



Published in final edited form as:

J Thorac Oncol. 2021 April ; 16(4): 619–629. doi:10.1016/j.jtho.2020.12.005.

Accuracy and Reproducibility of Intra-Operative Assessment on Tumor Spread Through Air Spaces (STAS) in Stage 1 Lung Adenocarcinomas

Julian A. Villalba^{1,2}, Angela R. Shih^{1,2}, Treah May S. Sayo^{1,3}, Keiko Kunitoki⁴, Yin P. Hung^{1,2}, Amy Ly^{1,2}, Marina Kem¹, Lida P. Hariri^{1,2}, Ashok Muniappan^{2,5}, Henning A. Gaissert^{2,5}, Yolonda L. Colson^{2,5}, Michael D. Lanuti^{2,5}, Mari Mino-Kenudson^{1,2,*}

1. Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts, USA

2. Harvard Medical School, Boston, Massachusetts, USA

3. Lung Center of the Philippines, Metro Manila, Philippines

4. Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

5. Division of Thoracic Surgery, Massachusetts General Hospital, Boston, Massachusetts, USA

Abstract

Introduction: Tumor spread through air spaces (STAS) is associated with worse prognosis in early-stage lung adenocarcinomas, particularly in sublobar resection. Intraoperative consultation for STAS has been advocated to guide surgical management. However, data on accuracy and reproducibility of intraoperative assessment of STAS remains limited. We evaluated diagnostic yield, inter-observer agreement (IOA), and intra-observer agreement (ITA) for STAS detection on frozen section (FS).

Methods: A panel of three pathologists evaluated stage 1 lung adenocarcinomas ($n=100$) for the presence/absence of STAS and artifacts as reference. Five pulmonary pathologists independently reviewed all cases in two rounds, detecting STAS and/or artifacts in FS and the corresponding

*To whom correspondence should be addressed: Mari Mino-Kenudson, M.D., Department of Pathology, Massachusetts General Hospital, 55 Fruit St, Warren 122, Boston, MA 02114, Telephone: (617)-726-8026, Fax: (617)-726-7474, mminokenudson@partners.org.

Author Contributions:

Julian A. Villalba: Conceptualization, Methodology, Investigation, Data Curation, Formal analysis, Visualization, Writing - Original Draft preparation, Writing - Review & Editing.

Angela R. Shih: Investigation, Formal analysis, Validation, Visualization, Writing - Original Draft preparation, Writing - Review & Editing.

Treah May S. Sayo: Investigation, Data Curation, Writing - Review & Editing.

Keiko Kunitoki: Formal analysis, Validation, Writing - Review & Editing.

Yin P. Hung: Investigation, Writing - Review & Editing.

Amy Ly: Investigation, Writing - Review & Editing.

Marina Kem: Resources, Writing - Review & Editing.

Lida P. Hariri: Investigation, Writing - Review & Editing.

Ashok Muniappan: Investigation, Writing - Review & Editing.

Henning A. Gaissert: Investigation, Writing - Review & Editing.

Yolonda L. Colson: Investigation, Writing - Review & Editing.

Michael D. Lanuti: Investigation, Writing - Review & Editing.

Mari Mino-Kenudson: Conceptualization, Methodology, Project administration, Investigation, Data Curation, Formal analysis, Visualization, Validation, Writing - Original Draft preparation, Writing - Review & Editing, Supervision.

permanent (FSP) and non-frozen permanent (NFP) sections, with a consensus conference between rounds.

Results: FS showed a low sensitivity (44%), high specificity (91%), relatively high accuracy (71%), and overall ROC/AUC of 0.67 for detecting STAS. The average ITA was moderate for both STAS (κ_{mean} :0.598) and artifact (κ_{mean} :0.402) detection on FS. IOA was moderate for STAS ($\kappa_{\text{round-1}}$:0.453; $\kappa_{\text{round-2}}$:0.506) and fair for artifact ($\kappa_{\text{round-1}}$:0.300; $\kappa_{\text{round-2}}$:0.204) detection on FS. IOA for STAS improved in FSP and NFP, while ITA was similar across the section types. Upon multivariable logistic regression, the only significant predictor of diagnostic discordance was the presence of artifacts.

Conclusion: FS is highly specific but not sensitive for STAS detection in stage 1 lung adenocarcinomas. IOA on STAS is moderate in FS, and improved only marginally after a consensus conference, raising concerns regarding global implementation of intraoperative assessment of STAS and warranting more precise criteria for STAS and artifacts.

Keywords

Tumor Spread Through Air Spaces (STAS); Lung adenocarcinoma; Frozen section; Reproducibility; Diagnostic yield

INTRODUCTION

Tumor clusters occupying air spaces in lung cancer were first identified in 1980.[1] The term “spread through air spaces” (STAS) was recently coined by *Kadota* et al,[2] who described STAS as isolated clusters of tumor cells in micropapillary clusters, solid nests, and/or single cells, spreading within air spaces beyond the edge of the main tumor. This original definition acknowledged that artifacts derived from tissue processing may mimic STAS,[2] with current ongoing debate regarding whether STAS represents a true biological phenomenon or an artefactual process.[3] Nevertheless, a compelling body of scientific evidence associates the presence of STAS in lung adenocarcinoma with lower recurrence-free and overall survivals.[4] STAS has recently been introduced as a novel mechanism of air space invasion that is associated with worse prognosis, and is recognized as an exclusion criterion in adenocarcinoma in situ (AIS) and minimally invasive adenocarcinoma (MIA).[5]

Low-dose computed tomographic screening programs have enhanced detection of clinical stage 1A non-small cell lung cancer (NSCLC),[6][7] which are typically treated with lobectomy.[8] Sublobar resection (wedge resection and segmentectomy), however, has been utilized as an acceptable surgical alternative to preserve lung function[9] in both low- and high-risk patients with early-stage NSCLC,[10] although the evidence is still insufficient. While non-inferiority randomized trials comparing overall survival of lobar and sublobar resection in patients with small-sized (diameter \leq 2 cm) peripheral non-small cell lung cancers are ongoing (CALGB/Alliance 140503, JCOG0802/WJOG4607L), and will provide crucial data regarding this topic,[11][12] the non-anatomical nature of sublobar resection makes STAS an important variable.[7]

Recent retrospective clinical evidence suggests that lobectomy in patients with STAS-positive T1 lung adenocarcinoma may be associated with better survival outcomes than

sublobar resection.[13] Therefore, there is a growing consideration that frozen section (FS) with assessment of STAS may inform intra-operative surgical management (lobar versus sublobar resection).[9][13] Data on the accuracy and reproducibility of FS for detecting STAS intraoperatively remains limited to date. In this study, we evaluated the diagnostic yield and accuracy of FS for intraoperative detection of STAS. We also assessed the inter-observer agreement (IOA) and intra-observer agreement (ITA) among pathologists evaluating STAS in FS and identified factors associated with low agreement. To clarify the potential sources of inter- and intra-observer disagreement, we assessed IOA and ITA on STAS in frozen-section permanent (FSP: a formalin-fixed paraffin-embedded [FFPE] section corresponding to the frozen section) and non-frozen permanent (NFP: a FFPE section that had not been processed in FS) slides that provide better-preserved morphology.

MATERIAL AND METHODS

Study Population

This retrospective study was approved by the Institutional Review Board at Massachusetts General Hospital (MGH). The surgical pathology case files at Massachusetts General Hospital were reviewed to identify patients treated with resection for lung adenocarcinoma between January 1, 2010, and December 31, 2015. A panel of three pathologists selected 100 consecutive stage 1 lung adenocarcinomas that met the following inclusion criteria: 1) FS of the main tumor was obtained for intraoperative consultation (Supplementary Method 1); 2) all FS, FSP and all NFP slides were available for review, and 3) tumor slides (including FS, FSP and NFP sections) had adequate adjacent non-neoplastic lung parenchyma for STAS evaluation, as defined by lung parenchyma surrounding at least one-third of the entire circumference of the tumor with ample non-neoplastic parenchyma between the tumor edge and the tissue edge. During the study period, 734 patients with stage 1 lung adenocarcinoma underwent resection. Of those, the following cases were excluded from this study: 106 cases without FS assessment on the main tumor; 157 cases with missing or poorly-preserved FS slides; and 371 cases without adequate adjacent lung parenchyma in the FS, FSP and/or NFP slides. The remaining 100 cases formed the study cohort and were distributed evenly throughout the study period without marked temporal skew.

Pathologic stage was determined based on the eighth edition of the American Joint Committee on Cancer (AJCC) Staging Manual.[14] Demographic, clinical, radiological, pathological, and post-operative follow-up data were collected from the patients' electronic medical records.

Histologic Evaluation: Reference Diagnoses

The histological slides of each case were evaluated by a panel of three pathologists (JAV, TS and MMK). The panel recorded the presence of STAS and artefactual clusters as reference. STAS was defined according to published criteria[2] as tumor cells within air spaces in the lung parenchyma beyond the edge of the main tumor, and comprised any of three morphological patterns: micropapillary structures, solid nests, and/or single cells. Artefactual clusters were defined as follows: clusters of cells randomly scattered over tissue and/or at the edges of the tissue section; clusters of cells with jagged edges suggestive

of tumor fragmentation or edges of a knife cut during specimen processing; linear strips of cells that were lifted off of alveolar walls; and isolated group(s) of tumor cells distant from the main tumor without a continuum of airspaces containing intraalveolar tumor cells back to the tumor edge (Figure 1). The panel simultaneously reviewed all slides including FS and FSP slides and discussed in real-time using a multi-headed microscope to render a final integrated diagnosis in each case. Additional clinical and morphologic assessments included: tumor size; stage (pT and pN); invasive component size; extent of surgical procedure; predominant histologic pattern; percentages of histological patterns; tumor grade (Supplementary Method 2);[15] presence of pleural, lymphatic and vascular invasion and tumor necrosis; and quantity of STAS clusters.

Histologic Evaluation: Pathologist Observers

FS and FSP slides as well as one NFP slide from all study cases were independently reviewed by five pulmonary pathologists (ARS, TS, YPH, AL, and MMK) blinded to clinicopathologic data four months after initial review by the panel. Each observer independently reviewed separate sets of FS, FSP and NFP slides (each differently arranged by JAV), and recorded the presence of STAS and artefact per published criteria in two sequential rounds separated by at least six weeks and by an intervening consensus conference. The observers were blinded to one another and to the final integrated diagnosis rendered by the panel. During each round the observers were asked to dichotomize cases as having STAS or no STAS. After generating the binary output, they were permitted to re-classify cases as “equivocal-STAS” if they felt the cases did not meet the published criteria. Additional information on histologic evaluation and data collection is available in the Supplementary Method 3.

Statistical analysis

Intraoperative diagnostic accuracy was determined by comparing the diagnosis of STAS made on the FS slide with the final integrated diagnosis based on review of all permanent section slides by the panel. Receiver Operator Curves (ROC) were used to determine the sensitivity, specificity, likelihood-ratio, negative predictive value (NPV), and positive predictive value (PPV) of the FS slide for detection of STAS.

IOA was evaluated in both rounds of slide evaluation with *Fleiss' kappa* coefficients and was calculated by comparing the agreement between observers in identifying the presence of STAS or artefactual clusters in FS, FSP, and NFP slides. ITA was similarly assessed. Statistical analysis was performed using Graph Pad Prism 7.0 (GraphPad Software, Inc. La Jolla, CA), as well as online IOA calculators.[16]

We performed a root cause analysis (RCA) to identify possible variables associated with inter-observer disagreement (Supplementary Method 4). All cases were classified into two categories: a full-agreement group (in which all 5 observers made the same diagnosis) and a controversy group (in which at least one observer disagreed) for each of the two rounds of FS evaluation. Potential variables were analyzed based on the difference in proportion between the two groups by univariate and multivariable logistic regression models. All

multivariable logistic regression analyses were performed with R statistical computing software (version 3.6.1). Additional information is available in the Supplementary Method 4.

RESULTS

Patient Characteristics–Clinicopathologic features

The study patient cohort (n=100) had a mean age of 69 years (standard deviation [SD]: 10 years), and 67% were female. Among these patients, 57 patients underwent limited resection (wedge [n=47], segmentectomy [n=10]); the remainder 43 patients, lobectomy or other anatomical resections (wedge resection with completion lobectomy [n=15]).

STAS Diagnosis

The panel of three pathologists identified STAS clusters in 43 (43%) tumors ranging from 0.4 cm to 3.5 cm in overall dimension, 0.1 cm to 2.8 cm in invasive size, and across different tumor grade categories and different histologic subtypes of adenocarcinoma (Table 1). STAS positive tumors had significantly larger entire and invasive tumor sizes ($p=0.04$ and $p=0.0001$, respectively), higher tumor grade ($p=0.0001$), and higher likelihood of lymphatic vessel invasion ($p=0.01$) than STAS negative tumors. STAS positive tumors were also more likely to harbor *KRAS* mutations ($p=0.002$) (Table 1). STAS was present in 19 of 43 (44.2%) FS slides and in 26 of 43 (60.4%) FSP slides. Of those, 2 cases (4.7%) showed STAS only in FS and FSP slides, and 5 cases (11.6%) showed STAS only in FSP slides, but not in any other histology sections.

The five observers independently evaluated FS, FSP and NFP slides from 100 lung adenocarcinomas as previously described. There was a high variability in the prevalence of STAS in FS, FSP and NFP slides reported by all observers in both the first round (range, FS: 20–44%; FSP: 23–48%; NFP: 33–60%) and the second round (range, FS: 28–41%; FSP: 22–46%; NFP: 35–58%). The identification of artifact in FS, FSP, and NFP slides was also variable in round 1 (FS: 40–68%; FSP: 54–77%; NFP: 29–62%) and round 2 (FS: 27–79%; FSP: 47–74%; NFP: 30–58%) (Supplementary Figure S1). In round 1, two or more observers categorized 16%, 19% and 16% of cases as equivocal-STAS in FS, FSP and NFP slides, respectively. In round 2, 36%, 34% and 21% of cases were categorized as equivocal-STAS in FS, FSP and NFP slides, respectively. Details on equivocal-STAS cases in FS are available in the Supplementary Result 1 and Supplementary Table 1.

Diagnostic Accuracy of Intraoperative Consultation in STAS Diagnosis

The overall sensitivity and specificity of the FS were 44.2% (95% confidence interval [CI]: 29.1% to 60.1%) and 91.2% (95% CI: 80.7% to 97.1%), respectively. The overall PPV was 79.2% (95% CI: 60.7% to 90.4%), and NPV was 68.4% (95% CI: 62.1% to 74.1%). FS with STAS clusters had a positive likelihood ratio of 5.0 (95% CI: 2.0 to 12.4); whereas FS lacking STAS clusters had a negative likelihood ratio of 0.6 (95% CI: 0.5 to 0.8). Overall, most FS slides would be correctly classified, which is reflected in a 71% (95% CI, 61.1% to 79.6%) accuracy, and a ROC area under the curve (AUC) of 0.67 (95% CI, 0.56–0.78) (Figure 2A).

Regarding the performance of FS in the diagnosis of STAS across five pathologists, sensitivity ranged from 35% to 77%, specificity ranged from 77% to 91% (Supplementary Table S2), and ROC AUC ranged 0.63 to 0.80 (Figure 2B) in the first round of evaluation. This variability decreased in the second round of evaluation, with sensitivity ranging from 54% to 70%, specificity ranging from 81% to 91%, and ROC AUC ranging 0.71 to 0.74 (Figure 2C).

Intra-observer Agreement on STAS Diagnoses

Intra-observer data were assessed for each pathologist based on round 1 and 2 interpretations of the same cases in FS, FSP and NFP slides. The intra-observer concordance rates and the average *Cohen's kappa* (κ) statistics for the intra-observer analysis among the observers are shown in Table 2 and Supplementary Figure S2.

Cases interpreted as STAS positive on FS in round 1 were likely to receive the same diagnosis when interpreted by the same pathologist in round 2 (mean intra-observer concordance rate 77%). Pathologists' reproducibility when rendering the same STAS assessment twice on FS slides was slightly higher for cases initially interpreted as STAS negative (mean intra-observer concordance rate 85%). According to the Landis and Koch classification,[17] the average intra-observer concordance for the diagnosis of STAS (mean concordance rate, FS: 82.4%; FSP: 84.4%; NFP: 81.2% and mean κ , FS: 0.598; FSP: 0.654; NFP: 0.622) was moderate for FS and substantial for FSP and NFP. However, intra-observer concordance for the detection of artifacts (mean concordance rate, FS: 71.0%; FSP: 76.4%; NFP: 74.6% and mean κ , FS: 0.402; FSP: 0.497; NFP: 0.459) were consistently lower than the ones for STAS across the section types (Table 2).

Inter-observer Agreement on STAS Diagnoses

The IOA among the five pathologists for STAS evaluation was moderate (κ :0.453) in FS, slightly higher (κ :0.477) in FSP and highest (κ :0.585) in NFP in round 1, all of which slightly increased in round 2 after the consensus conference (Table 3). Conversely, the IOA for the detection of artefactual clusters among the observers was fair in round 1 across the section types (κ :0.300, 0.377 and 0.303 in FS, FSP and NFP, respectively) and remained low in round 2 of evaluation despite the intervening consensus conference. More detailed results with pair-wise comparison are available in the Supplementary Table S3.

Association between Clinicopathologic Factors and Inter-observer Agreement

In the first round of evaluation, 52% of cases showed full agreement among all the observers, most of which (42/52; 81%) still had full agreement after the second round of evaluation. Most of the cases that showed no variation in IOA between rounds ($R1-R2=0$) were called STAS negative by all the observers (34/42; 81%) (Figure 3).

We performed univariate analyses of variables including those that may have led to interobserver disagreement on STAS diagnosis in FS slides for each round (Supplementary Table S4). Prevalence of these variables, as either categorical (0 vs. 1) or continuous data, was compared between full-agreement and controversy groups in the univariate analysis that showed multiple variables to be significantly associated with interobserver discordance.

Overall, larger invasive size, higher histologic grade, presence of micropapillary pattern (5%), larger quantity of STAS clusters, presence of artefactual clusters (detected by three or more observers), and final integrated diagnosis of STAS determined by the panel were associated with discrepant diagnosis on STAS in FS among the observers.

In multiple multivariable logistic regression models that included all the variables significantly associated with the interobserver discordance by univariate analyses, only the presence of artefactual clusters recorded by three or more observers showed a significant independent association with increased diagnostic disagreement on STAS in both rounds (Supplementary Table S5). The odds of disagreement among slides with artefactual clusters (identified by three or more observers) were 5.84 and 7.71 times higher than those without artefactual clusters ($p<0.001$ and $p<0.001$) in rounds 1 and 2, respectively.

DISCUSSION

In this study, STAS clusters in lung adenocarcinoma could be recognized with a high degree of specificity on FS,[13][18] a finding that may inform intraoperative treatment decisions regarding extent of resection. However, FS appeared insensitive for detecting STAS, due in part to sampling error as shown in our stage I lung adenocarcinoma cohort, in which adequate adjacent lung parenchyma to appropriately evaluate STAS was not present in the majority of FS slides; as such, the absence of STAS clusters at intraoperative consultation does not necessarily indicate the absence of STAS after complete tumor assessment. These findings are consistent with *Waltz et al.*, who noted low sensitivity (47.9%), and high specificity (100%) for the detection of STAS at FS in a smaller study of stage 1 and 2 adenocarcinoma.[18] STAS assessment in the intraoperative setting should thus be interpreted with caution when thoracic surgeons consider the clinical utility of proceeding with completion lobectomy. Pathologists should also be aware and communicate the low sensitivity of FS in detecting STAS to their surgeons, and re-review the FS protocol to ensure that ample adjacent lung parenchyma be included along with tumor in the section.

We found that overall ITA and IOA were moderate-to-substantial for STAS and fair-to-moderate for artifact detection on FS, FSP, and NFP among pulmonary pathologists. The mean ITA for STAS detection in FS (range:78–88%), FSP (range:79–91%) and NFP (range:73–91%) was relatively high, which were reflected in moderate-to-substantial kappa values for overall ITA (FS κ :0.598; FSP κ :0.654; NSP κ :0.622). Most cases (42/52; 81%) that showed full IOA in FS in the first round were consistently called STAS negative or positive by all observers with 100% agreement in the second round (Figure 3). Only a small number of cases (19%) with perfect agreement in the first round showed lower agreement in the second round. Most discrepancy in STAS diagnosis in the second round occurred in cases flagged in round 1 as exhibiting a low IOA. Notably, this study is from a tertiary care center with over 300 lung cancer resections annually, and all lung cancer specimens, intraoperative consultations for lung resections, and the study sets herein were reviewed by pulmonary pathologists. ITA and IOA on STAS diagnosis in FS would therefore not likely be better when generalized to other less subspecialized practices.

In an effort to identify factors contributing to discrepancy, a consensus conference was held amongst the panel as well as all observers, with targeted discussion of discrepant cases with low IOA and difficult cases designated equivocal-STAS by more than two observers in the first round. Targeted discussion of complex cases at consensus conference was insufficient to change observer opinions, with only marginal improvement of IOA across section types; however, IOA was higher in FSP and highest in NFP, likely due to better quality NFP sections. While some observers continued to confidently interpret cases relatively uninfluenced by consensus discussion, others may have modified their diagnostic thresholds, resulting in a somewhat lower ITA and an increase in number of cases classified as equivocal-STAS. Currently, the clinical impact of variable ITA on intraoperative patient management remains unknown and requires further investigations.

The relatively high ITA on the diagnosis of STAS and the marginal improvement of IOA after the consensus conference across the section types suggests that discrepant diagnosis is largely based on a true difference in interpretative opinion. In contrast, *Eguchi et al* reported a modestly higher agreement (categorized as substantial) on STAS diagnosis in 48 FS slides among thoracic pathologists using the Gwet's AC1 statistic.[13] The difference in agreements between the two studies may be attributed to the differences in the study cohort size, statistical methods applied, and possible selection bias, as their study did not specify if cases were consecutive. Of note, the Fleiss' kappa statistic used in the current study is a statistical measure for assessing the reliability of agreement among multiple raters,[19] and significant differences in coefficients among studies using different estimators have been previously reported and may impact direct comparison of different agreement scales.[20] Further, it may raise the possibility of interinstitutional differences in recognition of STAS and FS procedure as additional complicating factors in consistent diagnosis.

Our observers reported a high prevalence of artefactual tumor clusters in our samples (Supplementary Figure S1), although it was unclear whether the high prevalence was due to procedural manipulation of fresh (unfixed) tumor tissue at the time of frozen section and/or surgical manipulation in the operating room.[3][4] For example, tissue extracted during minimally invasive resection through small incisions from the pleural space is compressed, potentially dislodging individual tumor cells or cell clusters. Nevertheless, to our knowledge, this is the first study to report the prevalence of artefactual clusters in lung adenocarcinoma resections using published criteria.[2] We found fair-to-moderate ITA and IOA for the detection of artefactual clusters even after a consensus conference, highlighting the fact that observers have lower confidence on the detection of artificially detached tumor cell clusters compared to STAS across the section types. Even though three-dimensional (3D) reconstruction studies[21][22] have demonstrated the biological nature of STAS, not all "free-floating" clusters on 2D sections may in fact correspond to this phenomenon.[4] The presence of detached extraneous tissue derived from the same patient sample is a well-defined artefactual phenomenon in surgical pathology.[23] The term "spreading through a knife surface" (STAKS) has been defined as a process secondary to lung tissue dissection by a knife along the plane of sectioning and manipulation with processing, which can lead to an artefactual displacement of tumor cells. Depending on their sizes, loose extraneous tissue fragments (also known as "floaters") may be displaced into airways, alveolar ducts, and/or alveoli, mimicking STAS.[3][4] Our data suggest a critical need to minimize the artefactual

clusters and to further refine and standardize their histopathological definitions, in order to more consistently and reliably distinguish artifact from bona fide STAS.

Interestingly, 81% of the cases (42/52) that showed full agreement in the first round of FS assessment were called STAS negative, and of those cases, 81% (34/42 cases) were re-assessed as STAS negative by all observers in the second round. The STAS negative tumors were more likely to be lepidic predominant with a small invasive size and lacked micropapillary or high-grade patterns. In this context, the high reproducibility of STAS negative diagnoses could be attributed to an inherent observer bias that low-grade tumors are less likely to harbor STAS. Given that these grade 1 tumors were also found to have less artifacts compared to grade 2–3 tumors in FS (data not shown), another contributing factor could be that the low-grade tumors may be biologically cohesive and therefore resistant to ex vivo artifacts by knife cuts.[4]

Furthermore, we found an observer-dependent variability in the diagnostic yield of FS for STAS detection. This was reflected in a variable ROC AUC among our pathologist observers (range: 0.63–0.80) in the first round of evaluation and is similar to those previously described. *Eguchi* et al. also reported a high variability in the sensitivity (range: 52–86%) and specificity (74–100%) of STAS detection in 48 FS slides among five different observers. [13] Nonetheless, it is important to note that the specificity of FS to detect STAS remained high across the observers herein (80%–90%), while the presence of artifacts hampered IOA in our study. Besides, 60% of cases that were unanimously interpreted as STAS positive in FS exhibited ≥ 10 tumor clusters in the FS slide (data not shown), implying that the pathologist, irrespective of experience, could identify STAS clusters even in the background of extensive artifact, as long as a large quantity of tumor clusters fulfilling criteria for STAS were present.

While the question of whether variability in recognition of STAS at the time of FS would have a direct impact on patient care and outcomes remains to be addressed, proper communications between the pathologists and surgeons is paramount, particularly in conveying degrees of certainty/uncertainty in STAS assessment.[24][25] Given that the vast majority of cases categorized as equivocal-STAS by multiple pathologists led to disagreement (data not shown), the use of a three-tier system (STAS vs. equivocal-STAS vs. No STAS) instead of a two-tier system (STAS vs. No STAS) may be considered as an alternative to convey the information including diagnostic confidence. Additionally, clear communication of other high-grade features that also portend poor prognosis is critical in guiding proper intraoperative management (appropriate nodal staging), particularly when there is uncertainty regarding the presence of STAS. Finally, surgeons may need to explain to their patients that not all information required in determining when to avoid sublobar resection is available on FS.

In conclusion, we demonstrate that FS is a highly specific but not a sensitive method for STAS detection in stage 1 lung adenocarcinoma, and interpretation is limited by overall moderate ITA and IOA for STAS detection in FS slides. As current accepted definitions for STAS and artefactual clusters are variably interpreted by pathologists, more precise criteria should be established and standardized, possibly via web-based or in person tutorials,

before the assessment of STAS can be implemented globally in the intraoperative setting to aid surgical decision making. Meanwhile, we recommend candid acknowledgement of the inherent limitations of FS and of the pathologists' ability to reliably report STAS intraoperatively, considering a technique prone to produce prevalent artifactual clusters, as well as a low sensitivity and IOA. Future development of objective and reliable digital techniques (such as whole slide imaging platforms and artificial intelligence) may aid pathologists' visual detection of STAS to improve accuracy and reproducibility of STAS assessment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENT

The authors thank Mayerling Dada, Jaimie Barth and Adriana Alvarez for assisting with administrative tasks and with the biobank slide repository of the Department of Pathology at Massachusetts General Hospital.

Role of the funding source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of Interest and Sources of Funding:

LPH reports grants from Boehringer Ingelheim and has received personal consulting fees from Boehringer Ingelheim, Pliant Therapeutics, and Biogen Idec, all of which are not related to this work. YLC reports grants and a patent from Canon USA, all of which are not related to this work. MMK is partially supported by the National Institutes of Health (R01CA240317). MMK has served as a compensated consultant for H3 Biomedicine and AstraZeneca and has received research (institutional) funding from Novartis, all of which are not related to this work. All other authors declare no competing interests.

REFERENCES

- [1]. Kodama T, Kameya T, Shimosato Y, Koketsu H, Yoneyama T, Tamai S. Cell incohesiveness and pattern of extension in a rare case of bronchioloalveolar carcinoma. *Ultrastruct Pathol.* 1980 Apr-Jun;1(2):177–88. PMID: 6262967 [PubMed: 6262967]
- [2]. Kadota K, Nitadori JI, Sima CS, Ujiie H, Rizk NP, Jones DR, 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 May;10(5):806–814. PMID: 25629637 [PubMed: 25629637]
- [3]. Thunnissen E, Blaauwgeers HJ, de Cuba EM, Yick CY, Flieder DB. Ex Vivo Artifacts and Histopathologic Pitfalls in the Lung. *Arch Pathol Lab Med.* 2016 Mar;140(3):212–20. [PubMed: 26927715]
- [4]. Shih AR, Mino-Kenudson M. Updates on Spread Through Air Spaces (STAS) in Lung Cancer. *Histopathology.* 2020 Jan 14. doi: 10.1111/his.14062. PMID: 31943337
- [5]. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol.* 2015 Sep;10(9):1243–1260. PMID: 26291008 [PubMed: 26291008]
- [6]. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med.* 2011 Aug 4;365(5):395–409. PMID: 21714641 [PubMed: 21714641]

- [7]. Cao C, Tian DH, Wang DR, Chung CD, Gossot D, Bott M. Sublobar resections-current evidence and future challenges. *J Thorac Dis.* 2017 Dec;9(12):4853–4855. PMID: 29312675 [PubMed: 29312675]
- [8]. Ginsberg RJ, Rubinstein LV. Randomized trial of lobectomy versus limited resection for T1 N0 non-small cell lung cancer. Lung Cancer Study Group. *Ann Thorac Surg.* 1995;60(3):615–623. doi:10.1016/0003-4975(95)00537-u. PMID: 7677489
- [9]. Kodama K, Higashiyama M, Okami J, Tokunaga T, Imamura F, Nakayama T, et al. Oncologic Outcomes of Segmentectomy Versus Lobectomy for Clinical T1a N0 M0 Non-Small Cell Lung Cancer. *Ann Thorac Surg.* 2016;101(2):504–511. doi:10.1016/j.athoracsur.2015.08.063. PMID: 26542438 [PubMed: 26542438]
- [10]. Cao C, Tian DH, Wang DR, Chung CD, Gossot D, Bott M. Sublobar resections-current evidence and future challenges. *J Thorac Dis.* 2017 Dec;9(12):4853–4855. PMID: 29312675 [PubMed: 29312675]
- [11]. Brown LM, Louie BE, Jackson N, Farivar AS, Aye RW, Vallières E. Recurrence and Survival After Segmentectomy in Patients With Prior Lung Resection for Early-Stage Non-Small Cell Lung Cancer. *Ann Thorac Surg.* 2016;102(4):1110–1118. doi:10.1016/j.athoracsur.2016.04.037. PMID: 27350237 [PubMed: 27350237]
- [12]. Altorki NK, Wang X, Wigle D, Gu Li, Darling G, Ashrafi AS, et al. Perioperative mortality and morbidity after sublobar versus lobar resection for early-stage non-small-cell lung cancer: post-hoc analysis of an international, randomised, phase 3 trial (CALGB/Alliance 140503). *Lancet Respir Med.* 2018;6(12):915–924. doi:10.1016/S2213-2600(18)30411-9. PMID: 30442588 [PubMed: 30442588]
- [13]. Nakamura K, Saji H, Nakajima R, Okada M, Asamura H, Shibata T, et al. A phase III randomized trial of lobectomy versus limited resection for small-sized peripheral non-small cell lung cancer (JCOG0802/WJOG4607L). *Jpn J Clin Oncol.* 2010;40(3):271–274. doi:10.1093/jjco/hyp156. PMID: 19933688 [PubMed: 19933688]
- [14]. Eguchi T, Kameda K, Lu S, Bott MJ, Tan KS, Montecalvo J, 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 Jan;14(1):87–98. PMID: 30244070 [PubMed: 30244070]
- [15]. Amin MB, Cancer. *AJCo AJCC Cancer Staging manual.* New York: Springer; 2016.
- [16]. Moreira A, Ocampo PS, Xia Y, Zhong H, Russell PA, Minami Y, et al. A Grading system for invasive pulmonary adenocarcinoma: a proposal from the IASLC pathology committee. *J Thorac Oncol* 2020, In press.
- [17]. The Chinese University of Hong Kong. Statistics Toolkit (STATTOOLS). URL: <http://www.obg.cuhk.edu.hk/ResearchSupport/StatTools/index.php> [Accessed on April 13, 2020]
- [18]. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977 Mar;33(1):159–74. PMID: 843571 [PubMed: 843571]
- [19]. 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 Jan;142(1):59–63. PMID: 28967802 [PubMed: 28967802]
- [20]. Gwet KL. Testing the Difference of Correlated Agreement Coefficients for Statistical Significance. *Educ Psychol Meas.* 2016 Aug; 76(4): 609–637. PMID: 29795880 [PubMed: 29795880]
- [21]. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med Res Methodol.* 2013 Apr 29;13:61. PMID: 23627889 [PubMed: 23627889]
- [22]. Onozato ML, Kovach AE, Yeap BY, Morales-Oyarvide V, Klepeis VE, Tammireddy S, et al. Tumor islands in resected early-stage lung adenocarcinomas are associated with unique clinicopathologic and molecular characteristics and worse prognosis. *Am J Surg Pathol.* 2013;37:287–94. PMID: 23095504 [PubMed: 23095504]
- [23]. Yagi Y, Aly RG, Tabata K, Barlas A, Rekhman N, Eguchi T, et al. Three-Dimensional Histologic, Immunohistochemical, and Multiplex Immunofluorescence Analyses of Dynamic

Vessel Co-Option of Spread Through Air Spaces in Lung Adenocarcinoma. *Thorac Oncol.* 2020 Apr;15(4):589–600. PMID: 31887430

- [24]. Gephardt GN, Zarbo RJ. Extraneous tissue in surgical pathology: a College of American Pathologists Q-Probes study of 275 laboratories. *Arch Pathol Lab Med.* 1996 Nov;120(11):1009–14. PMID: 12049100 [PubMed: 12049100]
- [25]. Lindley SW, Gillies EM, Hassell LA. Communicating diagnostic uncertainty in surgical pathology reports: disparities between sender and receiver. *Pathol Res Pract.* 2014 Oct;210(10):628–33. PMID: 24939143 [PubMed: 24939143]
- [26]. Galloway M, Taiyeb T. The interpretation of phrases used to describe uncertainty in pathology reports. *Patholog Res Int.* 2011;2011:656079. PMID: 21876845 [PubMed: 21876845]

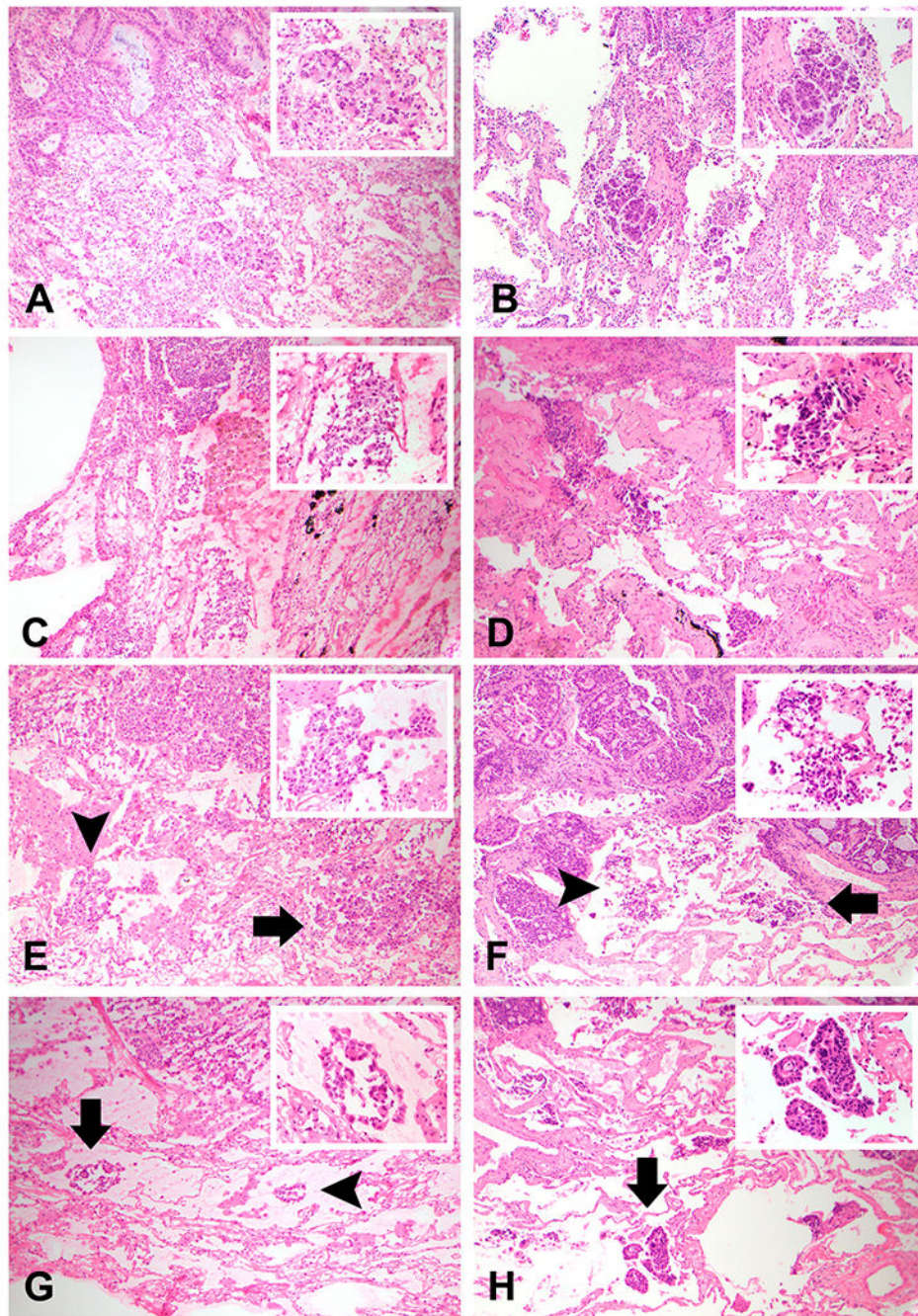


Figure 1. Comparison of STAS and Artifact on Frozen Section and Frozen Section Permanent Slides.

Examples highlighting the morphologic spectrum of STAS and artifact on Frozen Section (FS) and FS permanent (FSP) are shown, including foci considered to be equivocal for STAS. A FS showing a cohesive small tumor cluster spatially distant from the edge of the tumor (A) is confirmed on FSP (B) to be STAS in micropapillary clusters and small solid nests. Alternatively, FS showing an irregular tumor cluster with jagged and sharp contours (C) is confirmed to be artifact on FSP (D), which further highlights the irregular contours and dyshesive morphology. A third FS (E) shows a single focus of tumor that appears to be

micropapillary or solid STAS on the right (arrow), but also a more jagged focus of tumor on the left (arrowhead), which is highlighted on higher power in the inset and in isolation; this is considered to be equivocal for STAS on FS. Examination of the corresponding FSP (F) confirms the micropapillary STAS (arrow), but also reveals that the tumor cluster determined to be equivocal is likely a mixture of true micropapillary STAS and dyshesive irregular artifact that is intermixed (arrowhead and inset). Similarly, a fourth FS (G) shows a focus of linear epithelial strips definitive for artifact on the right (arrowhead), along with an indeterminate cluster composed of linear strips and small epithelial tufts (arrow and inset). Examination of the corresponding FSP (H) confirms that the focus determined to be equivocal on FS as definitive for STAS (arrow and inset).

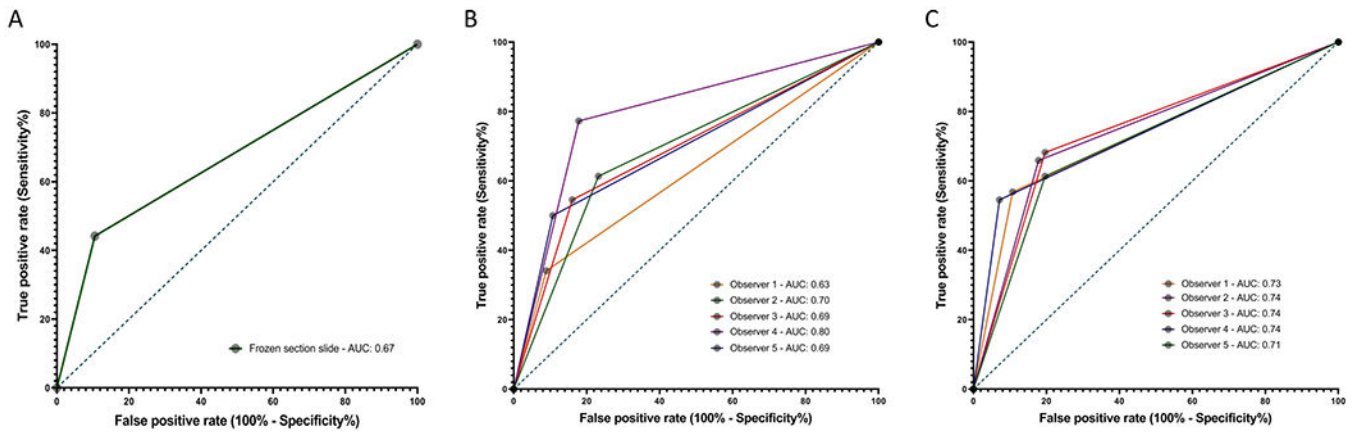


Figure 2. Receiver-Operating-Characteristic (ROC) Curves of Frozen section for STAS detection.

Diagnostic performance was determined by comparing the final integrated diagnosis based on review of all histology slides by the panel of three pathologists with: (A) the diagnoses obtained in Frozen section slides by the panel; (B) the individual diagnoses obtained by each of the 5 observers in Frozen section slides in the first round, and (C) those in the second round of evaluation.

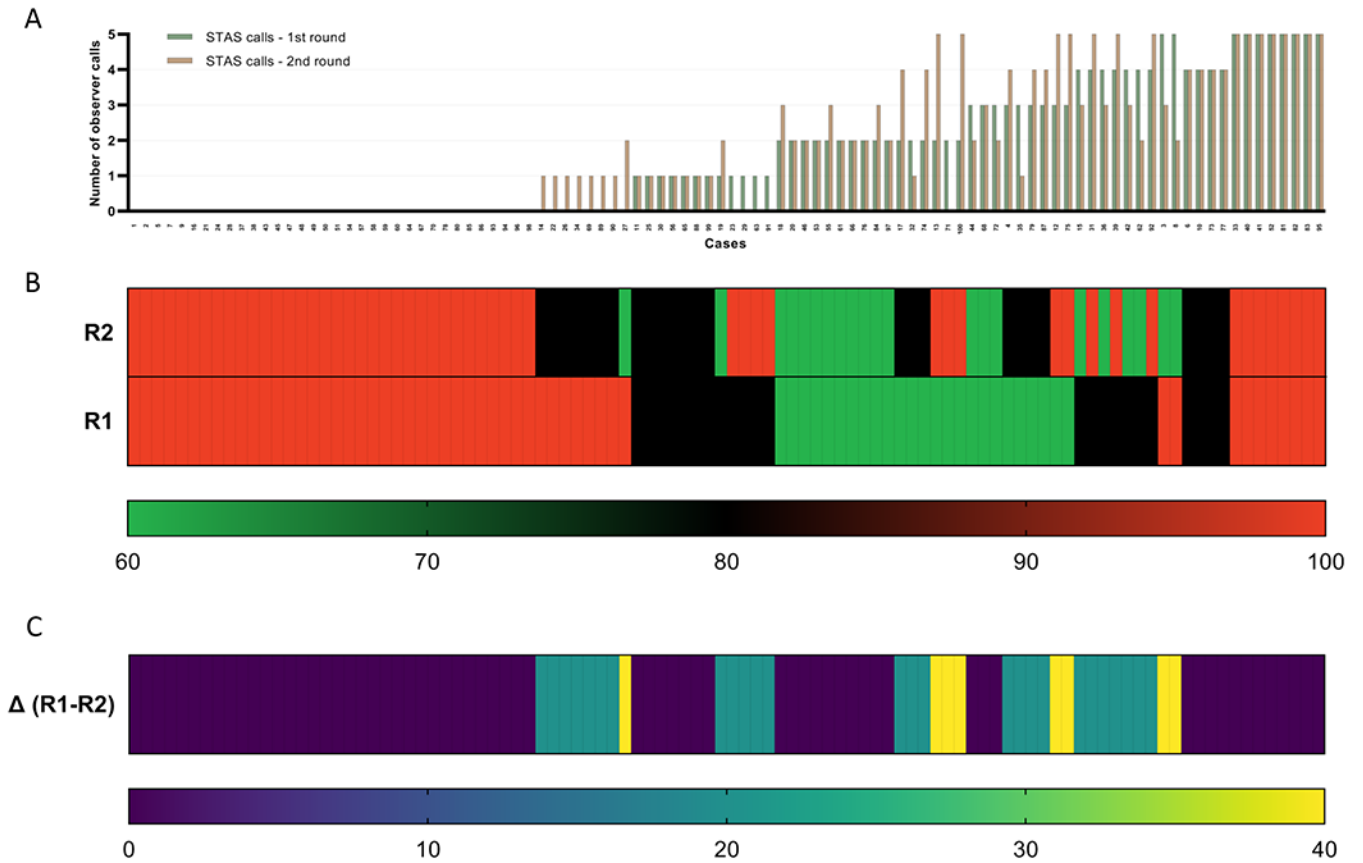


Figure 3. Variability in diagnosis of STAS across observers in Frozen section.
 (A) The number of observers identifying STAS for each case is presented for round 1 (R1), and round 2 (R2). The green bar represents the STAS calls in the round 1; the orange bars represent the STAS calls in the round 2. (B) Inter-observer concordance rate for each case is presented in the heat map for both rounds. Each column represents an individual case. The bottom row represents R1; the top row R2. The scale bar below (thin rectangle) indicates the level of inter-observer concordance. Red denotes full inter-observer concordance (100%); green, poor inter-observer concordance (60%) among the five observers. (C) Differences in inter-observer concordance rates between both rounds of evaluation (R1-R2) are presented in the heat map. Each column represents an individual case. The scale bar below (thin rectangle) indicates the degree of difference in inter-observer concordance between both rounds of evaluation. Yellow denotes a higher difference in concordance between rounds (R1-R2: 40%); blue, no difference (R1-R2: 0%). Cases are displayed in panels B and C to match the same sequence in which they are displayed in panel A.

Table 1:

Clinicopathological features in tumors with and without tumor spread through air spaces (STAS)

Clinicopathological characteristic of the resections		STAS (-) tumors N=57 n (%)	STAS (+) tumors N=43 n (%)	<i>p</i> -value ¹⁾	
Age at surgery (years; mean ± SD)		69.9 ± 9.8	67.7 ± 10.5	0.3	
Gender	Female / Male	42 (74) / 15 (26)	25 (54) / 18 (42)	0.1	
Tumor size	Entire size (cm; mean ± SD)	1.5 ± 0.6	1.7 ± 0.6	0.04	
	Invasive size (cm; mean ± SD)	0.8 ± 0.7	1.5 ± 0.6	<0.0001	
Type of resection				0.5	
	Lobectomy and other anatomical resections	13 (23)	15 (35)		
	Wedge resection + completion lobectomy	9 (16)	6 (14)		
	Segmentectomy	7 (12)	3 (7.0)		
	Wedge resection	28 (49)	19 (44)		
Morphologic patterns	Predominant histologic pattern			<0.0001	
		Lepidic	29 (51)	3 (7.0)	
		Acinar	16 (28)	11 (25.5)	
		Papillary	3 (5.3)	5 (11.5)	
		Micropapillary	2 (3.5)	8 (19)	
		Solid	2 (3.5)	9 (21)	
		Complex glands ²⁾	5 (8.8)	7 (16)	
		Lepidic 5%	43 (75)	9 (21)	<0.0001
		Micropapillary 5%	5 (8.8)	17 (40)	0.0006
		Solid 5%	10 (18)	22 (51)	0.0008
	Complex glands 5%	13 (23)	24 (56)	0.0009	
Other histological features	Tumor grade			0.0001	
		Grade 1	27 (48)	2 (4.7)	
		Grade 2	15 (26)	6 (14)	
		Grade 3	15 (26)	35 (81)	
		Lymphatic vessel invasion+	7 (12)	11 (26)	0.01
		Blood vessel invasion+	5 (8.8)	4 (9.3)	0.2
		Pleural invasion +	2 (3.5)	4 (9.3)	0.4
		Tumor necrosis +	8 (14)	13 (30)	0.08
Mutational status	<i>EGFR</i> (n=95)	17 (30)	7 (16)	0.1	
	<i>KRAS</i> (n=97)	11 (19)	20 (43)	0.002	
	<i>ALK</i> rearrangement (n=94)	1 (1.8)	1 (2.3)	0.2	

Data are *n* unless otherwise specified¹⁾ *p*-values were calculated by the t-test, and the chi-square test²⁾ including cribriform pattern

Table 2:

Intra-observer concordance of pathologists' interpretations of STAS and artifact in first and second rounds of evaluation

STAS	FS % Agreement	Kappa	FSP % Agreement	Kappa	NFP % Agreement	Kappa
Observer 1	81	0.508	79	0.567	73	0.488
Observer 2	82	0.554	81	0.455	86	0.713
Observer 3	78	0.531	83	0.649	91	0.816
Observer 4	83	0.651	91	0.819	82	0.648
Observer 5	88	0.748	88	0.737	74	0.443
Mean Intra-observer Concordance (%) (95% CI)	82.4 (79.2 – 85.6)		84.4 (78.7 – 90.1)		81.2 (75.5 – 86.9)	
Mean Cohen's Kappa ± SE		0.598 ± 0.100		0.654 ± 0.142		0.622 ± 0.155
Artifact	FS % Agreement	Kappa	FSP % Agreement	Kappa	NFP % Agreement	Kappa
Observer 1	63	0.286	69	0.389	75	0.399
Observer 2	73	0.460	79	0.561	82	0.579
Observer 3	65	0.314	72	0.431	74	0.470
Observer 4	77	0.419	81	0.487	72	0.451
Observer 5	77	0.532	81	0.617	70	0.396
Mean Intra-observer Concordance (%) (95% CI)	71.0 (62.8 – 79.2)		76.4 (68.2 – 84.6)		74.6 (66.4 – 82.3)	
Mean Cohen's Kappa ± SE		0.402 ± 0.102		0.497 ± 0.093		0.459 ± 0.074

FS: frozen section; FSP: frozen section permanent; NFP: non-frozen section permanent; CI: confident interval

Table 3.

Inter-observer agreement on interpretations of STAS and artifact in frozen (FS), corresponding permanent (FSP) and non-frozen permanent (NFP) sections among five pathologists

	Round	STAS		Artifact	
		Mean Inter-observer Concordance rate (%) (95% CI)	Fleiss' Kappa (mean κ \pm SE)	Mean Inter-observer Concordance rate (%) (95% CI)	Fleiss' Kappa (mean κ \pm SE)
FS	1	75.8 (72.9-78.7)	0.453 \pm 0.032	65.0 (62.5-67.5)	0.300 \pm 0.032
	2	77.4 (74.5-80.3)	0.506 \pm 0.032	60.2 (54.3-66.1)	0.204 \pm 0.032
FSP	1	76.6 (72.4-80.8)	0.477 \pm 0.032	70.6 (67.3-73.9)	0.377 \pm 0.032
	2	79.6 (76.8-82.4)	0.571 \pm 0.032	69.4 (63.1-75.7)	0.368 \pm 0.032
NFP	1	79.6 (75.9-83.3)	0.585 \pm 0.032	65.6 (60.6-70.6)	0.303 \pm 0.032
	2	82.6 (78.1-87.1)	0.646 \pm 0.032	67.6 (63.3-71.9)	0.331 \pm 0.032

CI: confidence interval; SE: standard error