

RESEARCH

Open Access



Spatial analyses revealed S100P + TFF1 + tumor cells in spread through air spaces samples correlated with undesirable therapy response in non-small cell lung cancer

Guangyu Fan^{1†}, Tongji Xie^{1†}, Mengwei Yang^{1†}, Lin Li², Le Tang¹, Xiaohong Han^{3*} and Yuankai Shi^{1*} 

Abstract

Spread through air spaces (STAS) is a recognized aggressive pattern in lung cancer, serving as a crucial risk factor for postoperative recurrence. However, its phenotype and related spatial structure have remained elusive. To address these limitations, we conducted a comprehensive study based on spatial data, analyzing over 30,000 spots from 14 non-STAS samples and one STAS sample. We observed increased proliferation activities and angiogenesis in STAS, identifying S100P as a potential biomarker for STAS. Furthermore, our investigation into the heterogeneity of STAS tumor cells revealed a subset identified as S100P + TFF1 +, exhibiting a negative impact on patients' survival in public datasets. This subtype exhibited the highest activities in the TGF β and hypoxia, suggesting its potential pro-tumor role within the tumor microenvironment. To assess the role of S100P + TFF1 + tumor cells in therapy response, we included data from two clinical trial cohorts (BPI-7711 for EGFR-TKI therapy and ORIENT-3 for immunotherapy). The presence of S100P + TFF1 + tumor cells correlated with worse responses to both EGFR-TKI therapy and immunotherapy. Notably, TFF1 emerged as a serum marker for predicting EGFR-TKI response. Cell-cell communication analysis revealed that the TGF β signaling pathway was the most activated in S100P + TFF1 + tumor cells, with TGFB2-TGFBR2 identified as the main ligand-receptor pair. This was further validated by multiplex immunofluorescence performed on twenty NSCLC samples. In summary, our study identified S100P as the biomarker for STAS and highlighted the adverse role of S100P + TFF1 + tumor cells in survival outcomes.

Keywords Spread through air spaces, Non-small cell lung cancer, Spatial patterns, Spatial transcriptomics

[†]Guangyu Fan, Tongji Xie and Mengwei Yang have contributed equally to this work.

*Correspondence:

Xiaohong Han
hanxiaohong@pumch.cn
Yuankai Shi
syuankai@cicams.ac.cn

¹ Department of Medical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing Key Laboratory of Clinical Study on Anticancer Molecular Targeted Drugs, No. 17 Panjiayuan Nanli, Chaoyang District, Beijing 100021, China

² Department of Pathology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100021, China

³ Clinical Pharmacology Research Center, Peking Union Medical College Hospital, State Key Laboratory of Complex Severe and Rare Diseases, NMPA Key Laboratory for Clinical Research and Evaluation of Drug, Beijing Key Laboratory of Clinical PK & PD Investigation for Innovative Drugs, Chinese Academy of Medical Sciences & Peking Union Medical College, No. 1, Shuaifuyuan, Dongcheng District, Beijing 100730, China



Introduction

Lung cancer stands as one of the most prevalent malignancies, with non-small cell lung cancer (NSCLC) being the predominant form [1]. For early-stage lung cancer, surgical resection is typically recommended. However, a considerable percentage, approximately 20–40% of NSCLC patients, experience a relapse following surgery. Prognosis in such cases is influenced by various pathological factors including lymphatic invasion, pleural invasion and vascular invasion. Recent research has notably highlighted the role of spread through air spaces (STAS) in the postoperative recurrence of early-stage NSCLC [2, 3].

Defined by the WHO in 2015, STAS represents an aggressive pattern in lung cancer where micropapillary clusters, solid nests, or single cells extend beyond the tumor's edge, occupying the air spaces of the surrounding lung tissue [4]. The 2021 WHO classification explicitly denotes STAS as a histological feature bearing prognostic significance. Studies propose that patients with STAS-positive early stage NSCLC who undergo lobectomy exhibit a more favorable prognosis than those treated with sublobectomy [5, 6]. Consequently, many surgeons advocate for lobectomy in such cases. Yet, the challenge lies in the microscopic examination of the entire surgical specimen to assess STAS, limiting its practicality as a prognostic tool in clinical decision-making, despite its evident clinical relevance [7, 8]. Furthermore, uncovering the molecular determinants of STAS in lung adenocarcinoma (LUAD) could lead to the identification of novel therapeutic targets and better patient outcomes.

Traditional bulk-level transcriptomics data fail to discern STAS from other cancer cells, while single-cell RNA sequencing lacks spatial context, impeding the exploration of correlations between the local environment and specific cell–cell interactions. Recent advancements in spatial transcriptomics (ST) technologies have introduced potent tools to delineate the precise spatial distribution of genes, facilitating an understanding of how tumor intrinsic features interact with crucial cell types in the context of tumor development and response to therapy [9, 10]. By preserving tissue architecture, ST allows an examination of STAS's molecular and cellular composition, as well as their subtypes.

In this study, we harnessed ST to characterize the phenotype of STAS and its heterogeneity in NSCLC. We identified S100P as the potential biomarker for STAS. The S100P + TFF1 + subtype of STAS tumor cells exhibited an adverse role in patients' survival and was associated with worse outcomes in EGFR-TKI therapy and immunotherapy. Studies based on spatial transcriptomics data revealed that the TGF β signaling pathway was most activated in S100P + TFF1 + tumor cells, pointing toward

an immune-resistant microenvironment. In conclusion, our study sheds light on the roles of S100P as the marker of STAS and provides insights into the unique spatial structures of STAS that foster immune resistance.

Results

The identification of tumor cells in ST data

The methodology employed in this study is depicted in Fig. 1A. STAS is a recently identified histologic feature of cancer invasion that presents challenges for diagnosis on frozen samples. Currently, STAS can only be identified on final pathology, and its presence is associated with poorer overall and disease-free survival [11]. In this investigation, we obtained 15 formalin-fixed paraffin-embedded (FFPE) samples from NSCLC patients, with one of these samples exhibiting the STAS phenotype. Through the hematoxylin and eosin (HE) image of STAS sample, we could discrete multiple clusters of cancer cells in the air spaces. Through this specimen, we conducted a comprehensive exploration of potential biomarkers and the TME characteristics associated with STAS.

First, we decided to identify the STAS clusters from normal lung tissues. The scatter feature of STAS made manually annotation inaccurate and laborious. Moreover, the task of differentiating between malignant and normal epithelial cells based on the gene expression profile in ST data is a significant challenge. In order to tackle this intricate issue, the inferCNV study was conducted to precisely distinguish cancerous cells from other types of cells by examining their copy number variation patterns. This procedure consisted of two clustering stages.

The first phase aimed to find reference cells for the inferCNV pipeline, based on which inferCNV analysis inferred CNV patterns of malignant cells. Initially, all spots in this STAS sample were segmented into 12 clusters based on gene expression patterns (Fig. 1B). The "immune score" was determined for each spot by evaluating a collection of immune-related signatures, which represents the average value of immune features within each spot. The signatures comprised of pan-immune markers (PTPRC), pan-T cell markers (CD2, CD3D, CD3E, CD3G), B cell markers (CD79A, MS4A1, CD79B), and myeloid cell markers (CD68, CD14). The cluster 9, which had the greatest immunological score, was used as the reference for the inferCNV analysis (Fig. 1C).

The primary objective of the second clustering phase was to differentiate malignant cells from other cell types by analyzing copy number variation (CNV) patterns. The hierarchical clustering algorithm, which utilizes tree partitioning, allocated every spot with the exception of the reference cluster, into eight distinct clusters (Fig. 1D). Clusters T1 to T5, which had very high CNV scores, were classified as malignant clusters, whereas

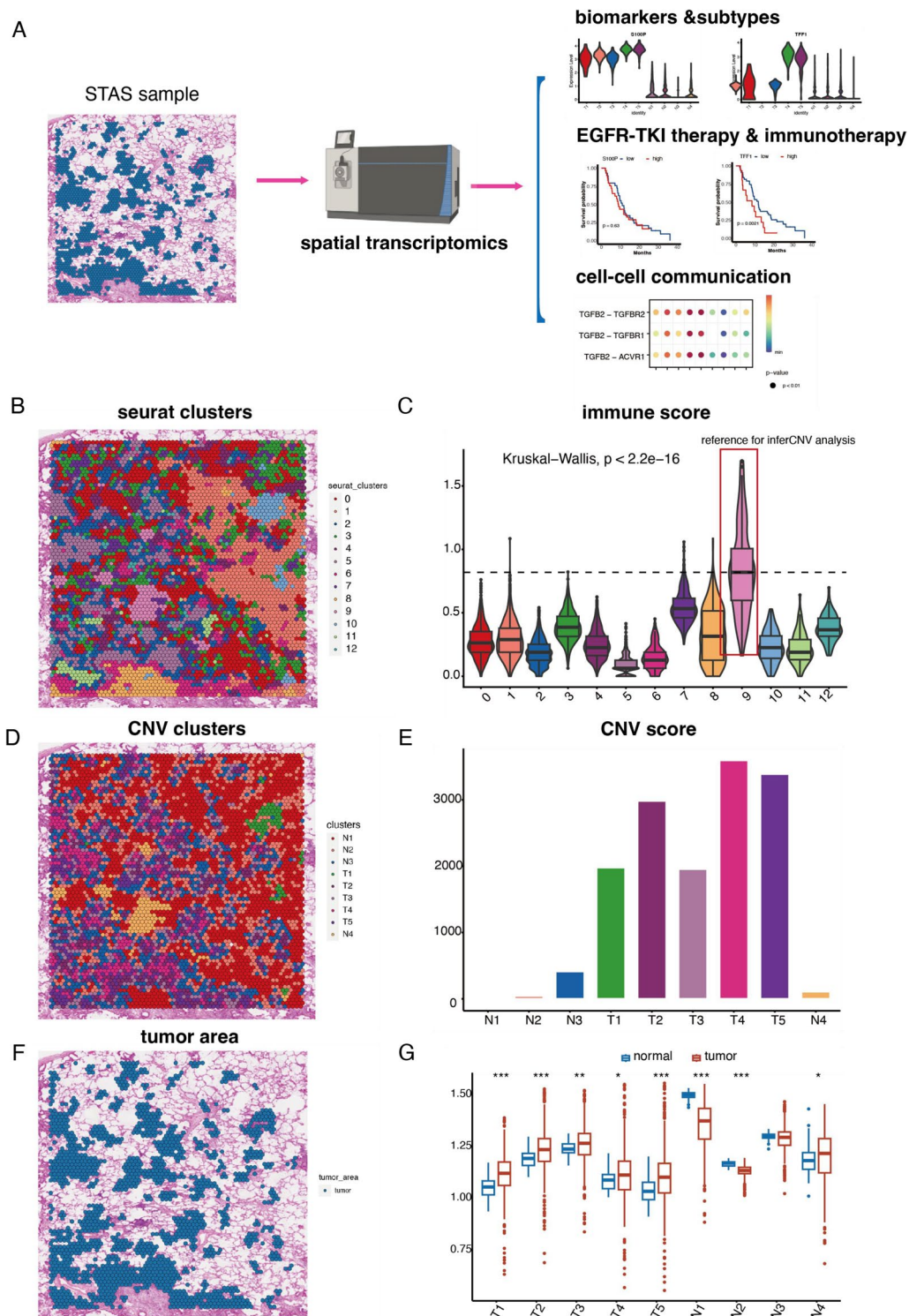


Fig. 1 The identification of malignant cells in ST data. **A** The methodology employed in this study. **B** Clustering of 4965 spots in spread through air spaces (STAS) sample into 12 distinct clusters. **C** Distribution of immune score in the 12 clusters. **D** Hierarchical clustering assigning all spots, except the reference cluster, into eight clusters. **E** Bar charts showing the distribution of copy number variation (CNV) score in the nine clusters. **F** The plot depicted the identified tumor area in STAS sample. **G** The distribution of the signature of each cluster in tumor and non-tumor samples in the bulk transcriptomics data

the other clusters had much lower CNV scores and were labeled as non-tumor regions (Fig. 1E). The accuracy of these annotations was confirmed by consulting two separate pathologists who examined the HE histological data. T1 to T5 represented dispersed tumor regions, while the remaining clusters were mainly composed of normal epithelial cells, fibroblasts, and a combination of immune cells (Fig. 1F). In addition, we acquired the top 50 marker genes for each cluster and evaluated these signatures in the bulk transcriptomics data. Signatures obtained from tumor clusters exhibited a higher presence in tumor tissues, hence verifying the precise detection of tumor regions (Fig. 1G). The tumor regions of additional NSCLC samples were detected using the identical methodology.

S100P was identified as the biomarker of STAS

In order to identify potential STAS markers, we compared the gene expression patterns of tumor cells in the STAS sample with those in other 14 non-STAS samples. As a result, we created a list of genes that showed differential expression (DEGs) (Fig. 2A). One of the DEGs in STAS is S100P, which belongs to the S100 protein family and has increased expression. GABRP was highly expressed in STAS, sustaining the stemness of cancer cells through EGFR signaling in triple-negative breast [12]. LGALS4 also showed higher expression in STAS, modulating disease progression in colorectal, gastric, pancreatic, hepatocellular, tongue, and breast cancer [13]. Additionally, MUC6 and MUC5AC, members of glycoproteins synthesized by epithelial cells, exhibited upregulation in STAS.

Pathway analysis unveiled increased proliferation activities in STAS compared to non-STAS tumor regions, characterized by upregulation of E2F targets, spliceosome and G2M checkpoint (Fig. 2B). Active angiogenesis and ECM-receptor interaction were observed in STAS. Concurrently, there were indications of involvement in metabolism changes, including glycosylation, arachidonic acid metabolism, and retinol metabolic process. Conversely, immune-related pathways (complement, inflammatory response, and type I interferon) were suppressed in STAS.

In addition, we examined the function of transcription factors (TFs) in increasing the aggressive characteristics of STAS. We employed Dorothea to investigate potential differences in regulon activity between STAS and other tumor samples (Fig. 2C). Figure 1F depicts the expression patterns of the 20 TFs that exhibit the most diverse activity in cellular populations. The notable findings revealed increased regulon activities of FOXJ2, which triggered epithelial-mesenchymal transition in NSCLC [14]. Furthermore, TFAP4 exhibited heightened activities and is

linked to the process of trans-differentiation, when an adenocarcinoma transforms into a small cell neuroendocrine state in lung cancer [15]. KLF3, regulating cell proliferation, migration, and therapy resistance, exhibited heightened regulon activities in STAS [16].

We hypothesized that ideal STAS biomarkers should be specifically expressed in tumor cells. Hence, by utilizing extensive single-cell datasets (including GSE148071, GSE127465, GSE143423, and EMTAB6149), we identified specific marker genes that were unique to tumor cells in comparison to immune cells and normal epithelial cells. This was done by applying criteria of an average log₂ fold change greater than 0.25 and a p-value less than 0.05, respectively. By intersecting these three gene lists, we have discovered possible STAS biomarkers. Among them, S100P exhibits the highest expression in the STAS cluster, as shown in Fig. 2D. Figure 2E shows that S100P had higher expression levels in tumor cells compared to immune cells and normal epithelial cells.

We subsequently examined the spatial expression patterns of S100P. In the STAS sample, S100P showed the highest expression within STAS clusters (Fig. 2F and G). Conversely, in the remaining fourteen samples without STAS, S100P displayed lower expression levels (Fig. 2F and G). To validate S100P in protein level, we performed multiple immunofluorescence (MIF) staining in 20 NSCLC patients from our cohort. In addition, we employed panCK to classify tumor cells, enabling us to differentiate them from the adjacent microenvironment. Consistent with transcriptomics analysis, we observed the main expression of S100P in tumor cells (Fig. 2H).

The presence of S100P + TFF1 + tumor cells is correlated with worse prognosis in NSCLC

Subsequently, we decided to investigate the heterogeneity of STAS clusters. Figure 3A displays the expression patterns of classical cell type marker genes for each CNV cluster. Tumor clusters (T1 to T5) featured marker genes like EPCAM, KRT8 and KRT19, which are tumor-specific epithelial markers. Normal areas (N1 to N4) expressed many immune cells related markers, including B cells, T cells, myeloid cells and fibroblasts. These results consistent with the identification of tumor areas.

Figure 3B provides a visual representation of the expression profiles of the top 5 marker genes for each cluster. T1 exhibited metalloproteinases such as MMP9 and MMP12, involved in extracellular matrix remodeling and COMP a marker of cancer-associated fibroblast. T2 and T3 showed similar expression patterns, expressing MMP11 and epithelial markers KRT17 and MUC6. T4 and T5 exhibited marker genes such as TFF1, TFF2, and TFF3, which are part of the trefoil factor family peptides. These genes play a vital role in processes like as

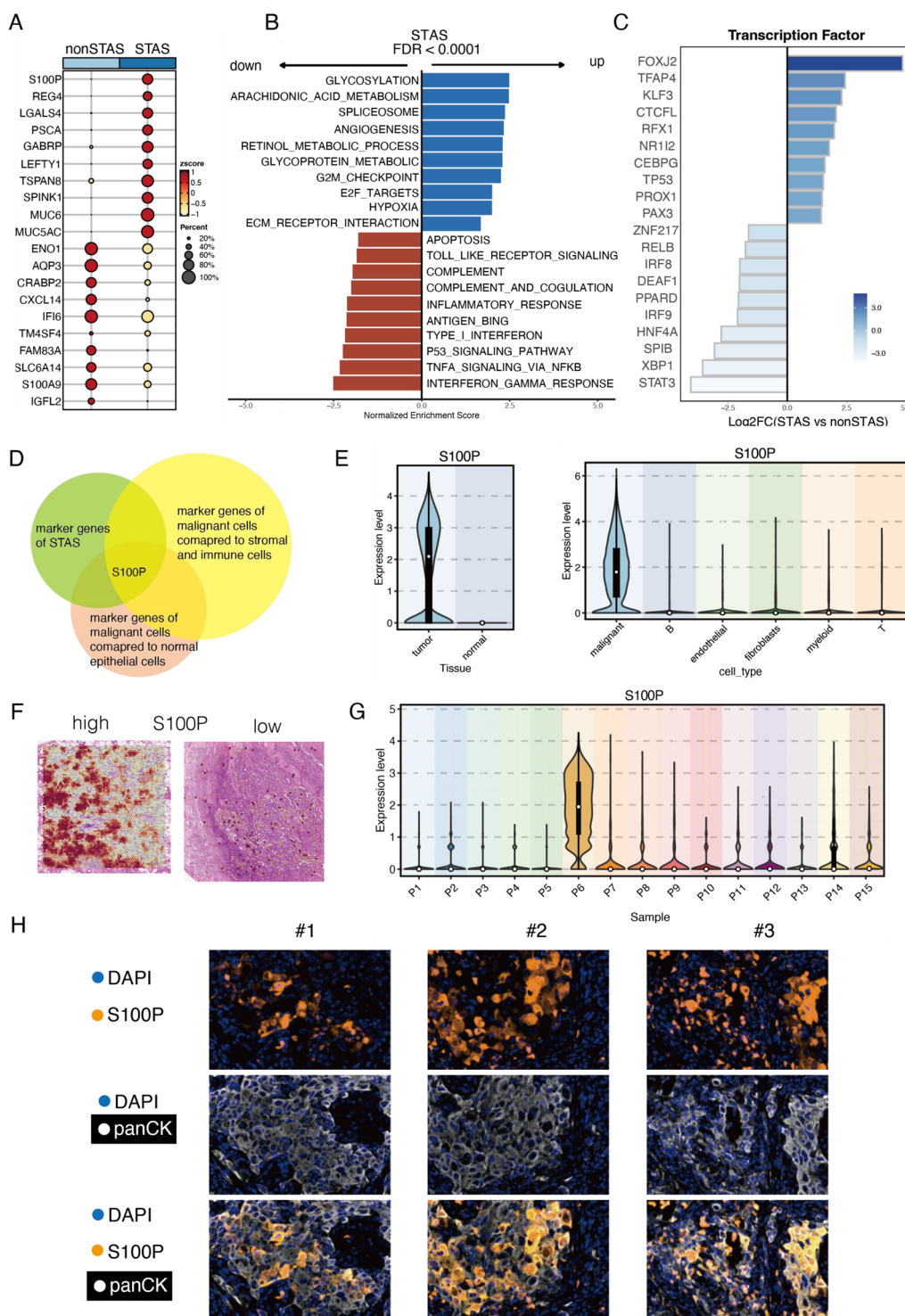


Fig. 2 S100P was identified as the biomarker of spread through air spaces (STAS). **A** The differentially expressed genes (DEGs) of tumor cells in STAS sample compared to non-STAS samples. **B** Pathway analysis unveiled distinct biological activities in STAS and non-STAS samples. **C** The distinct transcription factors in STAS and non-STAS samples. **D** S100P was identified as the biomarker of STAS. **E** S100P exhibited elevated expression in tumor cells compared to normal epithelial cells (left) and immune cells and stromal cells (right). **F** The spatial expression patterns of S100P in STAS and non-STAS samples. **G** The expression patterns of S100P in STAS and non-STAS samples. **H** The multiplex immunofluorescence (MIF) performed in 20 non-small cell lung cancer (NSCLC) patients demonstrated the main expression of S100P in tumor cells

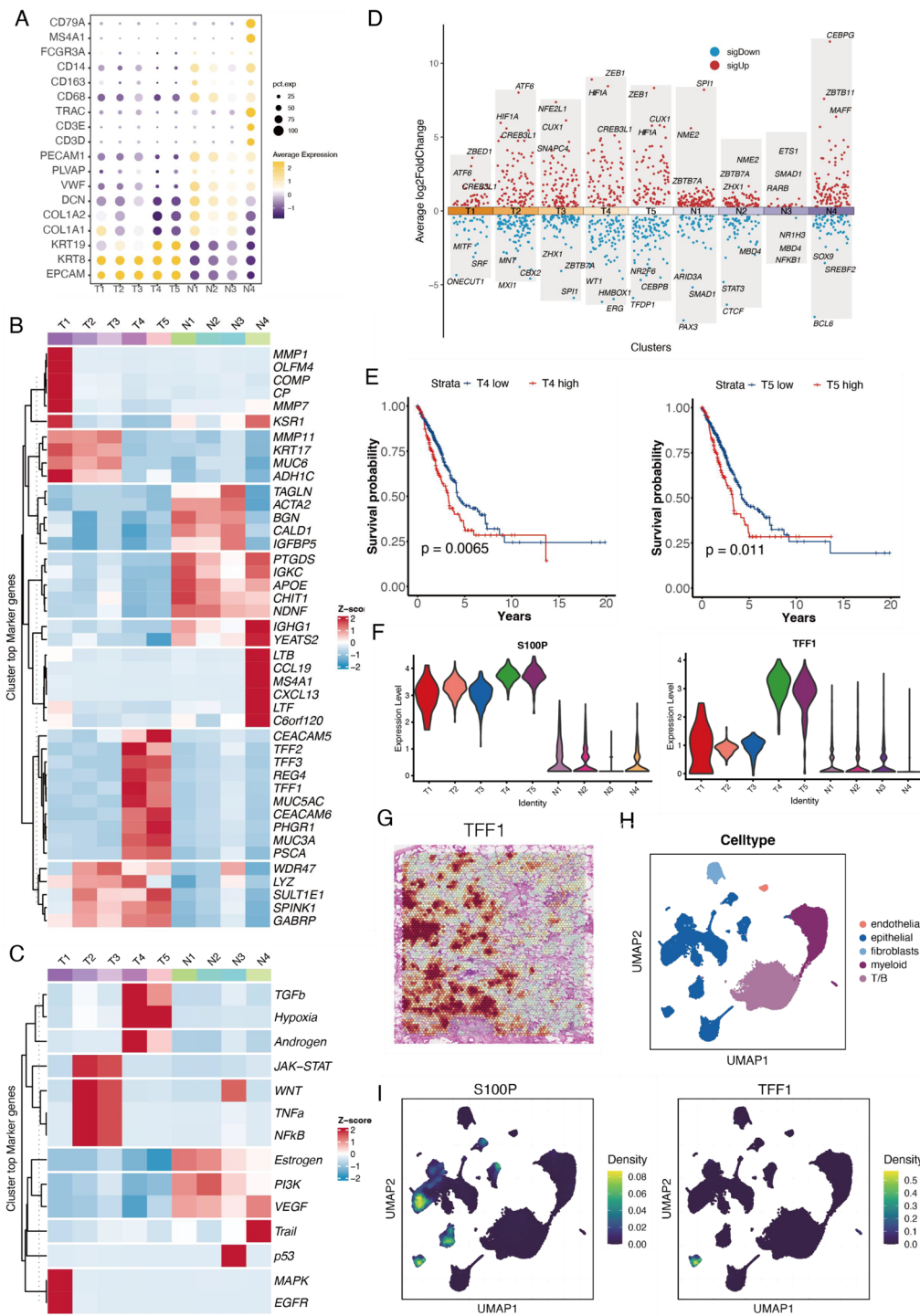


Fig. 3 S100P+TFF1 + tumor cells in spread through air spaces (STAS) sample correlated with worse prognosis in non-small cell lung cancer (NSCLC). **A** The expression profiles of classical cell types marker genes for copy number variation (CNV) clusters. **B** The expression profiles of the top 5 marker genes for each cluster. **C** The distinct pathway activities among these clusters. **D** The top3 up-regulated and down-regulated transcription factors among these clusters. **E** The signature of T4 and T5 showed adverse role in patients' survival time. **F** Distribution of S100P and TFF1 in CNV clusters. **G** The spatial distribution of TFF1 in STAS sample. **H** The single-cell dataset was divided into five main clusters: epithelial cells, myeloid cells, fibroblasts, endothelial cells, and T/B cells. **I** Expression patterns of S100P and TFF1 in single-cell level

angiogenesis, proliferation, antiapoptotic characteristics, and differentiation [17]. Additionally, N1, N2 and N3 displayed B cell marker IGKC, CAF marker ACTA2 and epithelial markers KRT17 and MUC6, indicating the infiltration of CAF and B cell in normal lung tissues. N4 had notable B cells infiltration, expressing markers (IGHG1 and MS4A1) and chemokines (CXCL13 and CCL19).

Pathway analysis using the PROGENy R package unveiled distinct pathway activities among these clusters (Fig. 3C). T1 exhibited high levels of MAPK and EGFR activities. T2 and T3 exhibited activated WNT, TNFA and NFKB activities. T4 and T5, notably, displayed the highest TGFb and hypoxia pathway activities, suggesting its potential pro-tumor role within TME. Normal clusters demonstrated heightened PI3K activities, VEGF and Trail activities. There is mounting evidence supporting the notion that tumor cells undergo reversible transitions between transcriptional states, driving metastasis and therapy resistance. To elucidate these transcriptional states across tumor cells, we initially analyzed the ST data using Dorothea to identify potential TFs (Fig. 3D). HIF1A, a crucial component in the hypoxia-induced pathway, was the most significantly over-represented transcription factor in T4 and T5 [18]. ZEB1, well-known epithelial-mesenchymal transition transcription factors, exhibited heightened regulon activities in T4 and T5 [19].

Subsequently, we calculated the top 50 DEGs of each cluster as the signature and explored their role in clinical outcome. The signature of T4 and T5 showed adverse impact on patients' survival in public datasets, indicating their aggressive phenotypes (Fig. 3E). TFF1 was marker gene of T4 and T5 and could be combined with the STAS marker S100P as the marker for a subset of aggressive STAS tumor cells (Fig. 3F). The spatial distribution of TFF1 was illustrated in Fig. 3G. In order to confirm this discovery at the level of individual cells, we utilized many datasets (including GSE148071, GSE127465, GSE143423, and EMTAB6149) and found five primary groups: tumor cells, fibroblasts, myeloid cells, T/B cells and endothelial cells [20, 21] (Fig. 3H). The distribution of TFF1 mainly in S100P + tumor cells validated its role as a subtype marker for the subpopulation of tumor cells that may lead to STAS (Fig. 3I).

S100P + TFF1 + tumor cells correlated with worse EGFR-TKI therapy and immunotherapy

Due to the lack of specific clinical information on STAS in the other samples from clinical trials and databases, we were unable to definitively classify them as STAS or non-STAS samples. Given the aggressive nature of STAS, we hypothesized that S100P + TFF1 + may represent a subtype of aggressive lung cancer cells associated with poor prognosis and reduced treatment efficacy.

Therefore, in the next step, we used S100P and TFF1 as markers to detect whether this subtype of cells exists and to explore what influence it may have. Analysis of public datasets showed that the expression of TFF1 was higher in tumor cells compared to immune cells and normal epithelial cells in Fig. 4A and B. The expression of TFF1 in tumor cells was validated in 20 NSCLC patients using MIF labeling, with panCK used to identify the tumor cells (Fig. 4C). The correlation coefficient between S100P and TFF1 in the bulk transcriptomics data of NSCLC was found to be 0.51, confirming their underlying correlation (Fig. 4D). In order to validate this association at the protein level, we performed MIF labeling of S100P and TFF1 on a specific group of 20 patient samples with NSCLC. We observed that S100P and TFF1 were found together in the tumor cells of NSCLC, as shown in Fig. 4E. Survival analysis of public datasets showed patients with high levels of either S100P or TFF1 had shorter overall survival (OS) time, and those with high levels of both S100P and high TFF1 displayed the worst clinical outcome (Fig. 4F). In addition, we conducted immunohistochemistry (IHC) analysis of S100P and TFF1 on a group of 70 NSCLC patients that we have in our own research facility. Figure 4G shows that tumor cells had higher expression levels of S100P and TFF1 relative to normal epithelial cells.

We conducted a thorough analysis using multi-omics data to investigate the therapeutic consequences of S100P + TFF1 + tumor cells in NSCLC patients who were treated with third EGFR-TKI therapy. At first, data from single cells were collected from 49 clinical biopsies of 30 patients with metastatic lung cancer. The biopsies were taken before and throughout targeted therapy. Figure 5A depicted the distribution of cells in non-responders (NR) and responders (R). Both S100P and TFF1 exhibited increased expression levels in patients with NR (Fig. 5B). Next, we assessed the levels of S100P and TFF1 in a prospective cohort participating in an open-label, single-arm, phase I/IIa clinical trial called BPI-7711 (NCT03386955). This group consisted of 186 individuals with locally progressed or metastatic NSCLC. Among the 186 patients included in the investigation, 57 individuals (30.65%) did not experience any clinical improvement, while 129 patients (69.35%) derived benefits from the medication. The enzyme-linked immunosorbent assay (ELISA) kit detected a substantial increase in S100P and TFF1 levels in NR compared to R, as shown in Fig. 5C. Furthermore, the study found a notable negative correlation between TFF1 and OS in patients treated with EGFR-TKI, as shown in Fig. 5D. These findings indicate that TFF1 has the potential to be used as a biomarker for predicting treatment outcomes.

Subsequently, we investigated the capacity of S100P + TFF1 + tumor cells to anticipate the efficacy of

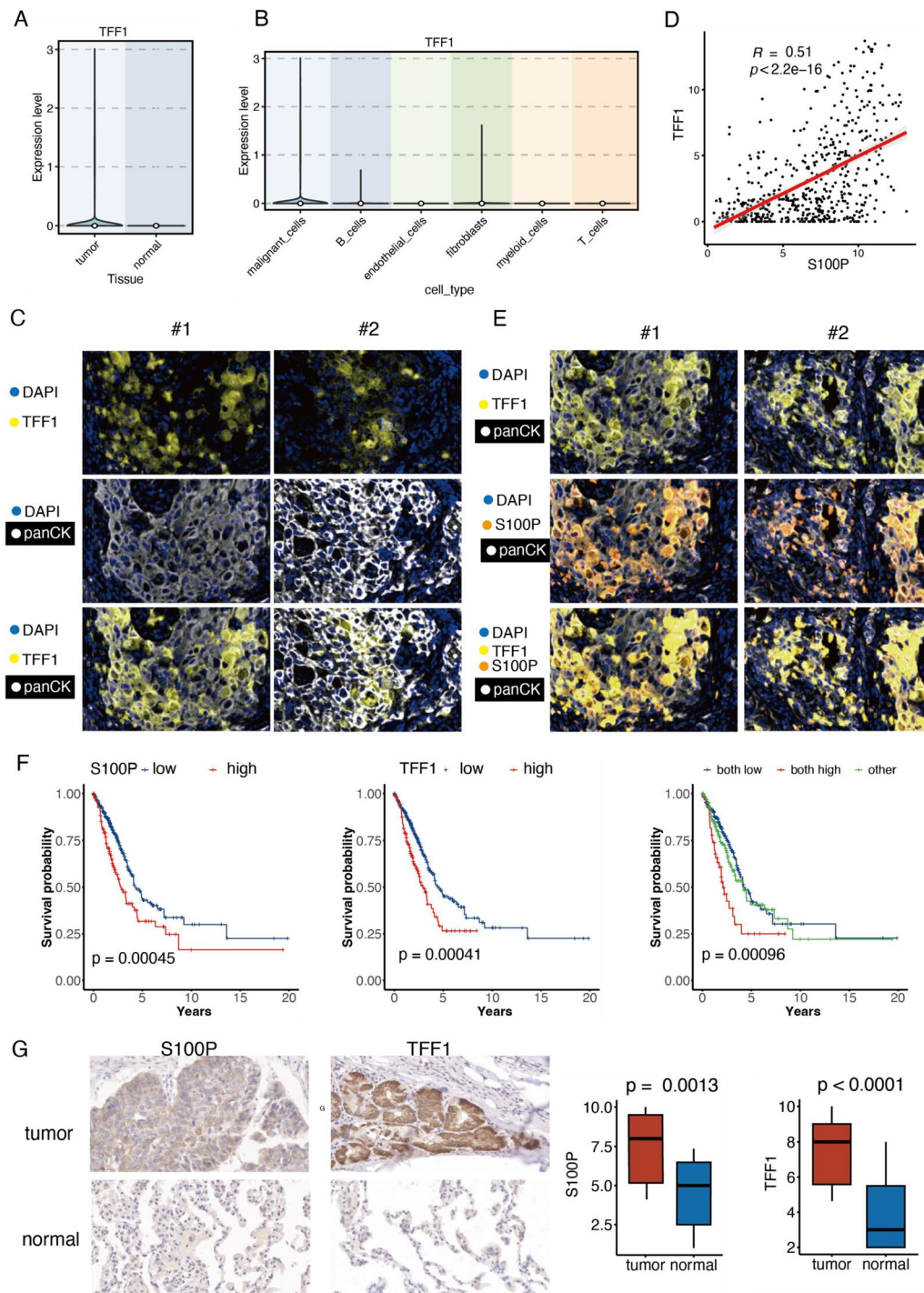


Fig. 4 Correlation between TFF1 and S100P. **A** TFF1 displayed elevated expression in tumor cells compared to normal epithelial cells. **B** TFF1 exhibited increased expression in tumor cells compared to immune and stromal cells. **C** Multiplex immunofluorescence (MIF) performed on 20 non-small cell lung cancer (NSCLC) patients demonstrated the primary expression of TFF1 in tumor cells. **D** Significant correlation between TFF1 and S100P. **E** MIF performed on 20 NSCLC patients demonstrated the co-location of S100P and TFF1 in tumor cells. **F** Patients with high levels of both S100P and TFF1 had shorter overall survival (OS) time, and those with high S100P and high TFF1 displayed the worst clinical outcome. **G** Immunohistochemistry demonstrated elevated expression of S100P and TFF1 in tumor cells compared to normal epithelial cells

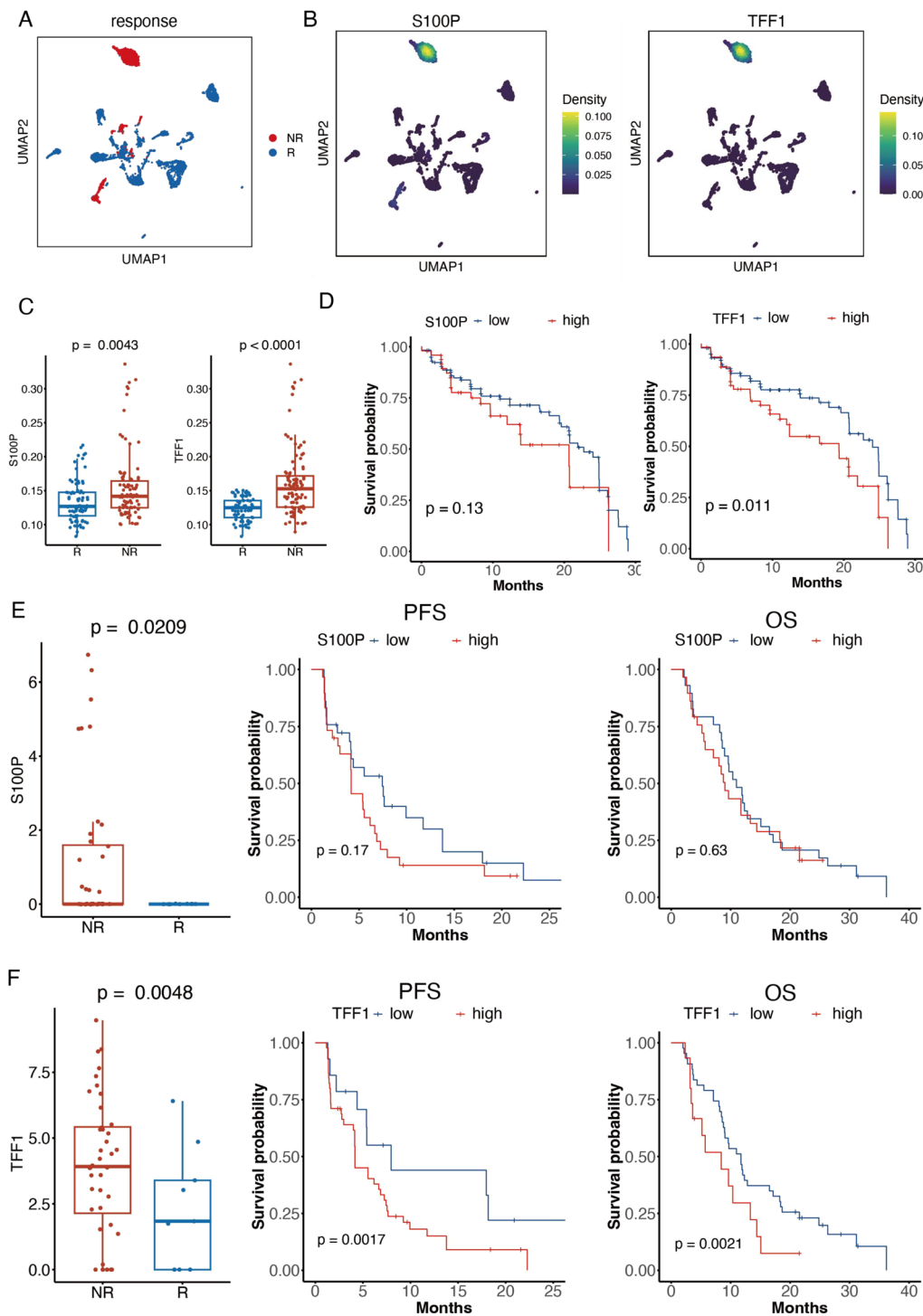


Fig. 5 S100P+TFF1 + tumor cells correlated with worse EGFR-TKI therapy and immunotherapy response. **A** Distribution of cells in non-responders (NR) and responders (R) receiving EGFR-TKI therapy. **B** Distribution of S100P and TFF1 in NR and R receiving EGFR-TKI therapy. **C** S100P and TFF1 measured by enzyme-linked immunosorbent assay (ELISA) kit showed the most significant upregulation in NR compared to R. **D** TFF1 showed a significant adverse association with worse clinical outcomes in EGFR-TKI-treated patients. **E** Distribution of S100P and TFF1 in NR and R receiving immunotherapy. **F** Role of S100P and TFF1 in predicting the therapeutic results of immunotherapy

immunotherapy in patients. The transcriptomics data of an open-label, randomized trial (ORIENT-3) that was conducted in 39 centers across China (NCT03150875) was incorporated [22]. This cohort was composed of 61 patients who had failed first-line chemotherapy and were diagnosed with late stage NSCLC. In immunotherapy-treated patients, S100P and TFF1 were both more enriched in NR (Fig. 5E and F). Additionally, we discovered that patients with elevated TFF1 experienced substantially worse therapeutic outcomes (Fig. 5E and F).

TGF β signaling in self communication of S100P + TFF1 + tumor cells

We further investigated cell–cell interactions in the STAS sample. T4 and T5 demonstrated the maximum level of activity among all clusters, with the highest number of incoming and outgoing interactions (Fig. 6A). In general, tumor clusters exhibited a greater number of cell–cell communication activities than normal clusters. We then concentrated on the signaling patterns of each cluster, both inbound and outbound. Tumor clusters and normal clusters exhibited distinct signaling pathways (Fig. 6B and C). The TGF β signaling pathway was the most active, which was consistent with the upregulated TGF β pathway in the pathway analysis. Pathways such as MIF, GRN, EGF, and PDGF, known to support cancer progression in various cancer types, were also identified. Many immune response-related signaling pathways, such as chemokines (CCL and CXCL) and complement, were observed to be enhanced in normal clusters. In tumor cells, the TGF β signaling drives tumorigenesis by inducing EMT, metastasis, angiogenesis, autophagy, and immune suppression. Therefore, we mainly focused on this signaling.

TGF β 2 released from T4/T5 interacted with TGF β 1, TGF β 2, and ACVR1 most on tumor cells rather than in the normal area (Fig. 6D). This ligand–receptor pair exhibited the most strength in the self-communication of T4/T5 clusters. T4 and T5 were identified as the main senders of the TGF β signaling pathway. Meanwhile, they also acted as receivers, mediators, and influencers of the TGF β signaling (Fig. 6E). Three tumor clusters, T1, T2, and T3, were also influencers of this signaling. These findings consistently supported the central role of T4 and T5 in the TGF β pathway in STAS, demonstrating the pro-tumor effect of self-communication of this subtype. To validate this finding at the protein level, MIF was performed on 20 tumor samples of NSCLC patients using panCK antibody to annotate tumor cells, and S100P and TFF1 to annotate this subtype of STAS tumor cells. The findings confirmed that TGF β 2 functions as a secretory protein originating from the S100P + TFF1 + tumor cells. Its role is to activate the TGF β 2 receptor on the same tumor cells, hence confirming the unique relationship

between this ligand–receptor pair (Fig. 6F). The findings consistently confirmed our ST-level results, highlighting the interaction between TGF β 2 and TGF β 2 on S100P + TFF1 + tumor cells. These results highlight the crucial importance of TGF β signaling in the self-communication of S100P + TFF1 + tumor cells.

To elucidate the impact of S100P + TFF1 + tumor cells on TME, we utilized datasets from 31 NSCLC samples available in the GEO database, specifically GSE148071, GSE127465, GSE143423, and EMTAB6149. First, we calculated the proportion of S100P + TFF1 + tumor cells within the tumor cell population and divided all samples into high and low groups based on the median proportion of these cells. We then compared the proportions of various immune and stromal cells between the high and low groups. Our analysis revealed distinct landscapes of TME components between the two groups (Figure S1). Specifically: The proportion of T cells was significantly lower in the high group (mean=0.27) compared to the low group (mean=0.44). There was a higher infiltration of myeloid cells in the high group (mean=0.28) compared to the low group (mean=0.19). Fibroblasts exhibited greater infiltration in the high group (mean=0.08) compared to the low group (mean=0.02) (Figure S1). No significant differences were observed in the proportions of B cells, endothelial cells, and tumor cells between the two groups. To further explore cell–cell communication within the TME, we performed CellChat analysis focusing on the interactions between T cells/myeloid cells and S100P + TFF1 + tumor cells. Our findings highlighted the TGF β signaling pathway as a key interaction mechanism. Specifically, the TGF β 2–TGF β 2 ligand–receptor pair was predominantly involved in cell–cell communication. TGF β has been reported to inhibit immune function, which is consistent with our observation of decreased T cells and increased myeloid cells in the TME of high S100P + TFF1 + tumor cell samples (Figure S2).

Discussion

STAS has been implicated in facilitating tumor advancement, dissemination, and resistance to treatment. Nevertheless, the examination of STAS and its spatial arrangement in NSCLC is still restricted. This work aimed to analyze and describe the specific characteristics of STAS in NSCLC. As a result, we successfully identified S100P as a biomarker for STAS. The presence of the S100P + TFF1 + subtype of STAS tumor cells had a negative impact on survival time and was linked to poorer results in both EGFR-TKI therapy and immunotherapy. Analysis of spatial transcriptomics revealed that the TGF β signaling pathway had the highest level of activation in S100P + TFF1 + tumor cells. The results of our research offer vital knowledge on STAS, which

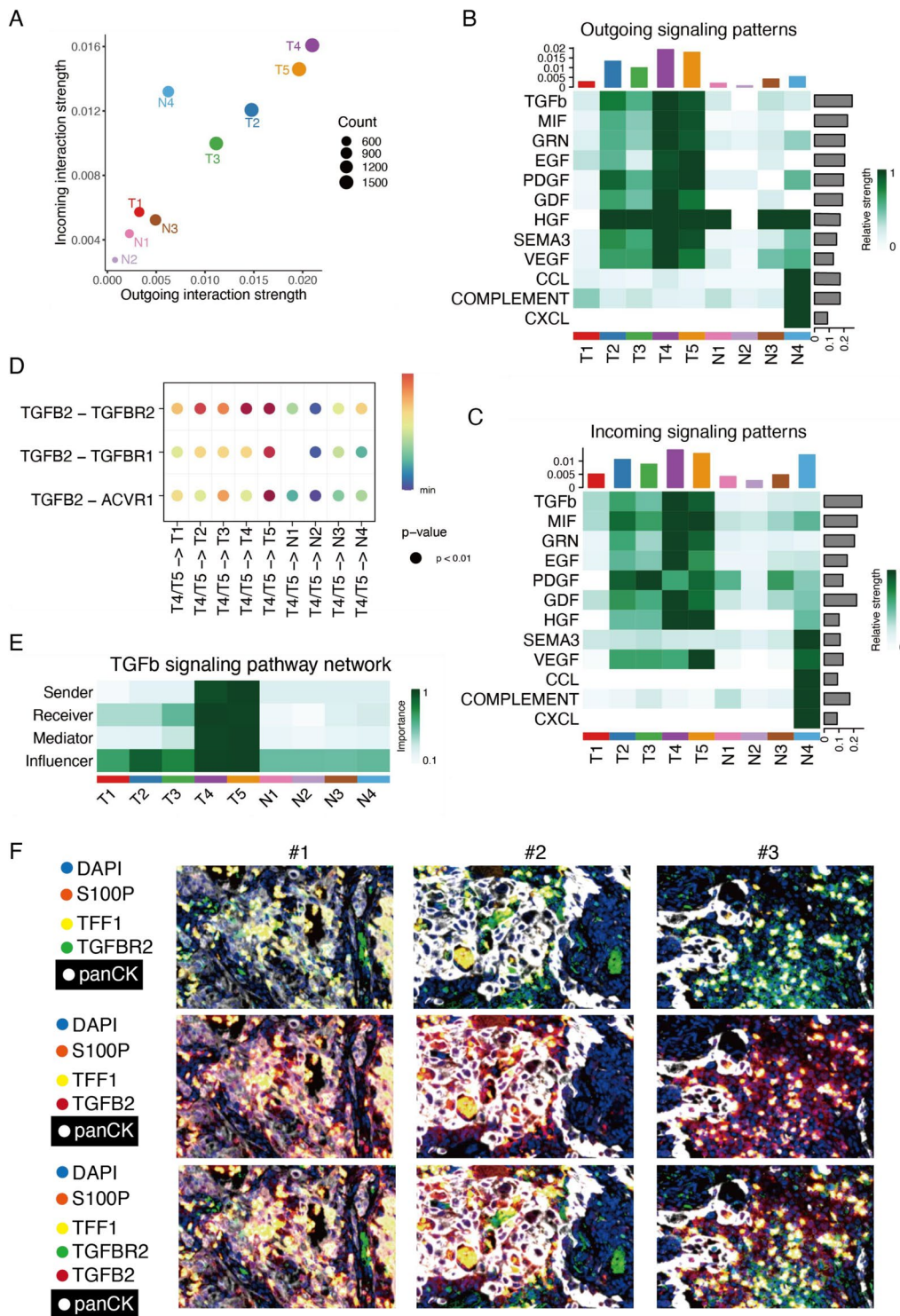


Fig. 6 TGFb signaling in self-communication of S100P+TFF1+tumor cells. **A** Role of spread through air spaces (STAS) tumor subclusters in cell-cell communication. **B** Outgoing signaling pattern of all cell clusters. **C** Incoming signaling pattern of all cell clusters. **D** Ligand-receptor pairs in TGFb signaling pathway signaled from T4/T5 to other clusters. **E** TGFb signaling pathway network. **F** MIF image revealing the TGFb2-TGFR2 ligand-receptor pair in the self-cell-cell communication of S100P+TFF1+tumor cells

can be used to direct future experiments and identify biomarkers.

Initially, we examined the characteristics of tumor cells in the STAS sample and discovered that S100P could serve as a promising biomarker for STAS. The expression level of S100P mRNA is directly associated with the activation state of the PI3K/AKT pathway, which is a well-known mechanism implicated in facilitating the migration, invasion, proliferation, and resistance to therapy in different types of malignancies. S100P+ epithelial cells, which are linked to negative outcomes, are more abundant in advanced stages [23]. In vitro, it promotes cell proliferation, migration, and invasion. In vivo experiments have shown that elevated S100P expression significantly triggers cancer metastasis to the liver [24]. Increased S100P expression is strongly correlated with the metastatic spread of colorectal cancer and is associated with shorter metastasis-free survival periods [25]. Additionally, S100P plays a role in activating the transcription of SLC2A5, thereby promoting cancer cell dissemination in colorectal cancer [26]. S100P shows potential as a biomarker for the immunosuppressive microenvironment [27]. It is also identified as a distinctive indicator for intrahepatic cholangiocarcinoma, a cancer type characterized by a notable decrease in CD4 T cells, alongside an increase in CCL18 TAM and PD1CD8 T cells [28].

The pro-tumor role S100P plays in NSCLC is similar to that in other cancers by enhancing cell migration, invasion and metastasis. In vitro, S100P overexpression in less invasive lung cancer cells increased these traits, while its knockdown in highly invasive cells reduced them and reversed EMT. In vivo, S100P knockdown prevented metastasis of highly metastatic cells. These effects are mediated through S100P's interaction with integrin α 7, activating FAK and AKT pathways [29]. Blocking FAK or inhibiting AKT reduces S100P-induced migration and ZEB1 expression. Additionally, genes involved in the regulation of S100P translation also promote NSCLC metastasis. RBMS1, a gene coding for an RNA-binding protein, promotes NSCLC metastasis by enhancing S100P translation, correlating with increased lymphnode metastasis and shorter survival [30]. S100P also stimulates tumor cell proliferation by binding to the receptor for advanced glycation end products (RAGE), activating MAP kinase and NF κ B pathways. RAGE is linked to metastasis and poor prognosis in various cancers. However, its role in NSCLC is complex, as it both inhibits growth through p21CIP1 suppressing CDK2 activity and promotes metastasis via ERK signaling, further accelerating tumor growth by inducing tumor-associated macrophage accumulation [31].

Moreover, our study explored the heterogeneity of STAS tumor cells and identified S100P+TFF1+tumor cells in the STAS sample as a subtype associated with adverse survival outcomes, as validated by public datasets. Subsequent validation using samples from clinical trials confirmed that this subtype of S100P+TFF1+tumor cells is linked to poorer therapeutic responses and outcomes in both EGFR-TKI therapy and immunotherapy. TFF1, a constituent of the trefoil factor family peptides, has a vital function in preserving the integrity of mucous membranes and facilitating the restoration of epithelial tissue in different organs [32, 33]. Previous research using biochemical and genetic animal models has indicated that TFFs have tumor suppressor roles. However, current experimental and clinical investigations have provided compelling data suggesting that TFFs actually play a role in promoting the development of many solid tumors. TFF1 acts as a tumor suppressor in hepatocellular carcinoma by decreasing the amounts of nuclear β -catenin [34]. The absence of TFF1 in individuals with gastric cancer was linked to increased tumor invasiveness and poorer patient survival, especially in those who received curative surgery without additional treatment [35]. Furthermore, TFF1 was discovered to play a role in promoting the development of breast cancer by enhancing the levels of expression of cell cycle-regulatory molecules and transcription factors [36]. The investigation of the pathway network revealed that the overexpression of TFF1 controls the transcription factor FOXA2 in the luminal A subtype, which is a mechanism that contributes to the unique response to chemotherapy in this subtype [37]. In NSCLC, the functional role of TFF1 is also complex and context-dependent. Overexpression of TFF1 in NSCLC cells drove cell cycle transition, increased the proportion of cells in the S-G2/M phases, while simultaneously enhanced the apoptosis, resulting in a 19 to 25% decrease in proliferation and a 71 to 82% decrease in migration. These effects were restored by transfection with TFF1 siRNA [38]. Contrarily, in a KRAS-mutated NSCLC cell line, TFF1 knockdown inhibited cell proliferation and induced apoptosis [39].

As a plasma biomarker of cancer cells, TFF1 has the potential to predict prognosis and response of immunotherapy in a noninvasive way. When combined with pathological detection of S100P and TFF1 expression, it enhances the accuracy of evaluating patients' condition and facilitates the selection of effective treatment. Furthermore, conducting a plasma test prior to tissue biopsy could minimize unnecessary invasive procedures for patients. Future studies should encompass patients at different stages of treatment, different endpoints and cancer staging, incorporating not only IHC but also plasma

protein as experimental indices to validate the predictive value of these biomarkers.

In addition, our investigation revealed the intercellular communication of tumor cells that express both S100P and TFF1. Analysis of spatial transcriptomics showed that the TGF β signaling pathway had the highest level of activation in S100P+TFF1+tumor cells. The primary ligand-receptor pair responsible for cell-cell communication was TGFB2-TGFBR2. The TGFB2-TGFBR2 axis is involved in the advancement of cancer in solid tumors, indicating an essential pathway for intercellular communication between cancer cells and other constituents. TGF β exerts pro-tumorigenic effects through several key pathways, including the suppression of immunological function, stimulation of angiogenesis/lymphangiogenesis, and induction of EMT [40, 41]. TGF β suppresses numerous elements of both the innate and adaptive immune systems, hence establishing a conducive environment for tumor proliferation [42]. EMT, is a vital biological process in which cells derived from epithelial tissue acquire the traits of mesenchymal cells. EMT plays a crucial role in embryonic development and the healing of wounds [43]. TGF β upregulates various EMT-transcription factors such as SNAIL, resulting in a reduction in the expression of epithelial genes and an increase in the expression of mesenchymal genes [44]. The activation of the TGF-beta pathway in S100P+TFF1+tumor cells indicated that either S100P or TFF1 exerts functions through this signaling cascade, making this pathway holds the potential to be therapy target. Extensive anti-cancer interventions targeting TGF- β have been researched, including neutralizing antibodies, TGF- β inhibitors, ligand traps, vaccines, and other approaches. Some of them have passed clinical satge, such as Fresolimumab (GC1008), Galunisertib (LY2157299), Trabedersen (AP12009) [45]. In other studies, TFF1 was found to suppress EMT through inhibition of the TGF-beta pathway in gastric cancer as a tumor suppressor [46]. In breast carcinoma cells, TFF1 and TGF- β serve as downstream genes of estrogen receptor (ER) and mediate many growth effects of estrogen, which could be inhibited by curcumin [47].

Our study possesses numerous significant strengths in comparison to prior research. This study is the first to thoroughly describe the characteristics of STAS tumor cells in NSCLC and investigate their spatial arrangement. We have effectively determined and confirmed the specific subtype of STAS tumor cells, which allows us to identify and focus on the pro-tumor element inside the tumor microenvironment. It is crucial to recognize certain limitations when examining the role of S100P+TFF1+tumor cells in EGFR-TKI therapy and immunotherapy utilizing two real-world cohorts. However, this study has certain limitations. Due to the limited

clinical information and patient samples, we identified only one case of STAS, and the initial analysis of STAS spatial data was based on this single case. Nonetheless, the S100P+TFF1+subtype identified in our study was present in many samples from clinical cohorts and datasets. Given the uncertainty regarding the sampling location and the finding that S100p serves as a biomarker for STAS, these S100P+TFF1+samples may represent STAS or a mixture of primary lesions and STAS. Besides, despite identifying cell-cell communication involving S100P+TFF1+tumor cells, we did not determine the existence of a specific axis or signaling pathway between S100P and TFF1. Furthermore, the mechanism by which TFF1 functions in tumors via the TGF- β signaling pathway remains poorly understood, presenting significant obstacles to the identification of useful and effective drug targets. In conclusion, the specific functions of S100P and TFF1 in NSCLC remain insufficiently explored. Their interactions with tumor cells vary across cancer types, in vitro or in vivo environments, signaling pathways, and expression levels. In this study, we have presented our finding on the roles of these two markers in the tumorigenesis of NSCLC patients. Further experiments and studies with larger cohorts are necessary to elucidate the detail roles of S100P, TFF1 and S100P+TFF1+tumor cells in NSCLC.

Conclusions

This study demonstrated the phenotype of STAS in NSCLC, identifying S100P as its biomarker. S100P+TFF1+tumor cells as the subtype of STAS tumor cells exhibited the adverse role in survival time and associated with worse EGFR-TKI therapy and immunotherapy.

Materials and methods

Patient samples

Seventy pre-treatment patients at the Cancer Hospital, Chinese Academy of Medical Science in Beijing, China, provided FFPE NSCLC samples. The collection of these samples followed institutional ethical procedures and required informed agreement from the patients. The procedure obtained approval from the Ethics Committee of Institut Curie (No.23/262-4004). Out of the samples, 70 were employed for IHC, while spatial transcriptomic sequencing was performed on 15 selected samples.

Data and materials

Single-cell data from GSE148071, GSE127465, GSE143423, and EMTAB6149 were obtained from the GEO database [20, 21]. The clinical data and metadata that matched were obtained from the original trials. Furthermore, the researchers acquired the single-cell data

of 49 clinical samples taken from 30 patients with metastatic lung cancer both before and during EGFR-TKI targeted therapy [48]. The response status of these patients may be obtained from the original study. To conduct a comprehensive study, we obtained mRNA expression data and clinical information for patients with NSCLC from The Cancer Genome Atlas (TCGA).

Spatial transcriptomics sequencing

We acquired eight FFPE tissue blocks from individuals with cancer. The samples were mounted on IHC slides using FFPE sections that were five micrometers thick. Subsequently, the slides were subjected to incubation at a temperature of 42 °C for a duration of 2 h, followed by air drying at ambient room temperature. Subsequently, the slides were subjected to a further drying process for a duration of 3 h at a temperature of 60 °C. The H&E staining procedure utilized Hematoxylin (Dako, Part number S330930-2) and Eosin (Sigma-Aldrich, Product number HT110216). The staining duration was modified based on the specific tissue being stained. Approximately 100 µl of 85% glycerol (ThermoFisher, Catalog number 15514011) was poured, coverslips were placed on top, and tissue imaging was conducted. A beaker filled with Milli-Q water was employed to eliminate the coverslips.

The Visium slide was inserted into a cassette. Each well was treated with 100 µL of 0.1 N HCl (Sigma-Aldrich, Product number H1758) and incubated at 42 °C for 15 min. After the HCl was extracted, the decrosslinking buffer was introduced. The slide was subjected to incubation at a temperature of 95 °C for a duration of 1 h. The Pre-hybridization stage was performed in accordance with the instructions provided in The Visium Spatial Gene Expression for FFPE reagent kit (10×Genomics, User Guide CG000407 Rev C, human transcriptome Product number 1000338). Each well was supplemented with 100 µL of Pre-hybridization mix and then incubated at room temperature for a duration of 15 min. Following the incubation period, the Pre-hybridization mix was extracted, and 100 µL of Hybridization mix was introduced. The Visium slide was subjected to incubation with the Hybridization mix for the duration of one night at a temperature of 50 °C.

The user followed the instructions provided in the user guide of "Visium Spatial Gene Expression for FFPE reagent kit" (10×Genomics, User Guide CG000407 Rev C, mouse transcriptome Product number 1000339, human transcriptome Product number 1000338) for the remaining steps of library preparation, which include probe ligation, probe release and extension, probe elution, and FFPE library construction. The completed libraries underwent sequencing using the Novaseq6000 platform

from Illumina. The length of read 1 was 28 base pairs, while the length of read 2 was 91 base pairs.

Pathological annotations for HE images

Each spots within the Visium sections was separately annotated by two pathologists, Lin Li and Tongji Xie. The pathologists classified the spots into histological classifications, such as normal hepatocytes, tumor cells, stromal cells, and immune cells, using a coverage criterion of >50% specific to each cell type.

Clustering analysis of spatial transcriptomics

The gene-spot matrices obtained from the ST data were analyzed using the R Seurat tool. Normalization was accomplished using the SCTransform function in Seurat. Clustering analysis was performed inside each sample using tools such as FindVariableFeatures, FindNeighbors, and FindCluster.

Identification of malignant cells in spatial analysis

Spot scoring was performed using a collection of immune-related signatures that included pan-immune markers (PTPRC), pan-T cell markers (CD2, CD3D, CD3E, CD3G), B cell markers (CD79A, MS4A1, CD79B), and myeloid cell markers (CD68, CD14). The mean of these features was assigned as the immunity score for each spots. The cluster with the highest median immunological score was selected as the reference for inferCNV, based on the results of clustering. The inferCNV analysis was performed using the following parameters: `cut-off=0.1`, `cluster_by_groups=FALSE`, `denoise=TRUE`, `HMM=TRUE`, `analysis_mode="subclusters,"` and `tumor_subcluster_partition_method="random_trees."` The HMM_type is "i6." The Hidden Markov Model was used to evaluate the levels of CNV within spots. In order to differentiate between malignant and non-malignant spots, a hierarchical clustering analysis was performed using the inferCNV package with the random trees approach. This analysis resulted in the division of all observed spots into 8 distinct clusters. The spots that were used as a point of reference were clearly identified as "reference." During inferCNV analysis, a gene state of 3 signifies the absence of CNV variation, a state more than 3 indicates CNV amplification, and a state less than 3 indicates CNV deletion. The CNV score for each gene was determined by subtracting 3 from the absolute value of the gene state. The cluster CNV score was calculated by adding up the CNV scores for all genes. The identification of the tumor cluster was established using the utilization of CNV scores and pathological annotations.

Differential expression analysis and gene set enrichment analysis

We employed the FindMarkers function from the Seurat package, utilizing the MAST approach for differential expression analysis, to detect the DEGs between distinct groups. The run was conducted using a cutoff log fold change of 0.25. We utilized the GSEA function in the R package fgsea to assess the enrichment of cancer hallmark and Biological Process Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) genesets.

Transcription factor analysis

Our goal was to examine the distinct functionality of TFs. The Dorothea resource, which includes signed TF-target interactions, was used to infer TF activity [49]. In order to create TF regulons, we employed the 'dorothea regulon human' wrapper function from the 'dorothea' library and chose high-confidence TFs at levels 'A', 'B', and 'C'. The run_viper function was subsequently utilized to compute the activities of the regulons. Regulons were formed by analyzing the mRNA expression levels of each TF and its direct targets. We utilized the run_viper function to integrate the VIPER method with DoRothEA and estimate TF activity based on the Dorothea regulons.

MIF

The MIF panel, consisting of panCK (abcam, ab234297), S100P (abcam, ab124743), TFF1 (abcam, ab92377), TGFB2 (abcam, ab53778), and TGFBR2 (abcam, ab61213), was conducted following the instructions provided by the manufacturer (Akoya, 5-Color Multiple IHC Kit). In summary, sections of a formalin-fixed FFPE block were briefly treated with xylene to remove the paraffin and then soaked in ethanol to restore moisture. Following microwave antigen retrieval in a heated citric acid buffer with a pH of 6.0 for a duration of 10 min, the activity of endogenous peroxidase was inhibited using a 3% H₂O₂ solution for 10 min. Additionally, any nonspecific binding sites were blocked using goat serum for 10 min. The primary antibodies were placed in a humidified chamber at room temperature and incubated for 1 h. After that, they were treated with the appropriate secondary horseradish peroxidase-conjugated polymer. Each target was seen using a 1:100 dilution of fluorescein TSA Plus. Subsequently, the slide was subjected to microwave antigen retrieval in a heated citric acid solution (pH 6.0) to eliminate excess antibodies before proceeding to the next stage. Ultimately, the nuclei were then seen using DAPI, and the slices were covered with antifade mounting media.

Immunohistochemistry

Seventy samples of NSCLC were analyzed using IHC. The pathology department determined the histologic stage of all NSCLC tissues. The dewaxed slices were subsequently exposed to particular primary antibodies (S100P: abcam ab124743, TFF1: abcam ab92377) at a temperature of 4°C for the duration of one night. This was followed by incubation with a biotinylated secondary antibody (Proteintech, Wuhan, China) at room temperature for a period of 1 h. Positive staining was seen by utilizing DAB chromogenic reagent, and each section was subsequently counterstained with hematoxylin. Each sample was given a grade based on the level of staining intensity (0 = no staining; 1 = weak staining; 2 = moderate staining; and 3 = strong staining) and the percentage of stained cells (0 = 0%; 1 = 1–25%; 2 = 25–50%; 3 = 50–75%; 4 = 75–100%). The final score was determined by multiplying the staining intensity with the positive area score, which ranged from 0 to 12. Two expert pathologists, who were unaware of the clinical data, independently assessed the IHC results of the samples.

ELISA

ELISA is a method used to measure the levels of proteins in serum. Before treatment, around 5 mL of peripheral blood samples were obtained from the test individuals using sterile tubes without anticoagulants. The centrifugation process was performed on all samples at a speed of 4000 revolutions per minute for a duration of 10 min, maintaining a constant temperature of the room. Ultimately, the serum samples were preserved at a temperature of -20 °C until they were examined. The levels of Serum S100P and TFF1 were quantified using an Enzyme-Linked Immunosorbent Assay (ELISA) on a TECAN Freedom EVOLyzer-2 150 platform manufactured by Tecan in Männedorf, Switzerland. The S100P and TFF1 ELISA kit were acquired from Shanghai Tongwei Biotechnology Co. Ltd. (Shanghai, China). The absorbance of S100P and TFF1 was measured at 450 nm, as instructed by the manufacturer. Concentrations were then determined using a suitable calibration curve.

Survival analysis

The R package survival was utilized to perform survival analysis. The Cox proportional hazards model was utilized to compute the hazard ratio, along with a 95% confidence interval. Additionally, the survfit function was implemented to construct Kaplan–Meier survival curves. The "maxstat.test" function from the R package maxstat was used to dichotomize cell population infiltration or gene expression by testing all possible cutting points to find the highest rank statistic. The patients were

separated into two groups based on the selected maximal logarithm statistics. Subsequently, the two-sided log-rank test was utilized to compare the Kaplan–Meier survival curves. The Chi-squared test was utilized to examine the comparison of the response rate to immunotherapy treatment among different groups.

Collection of plasma samples from patients undergoing third-line therapy with EGFR-TKI

186 plasma samples were collected from 186 NSCLC patients who took part in the BPI-7711 phase I (NCT03386955) and phase IIa (NCT03812809) clinical trials⁴². These patients had non-small cell lung cancer (NSCLC) that was either locally progressed or had spread to other parts of the body (metastatic/recurrent). They also had a verified mutation in the EGFR gene called T790M. These patients had either suffered disease progression after being treated with first- or second-generation EGFR TKI therapy, or they had the T790M mutation from the beginning. The clinical trial database (<https://clinicaltrials.gov/>) provided comprehensive information regarding the clinical studies. The collection of blood samples was conducted using EDTA tubes. After centrifuging at 16,000 times the force of gravity (16,000 g) and a temperature of 4 °C for a duration of 10 min, the plasma was isolated and kept at a temperature of – 80 °C until it was needed.

The effectiveness of the treatment was assessed by oncologists and radiologists through clinical and radiological tests. Clinical responses were classified as complete response, partial response, stable disease, or advancing disease based on the Response Evaluation Criteria in Solid Tumours (RECIST) version 1.1. R was defined as the group of patients who achieved either complete response or partial response, while NR referred to those who experienced either advancing disease or stable disease. All experiments were approved by the Research Ethics Committee and conducted in accordance with the Declaration of Helsinki.

Dataset of the ORIENT-3 study

The ORIENT-3 phase 3 trial was carried out at 39 centers throughout China. It was an open-label, randomized controlled trial. Approval from the Ethics Committee was received from all participating centers, and written informed permission was obtained from all patients. The study followed the criteria of Good Clinical Practice and the Declaration of Helsinki. The trial was registered on ClinicalTrials.gov with the identifier NCT03150875. The primary endpoint was OS, which was defined as the time

from randomization to death from any cause within the whole study set. Patients who received anti-PD-1/PD-L1 medication before their disease progressed after randomization were not included in the entire analysis set of the docetaxel arm.

Transcriptome sequencing

Out of the 157 patients that were sequenced, 86 were in the sintilimab group and 71 were in the docetaxel group. Among these patients, 110 samples had available archival tumor tissue samples and validated RNA sequencing data. Specifically, 61 samples were from the sintilimab group and 49 samples were from the docetaxel group. These samples were included in the downstream analysis. The RNeasy FFPE Kit (Qiagen, Hilden, Germany) was used to extract RNA from FFPE baseline tumor tissues.

Analysis of intercellular communication

The R package CellChat was employed to analyze communication relationships and identify communicating molecules. CellChatDB.human enabled the examination of primary signaling inputs and outputs across all cell clusters. The netAnalysis_signalingRole_scatter function was used to determine the role of all cell types in the cell–cell communication network.

Quantitative analysis

The Mann–Whitney U test was conducted to examine the disparities between the two groups. The Spearman's correlation test was employed to evaluate the associations between two variables. A two-tailed P-value of 0.05 was deemed to be statistically significant. The entire data processing, statistical analysis, and charting operations were conducted using R 4.1.0.

Abbreviations

CNV	Copy number variation
DEGs	Differential expressed genes
ECM	Extracellular matrix
ELISA	Enzyme-linked immunosorbent assay
EMT	Epithelial-to-mesenchymal transition
FFPE	Formalin-fixed paraffin-embedded
GEO	Gene Expression Omnibus
GSEA	Gene set enrichment analysis
IHC	Immunohistochemistry
KEGG	Kyoto Encyclopedia of Genes and Genomes
GO	Gene Ontology
NR	Non-responders
MIF	Multiplex immunofluorescence
OS	Overall survival
RECIST	Response Evaluation Criteria in Solid Tumours
R	Responders
ST	Spatial transcriptomics
TCGA	The Cancer Genome Atlas
TF	Transcription factor
TME	Tumor microenvironment

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-024-05722-6>.

Supplementary Material 1. Figure S1. The proportion of all cell types in samples with high/low proportion of S100P + TFF1 + tumor cells. Figure S2. The ligand-receptor pairs in the cell–cell communication between S100P + TFF1 + tumor cells and stromal cells/myeloid cells/T cells in the TME. Table S1. The marker gene of STAS tumor cells. Table S2. The list of marker genes of T4/T5

Acknowledgements

Not applicable.

Author contributions

Conception/design: YKS, XHH; Methodology: GYF, TJX; Formal analysis: GYF, TJX; Writing–manuscript writing: GYF, TJX, MWY, LL, LT; Writing–revising & editing: GYF, MWY, TJX, XHH, YKS.

Funding

National Science and Technology Major Project for New Drug Development [2017ZX09304015, 2019ZX09201-002] and the National High Level Hospital Clinical Research.

Availability of data and materials

The spatial transcriptomics data in our study is available upon request.

Declarations

Ethics approval and consent to participate

All samples were collected with the approval of the ethics committee of the Cancer Hospital of the Chinese Academy of Medical Sciences (No. 23/262-4004) and following the principles outlined in the Declaration of Helsinki.

Consent for publication

All authors agree to publish.

Competing interests

The authors declare that they have no competing interests.

Received: 25 February 2024 Accepted: 2 October 2024

Published online: 09 October 2024

References

- Sung H, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021. <https://doi.org/10.3322/caac.21660>.
- Shiono S, Endo M, Suzuki K, Yanagawa N. Spread through air spaces affects survival and recurrence of patients with clinical stage IA non-small cell lung cancer after wedge resection. *J Thorac Dis*. 2020. <https://doi.org/10.21037/jtd.2020.04.47>.
- Kadota K et al. 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*. 2015. <https://doi.org/10.1097/jto.0000000000000486>.
- Dai C, et al. Tumor spread through air spaces affects the recurrence and overall survival in patients with lung adenocarcinoma >2 to 3 cm. *J Thorac Oncol*. 2017. <https://doi.org/10.1016/j.jtho.2017.03.020>.
- Li J, Wang Y, Li J, Cao S, Che G. Meta-analysis of lobectomy and sublobar resection for stage I non-small cell lung cancer with spread through air spaces. *Clin Lung Cancer*. 2022. <https://doi.org/10.1016/j.clc.2021.10.004>.
- Eguchi T, et al. 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*. 2019. <https://doi.org/10.1016/j.jtho.2018.09.005>.
- Zhou F, et al. Assessment of the feasibility of frozen sections for the detection of spread through air spaces (STAS) in pulmonary adenocarcinoma. *Mod Pathol*. 2022. <https://doi.org/10.1038/s41379-021-00875-x>.
- Walts AE, Marchevsky AM. Current evidence does not warrant frozen section evaluation for the presence of tumor spread through alveolar spaces. *Arch Pathol Lab Med*. 2018. <https://doi.org/10.5858/arpa.2016-0635-OA>.
- Stahl PL et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*. 2016. <https://doi.org/10.1126/science.aaf2403>.
- Vickovic S et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nat Methods*. 2019. <https://doi.org/10.1038/s41592-019-0548-y>.
- Kadota K, et al. 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*. 2019. <https://doi.org/10.1097/PAS.0000000000001285>.
- Li X, et al. GABRP sustains the stemness of triple-negative breast cancer cells through EGFR signaling. *Cancer Lett*. 2021. <https://doi.org/10.1016/j.canlet.2021.04.028>.
- Kaur S, Singh J, Kaur M. Multifaceted role of galectin-4 in cancer: a systematic review. *Eur J Clin Invest*. 2023. <https://doi.org/10.1111/eci.13987>.
- Yang Q, et al. Effects of FOXJ2 on TGF- β 1-induced epithelial-mesenchymal transition through Notch signaling pathway in non-small lung cancer. *Cell Biol Int*. 2017. <https://doi.org/10.1002/cbin.10680>.
- Chen C-C, et al. Temporal evolution reveals bifurcated lineages in aggressive neuroendocrine small cell prostate cancer trans-differentiation. *Cancer Cell*. 2023. <https://doi.org/10.1016/j.ccell.2023.10.009>.
- Zhu J, et al. Pan-cancer analysis of Krüppel-like factor 3 and its carcinogenesis in pancreatic cancer. *Front Immunol*. 2023. <https://doi.org/10.3389/fimmu.2023.1167018>.
- Werner Rönnerman E, et al. Trefoil factor family proteins as potential diagnostic markers for mucinous invasive ovarian carcinoma. *Front Oncol*. 2022. <https://doi.org/10.3389/fonc.2022.1112152>.
- Abou Khouzam R, et al. Hypoxia as a potential inducer of immune tolerance, tumor plasticity and a driver of tumor mutational burden: impact on cancer immunotherapy. *Semin Cancer Biol*. 2023. <https://doi.org/10.1016/j.semcancer.2023.11.008>.
- Ebrahimi N, et al. Harnessing function of EMT in cancer drug resistance: a metastasis regulator determines chemotherapy response. *Cancer Metastasis Rev*. 2024. <https://doi.org/10.1007/s10555-023-10162-7>.
- Wu F, et al. Single-cell profiling of tumor heterogeneity and the micro-environment in advanced non-small cell lung cancer. *Nat Commun*. 2021. <https://doi.org/10.1038/s41467-021-22801-0>.
- Zilionis R, et al. Single-cell transcriptomics of human and mouse lung cancers reveals conserved myeloid populations across individuals and species. *Immunity*. 2019. <https://doi.org/10.1016/j.immuni.2019.03.009>.
- Shi Y et al. Sintilimab versus docetaxel as second-line treatment in advanced or metastatic squamous non-small-cell lung cancer: an open-label, randomized controlled phase 3 trial (ORIENT-3). *Cancer Commun Lond Engl*. 2022. <https://doi.org/10.1002/cac2.12385>.
- Lu H, et al. Single-cell RNA-sequencing uncovers the dynamic changes of tumour immune microenvironment in advanced lung adenocarcinoma. *BMJ Open Respir Res*. 2023. <https://doi.org/10.1136/bmjresp-2023-001878>.
- Schmid F, et al. Calcium-binding protein S100P is a new target gene of MACC1, drives colorectal cancer metastasis and serves as a prognostic biomarker. *Br J Cancer*. 2022. <https://doi.org/10.1038/s41416-022-01833-3>.
- Ismail TM, Gross SR, Lancaster T, Rudland PS, Barraclough R. The role of the C-terminal lysine of S100P in S100P-induced cell migration and metastasis. *Biomolecules*. 2021. <https://doi.org/10.3390/biom11101471>.
- Lin M, et al. S100P contributes to promoter demethylation and transcriptional activation of SLC2A5 to promote metastasis in colorectal cancer. *Br J Cancer*. 2021. <https://doi.org/10.1038/s41416-021-01306-z>.
- Hao W, Zhang Y, Dou J, Cui P, Zhu J. S100P as a potential biomarker for immunosuppressive microenvironment in pancreatic cancer: a bioinformatics analysis and in vitro study. *BMC Cancer*. 2023. <https://doi.org/10.1186/s12885-023-11490-1>.

28. Song G, et al. Single-cell transcriptomic analysis suggests two molecularly subtypes of intrahepatic cholangiocarcinoma. *Nat Commun*. 2022. <https://doi.org/10.1038/s41467-022-29164-0>.
29. Ya-Ling H, et al. S100P interacts with integrin $\alpha 7$ and increases cancer cell migration and invasion in lung cancer. *Oncotarget*. 2015. <https://doi.org/10.18632/oncotarget.4987>.
30. Yu S, et al. RBMS1 coordinates with the m(6)A reader YTHDF1 to promote NSCLC metastasis through stimulating S100P translation. *Adv Sci*. 2024. <https://doi.org/10.1002/advs.202307122>.
31. Mei-Chih C, et al. RAGE acts as an oncogenic role and promotes the metastasis of human lung cancer. *Cell Death Dis*. 2020. <https://doi.org/10.1038/s41419-020-2432-1>.
32. Hoffmann W. Trefoil factors TFF (trefoil factor family) peptide-triggered signals promoting mucosal restitution. *Cell Mol Life Sci CMLS*. 2005. <https://doi.org/10.1007/s00018-005-5481-9>.
33. Jahan R, et al. Odyssey of trefoil factors in cancer: diagnostic and therapeutic implications. *Biochim Biophys Acta Rev Cancer*. 2020. <https://doi.org/10.1016/j.bbcan.2020.188362>.
34. Ochiai Y, et al. Trefoil factor family 1 inhibits the development of hepatocellular carcinoma by regulating β -catenin activation. *Hepatol Baltim Md*. 2020. <https://doi.org/10.1002/hep.31039>.
35. Soutto M, et al. Activation of β -catenin signalling by TFF1 loss promotes cell proliferation and gastric tumorigenesis. *Gut*. 2015. <https://doi.org/10.1136/gutjnl-2014-307191>.
36. Amiry N, et al. Trefoil factor-1 (TFF1) enhances oncogenicity of mammary carcinoma cells. *Endocrinology*. 2009. <https://doi.org/10.1210/en.2009-0066>.
37. Buache E, et al. Deficiency in trefoil factor 1 (TFF1) increases tumorigenicity of human breast cancer cells and mammary tumor development in TFF1-knockout mice. *Oncogene*. 2011. <https://doi.org/10.1038/nc.2011.41>.
38. Kentaro M, et al. TFF-1 functions to suppress multiple phenotypes associated with lung cancer progression. *Onco Targets Therapy*. 2021. <https://doi.org/10.2147/OTT.S322697>.
39. Daisuke M, et al. Reciprocal expression of trefoil factor-1 and thyroid transcription factor-1 in lung adenocarcinomas. *Cancer Sci*. 2020. <https://doi.org/10.1111/cas.14403>.
40. Batlle E, Massagué J. Transforming growth factor- β signaling in immunity and cancer. *Immunity*. 2019. <https://doi.org/10.1016/j.immuni.2019.03.024>.
41. Treford CB, Dagnino L, Di Guglielmo GM. Transforming growth factor- β in tumour development. *Front Mol Biosci*. 2022. <https://doi.org/10.3389/fmolb.2022.991612>.
42. Moo-Young TA et al. Tumor-derived TGF- β mediates conversion of CD4⁺Foxp3⁺ regulatory T cells in a murine model of pancreas cancer. *J Immunother*. 2009. <https://doi.org/10.1097/CJI.0b013e318189f13c>.
43. Chaffer CL, San Juan BP, Lim E, Weinberg RA. EMT, cell plasticity and metastasis. *Cancer Metastasis Rev*. 2016. <https://doi.org/10.1007/s10555-016-9648-7>.
44. Stemmler MP, Eccles RL, Brabletz S, Brabletz T. Non-redundant functions of EMT transcription factors. *Nat Cell Biol*. 2019. <https://doi.org/10.1038/s41556-018-0196-y>.
45. Byung-Gyu K, Ehsan M, Sung Hee C, James JIH, James JD. Novel therapies emerging in oncology to target the TGF- β pathway. *J Hematol Oncol*. 2021. <https://doi.org/10.1186/s13045-021-01053-x>.
46. Da-Young L, Moon-Young S, Eun-Hee K. Trefoil factor 1 suppresses epithelial-mesenchymal transition through inhibition of TGF- β signaling in gastric cancer cells. *J Cancer Prev*. 2021. <https://doi.org/10.15430/JCP2021.26.2.137>.
47. Zhi-Ming S, et al. Curcumin exerts multiple suppressive effects on human breast carcinoma cells. *Int J Cancer*. 2002. <https://doi.org/10.1002/ijc.10183>.
48. Maynard A et al. Therapy-induced evolution of human lung cancer revealed by single-cell RNA sequencing. *Cell*. 2020. <https://doi.org/10.1016/j.cell.2020.07.017>.
49. Holland CH, et al. Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biol*. 2020;21:36.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.