the 1-year DFS in the STAS$^{+/-}$ subgroup of patients

with stage II lung cancer was 90.1% and 95.1% ($p = 0.414$, Figure 5D), and the STAS$^{+/-}$ subgroup of patients with stage III lung cancer had a 1-year DFS of 80.7% and 82.9% ($p = 0.728$, Figure 5E), respectively. The 1-year DFS of the SP population did not differ from that of the SN patients in this study.

FIGURE 5
(A) Analysis of the relationship between prognosis and clinical characteristics in STAS-positive and -negative groups. One-year disease-free survival (DFS) for this cohort. **(B)** DFS based on overall patients grouped according to STAS presence or absence. **(C)** DFS based on stage I patients. **(D)** DFS based on stage II patients. **(E)** DFS based on stage III patients. STAS, propagation through the air space; HR, hazard ratio; CI, confidence interval.

Discussion

STAS is known as an indicator of poor prognosis for lung cancer and is associated with a high possibility of early local recurrence and segmentation. However, the molecular characteristics and the mechanisms of association of STAS patients remain incompletely understood. With this in mind, we compared the genetic profiles of

lung adenocarcinoma patients with and without STAS to elucidate the mutation profiles and biological relationships associated with STAS patients.

In this study, we found that the fusion rearrangement mutation detection rate is significantly higher in the SP population (25.3% in STAS$^+$ and 9.0% in STAS$^-$). EGFR mutation had significantly lower detection rates in the SP population than in the SN population (52.5%

in STAS⁺ and 69.7% in STAS⁻). We also found that EGFR L858R (19.5% in STAS⁺ and 37.6% in STAS⁻) and ERBB2 20ins (1.8% in STAS⁺ and 5.9% in STAS⁻) of mutational hotspots of derived genes in the SP group were dramatically lower than those in the SN group. The gene with a significantly higher detection rate in the SP population than in the SN population was TP53 (49.8% in STAS⁺ and 34.8% in STAS⁻). In the SP group, the mutated hotspot of derived genes was ALK fusion (13.1% in STAS⁺ and 2.3% in STAS⁻). Other gene mutations, such as KRAS, ROS1, and BRAF, were not differently expressed between the SP and SN populations. Similarly, several studies have demonstrated that STAS is correlated with a low mutation rate of EGFR (Warth et al., 2015; Lee et al., 2018; Jia et al., 2020). At the same time, STAS was also found to be more frequently observed in ALK-rearranged lung cancer (Lee et al., 2018; Kadota et al., 2019; Jia et al., 2020). In addition, Tian et al. (2021) found that the occurrence of STAS is often accompanied by the increased expression of TP53. Importantly, most studies also found that KRAS (Warth et al., 2015; Lee et al., 2018; Zeng et al., 2020) is not related to STAS, which is proved in the current study. However, some studies propose a totally opposite conclusion. Several research studies suggest that there is no correlation between STAS and EGFR (Toyokawa et al., 2018a; Toyokawa et al., 2018c; Zhang Z. et al., 2020). Meanwhile, two other studies reported that high mutation rates of ROS1 (Jia et al., 2020) and BRAF (Warth et al., 2015) were associated with the occurrence of STAS. Ethnic diversity and different testing approaches can cause contradictions.

We also explored the gene mutations in some important tumor-related signaling pathways in SP and SN populations, including the P53 cell cycle signaling pathway and Wnt and ERBB signaling pathways. It is well known that genomic rearrangements associated with ERBB networks represent one of the major driver mutations and potential targets for establishing novel ADC treatment modalities (Trombetta et al., 2017; Kruspig et al., 2018). Many studies have shown that, as important genes constituting the ERBB network, EGFR (da Cunha Santos et al., 2011) and ERBB2 (20ins) (Xiang et al., 2022) are significantly related to the development and invasion of lung cancer. At the same time, EGFR has also been found to be closely related to STAS (Warth et al., 2015; Lee et al., 2018). Similarly, we have observed a lower detection rate of EGFR and ERBB2 (20ins) in the SP population, which is the major discovery in the ERBB signaling pathway. In addition, TP53 is a major tumor suppressor gene and the most frequently inactivated gene in cancer and plays an important role in both TP53 cell cycle signaling and Wnt signaling pathways. Prevention and suppression of TP53 mutation are one of the approaches to be considered in the treatment of tumors (Bykov et al., 2018). In this study, we have discovered a higher TP53 mutation (missense) rate in the SP population in the analysis of not only the P53 cell cycle signaling pathway but also the Wnt signaling pathway. Similarly, TP53 has been previously proved to be a common alteration in SP adenocarcinomas (Cardona et al., 2018). However, we cannot deduce that the P53 cell cycle signaling pathway and Wnt signaling pathway are different between SP and SN patients according to the results of TP53 mutation, as only a minority of these signaling pathway-related genes have been detected in this research. Collectively, the alteration of all p53 cell cycle, Wnt, and ERBB signaling pathway-related genes is not completely understood in this article. Therefore, more studies are needed to better understand STAS.

Furthermore, we analyzed the mutual exclusivity and co-occurrence of genes in both groups and found that in the SP population, EGFR excluding driver genes, including ALK, KRAS, ROS1, ERBB2, EGFR, and STK11, was mutually exclusive, while PIK3CA co-occurred with CTNNB1/PTEN, and KRAS co-occurred with STK11. In the SN population, EGFR and driver genes, including ALK, KRAS, ROS1, ERBB2, and BRAF, were mutually exclusive. TP53 co-occurred with CDKN2A, RB1, and APC, and SMAD4 co-occurred with CTNNB1. It is known that analysis of mutually exclusive co-occurrence of genes aids the exploration of the pathogenic pathway of tumors and provides essential references for the localization of driver genes (Sanchez-Vega et al., 2018; Scheffler et al., 2019). Previous large-scale genome sequencing data have indicated that mutations driving oncogenes, such as KRAS versus EGFR, are often mutually exclusive (Gainor et al., 2013), which is consistent with our results. Interestingly, Cisowski and Bergo (2017) suggested in a mouse lung cancer model that the co-occurrence of two tumor driver genes (KRAS and BRAF) is deleterious, leading to cell cycle exit, senescence, and death. It is reported that mutations in STK11 can be found in 19.8% of KRAS-mutated NSCLCs, which is associated with poor overall survival (Facchinetti et al., 2017; Scheffler et al., 2019). Based on our current results, STAS may detrimentally effect patients' overall survival as co-alterations of KRAS and STK11 occur more frequently in SP subjects. Although a previous study has proved that alterations in the PI3K pathway most likely occur as isolated events in lung cancer (Millis et al., 2019), the mutations of PTEN and PIK3CA, both PI3K pathway-related genes, also co-occur frequently in the SP population. However, the role of the co-occurrence of PTEN and PIK3CA in STAS is not clear yet.

We found there are differences between the two groups in terms of mutation, so we tried to use unsupervised clustering and dimensionality reduction to discover whether the overall mutation characteristics of different groups of STAS could be extracted. Unfortunately, mutation characteristics are affected by many factors, and unsupervised clustering could not accurately extract the mutation characteristics of the two groups of STAS. The reason why STAS cannot be distinguished is that, on the one hand, lung adenocarcinoma is a well-defined driving type, and the molecular typing induced by driving is more significant than the STAS classification. On the other hand, the NMF model has been used in the application of mutation spectrum dimensionality reduction via limited numerical transformation (Jiang et al., 2022).

Numerous studies have revealed that STAS is an independent prognostic indicator for poor outcomes (Kadota et al., 2015; Shiono and Yanagawa, 2016; Toyokawa et al., 2018d; Zhang B. et al., 2020; Ding et al., 2020). The 1-year DFS did not differ significantly between STAS-positive and -negative patients in our study. Because of the relatively short follow-up period, the outcome measures were limited. We only analyzed the 1-year DFS of these subjects in this study. Further multicentric prospective studies are required to provide more convincing results, especially studies analyzing the prognostic effects of genomic alterations when tumor spread through the air space is taken into account.

The limitations of our study were as follows: first, the relatively small sample size and the small size of the NGS panel limited the analysis on less common genomic alterations and STAS associations. Second, the PSM approach has its intrinsic limitations. For instance,

there may have been other clinical features that were not included in this regression model.

Conclusion

In summary, the present study analyzed the genomic characteristics of STAS-positive and -negative populations in a relatively large population cohort of lung adenocarcinoma patients. It demonstrated the global genetic mutation profiles of STAS-positive and -negative lung adenocarcinoma patients.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: NGDC. CNCB, PRJCA013869

Ethics statement

The studies involving human participants were reviewed and approved by the Shanghai Chest Hospital Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

Author contributions

RY collected the data, provided statistical analysis and data analysis, and wrote the original draft. RZ assisted in collecting the data. YY, YH, and SL conceived and designed the study. All authors contributed to the article and approved the submitted version.

Funding

Acknowledgments

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1101443/full#supplementary-material

SUPPLEMENTARY FIGURE S1
(A) Gene mutations detected greater than 1% in the STAS-positive patients. **(B)** Gene mutations detected greater than 1% in the STAS-negative patients. STAS, propagation through the air space.

SUPPLEMENTARY FIGURE S2
(A) Detection rate of signaling pathways in overall lung adenocarcinoma patients grouped according to STAS presence or absence. The X-axis represents the signaling pathway names, and the Y-axis indicates the detection frequency. **(B)** Mutational spectrum of the P53 cell cycle signaling pathway of lung adenocarcinoma patients grouped according to STAS presence or absence. **(C)** Mutational spectrum of the Wnt signaling pathway of lung adenocarcinoma patients grouped according to STAS presence or absence. **(D)** Mutational spectrum of the ERBB signaling pathway of lung adenocarcinoma patients grouped according to STAS presence or absence. Gender, smoking status, disease stage, age, and type of surgery are annotated at the bottom of the printout. Each column represents a patient, and each row represents a gene. The left column shows the mutated genes. The right column shows the mutation rate of each gene and the number of mutation types for each gene. The top plot represents the total number of mutations carried by the patient. Different colors indicate different types of mutations.

References

Aly, R. G., Rekhtman, N., Li, X., Takahashi, Y., Eguchi, T., Tan, K. S., et al. (2019). Spread through air spaces (STAS) is prognostic in atypical carcinoid, large cell neuroendocrine carcinoma, and small cell carcinoma of the lung. *J. Thorac. Oncol.* 14 (9), 1583–1593. doi:10.1016/j.jtho.2019.05.009

Bykov, V. J. N., Eriksson, S. E., Bianchi, J., and Wiman, K. G. (2018). Targeting mutant p53 for efficient cancer therapy. *Nat. Rev. Cancer* 18 (2), 89–102. doi:10.1038/nrc.2017.109

Cardona, A. F., Zatarain-Barron, Z. L., Rubio, C., Martinez, S., Ruiz-Patino, A., Ricaurte, L., et al. (2018). Probable hereditary familial overlap syndrome with multiple synchronous lung tumors. *Lung Cancer* 124, 279–282. doi:10.1016/j.lungcan.2018.08.022

Cisowski, J., and Bergo, M. O. (2017). What makes oncogenes mutually exclusive? *Small GTPases* 8 (3), 187–192. doi:10.1080/21541248.2016.1212689

da Cunha Santos, G., Shepherd, F. A., and Tsao, M. S. (2011). EGFR mutations and lung cancer. *Annu. Rev. Pathol.* 6, 49–69. doi:10.1146/annurev-pathol-011110-130206

Ding, Y., Zhang, L., Guo, L., Wu, C., Zhou, J., Zhou, Y., et al. (2020). Comparative study on the mutational profile of adenocarcinoma and squamous cell carcinoma predominant histologic subtypes in Chinese non-small cell lung cancer patients. *Thorac. Cancer* 11, 103–112. doi:10.1111/1759-7714.13208

Eguchi, T., Kameda, K., Lu, S., Bott, M. J., Tan, K. S., Montecalvo, J., et al. (2019). Lobectomy is associated with better outcomes than sublobar resection in spread through air spaces (STAS)-Positive T1 lung adenocarcinoma: A propensity score-matched analysis. *J. Thorac. Oncol.* 14 (1), 87–98. doi:10.1016/j.jtho.2018.09.005

Facchinetti, F., Bluthgen, M. V., Tergemina-Clain, G., Faivre, L., Pignon, J. P., Planchard, D., et al. (2017). LKB1/STK11 mutations in non-small cell lung cancer patients: Descriptive analysis and prognostic value. *Lung Cancer* 112, 62–68. doi:10.1016/j.lungcan.2017.08.002

Falch, C. M., Sundaram, A. Y. M., Oystese, K. A., Normann, K. R., Lekva, T., Silamikelis, I., et al. (2018). Gene expression profiling of fast- and slow-growing non-functioning gonadotroph pituitary adenomas. *Eur. J. Endocrinol.* 178 (3), 295–307. doi:10.1530/eje-17-0702

Gainor, J. F., Varghese, A. M., Ou, S. H. I., Kabraji, S., Awad, M. M., Katayama, R., et al. (2013). ALK rearrangements are mutually exclusive with mutations in EGFR or KRAS: An analysis of 1,683 patients with non-small cell lung cancer. *Clin. Cancer Res.* 19 (15), 4273–4281. doi:10.1158/1078-0432.Ccr-13-0318

Han, Y. B., Kim, H., Mino-Kenudson, M., Cho, S., Kwon, H. J., Lee, K. R., et al. (2021). Tumor spread through air spaces (STAS): Prognostic significance of grading in non-small cell lung cancer. *Mod. Pathol.* 34 (3), 549–561. doi:10.1038/s41379-020-00709-2

Hu, S. Y., Hsieh, M. S., Hsu, H. H., Tsai, T. M., Chiang, X. H., Tsou, K. C., et al. (2018). Correlation of tumor spread through air spaces and clinicopathological characteristics in surgically resected lung adenocarcinomas. *Lung Cancer* 126, 189–193. doi:10.1016/j.lungcan.2018.11.003

Ikeda, T., Kadota, K., Go, T., Haba, R., and Yokomise, H. (2021). Current status and perspectives of spread through air spaces in lung cancer. *Thorac. Cancer* 12 (11), 1639–1646. doi:10.1111/1759-7714.13918

Jia, M., Yu, S., Yu, J., Li, Y., Gao, H., and Sun, P. L. (2020). Comprehensive analysis of spread through air spaces in lung adenocarcinoma and squamous cell carcinoma using the 8th edition AJCC/UICC staging system. *BMC Cancer* 20 (1), 705. doi:10.1186/s12885-020-07200-w

Jiang, A., Wang, J., Liu, N., Zheng, X., Li, Y., Ma, Y., et al. (2022). Integration of single-cell RNA sequencing and bulk RNA sequencing data to establish and validate a prognostic model for patients with lung adenocarcinoma. *Front. Genet.* 13, 833797. doi:10.3389/fgene.2022.833797

Kadota, K., Kushida, Y., Kagawa, S., Ishikawa, R., Ibuki, E., Inoue, K., et al. (2019). Limited resection is associated with a higher risk of locoregional recurrence than lobectomy in stage I lung adenocarcinoma with tumor spread through air spaces. *Am. J. Surg. Pathol.* 43 (8), 1033–1041. doi:10.1097/pas.0000000000001285

Kadota, K., Nitadori, J. I., Sima, C. S., Ujiie, H., Rizk, N. P., Jones, D. R., et al. (2015). Tumor spread through air spaces is an important pattern of invasion and impacts the frequency and location of recurrences after limited resection for small stage I lung adenocarcinomas. *J. Thorac. Oncol.* 10 (5), 806–814. doi:10.1097/jto.0000000000000486

Karim, N. A., Eldessouki, I., Yellu, M., Namad, T., Wang, J., and Gaber, O. (2017). A case study in advanced lung cancer patients with vimentin over expression. *Clin. Lab.* 63 (10), 1575–1579. doi:10.7754/Clin.Lab.2017.170201

Koboldt, D. C., Chen, K., Wylie, T., Larson, D. E., McLellan, M. D., Mardis, E. R., et al. (2009). VarScan: Variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* 25 (17), 2283–2285. doi:10.1093/bioinformatics/btp373

Kordiak, J., Bielec, F., Jabłoński, S., and Pastuszak-Lewandoska, D. (2022). Role of beta-carotene in lung cancer primary chemoprevention: A systematic Review with meta-analysis and meta-regression. *Nutrients* 14 (7), 1361. doi:10.3390/nu14071361

Kruspig, B., Monteverde, T., Neidler, S., Hock, A., Kerr, E., Nixon, C., et al. (2018). The ERBB network facilitates KRAS-driven lung tumorigenesis. *Sci. Transl. Med.* 10 (446), eaao2565. doi:10.1126/scitranslmed.aao2565

Lee, J. S., Kim, E. K., Kim, M., and Shim, H. S. (2018). Genetic and clinicopathologic characteristics of lung adenocarcinoma with tumor spread through air spaces. *Lung Cancer* 123, 121–126. doi:10.1016/j.lungcan.2018.07.020

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25 14, 1754–1760. doi:10.1093/bioinformatics/btp324

Liu, Y., Chen, D., Qiu, X., Duan, S., Zhang, Y., Li, F., et al. (2018). Relationship between MTA1 and spread through air space and their joint influence on prognosis of patients with stage I-III lung adenocarcinoma. *Lung Cancer* 124, 211–218. doi:10.1016/j.lungcan.2018.07.040

Lu, S., Tan, K. S., Kadota, K., Eguchi, T., Bains, S., Rekhtman, N., et al. (2017). Spread through air spaces (STAS) is an independent predictor of recurrence and lung cancer-specific death in squamous cell carcinoma. *J. Thorac. Oncol.* 12 (2), 223–234. doi:10.1016/j.jtho.2016.09.129

Ma, K., Zhan, C., Wang, S., Shi, Y., Jiang, W., and Wang, Q. (2019). Spread through air spaces (STAS): A new pathologic morphology in lung cancer. *Clin. Lung Cancer* 20 (2), e158–e162. doi:10.1016/j.cllc.2018.10.007

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20 (9), 1297–1303. doi:10.1101/gr.107524.110

Millis, S. Z., Jardim, D. L., Albacker, L., Ross, J. S., Miller, V. A., Ali, S. M., et al. (2019). Phosphatidylinositol 3-kinase pathway genomic alterations in 60,991 diverse solid tumors informs targeted therapy opportunities. *Cancer* 125 (7), 1185–1199. doi:10.1002/cncr.31921

Newman, A. M., Bratman, S. V., Stehr, H., Lee, L. J., Liu, C. L., Diehn, M., et al. (2014). FACTERA: A practical method for the discovery of genomic rearrangements at breakpoint resolution. *Bioinformatics* 30 (23), 3390–3393. doi:10.1093/bioinformatics/btu549

Pyo, J. S., and Kim, N. Y. (2022). Clinicopathological impact of the spread through air space in non-small cell lung cancer: A meta-analysis. *Diagn. (Basel)* 12 (5), 1112. doi:10.3390/diagnostics12051112

Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W. K., Luna, A., La, K. C., et al. (2018). Oncogenic signaling pathways in the cancer genome atlas. *Cell* 173 (2), 321–337.e10. doi:10.1016/j.cell.2018.03.035

Scheffler, M., Ihle, M. A., Hein, R., Merkelbach-Bruse, S., Scheel, A. H., Siemanowski, J., et al. (2019). K-Ras mutation subtypes in NSCLC and associated Co-occuring mutations in other oncogenic pathways. *J. Thorac. Oncol.* 14 (4), 606–616. doi:10.1016/j.jtho.2018.12.013

Shiono, S., Endo, M., Suzuki, K., Yarimizu, K., Hayasaka, K., and Yanagawa, N. (2018). Spread through air spaces is a prognostic factor in sublobar resection of non-small cell lung cancer. *Ann. Thorac. Surg.* 106 (2), 354–360. doi:10.1016/j.athoracsur.2018.02.076

Shiono, S., and Yanagawa, N. (2016). Spread through air spaces is a predictive factor of recurrence and a prognostic factor in stage I lung adenocarcinoma. *Interact. Cardiovasc Thorac. Surg.* 23, 567–572. doi:10.1093/icvts/ivw211

Tian, Y., Feng, J., Jiang, L., Ning, J., Gu, Z., Huang, J., et al. (2021). Integration of clinicopathological and mutational data offers insight into lung cancer with tumor spread through air spaces. *Ann. Transl. Med.* 9 (12), 985. doi:10.21037/atm-21-2256

Toyokawa, G., Yamada, Y., Tagawa, T., Kamitani, T., Yamasaki, Y., Shimokawa, M., et al. (2018). Computed tomography features of resected lung adenocarcinomas with spread through air spaces. *J. Thorac. Cardiovasc Surg.* 156 (4), 1670–1676. doi:10.1016/j.jtcvs.2018.04.126

Toyokawa, G., Yamada, Y., Tagawa, T., Kinoshita, F., Kozuma, Y., Matsubara, T., et al. (2018). Significance of spread through air spaces in resected lung adenocarcinomas with lymph node metastasis. *Clin. Lung Cancer* 19, 395–400. doi:10.1016/j.cllc.2018.04.002

Toyokawa, G., Yamada, Y., Tagawa, T., Kozuma, Y., Matsubara, T., Haratake, N., et al. (2018). Significance of spread through air spaces in resected pathological stage I lung adenocarcinoma. *Ann. Thorac. Surg.* 105 (6), 1655–1663. doi:10.1016/j.athoracsur.2018.01.037

Toyokawa, G., Yamada, Y., Tagawa, T., and Oda, Y. (2018). Significance of spread through air spaces in early-stage lung adenocarcinomas undergoing limited resection. *Thorac. Cancer* 9 (10), 1255–1261. doi:10.1111/1759-7714.12828

Trombetta, D., Rossi, A., Fabrizio, F. P., SpArAneo, A., Graziano, P., Fazio, V. M., et al. (2017). NRG1-ErbB lost in translation: A new paradigm for lung cancer? *Curr. Med. Chem.* 24 (38), 4213–4228. doi:10.2174/0929867324666170911170554

Upton, M. P., Hirohashi, S., Tome, Y., Miyazawa, N., Suemasu, K., and Shimosato, Y. (1986). Expression of vimentin in surgically resected adenocarcinomas and large cell carcinomas of lung. *Am. J. Surg. Pathol.* 10 (8), 560–567. doi:10.1097/00000478-198608000-00006

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38 (16), e164. doi:10.1093/nar/gkq603

Warth, A., Muley, T., Kossakowski, C. A., Goeppert, B., Schirmacher, P., Dienemann, H., et al. (2015). Prognostic impact of intra-alveolar tumor spread in pulmonary adenocarcinoma. *Am. J. Surg. Pathol.* 39 (6), 793–801. doi:10.1097/pas.0000000000000409

Xiang, C., Ji, C., Cai, Y., Teng, H., Wang, Y., Zhao, R., et al. (2022). Distinct mutational features across preinvasive and invasive subtypes identified through comprehensive profiling of surgically resected lung adenocarcinoma. *Mod. Pathol.* 35 (9), 1181–1192. doi:10.1038/s41379-022-01076-w

Yang, Z. R., Liu, M. N., Yu, J. H., Yang, Y. H., Chen, T. X., Han, Y. C., et al. (2020). Treatment of stage III non-small cell lung cancer in the era of immunotherapy: Pathological complete response to neoadjuvant pembrolizumab and chemotherapy. *Transl. Lung Cancer Res.* 9 (5), 2059–2073. doi:10.21037/tlcr-20-896

Yen, J. L., Garcia, S., Montana, A., Harris, J., Chervitz, S., Morra, M., et al. (2017). A variant by any name: Quantifying annotation discordance across tools and clinical databases. *Genome Med.* 9 (1), 7. doi:10.1186/s13073-016-0396-7

Yokoyama, S., Murakami, T., Tao, H., Onoda, H., Hara, A., Miyazaki, R., et al. (2018). Tumor spread through air spaces identifies a distinct subgroup with poor prognosis in surgically resected lung pleomorphic carcinoma. *Chest* 154 (4), 838–847. doi:10.1016/j.chest.2018.06.007

Zeng, Q., Wang, B., Li, J., Zhao, J., Mao, Y., Gao, Y., et al. (2020). Solid nodule appearance as a predictor of tumor spread through air spaces in patients with lung adenocarcinoma: A propensity score matching study. *Cancer Manag. Res.* 12, 8197–8207. doi:10.2147/cmar.S266750

Zhang, B., Zhang, L., Yue, D., Li, C., Zhang, H., Ye, J., et al. (2020). Genomic characteristics in Chinese non-small cell lung cancer patients and its value in prediction of postoperative prognosis. *Transl. Lung Cancer Res.* 9, 1187–1201. doi:10.21037/tlcr-19-664

Zhang, Z., Liu, Z., Feng, H., Xiao, F., Shao, W., Liang, C., et al. (2020). Predictive value of radiological features on spread through air space in stage cIA lung adenocarcinoma. *J. Thorac. Dis.* 12 (11), 6494–6504. doi:10.21037/jtd-20-1820